

Jorge Serrano Cobos

Hábitos de recuperación de información en motores de búsqueda sobre lectura, libro y bibliotecas en España (2004-2016)

Departamento

Ciencias de la Documentación e Historia de la Ciencia

Director/es

Tramullas Saz, Jesús

<http://zaguan.unizar.es/collection/Tesis>

Departamento de Ciencias de la Documentación

Hábitos de recuperación de información en motores de búsqueda sobre lectura, libro y bibliotecas en España (2004 – 2016)

TESIS DOCTORAL

Autor:

Jorge Serrano-Cobos

Director:

Jesús Tramullas Saz



Universidad Zaragoza

Zaragoza, Febrero de 2019

“Lee, o te perderás algo extraordinario.”

(Roland J. Greenen, Chicago Sun-Times,
1 de Mayo 1988, en su revisión de
“Falling Free”, de Lois McMaster Bujold)

Agradecimientos

A la paciencia de mi director de tesis, mi familia, amigos, compañeros de trabajo y universidad, en especial a la Dra. Ángeles Calduch-Losa, Vicent Ahuir Esteve, Fernando Alcina Sanchis, y a la lectura, la lectura...

SUMARIO

Índice de figuras.....	11
Índice de tablas	16
Resumen	17

Contenido

1. INTRODUCCIÓN.....	23
1.1. Objeto de estudio.....	25
1.2. Objetivos	26
1.3. Método.....	27
1.4. Limitaciones	30
1.5. Fases de la investigación	33
1.5.1. Google Keyword Planner.....	33
1.5.2. Selección de expresiones semilla de búsqueda	39
1.5.3. Eliminación de frases de búsqueda no aplicables al estudio	44
1.5.4. Clasificación manual de las expresiones identificadas.....	47
1.5.5. Extracción de datos históricos de Google Trends	57
1.5.6. Procesamiento e integración de los datos	62
1.5.7. Clasificación entre expresiones de búsqueda temporales y no temporales.....	65
2. ESTADO DE LA CUESTIÓN.....	68
2.1. Introducción al análisis de los hábitos lectores.....	70
2.1.1. Necesidades en el sector editorial	72

2.1.2.	Necesidades en el sector de Bibliotecas	78
2.2.	EL ESTUDIO DE LOS HÁBITOS LECTORES	88
2.2.1.	Qué es lectura, qué es lector.....	88
2.2.2.	Metodologías de investigación de los hábitos de lectura	92
2.2.3.	Hábitos lectores del no lector	101
2.2.4.	El <i>best seller</i>	103
2.2.5.	Tipos de lectores	105
2.2.6.	Motivaciones para la lectura.....	108
2.2.7.	De la lectura en papel a la lectura en libro electrónico	112
2.2.8.	Leer en internet.....	116
2.2.9.	Conclusiones parciales	119
2.3.	INFORMATION SEEKING	121
2.3.1.	Introducción	123
2.3.2.	Algunos modelos de <i>Information Seeking Behaviour</i>	129
2.3.2.1.	Modelo de Wilson	131
2.3.2.2.	Modelo de Belkin	132
2.3.2.3.	Modelo Sensemaking	134
2.3.2.4.	Modelo de Ellis	136
2.3.2.5.	Modelo <i>Berrypicking</i> de Marcia Bates	138
2.3.2.6.	Modelo <i>Information Search Process</i> de Kulthau.....	140
2.3.2.7.	Modelo de Marchionini.....	142
2.3.3.	De <i>Information Seeking Behaviour</i> a <i>User Search Behaviour</i>	145
2.3.4.	<i>Browsing versus Searching</i>	147
2.3.4.1.	Nueva disciplina: Analítica Digital	148
2.3.4.2.	Casos de estudio.....	150
2.3.4.3.	Del análisis de un caso al análisis de la Red: la Cibermetría	154
2.3.5.	<i>User Search Behaviour</i>	157
2.3.5.1.	Multiplicidad de variables de estudio en una búsqueda <i>viasearching</i>	157

2.3.5.2.	Interfaz de búsqueda	159
2.3.5.3.	Interacciones durante la búsqueda.....	164
2.3.5.4.	Aspectos cognitivos del usuario	170
2.3.5.5.	Idioma y lenguaje en las expresiones de búsqueda.....	174
2.3.5.6.	<i>Query Intent</i> , Intención de Búsqueda	177
2.3.6.	Análisis longitudinal de la búsqueda	189
2.3.7.	<i>User search behaviouren</i> el sector de la lectura	195
2.3.8.	Recapitulación	201
2.4.	Conclusiones generales del marco teórico	203
3.	ANÁLISIS Y DISCUSIÓN DE RESULTADOS	205
3.1.	Introducción al análisis y discusión	207
3.2.	Discusión metodológica	209
3.2.1.	De <i>Google Keyword Planner</i> como herramienta para la obtención de expresiones de búsqueda relacionadas	210
3.2.2.	De la clasificación facetada de expresiones	213
3.2.3.	De <i>Google Trends</i> como fuente de datos cibernéticos de hábitos de búsqueda de lectura, libro y bibliotecas.....	223
3.2.4.	De la técnica de clasificación longitudinal de expresiones de búsqueda en temporales y atemporales	227
3.3.	Discusión de resultados.....	231
3.3.1.	Distribución de expresiones de búsqueda en las facetas	233
3.3.2.	Co-ocurrencia en la facetación.....	237
3.3.3.	Temporalidad y atemporalidad en facetas y aspectos.....	242
3.3.4.	Índices de búsqueda en la serie histórica	244
3.3.5.	Análisis de evolución general - temporales versus atemporales	248
3.3.6.	Análisis específico de evolución de expresiones temporales	254
3.3.7.	Análisis específico de evolución de expresiones atemporales	262
3.4.	En resumen.....	275

4.	CONCLUSIONES E INVESTIGACIÓN FUTURA	276
4.1.	Conclusiones y limitaciones por objetivo.....	278
4.2.	Conclusiones generales.....	290
4.3.	Futuras líneas de investigación	293
5.	BIBLIOGRAFÍA.....	298

ÍNDICE DE FIGURAS

Fig. 1-1 Ubicación de Google Keyword Planner en Google AdWords. Fuente: Google Adwords.	34
Fig. 1-2 Selección de opción en Google Keyword Planner. Fuente: Google Adwords.	35
Fig. 1-3 Campos a completar en el buscador de palabras clave de KeywordPlanner.....	36
Fig. 1-4 Palabras clave relacionadas con el término de búsqueda insertado, en este ejemplo, "diabetes". Fuente: Google AdWords.....	37
Fig. 1-5 Botón de descarga de palabras clave. Google AdWords.....	38
Fig. 1-6 Selección de descarga de palabras clave en formato .csv. Fuente: Google AdWords. ..	38
Fig. 1-7 Ordenación de expresiones de búsqueda en excel. Elaboración propia.	44
Fig. 1-8 Asignación de identificadores únicos para cada expresión de búsqueda. Elaboración propia.	45
Fig. 1-9 Ejemplo de gráfico de tendencias de búsqueda en Google Trends para "poemas de amor". Fuente: Google Trends.....	59
Fig. 1-10 Botón de descarga de datos de Google Trends en formato .csv. Fuente: Google Trends.....	60
Fig. 1-11 Proceso de importación de datos de Google Trends en formato .csv. Elaboración propia.	61
Fig. 1-12 Estructura de tablas de la base de datos de análisis de queries. Elaboración propia..	63
Fig. 2-1 Número de bibliotecas públicas en España (2011 - 2015). Fuente: Ministerio de Educación, Cultura y Deporte, 2018 . Elaboración propia.	78
Fig. 2-2 Promedio de préstamos por usuario en toda España. Fuente: Ministerio de Educación, Cultura y Deporte, 2016 . Elaboración propia.....	79
Fig. 2-3 Préstamos bibliotecarios en España (2011 - 2016). Fuente: Ministerio de Educación, Cultura y Deporte, 2016 . Elaboración propia.....	81
Fig. 2-4 Evolución de préstamos de libros digitales y audiolibros (2015-2017). Fuente: Ministerio de Educación, 2018 (elaboración propia).....	82
Fig. 2-5 Lectura total vs lectura digital. Federación de Gremios de Editores de España, 2012. .	89
Fig. 2-6 Personas que han leído libros por motivos no relacionados con la profesión o estudios en un trimestre según los temas de los libros leídos. Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015.	90
Fig. 2-7 ¿Cuál era la materia del último libro leído? Federación de Gremios y Editores de España, 2017.	91

Fig. 2-8 Prácticas de recomendación y obtención de lectura entre prestatarios de bibliotecas francesas. Fuente: RENARD, Hervé, 1995 (Cordón-García, 2006)	94
Fig. 2-9 Evolución del porcentaje de lectores frecuentes respecto a la población española. Fuente: Federación de Gremios de Editores de España, 2012.	95
Fig. 2-10 Evolución de la percepción del grado de lectura en España 1998-2016. Fuente: Centro de Investigaciones Sociológicas de España (CIS), 2016. (Elaboración propia)	96
Fig. 2-11 Ejemplo de pregunta en una encuesta sobre hábitos lectores. Fuente: Igarza y Monak Salinas, 2014.....	99
Fig. 2-12 Who doesn't read books in America? PEW Research, 2016.	102
Fig. 2-13 Novelas que más y menos gustan. Año 2004. Fuente: Federación de Gremios de Editores de España.....	104
Fig. 2-14 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.	106
Fig. 2-15 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.	108
Fig. 2-16 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.	109
Fig. 2-17 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.	113
Fig. 2-18 Criterio que tiene más en cuenta al elegir qué libro va a leer. Fuente: Centro de Investigaciones Sociológicas, 2016 (elaboración propia).....	117
Fig. 2-19 Tipos de variables, fuentes y técnicas en los Estudios de Necesidades (Núñez, 2004)	124
Fig. 2-20 Aspectos del proceso de búsqueda de la información estudios en la literatura científica en español, período 1990–2004 (González-Teruel, 2007)	125
Fig. 2-21 Tipos de estudios en HCI. Fuente: Dumais et al., 2014.	126
Fig. 2-22 Relaciones entre usuario, sistema de información y recursos de información. Fuente: Wilson, 1981.....	131
Fig. 2-23 Escalas y niveles de especificación de una necesidad de información. Fuente: Belkin, 1980.....	132
Fig. 2-24 Representación original del modelo Sensemaking. Fuente: Belkin, 1980.	134
Fig. 2-25 Modelo de Kuhlthau, Information Search Process. Fuente: Kuhlthau, 1989	140
Fig. 2-26 Interacción entre usuario y motor de búsqueda. Fuente: Xie, Yu y Cen, 2011.....	144
Fig. 2-27 Diversas formas de browsing en Renardus.com. Fuente: Koch et. al. (2004).....	150
Fig. 2-28 Best bets en BBC. Fuente Martin Belam, 2003.....	160
Fig. 2-29 Análisis de correlación de pares de términos en expresiones de búsqueda. Fuente: Silverstein et. al., 1999.....	162
Fig. 2-30 Moves operacionales y conceptuales. Fidel, 1985	166

Fig. 2-31 Moves para aumentar e incrementar precisión. Fidel, 1985.	167
Fig. 2-32 Distribución de palabras y términos usados en búsquedas. Fuente: Xie, Yu y Cen, 2011.....	174
Fig. 2-33 Actividades de búsqueda. Fuente: Marchionini, 2006.....	177
Fig. 2-34 Clasificación de intencionalidad de búsqueda. Fuente: Andrei Broder, 2002.	179
Fig. 2-35 Clasificación de intencionalidad de búsqueda. Fuente: Rose y Levinson, 2004.	180
Fig. 2-36 Clasificación de intencionalidad de búsqueda mediante aprendizaje semi-supervisado. Fuente: Li, Wang y Acero, 2008.	182
Fig. 2-37 Visualización del análisis con Eye-tracking de una página de respuestas (SERP) en Google. Fuente: Jakob Nielsen, 2006.....	183
Fig. 2-38 Visualización de análisis con Eye-tracking de búsquedas informacionales . Fuente: Van Gisbergen, Van der Most y Aelen, 2003.....	184
Fig. 2-39 Visualización de análisis con Eye-tracking de búsquedas transaccionales. Fuente: Van Gisbergen, Van der Most y Aelen, 2003.....	185
Fig. 2-40 Ejemplos de distintas composiciones de expresiones de búsqueda de tipo CQA. Fuente: Pinter, Reichart y Szpektor, 2016.....	186
Fig. 2-41 Beneficios y pérdidas de estrategia de inversión en bolsa utilizando Google Trends. Fuente: Preis et. al. (2013)	191
Fig. 2-42 Comparación de datos de Google Flu / CDC / Flu Near You. Fuente: Butler, 2013 ...	193
Fig. 2-43 Búsquedas navegacionales para Amazon y Barnes & Noble, con datos de comScore, 2012. Fuente: Baye, Santos y Wildenbeest, 2013.....	197
Fig. 2-44 Análisis longitudinal de búsquedas sobre títulos de libros en Google tras la salida del título. Fuente: Baye, Santos y Wildenbeest, 2013.	198
Fig. 3-1 Nota de Google Trends sobre cambios en asignación geográfica. Fuente: Google Trends (2018)	214
Fig. 3-2 SERP o Página de Resultados de Google ante la expresión “autores”. Fuente: Google.es (2018)	216
Fig. 3-3 Nota de Google Trends sobre cambios en asignación geográfica. Fuente: Google Trends (2018)	224
Fig. 3-4 Nota de Google Trends sobre cambios en sistema de recogida de datos. Fuente: Google Trends (2018)	224
Fig. 3-5 Evolución de índices de búsqueda de la expresión atemporal “biblioteca digital mundial”. Elaboración propia, sobre datos de Google Trends.	228

Fig. 3-6 Evolución de índices de búsqueda de la expresión atemporal “biblioteca digital mundial unesco”. Elaboración propia, sobre datos de Google Trends.....	229
Fig. 3-7 Facetas ordenadas por número de expresiones de búsqueda clasificadas en esa faceta. Elaboración propia.....	235
Fig. 3-8 Página de resultados de la búsqueda perteneciente a la faceta Plural “poesias del dia del padre cortas” realizada en octubre de 2018. Elaboración propia.....	236
Fig. 3-9 Grado de coincidencia en la facetación o co-ocurrencia en la facetación entre aspectos y facetas. Elaboración propia.....	238
Fig. 3-10 Comparativa de temporalidad o atemporalidad en las facetas.....	242
Fig. 3-11 Análisis global de facetas por frecuencias en la serie histórica.....	244
Fig. 3-12 Análisis de facetas por frecuencias en la serie histórica - Expresiones de búsqueda temporales.....	246
Fig. 3-13 Análisis de facetas por frecuencias en la serie histórica - Expresiones de búsqueda atemporales.....	247
Fig. 3-14 Serie histórica global (todas las facetas, temporales y atemporales) Elaboración propia.....	249
Fig. 3-15 Estacionalidad en toda la serie histórica (todas las facetas, temporales y atemporales) Elaboración propia.....	251
Fig. 3-16 Serie histórica (todas las facetas, atemporales) Elaboración propia.....	251
Fig. 3-17 Serie histórica (todas las facetas, temporales) Elaboración propia.....	252
Fig. 3-18 Serie histórica para la expresión de búsqueda “premio planeta 2010 ganador”. Elaboración propia.....	254
Fig. 3-19 Serie histórica para la expresión de búsqueda “naruto manga 407”. Elaboración propia.....	255
Fig. 3-20 Serie histórica para la expresión de búsqueda “naruto manga 300”. Elaboración propia.....	256
Fig. 3-21 Series históricas para cuatro expresiones de búsqueda relacionadas con Mario Vargas Llosa. Elaboración propia.....	257
Fig. 3-22 Serie histórica para la expresión “audiolibros megaupload”. Elaboración propia.....	258
Fig. 3-23 Serie histórica para la expresión “libro eclipse online”. Elaboración propia.....	259
Fig. 3-24 Serie histórica para la expresión “harry potter y el misterio del principe on line”. Elaboración propia.....	260
Fig. 3-25 Faceta Biblioteca- Evolución de queries atemporales.....	262
Fig. 3-26 Evolución de la expresión “biblioteca publica de zamora”. Elaboración propia.....	263

Fig. 3-27 Evolución de la expresión “biblioteca ugr”. Elaboración propia.	264
Fig. 3-28 Evolución de la expresión “biblioteca universidad de valencia”. Elaboración propia.	265
Fig. 3-29 Evolución de la expresión “frases celebres gabriel garcia marquez”. Elaboración propia.	266
Fig. 3-30 Evolución de la expresión “comics”. Elaboración propia.	267
Fig. 3-31 Evolución de la expresión “lobo comic”. Elaboración propia.....	267
Fig. 3-32 Evolución de la expresión “star wars comics”. Elaboración propia.	268
Fig. 3-33 Evolución de la expresión “libros eroticos”. Elaboración propia.	269
Fig. 3-34 Evolución de la expresión “frases principito”. Elaboración propia.	270
Fig. 3-35 Evolución de la expresión “lengua castellana y literatura”. Elaboración propia.	271
Fig. 3-36 Evolución de la expresión “libros novela romantica”. Elaboración propia.	272
Fig. 3-37 Evolución de la expresión “libros electrónicos”. Elaboración propia.....	272
Fig. 3-38 Evolución de la expresión “planeta editorial”. Elaboración propia.....	273

ÍNDICE DE TABLAS

Tabla 2-1 Préstamos por usuario en porcentajes, en las distintas regiones de España (2010-2016). Fuente: Ministerio de Educación, Cultura y Deporte, 2016. Elaboración propia.	80
Tabla 2-2 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.	97
Tabla 2-3 Criterio que tiene más en cuenta al elegir qué libro va a leer. Fuente: Centro de Investigaciones Sociológicas, 2016 (elaboración propia).....	110
Tabla 2-4 Fondos por clase de fondos en Bibliotecas españolas. Fuente: CULTURABase, 201 (Elaboración propia).....	114
Tabla 2-5 Modelo de Ellis de Information Seeking Behaviours. Fuente: Choo, Detlor y Turnbull, 2000.....	136
Tabla 2-6 "Web Moves" según Ellis. Fuente: Choo, Detlor y Turnbull, 2000.....	165
Tabla 2-7 Tipos de Web Moves. Fuente: Choo, Detlor y Turnbull, 2000.	168
Tabla 2-8 Seven Information Seeking and Using (ISU) Situations in the Workplace. Fuente Cheuk, 1998.....	172
Tabla 3-1 Volumen de búsqueda caso "Biblioteca de Madrid"	210
Tabla 3-2 Volumen de búsqueda caso "libro" versus "libros"	211
Tabla 3-3 Ejemplos de abreviaturas utilizadas en la muestra de expresiones de búsqueda....	215
Tabla 3-4 Cifras globales facetas y aspectos. Elaboración propia.....	233
Tabla 3-5 Porcentajes de Co-ocurrencia en facetación (en %)	239

RESUMEN

EN ESPAÑOL:

Este estudio ha tenido como objetivo principal, determinar si los procesos y expresiones de búsqueda de información usados por los usuarios en motores de búsqueda, pueden considerarse como indicadores válidos para el análisis y estudio de los hábitos de lectura y posible interés en otros contenidos ofrecidos por las bibliotecas en España (como videojuegos o películas).

Para ello se propone un modelo de análisis con el que caracterizar el lenguaje de búsqueda de información de los usuarios de internet que utilizan Google desde España como motor de búsqueda, durante el período 2004 - 2016, al recuperar información sobre la temática de el libro, la lectura y las bibliotecas, desde una perspectiva histórica. De esta forma, se pretende aportar otra dimensión de análisis a los estudios que hay sobre los hábitos lectores en general, y en España en particular.

La investigación tiene distintas áreas de aplicación del análisis del lector online, como son el apoyo a la indización y la clasificación bibliotecaria, la evaluación de colecciones y evaluación de la biblioteca, los estudios de necesidades de usuarios, la evaluación de OPACs, la analítica digital de sedes web bibliotecarias o de entidades de la industria del libro como editoriales, librerías online, metabuscadores o páginas web de autores y aficionados a la literatura en general, márketing bibliotecario y promoción de la lectura, márketing editorial, almetría y Cibermetría, SEO o ASEO (posicionamiento en buscadores académicos)

El análisis de los hábitos lectores tiene una larga tradición en el mundo offline, especialmente en España, donde el estudio de hábitos lectores es parte importante de la investigación estratégica en la industria del libro. Se han observado distintas metodologías, desde las encuestas y entrevistas a lectores y no lectores, el análisis de las ventas de los libros y la prensa, a los análisis de logs de préstamos en las bibliotecas.

Al entrar la lectura en e-book, y en plena era de internet, la lectura en papel ha sufrido una transformación, donde los usuarios leen por internet, y buscan su lectura (ya sea online, en e-book y/o en papel) a través de internet, especialmente utilizando motores de búsqueda, de los que en España el más utilizado desde principios de siglo hasta al menos su segunda década, es el buscador Google.

Es este cambio en las formas de localizar la lectura la que impulsa a investigar cómo se busca información sobre lectura en un buscador. Anteriormente se han investigado distintos aspectos de esas conductas con distintas técnicas, dentro del paradigma cognitivo, y especialmente dentro de la disciplina de Information Seeking, de difícil traducción al castellano. Tras consignar modelos de búsqueda por parte de los usuarios, como el modelo Berrypicking de Marcia Bates, el modelo de Ellis, el modelo de Marchionini, o el modelo de Information Search Process de Kulthau, entre otros, se han estudiado otros modificadores de las conductas de búsqueda, llegando a los estudios sobre User Search Behaviour (conductas de búsqueda de los usuarios en motores de búsqueda) especialmente en lo concerniente a desambiguación y expansión de búsquedas, análisis longitudinal de la búsqueda y de Query Intent, el Análisis de la Intención de Búsqueda.

Es precisamente en la combinación de las últimas subdisciplinas hacia donde se ha orientado este estudio. Para la investigación, en 2010 se obtuvieron de *Google Keywords Planner*, el log de búsquedas del motor de búsqueda, más de 30.000 expresiones de búsqueda (denominadas también como frases de búsqueda, queries, keywords o palabras clave), relacionadas con el libro, la lectura y las bibliotecas, segmentando la búsqueda de palabras clave en lenguaje español y de búsquedas realizadas desde España. Posteriormente se extrajo de *Google Trends* la serie de datos histórica de 2004 a 2016, para conformar un dataset con el que realizar un análisis longitudinal.

Las palabras clave fueron clasificadas en 27 facetas distintas de intención de búsqueda, contando también con aspectos modificadores y aspectos lingüísticos. Por tanto, no se clasificó en categorías mutuamente excluyentes, sino de forma que una expresión de búsqueda pudiera pertenecer a varias clases simultáneamente, por lo que se realizó un estudio del grado de co-ocurrencia entre las distintas facetas y los aspectos identificados.

Posteriormente se dividió las palabras clave, previamente clasificadas, en una nueva dimensión de análisis, según si era atemporales (tenían una larga vida en la serie histórica) o temporales, aquellas que nacían en algún momento de la serie, y tenían una vida más o menos corta.

Como resultado del análisis, se han estudiado las posibilidades de la facetación como mejora o complemento de otras técnicas de análisis de las intenciones de búsqueda (query intent analysis); se ha validado el modelo de estudio, de forma que sirva como corpus inicial de futuros análisis de los hábitos de lectura en España, a través del estudio de la demanda de

información en motores de búsqueda; se han descubierto subtipos de intenciones de búsqueda propias del sector de la lectura, dentro de las clasificaciones clásicas de intención de búsqueda (navegacional, informacional, transaccional); se han identificado facetas adicionales, distintas a las meramente temáticas, como modificadores y características del lenguaje, que sirvan para completar las facetas halladas desde una dimensión de análisis complementaria; se ha descubierto distintos patrones de uso, nuevas abreviaturas y formas de expresión de las necesidades de búsqueda de los usuarios mediante lenguaje natural, se han relacionado distintos media y/o formatos, así como, tras una selección mediante una muestra intencionada, de distintos ejemplos paradigmáticos de estas tendencias de búsqueda y sus posibles relaciones causales, observando los efectos producidos en la evolución de la demanda de información en torno a la lectura a través de la búsqueda de la misma en Google en España, durante el período 2004-2016.

Finalmente, y además de constatar su utilidad para completar otras técnicas de análisis de los hábitos lectores mediante una técnica inédita hasta la fecha en el sector del libro y bibliotecas, se ha observado cómo la demanda de información sobre lectura en España ha decaído de forma paulatina en la segunda década del siglo XXI, lo que coincide con otras investigaciones y datos de estudios de hábitos lectores, esta vez desde la perspectiva de la demanda online o a través de Internet.

PALABRAS CLAVE:

Hábitos de lectura, conductas de búsqueda de información, e-books, libro electrónico, google trends, lectores, análisis de logs, minería de búsqueda, information seeking, paradigma cognitivo, facetación de expresiones de búsqueda, categorización de palabras clave, análisis de la intención de búsqueda, recuperación de información, análisis de tendencias, análisis de la conducta de los usuarios, análisis longitudinal, altimetría, cibermetría, computación aplicada a las ciencias sociales

ABSTRACT IN ENGLISH:

The main objective of this study was to determine whether the search processes and expressions of information search used by users in search engines can be considered as valid indicators for the analysis and study of reading habits and possible interest in other content offered to users by libraries in Spain (like videogames or movies).

To this end, an analysis model is proposed, in order to characterize the information search language of Internet users who use Google from Spain in Spanish as a search engine, during the period 2004 - 2016, when retrieving information on the topics of books, reading and libraries, from a historical perspective. In this way, it was intended to provide another dimension of analysis to the studies on reading habits in general, and in Spain in particular.

The research has different areas of application for online readers analysis, such as support for indexing and library classification, evaluation of collections and evaluation of the library, studies of user needs, evaluation of OPACs, digital analytics of library web sites or entities of the book industry such as publishers, online libraries, metasearch engines or websites of authors and fans of literature in general, library marketing and promotion of reading, editorial marketing, altimetry and Cybermetrics, SEO or ASEO (positioning in academic search engines)

The analysis of reading habits has a long tradition in the offline world, especially in Spain, where the study of reading habits is an important part of strategic research in the book industry. Different methodologies have been observed, from surveys and interviews to readers and non-readers, the analysis of sales of books and the press, to the analysis of loan logs in libraries.

When entering the reading in e-book, and in the middle of the internet age, reading on paper has undergone a transformation, where users read online, and seek to read it (either online, in e-book and / or on paper)) through the internet, especially using search engines, of which in Spain the most used since the beginning of the century until at least its second decade, is the Google search engine.

It is this change in the ways of locating the reading that drives to investigate how to look for information on reading and what to read in a search engine. Previously, different aspects of these behaviors have been investigated with different techniques, within the cognitive paradigm, and especially within the discipline of Information Seeking Behaviour. After remembering search models such as the Berrypicking model by Marcia Bates, the Ellis model,

the Marchionini model, or the Kulthau Information Search Process model, among others, other search behavioral modifiers have been studied, coming to the studies on User Search Behavior (search behaviors of users but specifically using search engines) especially regarding the disambiguation and expansion of searches, longitudinal analysis of the search and query intent, the analysis of the intentionalities of users when searching for information.

It is precisely in the combination of the last subdisciplines where this study has been oriented. For the research were located in 2010, using *Google Keywords Planner's* search engine queries log, more than 30,000 search expressions (also known as search phrases, queries, or keywords), related to the book, the reading and libraries, segmenting the search for keywords in Spanish language and searches made from Spain. Subsequently, the historical data series from 2004 to 2016 was extracted from *Google Trends* to form a dataset with which to perform a longitudinal analysis.

The keywords were classified into 27 different facets of search intent, also having modifying aspects and linguistic aspects. Therefore, it was not classified into mutually exclusive categories, but in such a way that a search expression could belong to several classes simultaneously, so a study was made of the degree of co-occurrence between the different facets and the aspects identified.

Subsequently, the previously classified keywords were divided into a new dimension of analysis, according to whether it was prone to be "timeless" (they had a long life in the historical series) or "temporary", those that were born at some point in the series, and had a longer life or less short.

As a result of the analysis, the possibilities of faceting have been studied as an improvement or complement to other techniques of analysis of query intent; the model has been proved useful to serve as an initial corpus of future analysis of reading habits in Spain, through the study of the demand for information in search engines; subtypes of search intentions of the reading sector have been discovered, within the classic classifications of search intent (navigational, informational, transactional); additional facets have been identified, different from the merely thematic ones, as modifiers and characteristics of the language, which serve to complete the facets found from a dimension of complementary analysis; different usage patterns, new abbreviations and ways of expressing the users' search needs have been discovered through natural language, different media and / or formats have been related, as well as, after a selection through an intentional sample, of different paradigmatic examples of these search

trends and their possible causal relationships, noticing the effects produced in the evolution of the demand for information around reading through the search for it in Google in Spain, during the period 2004-2016.

Finally, and in addition to confirming its usefulness to complete other techniques for analyzing reading habits by means of a technique previously unpublished in the book and library sector, it has been observed how the demand for information on reading in Spain has gradually declined, in the second decade of the 21st century, a trend that coincides with other studies and data on reading habits, this time from the perspective of online demand or through the Internet.

KEYWORDS

reading habits, user search behaviour, e-books, google trends, information behaviour, patrons, readers, query log analysis, search mining, query categorization, information seeking, cognitive paradigm, search intent analysis, information search retrieval, trends detection, longitudinal study, altmetrics, cybermetrics, computational social science

1. INTRODUCCIÓN

1.1. Objeto de estudio

Decía Manuel Carrión Gútez (1981, 566) que “una biblioteca es -según la definición de la Unesco- una colección organizada de libros y publicaciones periódicas impresas o de otros documentos, sobre todo gráficos y audiovisuales, servidos por un personal encargado de facilitar el uso de ella por los lectores para información, investigación, enseñanza o recreo de los mismos”. Por tanto, el fin último de la biblioteca es servir al lector. Es el lector (para el mundo editorial) o el usuario (para el mundo bibliotecario) el que realmente da valor a la lectura, y por tanto los distintos actores del sector del libro (tanto del mundo editorial como del mundo bibliotecario) pueden tener un particular interés en conocer mejor sus cambiantes gustos, sus necesidades moldeadas por la vida y el contexto histórico de la sociedad en la que se mueven.

Si el lector es pues un objeto de análisis válido, en tanto actor imprescindible en el sector del libro, diríase necesario estudiar sus hábitos lectores. Los hábitos lectores hacen referencia a dos acepciones: primero, a “la repetición frecuente del acto de leer” y segundo, “el conjunto de destrezas y habilidades implicadas en esta actividad, adquiridas gracias a su iteración y al progresivo dominio de sus mecanismos” (Red de Universidades Lectoras, 2014)

En este trabajo se estudiará un apartado referente al análisis de estos hábitos lectores que no ha sido investigado todavía (Yubero y Larrañaga, 2013), en concreto el estudio de los hábitos de búsqueda en medios digitales (concretamente en el motor de búsqueda de Google) de información relacionada con el libro, la lectura y las bibliotecas, así como de otros contenidos producidos por las industrias culturales y ofrecidos en préstamo por las bibliotecas, a través del análisis de las expresiones o frases de búsqueda que utilizan los usuarios cuando buscan en Google.

1.2.Objetivos

El objetivo principal de este trabajo es determinar si los procesos y expresiones de búsqueda de información usados por los usuarios en sus procesos de búsqueda utilizando motores de búsqueda, pueden considerarse como indicadores válidos para el análisis y estudio de los hábitos de lectura y posible interés en otros contenidos ofrecidos por las bibliotecas en España (como videojuegos o películas).

En este sentido, se plantean asimismo una serie de objetivos específicos:

- Crear un corpus de expresiones o frases de búsqueda buscadas en Google, relacionadas con los hábitos de lectura y posible interés en otros servicios y contenidos ofrecidos por las bibliotecas en España, para que puedan realizarse distintos estudios de demanda online de información relacionada.
- Identificar subtipos de intenciones de búsqueda propias del sector de la lectura, distintas a las meramente temáticas, como modificadores y características del lenguaje, que sirvan para completar las clasificaciones clásicas de intención de búsqueda (navegacional, informacional, transaccional), desde una dimensión complementaria de análisis.
- Identificar relaciones entre los tipos de búsqueda y los elementos del lenguaje utilizados en la muestra.
- Conocer la evolución de la intención de lectura y posible interés en otros servicios y contenidos ofrecidos por las bibliotecas a través de la demanda de información en Google en España, durante el período 2004-2016.
- Identificar distintos tipos o formatos de contenidos (media), y observar las diferencias y semejanzas de las expresiones de búsqueda relacionadas con esos tipos o formatos, en la evolución de la demanda de información durante el período y segmentación señalados.

1.3.Método

Para el desarrollo del trabajo de investigación planteado, y tras consultar diversas metodologías de investigación relacionadas con la materia a estudiar (Cook y Reichardt, 1986; Hernández Meléndrez, 2006; Jansen, 2009; Taksá et. al., 2009; Jansen y Rieh, 2010; González Teruel y Barrios Cerrejón, 2012; Irazú Hernández et. al., 2012; Jun et. al., 2014; Igarza y Monak Salinas, 2014; González-Teruel, 2015) se ha seleccionado una metodología de investigación basada en el método exploratorio observacional.

A través de un estudio longitudinal retrospectivo, cualitativo – cuantitativo, utilizando estadística descriptiva (Hernández, 2006), y a partir de la observación de la evolución del uso de frases o términos de búsqueda en el motor de búsqueda Google (con datos extraídos de dos herramientas, *Google Keyword Planner* (Google, 2010) y *Google Trends*, 2017) previamente clasificadas mediante facetas o clases mutuamente no excluyentes, en forma de categorías de intención de búsqueda, aspectos modificadores, lingüísticos y temporales, y después sometidas a muestreo intencional en las series históricas para localizar casos paradigmáticos (García, 1999) se han generado nuevas formas de clasificación de las expresiones de búsqueda localizadas. De esta forma, se ha buscado aportar otra dimensión de análisis a los estudios que hay sobre los hábitos lectores en general, y de los de España en particular.

Esta metodología, aunque original en su conjunto, se inspira en el trabajo de otros autores, como Belkin (1980) que define e incorpora el concepto del “estado anómalo de conocimiento” (hoy día inserto en las necesidades de información expresadas por los usuarios en los motores de búsqueda); Kuhlthau (1988) que realiza estudios longitudinales para estudiar la recuperación de la información desde el punto de vista del usuario, y no del algoritmo, por lo que toma cuerpo el paradigma cognitivo; en las investigaciones longitudinales basadas en los datos de *Google Trends*, en particular *Google Flu* sobre la evolución de expresiones de búsqueda relacionadas con la gripe (Google.com, 2008) o la de Preis et. al. (2013) al estudiar los patrones de búsqueda en *Google Trends* sobre terminología referente.

También se basa especialmente en Broder (2002), que clasifica las intenciones de búsqueda de los usuarios que utilizan motores de búsqueda en tres grandes grupos, “informativa”, “navegacional” y “transaccional”. Y por último en Baeza-Yates et. al. (2006) y Jansen et. al. (2008), quienes ya hablaba de que en ciertas expresiones de búsqueda la clasificación era

considerada como “vaga” o “multifacetada”. La investigación se ha desarrollado a través de las siguientes fases:

1. **Selección de expresiones de búsqueda**

Se realizó un proceso de obtención de expresiones de búsqueda relacionadas con el libro, la lectura, las bibliotecas, así como de servicios ofrecidos por las bibliotecas, mediante la herramienta *GoogleKeywordPlanner*. Del conjunto inicial, se eliminaron frases de búsqueda no aplicables al estudio.

2. **Clasificación cualitativa de las expresiones identificadas**

Una vez delimitado el conjunto de expresiones de búsqueda a investigar, se clasificaron manualmente las mismas en torno a los tipos de intención de búsqueda antes mencionados para Broder (2002). Siguiendo a Baeza-Yates, Calderón y González-Caro (2006), esta clasificación se amplió explorando otros aspectos de las frases de búsqueda, como aspectos modificadores y aspectos lingüísticos, dando lugar a una clasificación que hacía uso de facetas (en la que una expresión de búsqueda puede ser agrupada en torno a varios criterios) en lugar de una clasificación monojerárquica (en la que cada expresión sólo podría pertenecer a una categoría de clasificación).

La clasificación facetada se origina con la clasificación colonada de Ranganathan (Garfield, 1984), que permite representar un concepto, una materia o un tema, desde distintos puntos de vista (facetas). En un proceso analítico-sintético, se identificaron posibles facetas o aspectos desde los que se podía clasificar a cada expresión de búsqueda, para después agrupar las frases de búsqueda en torno a las distintas facetas (Jansen y Booth, 2010). Por tanto, las palabras utilizadas en una frase de búsqueda pueden ser asociadas a una o varias facetas, debido a que una frase de búsqueda puede estar formada por una única palabra (por ejemplo “libro”, “biografía”) o por varias palabras (“diccionario de español”, “adios a las armas ernest hemingway”, etc.) En este último ejemplo, observamos que la frase se compone de un título de un libro (“adios a las armas”) y de un autor (“ernest hemingway”) por lo que la frase podría ser clasificada bajo ambas categorías.

3. **Extracción de datos históricos de *GoogleTrends***

A partir de la muestra de expresiones de búsqueda obtenida en el punto 1, se extrajeron datos de popularidad de búsqueda para cada expresión mediante *Google*

Trends, segmentando la petición de datos a usuarios de España, durante los años 2004 a 2016, para capturar toda la serie longitudinal de cada expresión de búsqueda. Estos datos se integraron con los obtenidos mediante *Google Keyword Planner* para cada expresión, en una base de datos, en la que además se realizó la clasificación facetada de las expresiones de búsqueda.

4. **Clasificación cuantitativa de las expresiones de búsqueda**

Una vez integrados los datos, se generó una nueva clasificación, en este caso de forma calculada en base a los datos y no manualmente, en torno a la temporalidad de la frecuencia de búsqueda de las expresiones buscadas. Se obtenía así otra dimensión, aspecto o faceta de análisis, agrupando las frases de búsqueda en “temporales” y “no temporales” o “atemporales”, siendo las expresiones de búsqueda o *queries* atemporales aquellas que mostraran un volumen aceptable de apariciones a lo largo del período estudiado, y no ofrecieran datos reseñables sólo en uno o pocos momentos del período de tiempo estudiado, que serían las “temporales”, cuya frecuencia parece depender de fenómenos aislados en el tiempo.

5. **Análisis y discusión de los resultados**

La hipótesis que se quería contrastar era que los datos longitudinales así obtenidos podían aportar información valiosa sobre las conductas de búsqueda explicitadas en forma de popularidad de las expresiones de búsqueda relacionadas con el ámbito de estudio. A tal efecto se generaron visualizaciones de datos longitudinales, en forma de series temporales, separando series de expresiones de tipo “temporal” y de tipo “atemporal”, de forma que ayudaran a describir e interpretar su comportamiento a lo largo del tiempo analizado, y permitieran descubrir patrones de uso comunes entre expresiones distintas, así como un análisis más detallado de las posibles relaciones causales de algunos de estos comportamientos a lo largo del período estudiado.

1.4.Limitaciones

El estudio, en virtud de las posibilidades de segmentación que ofrece *Google KeywordPlanner*, se circunscribe a los usuarios que buscan en Google desde España, sobre expresiones, frases o términos de búsqueda expresados en Español, (aunque como se verá en el siguiente apartado, aparecerán también expresiones en otros idiomas o que combinan el español con otros idiomas), durante el período 2004 – 2016, siendo 2004 el primer año del que Google ofrece información estadística relacionada de la frecuencia de búsqueda de las expresiones analizadas.

Como nota aclaratoria, en este estudio se utilizará en general “expresión de búsqueda”, para delimitar aquella cadena de caracteres, formada por una o varias palabras, que los usuarios utilizan para formular su necesidad de información en la caja de búsqueda de un motor de búsqueda de internet (en este caso, Google) También se podrá encontrar en este estudio, por aparecer así en la literatura al respecto, otras denominaciones o sinónimos del concepto, como “frase de búsqueda”, “cadena de búsqueda”, o “término de búsqueda”.

Asimismo, en la literatura anglosajona se suelen usar las expresiones *query*, *queryphrase* o *keyword*, resultando ésta un problema para la traducción, puesto que *keyword* para el mundo anglosajón tiene la misma connotación de “expresión de búsqueda”, cuando significa “palabra clave” (en singular, que para el castellano hablante implica en principio el uso de una sola palabra, no de una frase completa o de varias palabras) Este préstamo lingüístico ha provocado que buena parte de la literatura científica y técnica al respecto en español, también use esta forma como sinónimo de “expresión de búsqueda”, cuando ésta pueda estar formada, como se ha dicho, por una palabra (por ejemplo “autor”, o por varias palabras (por ejemplo, “autor alicia en el país de las maravillas”).

Para el estudio se ha utilizado la base de datos del log de búsquedas del propio motor de búsqueda de Google como buscador de referencia y fuente de información única para esta investigación, debido a que desde principios de siglo, acapara ya la mayoría del mercado mundial, y en la actualidad, el 95,35 % de la cuota de mercado español y 91,06 % a nivel mundial, en relación con los demás buscadores –Yahoo, Bing, Ask, etc.-, según la firma Stat Counter (Audemar, 2016) Esta cuota de penetración del mercado asegura la validez estadística de esta herramienta en cuanto a cobertura de la audiencia a investigar, formada por los

usuarios de Internet en España que utilizan un motor de búsqueda para expresar su necesidad de información.

El hecho de que la práctica totalidad de los usuarios de España que buscan en Internet utilice Google (Audemar, 2016), convierte el análisis de las frases de búsqueda en una herramienta de primer orden para complementar investigaciones cualitativas como encuestas y entrevistas, porque los usuarios no saben en general que son observados (manteniendo su anonimato) evitando así el sesgo del entrevistador y que éste influya en la percepción de las preguntas por parte del sujeto investigado.

Aunque otros motores de búsqueda permiten el análisis de las expresiones de búsqueda de sus usuarios, como los servicios Yandex *Wordstat* (Yandex, 2018), Baidu *Index* (Baidu, 2018) o Bing *Keyword Research* (Bing, 2018), Google tiene una herramienta similar y anterior, *Google Keyword Planner*, antes denominada *Google Keyword Tool* (Google, 2018), que contiene una base de datos histórica con millones de expresiones de búsqueda realizadas por miles de millones de usuarios cada día, y que será utilizada en una primera fase de la investigación.

Ha de tenerse en cuenta que esta herramienta está orientada a gestionar la publicidad basada en buscadores (*Google Adwords*) Aun así es relevante para esta investigación, puesto que permite descubrir expresiones que Google considera relacionadas con un término, frase o expresión de búsqueda para una segmentación dada (en este caso usuarios de Internet en España que se expresan en principio en español, aunque como se verá más adelante, también se obtienen expresiones relacionadas en otros idiomas o que contienen palabras de otros idiomas), por lo que, a día de hoy, no hay otra herramienta que sirva como fuente de información suficientemente exhaustiva, desde el punto de vista de la muestra, en relación al coste por uso (es gratuita).

Para la selección del conjunto de expresiones de búsqueda a analizar, se ha utilizado el sistema de recomendación de expresiones de búsqueda de esta herramienta, según la lógica que seguía en 2010 (año en el que se obtuvieron los términos de búsqueda a analizar) por dos motivos: para evitar posibles sesgos, y para comprobar la idoneidad de la herramienta en este aspecto. De esta forma, fue la herramienta la que aportó miles de expresiones de búsqueda relacionadas con las palabras “libro”, “lectura” y “biblioteca”, que sirvieron como semilla para obtener el consiguiente conjunto de frases de búsqueda, lo que se explicará en el capítulo dedicado a la extracción de datos.

Otra de las limitaciones del estudio viene dada de nuevo por la herramienta de extracción de expresiones de búsqueda. A pesar de que permite segmentar por país e idioma, aparecen expresiones que utilizan en todo o en parte palabras en distintos idiomas. Al ser un software propietario, que no explica cuál es su lógica al ofrecer expresiones relacionadas, no es posible delimitar si son frases usadas por usuarios que buscan desde España redactando en inglés u otros lenguajes sus necesidades de información, o las frases se incorporan debido a que se computan las empleadas por usuarios no españoles que están consultando al motor de búsqueda desde IPs de España.

Asimismo, se utiliza otra herramienta, *Google Trends*, que aporta datos cuantitativos sobre la frecuencia de búsqueda en un rango de fechas mayor que el que aporta *Google Keyword Planner*. Al permitir segmentar la petición de datos de la misma forma en ambas herramientas, se pudieron cruzar ambas para obtener la información necesaria. Nótese que ambos sistemas trabajan con usuarios anonimizados, y sólo ofrece información estadística del conjunto de usuarios que buscan desde España, por lo que, siendo ésta una investigación longitudinal, será imposible saber si lo que un usuario buscó en 2004, lo buscó también en 2016, sólo se obtendrán datos estadísticos del conjunto de la muestra analizada (usuarios anonimizados que buscaron desde España en Google, durante el período 2004 -2016) de la cual no sabremos si un mismo usuario repite expresiones de búsqueda a lo largo de los años, o si son varios usuarios los que buscan utilizando el mismo lenguaje en sus frases de búsqueda, o si se producen ambos tipos de acciones por parte de los usuarios.

Por último, estas herramientas trabajan con distintos umbrales de frecuencia de búsqueda mensual, por lo que en algunas expresiones sí se obtuvo información en ambos casos, y en otros no. Es decir, *Google Keywords Planner* ofrece como expresiones de búsqueda relacionadas expresiones con muy poca frecuencia de búsqueda mensual (incluso una decena o menos de repeticiones de esa frase en el buscador, para la segmentación pedida, al mes) Sin embargo, cuando se hacía la consulta a *Google Trends* para esa misma expresión, sólo aportaba datos históricos en ciertos casos, aquellos con frecuencia más alta, por lo que se la parte de investigación longitudinal sólo se ha podido llevar a cabo con una muestra de las expresiones de búsqueda inicialmente localizadas. Por tanto, mientras al investigar la muestra global, a través de la clasificación de las expresiones en torno a distintas facetas y aspectos, se obtiene una gran variedad de frases distintas, al investigar la muestra empleada en el análisis longitudinal, se trabaja con la muestra que ofrece una mayor frecuencia de búsqueda.

1.5.Fases de la investigación

A continuación, se describen en detalle los pasos dados en la consecución de la extracción y preparación de los datos necesarios para dar respuesta a los objetivos de estudio planteados, ejecutados a partir de la ayuda oficial de Google para su correcta segmentación (*Google Ads Help*, 2010).

1.5.1. Google Keyword Planner

Google KeywordPlanner (anteriormente conocida como *Google KeywordsTool* y denominada en español “Planificador de Palabras Clave”) está dirigida a los usuarios del sistema de gestión de publicidad en buscadores *AdWords*. Ayuda a los mismos a determinar por qué expresiones de búsqueda quieren pujar para aparecer entre los primeros puestos de las páginas de respuesta del motor de búsqueda (SERP o *SearchEngine Response Pages*) y a partir de ahí, asignar anuncios de texto ante las palabras clave elegidas, en determinados segmentos (país, lengua, horario, dispositivo, etc.)

Pero esta herramienta online también se puede usar con fines de análisis del lenguaje natural utilizado por los usuarios del motor de búsqueda, y a partir de ahí, para otras estrategias de Marketing online, como el Posicionamiento en Buscadores (SEO o *Search Engine Optimization*). Es muy útil para encontrar ideas de palabras clave por las que posicionar un sitio web o un contenido (pudiendo éste ser una página html, un pdf, un vídeo...) en el apartado de respuestas orgánicas o naturales de las SERPs, aunque lo que la hace más interesante para nuestro caso es que proporciona expresiones relacionadas con la palabra que incluimos como semilla, con el lenguaje natural de lo que se busca (el texto)

Asimismo, permite obtener cifras de volumen de búsqueda aproximada por cada mes, durante al menos el último año anterior al momento de la consulta, con lo que se podrá descubrir la evolución de la frecuencia de veces que se utiliza cada expresión de búsqueda en la segmentación aplicada en la investigación, durante ese período.

Recientemente *Keyword Planner* permite obtener hasta cuatro años de datos, por lo que se puede incluso descubrir la *seasonality* o estacionalidad de la misma, una posible pauta anual

de períodos de mayor y menor volumen de búsquedas sobre un determinado tema que permita predecir en qué momentos del año se va a buscar más sobre esa expresión de búsqueda. De esta forma se pueden establecer, planificar e integrar distintas acciones de marketing electrónico en esos momentos de mayor o menor frecuencia de búsqueda, aunque en el momento de realizar la extracción de datos, no se disponía de esta característica.

Para utilizar la herramienta es necesario previamente obtener una cuenta del correo electrónico de Google, *Gmail*. Posteriormente hay que darse de alta en el servicio de GoogleAdwords, y generar una cuenta, por la que no hay por qué pagar hasta que no se genera y pone en marcha una campaña de publicidad en buscadores o PPC (*Pay per Click*). Una vez se ha identificado con cuenta de usuario en *Google Adwords*, en primer lugar hay que acceder a la opción “Planificador de palabra clave” en el menú superior (Terrapin, 2010).

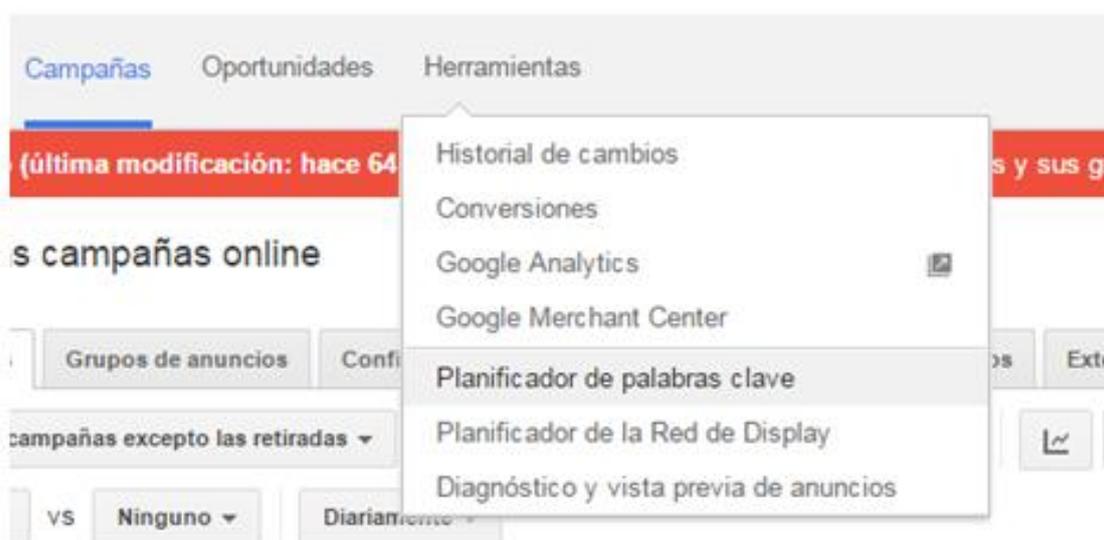


Fig. 1-1 Ubicación de Google Keyword Planner en Google AdWords. Fuente: Google Adwords.

Después se elige “buscar ideas para palabras clave o grupos de anuncios” y se escribe la palabra o frase que interese usar como semilla. Se puede también elegir otros parámetros para acotar la segmentación de expresiones de búsqueda que el sistema pueda aportarnos como relacionados, como la categoría del producto, el idioma o idiomas y el país o países, y se clicca en el botón “Obtener ideas” (ver ilustración).

Planificador de palabras clave

Planifique su próxima campaña de la red de búsqueda

¿Qué desea hacer?

- ▶ Buscar nuevas ideas para palabras clave y grupos de anuncios
- ▶ Obtener el volumen de búsquedas de una lista de palabras clave o incluirlas en grupos de anuncios
- ▶ Obtener previsiones de tráfico para una lista de palabras clave
- ▶ Multiplicar las listas de palabras clave para obtener nuevas ideas para palabras clave

Consejos sobre el Planificador de palabras clave

Encuentre nuevas palabras clave relacionadas con una frase, un sitio web o una categoría.

Campaña de la Red de Display? Pruebe el Planificador de palabras clave

Fig. 1-2 Selección de opción en Google Keyword Planner. Fuente: Google Adwords.

En un proceso de extracción de expresiones de búsqueda relacionadas con una temática (*keyword hunting*) se empieza por ingresar en el sistema palabras claves muy genéricas, compuestas de uno o dos términos, denominadas *semilla*, para que sea el sistema de Google el que nos provea de sugerencias temáticas, basadas no tanto en la relación semántica como en la coocurrencia de palabras buscadas en las mismas frases de búsqueda.

Es decir, si se usa como semilla “william shakespeare” nos daría como frases de búsqueda relacionadas (entre otras) “william shakespeare y romeo y julieta” o “william shakespeare hamlet resumen” por contener “william shakespeare”, aunque también parece realizar stemming a través del análisis de *n-grams*, ofreciendo asimismo expresiones como “williams shakespeare biografía”

🔍 Find new keywords and get search volume data

▼ Search for new keywords using a phrase, website or category

Enter one or more of the following:

Your product or service

Your landing page

Your product category

Targeting [?](#)

Customize your search [?](#)

Keyword filters

Keyword options

Show broadly related ideas

Hide keywords in my account

Hide keywords in my plan

Keywords to include

Date range [?](#)

Get ideas

Fig. 1-3 Campos a completar en el buscador de palabras clave de KeywordPlanner.

Como resultado, en la pestaña “Ideas para palabras clave” aparecerán las palabras que se han introducido junto con el volumen de búsquedas mensuales. Esto aportará una idea acerca de qué interesa más al usuario. Debajo de esa misma tabla se encuentra otra con una serie de palabras clave relacionadas con los términos de búsqueda que se habían indicado. Si se ordenan esos resultados por el promedio de búsquedas mensuales se podrán observar los términos más utilizados por los usuarios en relación al texto de búsqueda introducido, junto con la evolución del volumen de búsqueda (ver siguiente ilustración).

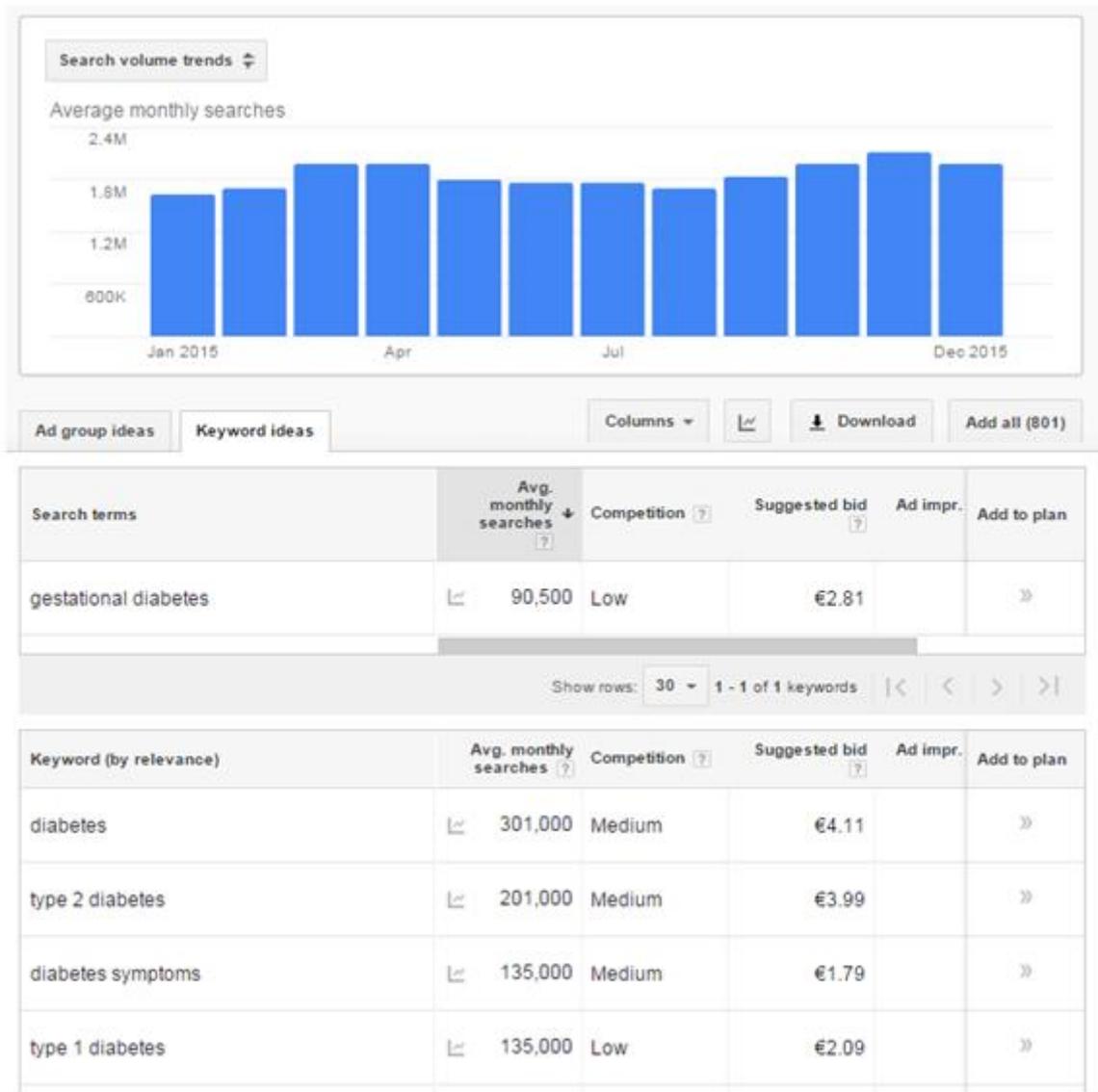


Fig. 1-4 Palabras clave relacionadas con el término de búsqueda insertado, en este ejemplo, "diabetes". Fuente: Google AdWords.

El sistema nos ofrecerá de cada expresión de búsqueda más indicadores, como la frecuencia media mensual de búsquedas de esa expresión de búsqueda durante los últimos 12 meses a partir del momento de realizar nuestro análisis (*Average Monthly Searches*) cuánta competencia hay para cada expresión en los anuncios de publicidad, y la puja sugerida para conseguir la primera posición de un anuncio de *AdWords* que quiera aparecer en primer lugar para el contexto seleccionado en la página de respuestas o *SERP* ante esa consulta. A continuación, se descargan las ideas palabras clave y la información estadística relacionada, para exportar los datos, clicando en el botón "descargar", en la parte superior del listado de resultados.



Fig. 1-5 Botón de descarga de palabras clave. Google AdWords.

Se elige la opción “CSV para excel”, y también la opción “segmentar estadísticas por mes”, lo que nos aportará datos de la frecuencia de búsqueda de cada frase o expresión de búsqueda durante cada mes de los últimos 12 meses.

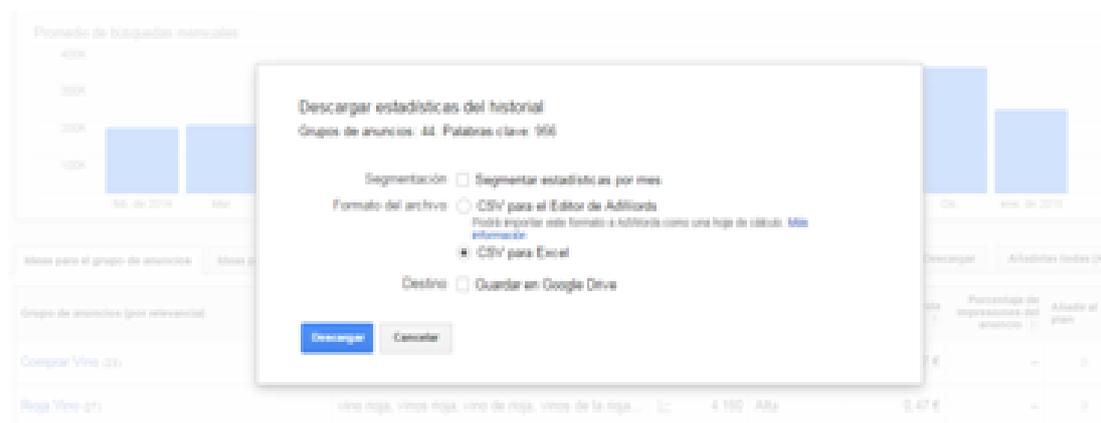


Fig. 1-6 Selección de descarga de palabras clave en formato .csv. Fuente: Google AdWords.

Este proceso se repite manualmente con cada una de las palabras semillas, extrayendo y descargando un archivo csv con todas las expresiones de búsquedas que ha ofrecido la herramienta. *Google Keyword Planner* ofrece un máximo de 800 expresiones de búsqueda relacionadas por cada palabra semilla, aunque no ocurre así siempre.

1.5.2. Selección de expresiones semilla de búsqueda

Para extraer la información que se precisa de *Google*, y así localizar las expresiones en lenguaje natural utilizadas por los usuarios del motor de búsqueda *Google* al buscar información sobre las temáticas analizadas, se ha empleado el sistema de descubrimiento de frases de búsqueda o *keywords* de *Google Adwords*, denominado *Keyword Planner*.

Para realizar la extracción de datos con esta herramienta en concreto, primero se ha de partir de un listado de palabras semilla. Una palabra semilla es una palabra o frase, una expresión de búsqueda, que *Keyword Planner* toma como fuente de información, a partir de la cual busca en su base de datos de frases de búsqueda, donde localiza otras expresiones de búsqueda con las que tenga relación, según su algoritmo. Por tanto, en lugar de incluir en el sistema, una a una todas las posibles variantes, se incluyen palabras genéricas (como “libro”, “lectura”, “leer”, etc) y el sistema las toma como referencias, para generar sus propios listados de expresiones relacionadas.

Desde un primer momento se desechó la idea de obtener datos de todos los títulos y autores posibles, y así extraer literalmente todo lo que se pudiera buscar sobre el tema, debido a la imposibilidad técnica de ser absolutamente exhaustivo en lo que se refiere a la producción mundial en materia de lectura. En el supuesto caso de querer extraer todas las expresiones de búsqueda relacionadas con todos los títulos de libros escritos hasta la actualidad, ni siquiera tomando como fuente de información la Biblioteca Nacional de España, ni todos los títulos de libros que hay en ella, ni todos los autores contenidos en su catálogo, para preguntar a Google si se busca (y cómo se busca) sobre todos esos títulos y autores, sería suficiente para abarcar completamente la producción mundial en literatura. Probablemente ni siquiera utilizando como fuente de información la Biblioteca del Congreso de los Estados Unidos, y eso sin contar que en las bibliotecas se consultan otros materiales como microformas, mapas, revistas, etc. Para que nos hagamos una idea de la enormidad de la tarea, Google estima que, excluyendo revistas (serials) había en 2010 129.864.880 libros en el mundo (Taycher, 2010), y continúan aumentando.

Pero incluso en el caso de conseguir generar un listado similar, la política de Google no permite la consulta de la frecuencia de búsquedas relacionadas con cada uno de esos millones de títulos o autores de forma programática y masiva, salvo que se utilice el servicio de pago, presupuesto imposible de asumir en esta investigación. Por otro lado, se tardarían quizá años

en hacerlo manualmente, puesto que habría que incluir cada palabra semilla y descargar el listado que la herramienta ofrece, una y otra vez (y de cada palabra semilla ofrece como máximo unas 800 expresiones relacionadas). Y si se hiciera con demasiada frecuencia, por ejemplo, muchas veces cada hora, Google prohibiría el acceso a *Keyword Planner* desde la IP que hiciera las consultas. Esto ocurre porque Google detecta el uso mecánico o excesivamente frecuente de su herramienta, por lo que habría que esperar periódicamente para volver a intentarlo.

A principios del siglo XXI Google puso a disposición del público una API (*Application Programming Interface*) para poder realizar consultas automáticas mediante programación, y así lanzar miles o centenares de miles de *queries* o consultas a la a su base de datos de expresiones de búsqueda. De esa forma se podían realizar estudios muy exhaustivos. Pero tiempo después, Google restringió su acceso y uso en general a clientes de publicidad en buscadores, y cobra en casos de mayor volumen como el que se necesitaría para abarcar tal cantidad de consultas a la herramienta (Google, 2018), por lo que, a efectos de una investigación global exhaustiva, ese camino está vedado.

Por tanto, para trabajar y extraer datos de *Google Keyword Planner* de forma adecuada según los términos de uso de la misma (Google, 2007), se ha de generar un listado de expresiones de búsqueda forzosamente limitado en número. A partir de este listado, se insertará en el sistema manualmente expresión a expresión. La herramienta devolverá para cada expresión, un listado de frases de búsqueda relacionadas con la insertada (así como ésta misma) junto con ciertos datos estadísticos de volumen o frecuencia de búsqueda de cada expresión.

Para no generar sesgos en la selección de expresiones de búsqueda que se utilizarían como semillas o frases iniciales para el estudio, se hizo una preselección de palabras semillas. Al localizar frases de búsqueda relacionadas con el libro, la lectura y las bibliotecas segmentando en la herramienta por España y en español, y para evitar sesgos, se inició una búsqueda en *Google Keyword Planner* utilizando esos tres términos precisamente (“libro”, “lectura”, “biblioteca”) como semillas.

Se tomó el listado de términos relacionados que la propia herramienta ofrecía para cada una de estas palabras (de unas 800 expresiones de búsqueda como máximo cada vez que se le pedía que aportara expresiones relacionadas a partir de una palabra semilla) De esa forma, en lugar de dirigir la extracción de expresiones hacia aquellas que se creyeran relacionadas a

priori, se eligió permitir que fuera *Google KeywordPlanner* el que aportara esas expresiones relacionadas, en virtud de su algoritmo de relación.

Precisamente se hizo así para evitar en lo posible la intervención humana y el sesgo del investigador, para poder descubrir nuevas expresiones que no se hubieran previsto y así aumentar la variedad de posibilidades de estudio en el lenguaje de búsqueda de los usuarios en torno al ámbito de la investigación, y de forma indirecta, para entender mejor los procesos de determinación de las prioridades de selección de expresiones de búsqueda relacionadas por parte de *Google Keyword Planner*.

De este modo se generó, a partir de cada palabra semilla, un nuevo conjunto de palabras o términos que se utilizaban a su vez como semillas, con las que realizar nuevas extracciones de expresiones de búsqueda por cada una de ellas, eliminando posteriormente duplicidades:

- audiolibro
- autor
- biblioteca
- capitulo
- certamen
- comentario
- comic
- concurso
- critica
- cuento
- edición
- editor
- editorial
- ensayo
- epopeya
- escritor
- fabula
- hemeroteca
- isbn
- lectura
- leer
- libreria
- libro
- literaria
- literario
- literatura

- manga
- musica
- novela
- opac
- pagina
- pelicula
- personaje
- poema
- poesia
- poeta
- premio
- publicacion
- publicar
- relato
- reseña
- resumen
- sinopsis
- sintesis
- tragedia
- tragicomedia
- verso
- videojuego

Además, a partir de la consulta en *Google Keyword Planner*, con la segmentación de idioma español y en España como país, de las expresiones de búsqueda relacionadas con “autora”, “autor” y “escritor” y “escritora”, se obtuvieron los nombres propios de escritores que aparecían mencionados entre esas expresiones relacionadas según la propia herramienta de Google. Para cada autor descubierto (recordemos que la extracción de expresiones de búsqueda se realizó en 2010), se hizo una nueva búsqueda de expresiones relacionadas, para ahondar en este aspecto concreto.

Así, se descubrieron frases de búsqueda que contuvieran menciones a ciertos autores que se utilizaron también como frases semilla para localizar expresiones relacionadas, de las que surgieron también menciones a otros autores. Se indican aquí los nombres tal y como aparecieron en las listas de expresiones relacionadas, aunque en ocasiones estén escritos incorrectamente o sin tildes, porque es como aparece en la fuente (*Google KeywordPlanner*):

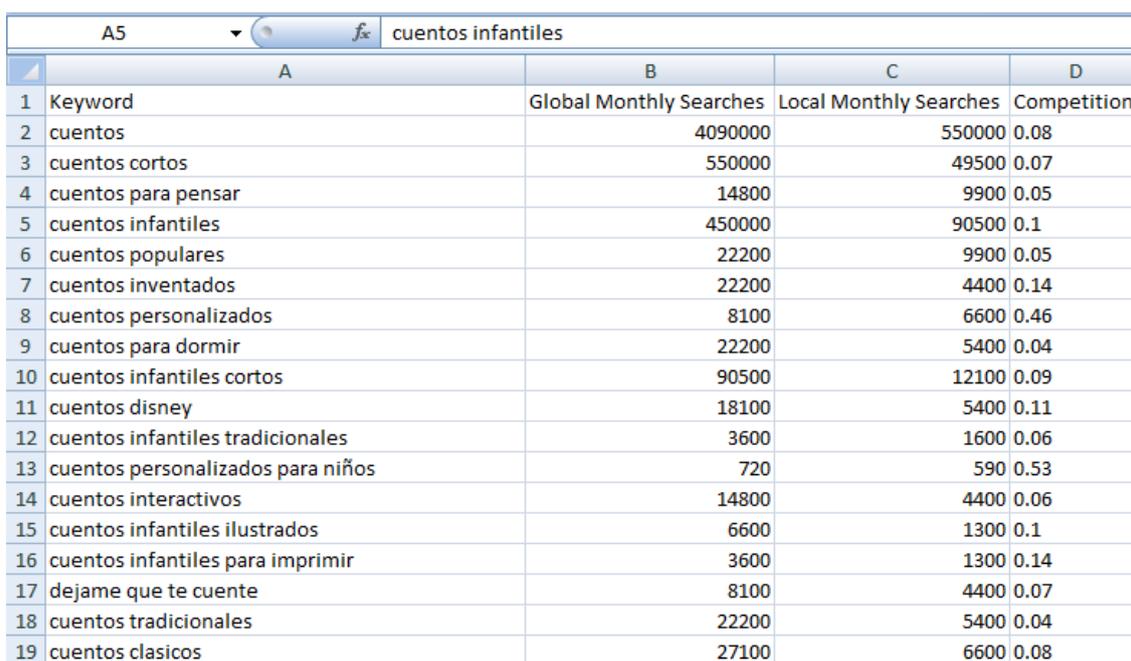
- Almudena Grandes
- Borges

- Bronte
- Cervantes
- Chejov
- Dickens
- Faulkner
- Fiodor Dostoievski
- Flaubert
- GarciaMarquez
- Goethe
- Hemingway
- Homero
- Kafka
- Lewis Carroll
- Machado
- Marcel Proust
- Nietzsche
- Ovidio
- J. K. Rowling
- Sartre
- Shakespeare
- Thomas Mann
- Tolstoi
- Truman Capote
- Vargas Llosa
- Verne
- Virginia Woolf

1.5.3. Eliminación de frases de búsqueda no aplicables al estudio

Una vez se han descargado los distintos archivos en formato csv, lo primero que se hizo fue unir todos esos archivos descargados en un único archivo utilizando la herramienta *Excel*, para unificar las expresiones, por lo que se deberán eliminar los duplicados de las expresiones de búsqueda, eliminando filas redundantes.

Aparecen palabras clave redundantes al integrar los archivos en uno, porque al utilizar palabras clave como semilla que ya están relacionadas en cierta medida, y en distintas sesiones, tras cada petición de elementos relacionados, se descarga un archivo distinto, y ofrecen algunas expresiones de búsqueda relacionadas que ya han aparecido anteriormente, por lo que hay que eliminar esos duplicados y así tener un listado único. Así se obtuvo un conjunto total de más de 37.500 expresiones de búsqueda diferentes.



	A	B	C	D
1	Keyword	Global Monthly Searches	Local Monthly Searches	Competition
2	cuentos	4090000	550000	0.08
3	cuentos cortos	550000	49500	0.07
4	cuentos para pensar	14800	9900	0.05
5	cuentos infantiles	450000	90500	0.1
6	cuentos populares	22200	9900	0.05
7	cuentos inventados	22200	4400	0.14
8	cuentos personalizados	8100	6600	0.46
9	cuentos para dormir	22200	5400	0.04
10	cuentos infantiles cortos	90500	12100	0.09
11	cuentos disney	18100	5400	0.11
12	cuentos infantiles tradicionales	3600	1600	0.06
13	cuentos personalizados para niños	720	590	0.53
14	cuentos interactivos	14800	4400	0.06
15	cuentos infantiles ilustrados	6600	1300	0.1
16	cuentos infantiles para imprimir	3600	1300	0.14
17	dejame que te cuente	8100	4400	0.07
18	cuentos tradicionales	22200	5400	0.04
19	cuentos clasicos	27100	6600	0.08

Fig. 1-7 Ordenación de expresiones de búsqueda en excel. Elaboración propia.

A partir de ahí ya se pueden ordenar, bien por orden alfabético de palabra clave, bien por la frecuencia de búsquedas. En este caso se optó por ordenarlas alfabéticamente. Después, se generó una columna nueva para asignar un número correlativo que sirviera como identificador único de cada palabra o expresión de búsqueda, y así después utilizar ese identificador único

para crear una base de datos en la que integrar la información obtenida para cada expresión de búsqueda

1	Id	KEYWORDS
2	1	Oscar ovidio
3	2	1964 borges
4	3	1964 borges analisis
5	4	1964 jorge luis borges
6	5	3 mujeres escritoras
7	6	6 personajes en busca de un autor
8	7	a christmas carol de charles
9	8	a la busca del tiempo perdido
10	9	a la busqueda del tiempo perdido
11	10	a los amigos borges
12	11	a mi manera autor
13	12	a puerta cerrada sartre
14	13	a puerta cerrada sartre descargar
15	14	a puerta cerrada sartre pdf
16	15	a puertas cerradas sartre
17	16	a que movimiento literario
18	17	a un gato borges
19	18	a un gato jorge luis borges
20	19	abel hernandez escritor
21	20	abigail borges

Fig. 1-8 Asignación de identificadores únicos para cada expresión de búsqueda. Elaboración propia.

Una vez preparada esta “lista maestra”, eliminadas ya las expresiones de búsqueda duplicadas, el siguiente paso de este punto fue la eliminación de frases de búsqueda no aplicables al estudio. Este paso era necesario para desambiguar expresiones que compartieran palabras del segmento temático de interés, pero que no estuvieran relacionadas (con la lectura o con servicios o formatos ofrecidos por las bibliotecas, como videojuegos o películas en préstamo) Así, se acometió un proceso selectivo de forma manual, en la que fueron segregadas un elevado número de expresiones de búsqueda. Por ejemplo:

- Relacionadas con la expresión de búsqueda “novela”, aparecieron términos como “telenovela” y a partir de éste, miríadas de palabras clave como “comprar telenovelas”, “capitulos telenovela marimar”, y un largo etcétera. Este tipo de

expresiones no tenían que ver ni semánticamente con el libro o la lectura de novelas, sino con series de televisión (Besancon, Rajman y Chappelier, 1999; Gong et. al., 2013).

- En otros casos eran palabras que se usaban junto con las palabras semilla, pero con un sentido distinto: un ejemplo viene de “autor”. Entre otras expresiones relacionadas, generaba otras como “homero”, palabra que vino de buscar “autor”, pero que a su vez aportó palabras clave que había que desambiguar y apartar, como “homero cantante peruano” u “homero grunge capitulo”, que hacían referencia a otro tipo de personajes (un músico y un personaje de serie de televisión, respectivamente).
- En otras ocasiones, y demostrando la acción de análisis de n-grams de Google, al buscar palabras clave como “borges”, se obtenían otras como “borgir”, y a partir de ella como relacionadas, otras como “dimmuborgir abracadabra”, donde DimmuBorgir es un grupo de música noruego. O “viajar a rusia por tu cuenta”, que Google entendía relacionada con “cuento”.

Este paso se realizó hasta cuatro veces manualmente a lo largo de todo el listado, para minimizar la tasa de aparición de expresiones de búsqueda que no estuvieran temáticamente relacionadas, eliminando más de 6.500 palabras clave de diversa índole, consolidando un conjunto de 30.925 frases de búsqueda.

1.5.4. Clasificación manual de las expresiones identificadas

Con la lista maestra ya expurgada de expresiones no verdaderamente relacionadas con el ámbito de estudio, se procedió a la clasificación manual de las expresiones identificadas en torno a sus intenciones de búsqueda potenciales (Stokes y Urquhart, 2013). Hay dos procedimientos para llevar a cabo este proceso:

- A. **Top Down (de arriba abajo):** se determinan unas categorías, y se busca que las distintas queries entren en la definición de cada una de ellas. Este proceso ha llevado a ciertos autores a generar categorías generales (Broder, 2002; Rose y Levinson, 2004), y a delimitar otras como “ambiguas” (Baeza-Yates, Calderón y González-Caro, 2006).
- B. **Bottom up (de abajo arriba):** se recorre la lista de expresiones de búsqueda, y se van descubriendo las características de cada búsqueda, generando distintas categorías conforme van apareciendo nuevas. Una vez recorrido el listado y desglosadas todas las características / categorías que hayan aparecido, se vuelve a tomar el listado, y ahora sí, se clasifica cada expresión en torno a una o varias de las clases delimitadas anteriormente. Esta técnica (Rosenfeld y Morville, 1998), aunque laboriosa, permite desglosar con alta granularidad y variedad, las características de las expresiones de búsqueda.

Para clarificarlo, tomemos un ejemplo de clasificación de tipo *bottom up*, esta pequeña lista de expresiones de búsqueda relacionadas con la palabra “cervantes”:

- autor miguel de cervantes
- avenida miguel de cervantes murcia
- bibliografía cervantes
- biblioteca cervantes
- biblioteca cervantes digital
- biblioteca miguel de cervantes burgos

Como se puede observar, aunque todas las expresiones de búsqueda contienen la palabra “cervantes”, cada una de ellas hace referencia a una intención distinta por parte del usuario, donde una sola palabra puede suponer una gran diferencia:

- “autor miguel de cervantes”: el usuario está buscando información sobre el conocido escritor.
- “avenida miguel de cervantes murcia”: el usuario está buscando una avenida que lleva por nombre el del escritor, y además utiliza una localización geográfica para especificar o acotar su localización.
- “bibliografía cervantes”: el usuario busca bibliografía sobre el autor miguel de cervantes, sin especificar su nombre completo.
- “biblioteca cervantes”: el usuario busca la biblioteca denominada con la marca cervantes. No especifica más qué tipo de biblioteca o dónde está.
- “biblioteca cervantes digital”: en este caso se busca la biblioteca, pero se especifica que busca un sitio web, mediante el uso de la palabra “digital”.
- “biblioteca miguel de cervantes burgos”: finalmente, se busca una biblioteca denominada con el nombre del autor, pero de nuevo se utiliza una ciudad para especificar que lo que busca es o su localización física, o información sobre esa biblioteca concreta, diferenciándola de otras posibles bibliotecas con el mismo nombre.

Por tanto, en este pequeño listado se han detectado distintas posibles categorías bajo las que clasificar las distintas expresiones de búsqueda, como “autor”, “biblioteca”, “lugar”, nunca antes determinadas en investigaciones anteriores (Jones et. al., 2008; Jun, Yoo, Choi, 2018) ... A partir de ahí, podremos ir localizando otras expresiones que se agrupen bajo estas categorías. Cuantas más expresiones se contemplen, más amplia la casuística, y más posibilidades de encontrar nuevas facetas mutuamente no excluyentes.

Hay que señalar que la facetación como estrategia para la definición de una ecuación de búsqueda no es nueva. Los sistemas de recuperación de información que permitían la búsqueda avanzada ya aportaban la posibilidad de ejecutar una facetación en esa expresión de la necesidad de información, como por ejemplo para buscar una palabra en un conjunto determinado de documentos, pero que fueran publicados en un determinado año, o en un determinado formato. Pero, inicialmente, este ejercicio de traducción de la necesidad del usuario al lenguaje de búsqueda era realizado por el bibliotecario o documentalista (Taylor, 2015)

Posteriormente llega el avance de los sistemas de recuperación de información hacia motores de búsqueda que tratan bien de adaptarse a los procesos de los usuarios en su contexto de uso

(Pharo y Järvelin, 2004; Sugiyama et. al., 2004), bien de adaptar al usuario al interfaz de búsqueda y a la lógica del motor de recuperación de información, a través del análisis de su conducta (Amitay, 2006; Jansen et al. 2007; Jones et. al. 2007; Jamali y Asadi, 2010). En esa evolución, es en mayor medida el propio usuario quien realiza esa labor, a través de ecuaciones de búsqueda que no utilizan comandos booleanos para conectar palabras, sino palabras dentro de frases (Feuer, 2007; Bendersky y Croft, 2009)

Este paso da lugar al uso de palabras que, en combinación con otras, puedan expresar necesidades distintas, dando como lugar un sistema de consulta mediante lenguaje natural que utiliza la facetación dentro de la propia expresión de búsqueda (Dou et. al., 2011; Saha Roy et. al., 2011) de forma directa o por reformulación (Mimiet.al. 2006), y no seleccionando operadores específicos o interactuando con funcionalidades propias del interfaz de búsqueda mediante clics (Beaulieu, 1997; Maniega-Legarda, 2002; Manning et. al., 2008; Stamou y Efthimiadis, 2010).

Por tanto, la clasificación de cada frase de búsqueda en distintas facetas mutuamente no excluyentes, implica una categorización cualitativa de las mismas, en torno a facetas y (en esta investigación) en torno a aspectos modificadores y lingüísticos. La distribución que genera la clasificación de las distintas expresiones no se define a priori, pues es la observación de los datos la que la determina, y lleva a la pregunta básica de si, como vemos en estos ejemplos, hay una relación entre las facetas y los aspectos descubiertos, entendiendo aquí relación como “una forma de agrupamiento entre elementos” (Camarero, 2002, 380)

Para el presente estudio, se eligió la técnica B, *bottom up*, por adaptarse mejor a las necesidades de la investigación, porque en lugar o además de intentar clasificar las palabras clave en las intenciones de búsqueda canónicas (informacional, navegacional y transaccional) se buscaba descubrir qué otras peculiaridades y características tenían las expresiones de búsqueda relacionadas con la temática a cubrir.

Para incluir mayor granularidad en el estudio y eliminar en lo posible el sesgo humano a la hora de clasificar, se separaron todas las palabras que conformaban cada frase de búsqueda. Se eliminaron duplicados, obteniéndose así un listado de palabras sueltas formado por 8.416 términos.

A continuación, este listado se clasificó manualmente, aumentando conforme se avanzaba el listado de clases a utilizar para su posterior clasificación. La identificación de clases surgió de la propia aparición recurrente de expresiones que indicaban una determinada clase, aportando

distintas características de cada expresión de búsqueda que ofrecían diferentes variables de análisis (siguiendo a autores como Hölscher y Strube, 2000; Gerhard Case y Given, 2002; Vundavalli y Hyderabad, 2008; Wolfram et. al., 2008) Una posterior clasificación manual de las frases completas, detectó un 3% de error en la clasificación, según los criterios que se explican a continuación.

En la gerneación de esta clasificación facetada, se identificaron características que podían ser utilizadas en cualquier sector, y otras más propias del sector de la lectura, lo que ya se había identificado en otros ámbitos, como la medicina o el comercio electrónico (Costa y Seco, 2008; Wang et. al., 2008; Teevan et. al., 2008 e Islamaj Dogan et. al., 2009) Estas clases definen diferentes características del comportamiento de los usuarios, al expresar en una consulta de búsqueda su necesidad de información en torno a las temáticas que ocupan el estudio¹, tal que:

- **Abreviaturas y siglas:** según la RAE, son una “representación gráfica reducida de una palabra o grupo de palabras, obtenida por eliminación de algunas de las letras o sílabas de su escritura completa”². La RAE dispone de un listado de abreviaturas, que como ella misma dice, “se trata de una lista necesariamente incompleta, ya que cualquier usuario de la lengua puede crear cuantas abreviaturas considere oportunas”³, lo que se ha comprobado en este estudio, pues al identificar las expresiones utilizadas como abreviaturas, se ha visto que los usuarios utilizan una gran diversidad de formas de abreviar. Por ejemplo, se usa “&” por “and” o “Y” (el operador booleano que indica que se buscan dos términos conjuntamente) Asimismo, hay multitud de formas de indicar el orden, por ejemplo “1arte”, “1o”, “1º”, “3d”, “3gp”, “3o”, “8va”, “9no”, etc. Asimismo, se han incluido aquí, expresiones como acrónimos (“ucm”, “uvigo”) nombres de formatos (“mp3”) o números romanos (“XVI”).
- **Ámbito:** en esta clase se han incluido aspectos que indicaran una pertenencia a alguna forma de agrupación. Así encontramos palabras que indican su inclusión o delimitación bajo una lengua (“alemán”, “castellano”) bajo un país (“venezolanas”, “colombiano”) una provincia (“andaluza”, “riojanos”) zonas geográficas más difíciles de especificar

1 **Nota:** muchos de los ejemplos de palabras clave se leen sin tildes a propósito, pues se transcriben tal cual se buscan en *Google* (así “cuando”, “vivio” y no “cuándo”, “vivió”)

2 <http://lema.rae.es/dpd/?key=abreviatura>

3 <http://www.rae.es/diccionario-panhispanico-de-dudas/apendices/abreviaturas>

(“andina”, agrupaciones culturales (“anglosajona”, “afrocolombiana”, “goticos”)
períodos (“postguerra”, “postmodernismo” “precolombina”) etc.

- **Autor:** aquí se han incluido palabras que identificaban claramente autores conocidos, tales como “alberti”, “asimov”, “quevedo”, “reverte”...
- **Biblioteca:** palabras que contienen la raíz “bibliotec” para clasificar aquellas frases que se clasifican dentro de búsquedas sobre las bibliotecas, tomadas éstas como entidades (como instituciones).
- **Ciencia:** palabras que se identificaban con terminología científica, como “antropología”, “matemáticas”, “sociología”, “sexualidad”...
- **Deporte:** se incluyen distintas expresiones sobre deporte y ejercicio, como “ajedrez”, “supercampeones”, “basquetbol”, “basket”...
- **Desambigüación:** aunque en realidad, la mayoría de las expresiones que se integran en otras clases, hay algunas que aportan una mayor carga semántica en la necesidad del usuario de desambiguar, de especificar su necesidad de información. Aquí se integran términos que también pueden verse en otras clases, como “ámbito”, y otras que denotan esa intención desde diversos puntos de vista, con ejemplos tales como “actores”, “artística”, “celebres”, “críticos”, “dadaístas”...
- **Entidad:** se incluyen términos que claramente denotan marcas de entidades públicas o empresas (“aenor”, “alfagrama”,...) entidades indicadas de forma más genérica (“agencia”, “gobierno”...) e incluso urls completas (www.poemas.del.alma.com)
- **Erótico:** aquí se clasificaron términos que explícitamente hacían uso de conceptos eróticos y/o pornográficos (“erotico”, “parafilias”...) como otros que no expresaban una terminología específicamente erótica, pero que en su contexto sí lo eran (“sukia”, “selen”...) Para localizar y desambiguar esas expresiones, se realizaron búsquedas en *Google.es* (en ventana de modo incógnito para evitar que *Google* personalizara o segmentara los resultados en función del histórico de búsquedas de un usuario logado, o de una cookie)

- **Error:** son palabras mal escritas, que pueden aparecer solas o dentro de frases de búsqueda donde el resto de las palabras sí pueden estar correctamente escritas, incluyendo palabras sin tilde (“basquetbol”, “abrahamadabra”, “saavedr”...)
- **Film:** hacen referencia a términos que derivan en búsquedas relacionadas con películas o series de televisión (“gags”, “nominaciones”, “trailers”...) y que surgen al haber incluido como palabras semillas algunas como “critica”, “capitulo”, “película”, “personaje”, “resumen”, “sinopsis”, “síntesis”.
- **Formato:** palabras que ayudan a desambiguar expresiones de búsqueda en las que se especifica el formato de lo buscado (“actas”, “album”, “audiolibros”, “cartoon”, “comics”...)
- **Género:** palabras que contienen una identificación del género de lo buscado (“actor” / “actriz”, “abuela” / “abuelo”) también en plural (“alemanas”, “andaluzas”...)
- **Idioma:** se incluyen aquí palabras en otros idiomas (inglés, latín, alemán, francés...) y préstamos lingüísticos que se usan en el castellano ya común (“bestseller”, “pizza”, “podcast”...) Hay que apuntar aquí dos anotaciones:
 - Que las expresiones que aparecen utilizando en todo o en parte palabras en distintos idiomas, lo hacen a pesar de haber segmentado la búsqueda de expresiones relacionadas en *Google KeywordPlanner* a Español de España. Es imposible saber si esas expresiones son usadas por hablantes españoles que escriben en inglés u otros idiomas en algunas de las sesiones de búsqueda que realizan en el buscador, o usuarios de otras nacionalidades que buscan desde IPs radicadas en España.
 - Por la misma circunstancia, no se han separado expresiones que denotan un español más común en otros países hispanohablantes de Hispanoamérica que no son España (“alegrijes”, “payadas”, “tepozteco”, etc.) No es posible saber si hay algún error en la forma que tiene Google de segmentar las búsquedas, y está tomando expresiones en español de Hispanoamérica realizadas desde IPs fuera de España, o si esas expresiones de búsqueda son buscadas por usuarios que utilizan como lengua materna el español de otros países de Hispanoamérica, pero buscan desde IPs geolocalizadas en España. Se asume esta segunda posibilidad, que está avalada por algunos experimentos manuales (aunque lamentablemente, sin relevancia estadística suficiente para demostrarlo) en los que, a lo largo de los años, *Google* ha ido presentando

para IPs de España resultados que antes eran desambiguados por geolocalización. Como ejemplo, la búsqueda “saco” antes producía resultados muy distintos en España (donde “saco” es una bolsa hecha de arpillera o similar) y en otros países hispanoamericanos, como Argentina (donde un “saco” es un traje de chaqueta) Hoy día es común encontrar resultados de ambas acepciones en búsquedas con este término de búsqueda realizadas desde España. Además, la inmigración suponía ya el 12% de la población de España en 2010 (Expansión, 2017), de la cual el 24,8% provenía de América del Sur (INE, 2017) por lo que el que aparezcan expresiones o usos latinoamericanos en este estudio, es de por sí un hecho interesante a analizar en el futuro.

- **Infantil:** se incluyen aquí tanto términos utilizados para desambiguar este tipo de información (“infantil”, “juvenil”) como otros que denotan una búsqueda sobre temáticas infantiles (“disney”, “sirenita”, “pitufos” ...)
- **Lugar:** se inscriben aquí palabras que hacen mención explícita de lugares geográficos (“alava”, “florenia”, “francia”, “yucatan”, “zaragoza” ...) No confundir con la clase **Ámbito:** “colombiano” se clasificará en la clase “Ámbito”, mientras que “Colombia” se clasificará dentro de la clase “Lugar”.
- **Nombre personal:** se han incluido en esta clase todos los nombres encontrados (“anna”, “horacio”, “richard”) y apellidos (“pattinson”, “santos” ...) que no pertenecieran solamente a la categoría “autor”, aunque en ocasiones estos nombres / apellidos también formaban parte del nombre de un autor, o podían funcionar como un personaje de una obra (“anibal”, “romulo”, etc.)
- **Personaje:** personajes de obras (“aladdin”, “aladino” ...) que a veces no eran obvios, y sólo una búsqueda de frases completas desambiguaban suficientemente (“alfombra”, que en este contexto hacía referencia a la alfombra de Aladino) También se incluyen aquí lugares míticos (“atlantis”, “camelot”...) y nombres de persona que pueden ser buscados tanto como autores, nombres personales, o personajes de obras de ficción o de no ficción (“bolivar”, “bonaparte” “conan”, “potter” ...)
- **Plural:** términos que contienen el plural (“salmos”, “tareas”, “vampiricos”, “zombies” ...)

- **Pregunta / frase:** palabras que indican que la frase en la que están incluidas, denotan una pregunta realizada en lenguaje natural (“cuando”, “cuantos”, “quienes”, “when”, “how” ...)
- **Tiempo:** incluye términos que expresan temporalidad de diferentes formas: años concretos (1974, 1408...) períodos de tiempo (“año”, “agosto”, “siglo” ...) y/o momentos más específicos (“hoy”, “alba”, “navideña” ...)
- **Tipo narrativo:** palabras que ayudaran a concretar y/o desambiguar tipologías de escritura (aforismos, anime, moralejas, narracion, verso...) pero también escuelas (“vanguardistas”, “alejandrinos”, “costumbrista” ...)
- **Título:** palabras usadas en títulos de obras (“araucana”, “aristogatos”, “principito”, “wuthering”, “visitadoras” ...)
- **Topic - Ficción:** palabras que hicieran referencia directa a la tipología de obras de ficción (fantastico, sobrenaturales...) o que se dedujeran que eran utilizadas en obras de ficción (“hiperboreos”, “hobbit”, “minotauro”, “quidditch”, “quijote” ...)
- **Topic - NO ficción:** palabras que hicieran referencia tanto directa a la tipología de obras de no ficción (“cientifico”, “termodinamica”) como a otros conceptos de los que se dedujera un ámbito de no ficción (“alcoholismo”, “asociacion”, “bibliografica” ...)
- **Transaccional:** palabras que harían referencia a la definición clásica de búsquedas transaccionales, es decir, aquellas de las que se deduce una intención por parte del usuario que ejecuta la búsqueda de iniciar una transacción o acción posterior al resultado de su búsqueda. En general se asocian a verbos (“bajar”, “descargar”, “comprar”, “leer”) pero también a sitios que implican una acción (“rapidshare”, o “megaupload”, que son sitios web de descargas de archivos) u otras palabras de las que se deduzca similar actitud (“juegos”, “torrents”...)
- **Videojuego:** son palabras que indican que la frase de búsqueda hace referencia a videojuegos (“naku”, “mugen”), plataformas de videojuegos (ps1, ps2, wii) o al formato en sí (“videojuego”)

Para explicar esta facetación, y siguiendo un ejemplo anterior, la expresión de búsqueda “biblioteca miguel de cervantes burgos” podría ser clasificada bajo las facetas “autor”,

“nombre personal”, “entidad”, “biblioteca” y “lugar”; y así sucesivamente, con los otros miles de expresiones de búsqueda.

Pero encontramos expresiones de búsqueda que pueden ser clasificadas en torno a dimensiones de análisis distintas de las facetas anteriormente identificadas. Por ejemplo, la frase “cervantes saavedra obras” puede ser clasificada no sólo dentro de la faceta “autor”, sino también dentro del aspecto lingüístico “plural”, “el viejo celoso cervantes” puede ser incluida en la clase “género”, otro aspecto lingüístico detectado. En otras frases, podemos ver aspectos modificadores, como en “la biblioteca virtual miguel de cervantes” donde la palabra “virtual” nos indica que puede clasificarse por el aspecto modificador “formato”, o frases como “cervantes 2009” o “librería cervantes oviedo horario”, que denotan un empleo de otro aspecto modificador, “tiempo”.

Estas clases o facetas se agruparon en torno a superjerarquías. Para constituir estas jerarquías que agruparan las anteriores clases, por un lado, se buscó seguir la clasificación canónica de expresiones de búsqueda iniciada por autores como Broder(2002) o Jansen et. al. (2008), a su vez basados en investigaciones anteriores como las de Byrne et. al. (1999) sobre clasificación de tareas de interacción o “*taskonomy*”, y Navarro-Prieto et. al. (1999) sobre estrategias cognitivas de búsqueda. Por otro lado, se tuvo en cuenta a autores citados por Broder, como Shneiderman et. al. (1997) y otros, como más tarde Jethava et. al. (2011), Fu (2014) o Figueroa (2015), quienes buscaron aspectos distintos presentes en las frases de búsqueda, tales como acrónimos, abreviaturas, u otros aspectos contextuales. Así, basándose en las facetas localizadas en el ámbito del conjunto de expresiones detectadas, la clasificación quedó establecida de la siguiente forma:

- NAVEGACIONALES
 - Biblioteca
 - Entidad
 - Lugar
- TRANSACCIONALES
 - Transaccionales
- INFORMACIONALES
 - Ámbito
 - Autor
 - Ciencia
 - Deporte
 - Desambiguación
 - Erótico
 - Film

- Infantil
- Nombre personal
- Personaje
- Pregunta
- Título
- Topic ficción
- Topic no ficción
- Videojuego
- ASPECTOS MODIFICADORES
 - Formato
 - Tiempo
 - Tipo narrativo
- ASPECTOS LINGÜÍSTICOS
 - Abreviatura
 - Error
 - Género
 - Idioma
 - Plural

1.5.5. Extracción de datos históricos de Google Trends

Una vez localizado y clasificado el conjunto de expresiones de búsqueda a estudiar, se procedió a obtener información estadística de la evolución histórica de la frecuencia de búsquedas en España de cada expresión. Para ello había dos opciones:

- A. Utilizar *Google Keyword Planner*, que aporta datos de volumen de búsquedas, anteriormente del último año anterior al momento de la petición de datos al sistema, en la actualidad de dos años.
- B. Utilizar *Google Trends*, que permite registrar datos de la evolución, desde 2004 hasta el momento de realizar la petición de datos.

Cabe plantearse entonces cuál de estas herramientas es la más adecuada para el análisis, o si el uso combinado de ambas produce mejores resultados. *Google Keyword Planner* ofrece datos numéricos de frecuencia de búsqueda, aunque según los expertos en SEO, de forma aproximada. *Google Trends* aporta información en forma de datos normalizados. Su forma de representar la frecuencia de búsqueda hace que los datos se escalen en un intervalo del 0 al 100, siendo 100 el momento de mayor volumen de búsquedas para esa expresión de búsqueda en todo el intervalo pedido al sistema (que puede ser de unos meses o en el presente caso, de toda la serie histórica)

En un primer momento se contempló la posibilidad de integrar ambas fuentes de datos, de forma que los datos de frecuencia se pudieran extrapolar a toda la serie, y así obtener volúmenes de frecuencia reales para todos los meses de la serie. La idea era que, si se tenía los valores de volumen de búsqueda para un mes (por ejemplo 1.000 búsquedas), sacados de *Google Keyword Planner*, y se tenía la posición en el intervalo (posición 100 por ejemplo) con ambos datos y los equivalentes de otros 11 meses del año en el que se hizo la toma de datos de *Keyword Planner*, se podrían extrapolar todos los datos. Se realizaron pruebas empíricas para utilizar esta técnica, que mostraron que las tendencias históricas de *Google Trends* y *Google Keyword Planner*, siguen tendencias que pueden resultar distintas, lo que fue confirmado por otros estudios (Palii, 2017; Churick, 2018). La diferencia estriba en el tratamiento de los datos:

- *Google Keyword Planner* está pensado para localizar expresiones de búsqueda con distintos grados de exactitud, para servir de base a campañas de publicidad en buscadores, por lo que los datos tienen en cuenta cuánto quieren invertir las empresas para pujar por posicionarse para la misma en los resultados de búsqueda (Palii, 2017).
- *Google Trends* elimina búsquedas duplicadas de los mismos usuarios (Churick, 2018), utilizando muestras aleatorias, para fijarse en la tendencia global de la misma, ofreciendo datos sólo de expresiones de búsqueda que tiene una frecuencia de búsqueda mínima (donde el límite mínimo o *threshold* para obtener datos de *Google Trends* se desconoce). Según su sección de ayuda, ofrece datos que “son una muestra imparcial de los datos de la Búsqueda de Google. Para recopilar los datos de *Trends* solo se utiliza un porcentaje de las búsquedas”; es más, los “datos que no son en tiempo real es una muestra aleatoria de los datos de la Búsqueda de Google que pueden obtenerse desde el año 2004 hasta 36 horas antes de la búsqueda” (*Google Support*, 2018).

Por tanto, y al no poder usarse con seguridad ambas fuentes para obtener un conjunto histórico estable y fiable, la pregunta pasó a ser cuál de ambas fuentes usar. Para obtener toda una serie histórica de *Google Keyword Planner*, al no tener acceso a una cuenta de *Google AdwordsAPI*, habría que plantear manualmente cada una de las más de treinta mil frases de búsqueda cada dos años en el sistema, para así descargar en formato .csv los datos de frecuencia, mientras que con *Google Trends*, sólo habría que hacer una vez la toma de datos.

Siendo más interesante a efectos del estudio, visualizar la evolución y la tendencia, y teniendo una importancia secundaria el volumen puro, la frecuencia de búsqueda, se optó por usar *Google KeywordPlanner* para extraer las expresiones de búsqueda, pero contar con *Google Trends* para capturar la información histórica, sin tener en cuenta los datos de meses anteriores que aportaba *Google KeywordPlanner*.

De esta forma se podía construir toda la serie histórica de cada expresión de búsqueda, debido a que la serie temporal que ofrecía *Google Trends* era mayor (desde 2004 hasta la actualidad, frente a sólo 12 meses anteriores a la fecha de la consulta). Nótese que en 2018 *Google*

KeywordPlanner comenzó a ofrecer hasta dos años de datos, pero aun así seguía siendo una serie menor que la que ofrece *Google Trends*.

Así pues, para la obtención de los datos históricos, se procedió a insertar en *Google Trends* (2017) cada palabra clave previamente obtenida en *Google KeywordPlanner*, eligiendo de entre las posibilidades de parametrización que ofrece la herramienta las siguientes opciones:

- **Segmentación por “término de búsqueda”** (para evitar que Google interprete la frase de búsqueda de forma semántica, como un concepto y no como una cadena de caracteres, que es lo que se precisa para eliminar sesgos en lo posible)
- **Segmentación geográfica**, restringiendo la búsqueda de datos para “España”
- **Segmentación temporal**, eligiendo la opción “2004 – hoy” para obtener toda la serie histórica, como se ve en el ejemplo con la expresión “poemas de amor”:

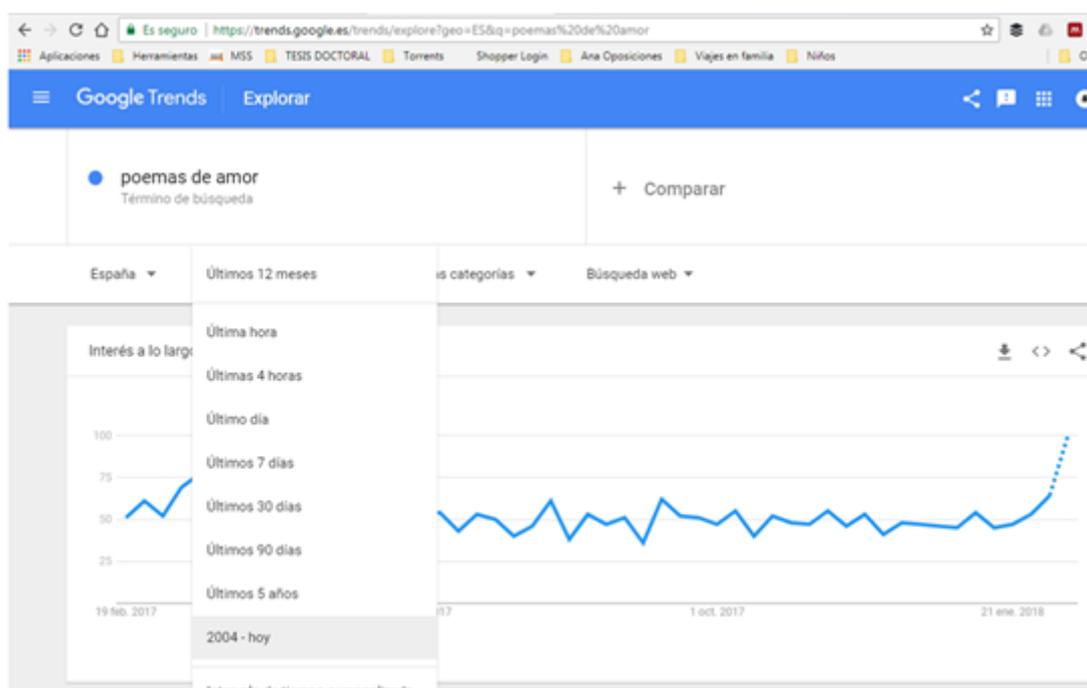


Fig. 1-9 Ejemplo de gráfico de tendencias de búsqueda en Google Trends para "poemas de amor". Fuente: Google Trends.

A continuación, se procedió a clicar en la opción de descarga de los datos de esa expresión de búsqueda en formato .csv, como se aprecia en la siguiente imagen (arriba a la derecha)

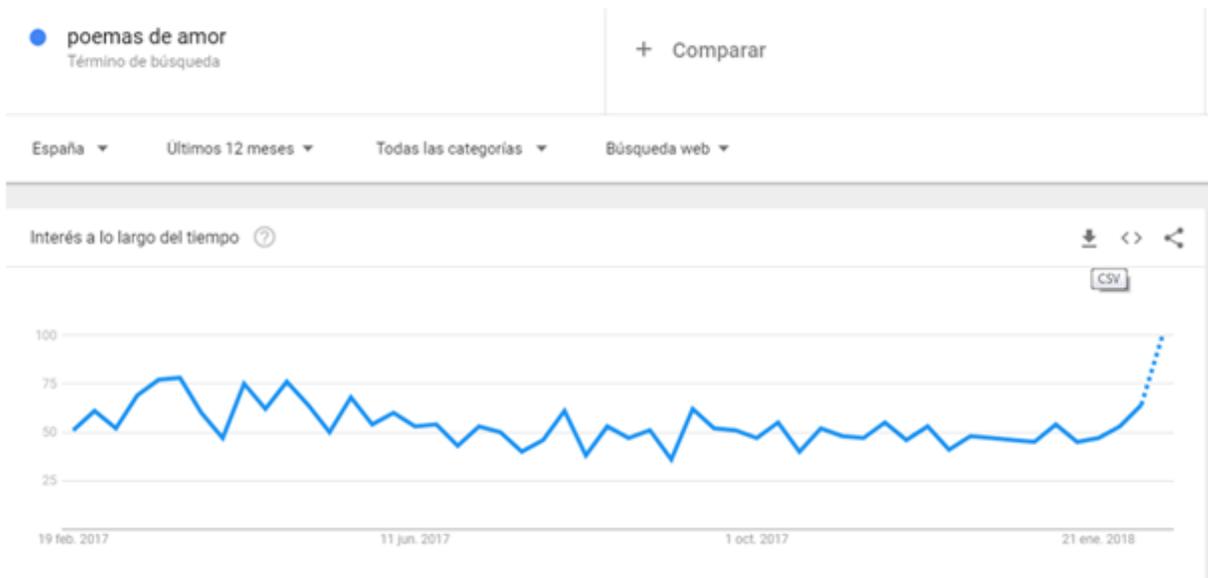


Fig. 1-10 Botón de descarga de datos de Google Trends en formato .csv. Fuente: Google Trends.

Finalmente, se importaron los datos desde csv, usando como separador la coma, de forma que se tuviera para cada mes de cada año de la serie, un valor, de 0 a 100.

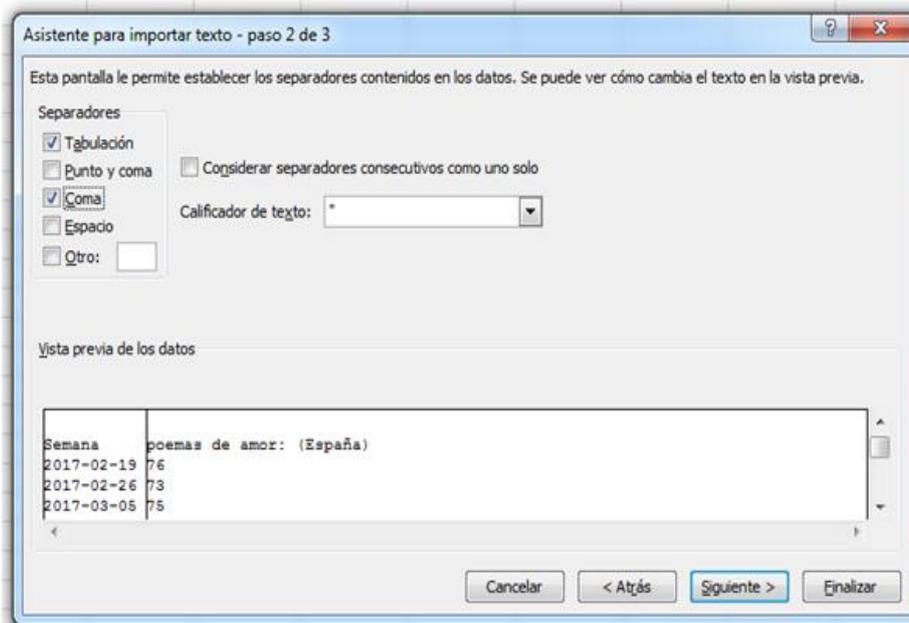


Fig. 1-11 Proceso de importación de datos de Google Trends en formato .csv. Elaboración propia.

1.5.6. Procesamiento e integración de los datos

Una vez clasificadas las expresiones de búsqueda y obtenidos los datos de *Google Trends*, se procedió a integrar los datos en una base de datos relacional, debido a que no era un conjunto de datos excesivamente grande como para necesitar procesamiento en tiempo real, y a que su estructura lógica permitía ajustar adecuadamente tanto los datos obtenidos como las tablas auxiliares para la creación de clases, al universo de discurso a representar y a los cálculos y extracciones de conjuntos de datos a realizar (Gajani, 2017).

Al constituirse con datos de dos fuentes distintas, se aplicó un identificador único (ID) para cada expresión de búsqueda, de forma que se pudieran centralizar en una única fuente combinada los datos, integrados así para que tuvieran una relación unívoca con cada expresión, eliminando la posibilidad de duplicados. Se creó una tabla de clases, de forma que las expresiones de búsqueda pudieran pertenecer a una o más de esas clases. Se ejecutó un script de búsqueda de coincidencias, de forma que para cada expresión o frase búsqueda, se localizara qué palabras de las que contenía la frase pertenecían a alguna clase, y se consignara a cuáles.

De ese modo, una expresión de búsqueda podía pertenecer a varias clases, en función de la pertenencia de las palabras de las que constaba la misma. Por ejemplo, “comprar novelas” pertenece tanto a la clase “tipo narrativo”, como a la clase “transaccional”, puesto que “novelas” se ha clasificado como una palabra que indica un “tipo narrativo”, y “comprar” como perteneciente a una búsqueda “transaccional”.

Además, se integraron aquí los datos obtenidos de *Google Trends* para cada expresión de búsqueda, de forma que de cada una se tuviera por un lado las clases a las que pertenecía, y de otro, la evolución histórica de frecuencia de búsqueda de los últimos años. Con la base de datos ya consolidada, se procedió a realizar el conjunto de análisis estadísticos.

Los datos de las 30.925 expresiones de búsqueda fueron recogidos e integrados en MySQL, un Sistema Gestor de Base de Datos que permite la inclusión de nuevas dimensiones de datos aplicadas en torno a las frases de búsqueda o *queries*. Así, se generaron distintas tablas.

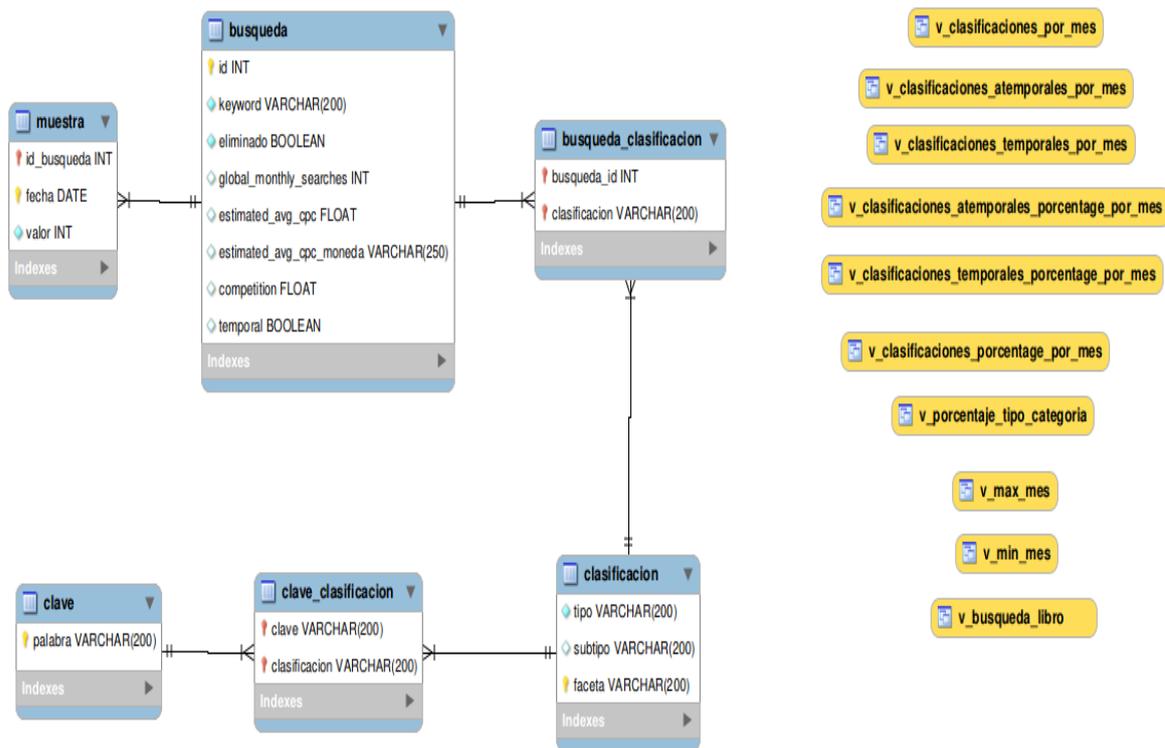


Fig. 1-12 Estructura de tablas de la base de datos de análisis de queries. Elaboración propia.

La composición de las tablas se estableció como sigue:

- **Búsqueda:** representa cada una de las frases que se utilizan para buscar. Se guardó la frase de búsqueda en un campo 'keyword', junto con otros datos que *Google KeywordPlanner* aporta sobre cada expresión que no se han utilizado en este estudio, como el nivel de competición o el CPC (*Cost per Click*) estimado.
- **Muestra:** contiene los datos recogidos de *Google Trends* para un subconjunto de las frases de búsqueda, puesto que *Google Trends* tiene un *threshold* o límite mínimo de frecuencias de búsqueda más alto que en el caso de *Google Keyword Planner*, y por tanto, no han aparecido datos en *Google Trends* para todas las expresiones de búsqueda que aparecen sugeridas en *Google Keyword Planner*. Para cada frase, se presenta la frecuencia relativa de búsqueda, con valores de 0 a 100, donde 100 es el mes en el que se ha alcanzado la mayor frecuencia de búsqueda para esa frase. Se han recogido los datos desde los años 2004 a 2016, ambos inclusive.
- **Clave:** Representa cada una de las palabras que forman las frases de búsqueda, las cuales se puede relacionar con una o varias clasificaciones.

- **Clasificación:** la estructura de clasificación establecida se subdivide en:
 - Tipo: Informativa, Navegacional, Transaccional, Aspectos Lingüísticos y Aspectos Modificadores
 - Subtipo: Desambiguación, Lenguaje Natural, Temática, Persona
 - Faceta: fue establecido un conjunto de facetas, agrupadas en torno a las clases de la clasificación clásica de intenciones de búsqueda (navegacional, informativa y transaccional) y a los aspectos, divididos a su vez en aspectos modificadores y aspectos lingüísticos.

- **Clave_clasificacion:** Relaciona cada una de las palabras con una o varias clasificaciones.

- **Busqueda_clasificacion:** La clasificación de una frase de búsqueda, que vendrá dada por el conjunto de facetas en las que se enmarcan las distintas palabras (claves) usadas en la frase de búsqueda.

1.5.7. Clasificación entre expresiones de búsqueda temporales y no temporales

Con los datos de *Google Trends* integrados en la base de datos, se buscó clasificar las expresiones de búsqueda atendiendo a un criterio totalmente distinto de los empleados anteriormente, la distribución de la frecuencia de búsqueda de las distintas frases de búsqueda a lo largo de la serie histórica, pero que pudiera ser escalable y automatizable (Hagen et. al., 2011). Se buscó determinar qué expresiones de búsqueda parecían atender a momentos concretos (que se han denominado búsquedas “temporales”), probablemente determinados por apariciones de nuevos títulos o eventos multitudinarios, y cuáles parecían tener una vida más estable a lo largo de la serie (denominadas aquí “atemporales”).

Esta dimensión de clasificación se combinaría con la clasificación por facetas, para conocer la evolución particular de cada agrupación de expresiones de búsqueda, teniendo en cuenta que, como hemos visto por la posible facetación de las mismas, una misma expresión de búsqueda puede aparecer clasificada por varias facetas y aspectos.

Para esta parte de la investigación no se pudo contar con toda la muestra de expresiones de búsqueda utilizada en la clasificación manual en torno a facetas y aspectos, derivada del análisis del lenguaje utilizado, debido a que *Google Trends* tiene un *threshold* o límite mínimo de frecuencia de búsqueda, a partir del cual no muestra datos. Si *Google Trends* no ofrecía datos, la frase no podía ser clasificada mediante esta técnica, lo que supuso eliminar de este estudio en particular a 23.675 de las frases de búsqueda inicialmente localizadas (30.925).

Para determinar si una frase de búsqueda era “temporal” o “atemporal” se generó un programa que fue contando en cada serie temporal de cada expresión de búsqueda, el número total de veces que aparece 0 (cero) como frecuencia de búsqueda en el total de celdas de los datos (el total de meses), y el número de valores distintos de 0. También se contó el número de segmentos (valores seguidos) de ceros, y de segmentos con valores distintos de 0.

El tamaño de cada segmento de ceros, y por consiguiente el número de ceros, se ve reducido si el segmento de ceros es menor al segmento de valores distintos de 0 anterior. De modo que el número de ceros del segmento se puede definir por:

$$\min(\text{tamaño_segmento}, \text{tamaño_segmento} * (\text{tamaño_segmento} / \text{tamaño_segmento_anterior}))$$

La penalización sólo afecta a los ceros, tanto al número de ceros como al número de segmentos de cero, dando mayor importancia a los valores mayores que 0. Por tanto, el programa decide si una frase de búsqueda es “temporal” si:

$$\text{nceros} * (\text{nsceros} / (\text{nsceros} + \text{ndsceros})) > \text{ndsceros} * (\text{ndsceros} / (\text{nsceros} + \text{ndsceros}))$$

Siendo:

nceros = Número de ceros

nsceros = Número de segmentos de ceros

ndceros = Número de valores distintos de cero.

ndsceros = Número de segmentos distintos de cero.

Las expresiones de búsqueda que no cumplieron esta condición, se clasificaron como “atemporales”. Así, del conjunto de expresiones de búsqueda de las que se pudo obtener datos en *Google Trends* (7.250), se clasificaron 6.280 frases de búsqueda como “atemporales”, lo que supone el 86,6% de la muestra de frases de búsqueda de las que se dispone de datos longitudinales, siendo consideradas “temporales”, el 13,4% de las frases (970).

Una vez realizada la clasificación cualitativa de las expresiones en torno a facetas e intenciones de búsqueda, y realizada la clasificación cuantitativa en torno a su línea temporal, el conjunto de expresiones de búsqueda ya quedó preparado para ser analizado a través de dos grandes dimensiones: por un lado, el estudio del uso del lenguaje explicitado en las facetas detectadas, y por otro lado, las semejanzas y o diferencias que arrojaran las expresiones clasificadas como temporales y atemporales. Siendo ésta una investigación exploratoria, se sondearon tanto posibles relaciones causales como casos que ejemplificaran o explicaran ciertas conductas, a través de un muestreo intencional, buscando descubrir nuevas variables de estudio que pudieran ser tenidas en cuenta.

2. ESTADO DE LA CUESTIÓN

2.1. Introducción al análisis de los hábitos lectores

Se ha dividido el estudio del marco teórico en dos grandes frentes: el primer bloque trata del análisis de los hábitos lectores en el sector del libro. En el segundo bloque, buscando salir de la “zona de confort” de los estudios clásicos de comportamiento lector, se dedica a conocer las investigaciones que hay sobre las conductas de búsqueda de los usuarios, para dilucidar cómo podían aplicarse al estudio de los hábitos de conducta de búsqueda en relación a la lectura, libros y otros formatos de productos culturales susceptibles de ser encontrados en bibliotecas.

Se tratará en este estudio no sólo del libro como objeto de la relación entre usuario y contenido usado sino también de otros formatos de contenido, como el comic, los videojuegos y películas. Esta cobertura de otros productos en el estudio se debe a que en las bibliotecas se han integrado desde tiempo atrás otros contenidos y productos culturales (comics, videojuegos, películas...).

Además, la estrategia *transmedia* de diversos tipos de empresa (editoriales, jugueteras, productoras cinematográficas en particular) ha hecho que los títulos de libros de éxito también se lleven a formatos distintos, como el gráfico o el cinematográfico, por lo que en algunos casos será preciso conectar los datos relacionados con la obra en sus distintos formatos, para entender mejor las relaciones causales de su éxito o fracaso. Por otro lado, la misma definición de libro se ha visto afectada por la aparición de los nuevos soportes y formatos de lectura, e incluso la comisión de cultura del Congreso de los Diputados, al aprobar la nueva *Ley de Depósito Legal* (que sustituía a la de 1971) el 21 de julio de 2011, adaptó el depósito del patrimonio bibliográfico, ante la aparición de los nuevos modos de publicación digital (Cordón-García, 2011).

En este sentido, la Real Academia de la Lengua atestigua esta transformación en cierto modo, pues admite entre sus acepciones que libro es “obra científica, literaria o de cualquier otra índole con extensión suficiente para formar volumen, que puede aparecer impresa o en otro soporte”, donde lo novedoso es “o en otro soporte” (Real Academia de la Lengua, 2017), contando también con la acepción de libro electrónico, del que dice que es tanto un “dispositivo electrónico que permite almacenar, reproducir y leer libros”, como un “libro en formato adecuado para leerse en un dispositivo electrónico”.

Como se verá, en los estudios será interesante diferenciar el libro en papel del que no lo es, para sopesar las diferencias y la evolución de esos hábitos de lectura. Pero incluso estableciendo esa diferencia en el propio informe, encontraremos estudios y estadísticas tan tardíos como la “Encuesta de hábitos y prácticas culturales 2014-2015” (Ministerio de Educación, Cultura y Deporte, 2015), en las que no es fácil saber si el encuestado es preguntado por la acepción de libro como volumen en papel, o el libro como concepto que englobe tanto al anterior, como a otros formatos de disfrute lector, como el libro electrónico.

Pero antes de profundizar en lo que investigaciones anteriores han averiguado sobre el comportamiento de los usuarios y lectores, resulta obligado dimensionar la necesidad que el sector del libro tiene de comprender mejor las conductas de búsqueda de información al respecto, a la luz de los últimos avances tecnológicos y cambios sociológicos que han traído. Para ello, hemos explorado las necesidades de dos grandes actores del sector del libro, el mundo editorial y el bibliotecario, cuyos intereses, aunque distintos, no son divergentes, y tienen sin duda el mismo objeto de investigación: el lector, o el usuario, según a quién se pregunte.

2.1.1. Necesidades en el sector editorial

España es la cuarta potencia editorial del mundo, con 34 empresas que en 2014 generaban dos tercios de la facturación editorial española (el 65,7%) quedando el otro tercio a dividir entre otras 855 editoriales (Cordón-García et. al. 2014). En los años 90 se editaban anualmente en España más de 50.000 nuevos libros, la mayoría compuesta de tiradas reducidas, siendo pocas las que superan los 5.000 ejemplares (Perez, 1994). Según la Federación de Gremios de Editores de España, el fondo editorial vivo español se situaba en 157.000 títulos en 1991 (entendiendo “vivo” como el formado por libros con capacidad de rotación comercial).

Esas prácticas de publicación han seguido similares tendencias en los últimos años, como “un progresivo acortamiento del ciclo de vida de los libros en el mercado; existencia de más títulos, pero también de menos copias (...) bajada de la tirada media; descenso de la edición impresa como consecuencia de la creciente apuesta por la edición digital; reducción del número de ejemplares vendidos” (Magadán-Díaz y Rivas-García, 2018). En concreto, y mencionando a los mismos autores, con datos de MECD (2017), “la tirada media se está reduciendo: en 2005, se situaba en los 4.619 ejemplares y en 2016 en 2.749, lo que ha supuesto una disminución de la tirada media en un 40,48% para el período considerado”.

Frente a los datos de publicación, veamos los datos de los lectores potenciales. Mientras que según la UNESCO, la población analfabeta era en 2006 sólo del 4,2% en España (Cordón-García, 2006), quedando en 2011 730.000 personas analfabetas en España, el 67%, mujeres, según el INE (Lucas 2014), más del 40% de la población mayor de 18 años se confesaba abiertamente como “no lectora”, o lo que es igual, que jamás coge un libro entre las manos (Cordón-García, 2006), aunque estos datos deben ser completados. En 2015 Fernández-Blanco, Prieto-Rodríguez y Suárez-Pandiello, basándose en datos de la *Encuesta de hábitos y prácticas culturales 2010-2011* (Ministerio de Educación, Cultura y Deporte, 2011) calculaban que entre los ciudadanos españoles que no contaban ni con educación primaria, el 73.55% de los hombres y el 61.1% de las mujeres podían ser considerados como no lectores. Estos porcentajes disminuían conforme aumentaba el nivel educativo, por lo que sólo el 4.12% de los hombres y el 1.87% de las mujeres con un grado universitario podían ser considerados no lectores.

Hoy día, “La gravísima situación del libro y la lectura en España hace urgente y necesaria una gran movilización social, liderada por los poderes públicos” según la Asociación de las Cámaras

del Libro de España (Manrique Sabogal, 2015). Para atajar este problema se crean iniciativas como el *Plan Integral para el Fomento del Libro y de la Lectura* (Asociación de las Cámaras del Libro de España, 2015).

Además, según las conclusiones de Millán, coordinador de los informes de la Lectura en España 2002, 2008 y 2017, auspiciados por la Fundación Germán Sánchez Ruipérez y la Federación de Gremios de Editores de España, "hace falta un plan de conocimiento de cómo está la lectura. Es decir, descubrir quiénes son los que leen más en detalle y qué leen. A las encuestas la gente responde con una pobreza extraordinaria. Se sabe por ejemplo que ha aumentado muchísimo la lectura digital, pero no hay datos reales", remata. (Koch, 2017)

El mundo editorial tiene una necesidad apremiante de entender al lector, bien para darle lo que necesita, bien para provocar en el lector la necesidad de leer lo que escritor, librero y editor busca vender. Los editores y libreros no tienen problemas al valorar el reto que supone comprender al lector, pues en palabras de Gil y Jiménez (2008, 167), "en el sector del libro español hace ya tiempo que asistimos a una situación de crisis, en el que no sólo nos cuesta imaginar y proponer posibles medidas de solución, sino que el mero hecho de plantearnos abordar las dificultades ya provoca graves anomalías". Cabría preguntarse por la independencia de estos estudios de los intereses económicos de las partes, aunque se observan tendencias similares en los datos de editoriales, asociaciones de libreros, y ministerio. Faltaría integrar estos distintos estudios, junto con otros independientes, como el presente, y/o incorporando a las editoriales académicas, de las que sólo hay información cualitativa de su prestigio y cuantitativa de su producción editorial (Giménez Toledo, 2017).

Entre los profesionales de la edición pueden existir diversos posibles enfoques en la reacción ante el conocimiento de los hábitos de lectura de los usuarios / lectores, a saber (entre otras):

- A. Variar las propuestas al ritmo de las modas cambiantes
- B. Cerrarse a cambiar nada y centrar el trabajo exclusivamente en el libro como materia de estudio
- C. Provocar los cambios y la demanda del usuario / lector
- D. Combinar A, B y C

El reto de entender al usuario lector, máxime en un mundo cada vez más dominado por el entorno del libro electrónico (García-Marco, 2008), y de aplicar medidas para mejorar la situación del sector del libro y por ende, el nivel cultural y científico de los usuarios / lectores, implica soluciones complejas (Larrañaga Rubio, 2008), que no dependen de un único actor, por

lo que incluso el Ministerio pregunta a las editoriales “¿hasta dónde llega la apuesta de estas editoriales hacia los libros en este formato (electrónico)? ¿Se ve esta apuesta reflejada en su facturación? ¿Cuáles son los rasgos más característicos de su oferta digital en la actualidad? ¿Qué razones impulsan su desarrollo? ¿Por qué canales están distribuyendo los libros en este formato? ¿Con qué modelos comerciales? ¿Qué medidas están adoptando para evitar la descarga no legal de contenidos? ¿Cuáles son sus expectativas respecto a este mercado?”. (Ministerio de Educación, Cultura y Deporte, 2016)

Pero la aplicación del conocimiento de los hábitos lectores, y por tanto de las estrategias y técnicas que se sucederán tras comprender mejor las variables motivacionales de los lectores / usuarios es extremadamente compleja, porque depende en gran medida de multitud de factores, entre otros:

Escalabilidad: quien está intentando entender al usuario lector puede ser:

- Un escritor, que puede querer aumentar su público lector, bien en torno a una lengua, a una cultura, una edad, un género, un país, o a varios.
- Un librero, que puede querer vender más ítems a su público, pudiendo ser éste local (su barrio) nacional o, en estos tiempos en los que Internet ha transformado la logística, internacional.
- Una editorial, que debe decidir su estrategia para aumentar las ventas de sus escritores y de los libreros con los que colabora, en todas las capas mencionadas.
- Una biblioteca, que puede servir a conjuntos de usuarios con segmentaciones diversas, a saber:
 - una biblioteca especial (como la de una cárcel) con un contexto específico, pero de temática variada
 - una biblioteca especializada, con una temática muy concreta
 - una biblioteca pública local, con un ámbito geográfico específico
 - una red de bibliotecas, que busca dar servicio a un espectro más amplio de usuarios
 - Etc, etc.

Los recursos: la Federación de Gremios de Editores de España (FGEE) considera que “el factor diferencial español tiene que ver con los recortes en el sector, con la “notable reducción de las consignaciones presupuestarias para el sistema bibliotecario, la retirada de las ayudas a las familias para la adquisición de libros y contenidos educativos y a la extrema situación de piratería que vive España” (La Vanguardia, 2015).

Los media: Internet, el libro electrónico, los videojuegos, el cine... La evolución audiovisual del contenido ha cambiado a la par que el continente (el dispositivo) y los nativos digitales

consumen la información de forma diferente que sus antepasados, puesto que “los contenidos se han desvinculado del soporte y, en consecuencia ... han aumentado las posibilidades de visibilidad y de difusión de la producción editorial, y han eliminado barreras geográficas y físicas” (Franganillo, 2008, 416) Es más, la producción editorial puede salir de las barreras de su medio original de creación hacia otros ámbitos, bien a través de creaciones *transmedia* (que utilizan diversos soportes digitales o no para mostrar partes de la trama) bien por su versionado en otros *media* como el de la televisión o el cine, lo que además puede producir un efecto de rebote por su mayor popularización, lo que se puede comprobar en casos como el del escritor James Patterson, el autor con más ingresos en todo el mundo (89 millones de dólares entre 2014 y 2015) siendo además el primer escritor del mundo en vender 1 millón de *e-books*, y sumando los ingresos obtenidos por derechos de autor procedentes de las obras adaptadas al cine o a la televisión (García Campos, 2015). Además, esa irrupción de nuevos formatos de consumo cultural está cambiando los hábitos de comprensión lectora, puesto que según investigadores como Liu (2005) en el contexto estadounidense o Mange et. al. (2013) en el noruego, detectan cómo los usuarios muestran distintos comportamientos e índices de comprensión lectora, al pasar del formato de libro clásico al libro digital.

El marketing: la creación de valor para el usuario lector es hoy día clave tanto para la supervivencia de editores, libreros, como de bibliotecas (no tanto para los escritores como tales, pues los hay, y muchos, que lo hacen literalmente por amor al arte, pues no obtienen suficientes ingresos como para vivir de ello) Siguiendo a Kotler (2012) hay 5 grandes temas en torno a “cómo crear valor para el usuario lector:

1. Creación de valor para los clientes a fin de capturar, a cambio, valor de ellos.
2. Creación y gestión de marcas fuertes que crean valor.
3. Gestión del rendimiento de marketing para recapturar valor.
4. Aprovechamiento de las nuevas tecnologías de marketing.
5. Marketing de una forma socialmente responsable.”

Los objetivos del marketing editorial y el marketing bibliotecario tienen al usuario / lector como punto de unión, por lo que en los últimos años ha aumentado el peso de los departamentos de marketing en los medios y grupos editoriales (Santiago Méndez, 2011). Esta persecución del valor en el marketing lleva a acciones de todo tipo (publicitarias, mediáticas...) y de todo ámbito (regional, generacional...) que pueden provocar la transformación de la percepción en el usuario lector de determinadas marcas, pudiendo tomar éstas la forma de un escritor, un personaje, una obra, una editorial, una empresa, una biblioteca...

Además, los movimientos promocionales de esas marcas llevarán a inundar de mensajes publicitarios todos los posibles canales de captación que puedan llegar al usuario lector, evolucionando desde los canales clásicos (como por ejemplo literalmente destacar un libro en una estantería de una librería o una biblioteca) a absolutamente todo lo que puedan captar nuestros sentidos, como publicidad interactiva, marketing en buscadores de Internet, cómics *transmedia*, vídeos virales, clubes de fans, y un largo etcétera, lo que hará destacar unas propuestas sobre otras, independientemente de su calidad e incluso de su idoneidad en términos de valor real para un determinado grupo de usuarios lectores, donde sólo importará su valor percibido.

Para analizar el comportamiento del consumidor de libros y sus hábitos de lectura y compra de libros a través del estudio del proceso de decisión y compra, las editoriales precisan de la combinación de la investigación cualitativa junto a la investigación cuantitativa, (Gómez-Tarragona, 2010) Este proceso, también denominado *buyer journey* o *customer journey* puede iniciarse por varios canales, en un entorno de consumo multimedia y multicanal (Dholakia et. al., 2010), donde destaca el canal del marketing en buscadores o *Search Engine Marketing* (SEM)

SEM es el conjunto de acciones de marketing que interactúa con el usuario cuando éste utiliza un motor de búsqueda de Internet para encontrar información pertinente, en lo que se denomina *search journey*. Este canal de marketing puede desglosarse en dos canales que no tienen por qué ser mutuamente excluyentes (Yang y Ghose, 2011) Por un lado está el canal de publicidad online, generado por anuncios publicitarios basados en conductas de búsqueda explicitadas en expresiones de búsqueda o *queries* (Mukherjee y Jansen, 2014) denominado *Pay Per Click* (PPC). Por otro lado, el que se genera de forma similar, al conseguir posicionar contenidos (sean textuales, de vídeo, imagen u otros *media*) entre los primeros resultados naturales u orgánicos de búsqueda ante una palabra clave buscada por el usuario (SEO o *Search Engine Optimization*) En este caso, este posicionamiento se produce sin pagar por ello al motor de búsqueda, como sí ocurre en el sistema de pujas del PPC, del cual tenemos un ejemplo en el servicio de publicidad online de Google Adwords.

Las editoriales deben estudiar al cliente potencial y sus hábitos, para delimitar la segmentación adecuada a sus intereses, entre ellos no sólo aumentar la compra de libros, sino la compra habitual y repetida de libros por parte de más lectores. Es decir, para rentabilizar una campaña de marketing, no sólo se debe aumentar el número de lectores, sino que éstos sean ávidos, y

repitan en sus compras de lectura, lo que hemos visto antes y se denomina *Customer Engagement* o “compromiso del cliente” (Bowden, 2009)

Éste es un concepto para el que aunque no hay un consenso claro en la definición y naturaleza del término “engagement” (Aguilera Moyano, 2016), sí se puede concluir, siguiendo a la literatura al respecto, que engloba y va más allá de la fidelización del usuario / cliente / lector en este caso, optimizando la participación y confianza de esos clientes comprometidos y leales, optimización basada en un marketing relacional que implique una experiencia de cliente interactiva que aporte valor mediante la co-creación entre marca (editorial / libro / biblioteca / autor) y lector, lo que empieza por identificar las necesidades de éste último (Brodie et. al., 2011).

Esta segmentación del usuario puede deberse a diversos criterios, según Fernández Marcial (2014):

1. “Criterios relativos al beneficio buscado por el consumidor.
2. Criterios relativos al comportamiento de consumo. Esto es, ¿qué cantidad consume y cuándo?
3. Criterios relativos a las características del consumidor. Aquí tienen cabida tanto las características sociodemográficas como las psicográficas (estilos de vida y personalidad). “

Después, las editoriales generan campañas de marketing que buscarán encontrarse con los lectores allá donde éstos acudan a informarse sobre qué libros comprar. Un ejemplo son las listas de *best sellers*, que son, según Fouces González (2004) “un pseudoconcepto descriptivo y prescriptivo a la vez. Es decir, describen una situación dada, en un momento dado: los libros que más se están vendiendo, pero al mismo tiempo, prescriben una situación, porque el hecho de incluir en la “lista de best seller” una novela que aún no ha alcanzado ese estatus, hace que ésta se convierta en un libro muy vendido”.

2.1.2. Necesidades en el sector de Bibliotecas

Si conocer los hábitos de los lectores ha sido una misión imperiosa para el mundo editorial, también lo es para el mundo bibliotecario desde sus inicios, máxime en un mundo conectado, dominado por la tecnología y en la que la lectura tradicional se mezcla con los formatos digitales de consumo, que como se verá más adelante, están ganando tracción.

En 2015, quedaban en España 4.610 bibliotecas públicas⁴ de diversa tipología y titularidad, un número que no para de descender, cerrándose más de 200 en el período 2010-2014 (Ministerio de Educación, Cultura y Deporte, 2014).

Número de bibliotecas públicas en España (2011-2015)

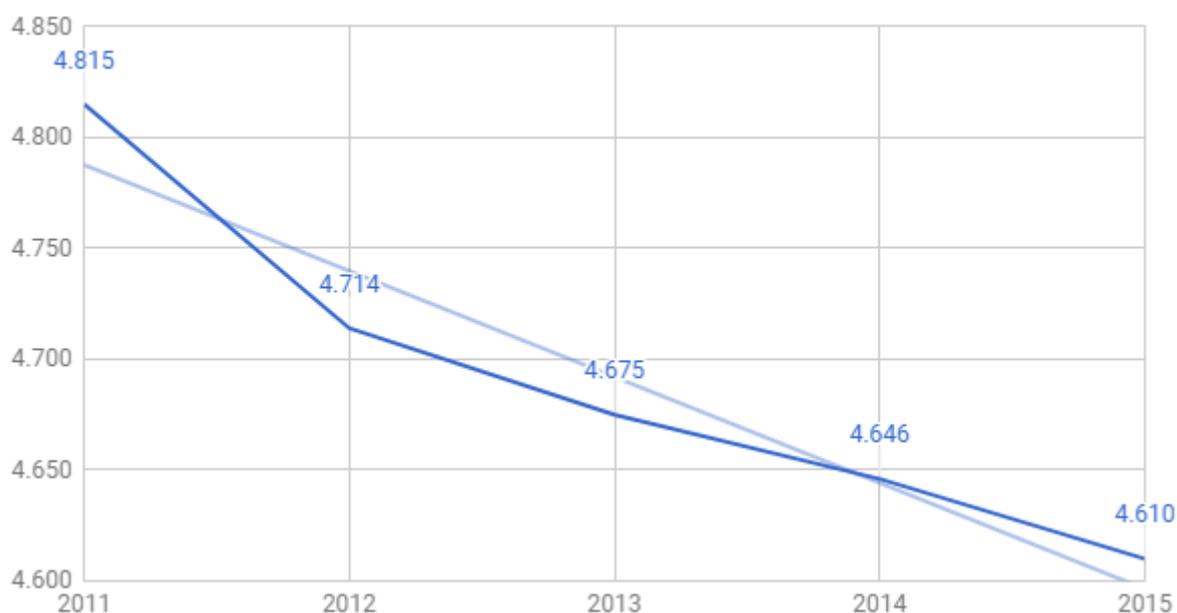


Fig. 2-1 Número de bibliotecas públicas en España (2011 - 2015). Fuente: Ministerio de Educación, Cultura y Deporte, 2018 . Elaboración propia.

Echando la mirada atrás, en 1991, “3,4 millones de personas confiesan acudir a leer a un centro público” (Perez, 1994, 146). Según la última Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015 (Subdirección General de Estadística y Estudios, Secretaría General Técnica, Ministerio de Educación, Cultura y Deporte, 2015) sólo el 25,6% de la población

4 Fuente: <http://www.cultura.gob.es/alziraweb/alziraweb.cmd?command=GetEvolucionIndicador&id=1>

española asiste a una biblioteca o accede a ella por Internet, y éstos lo hacen 10,4 veces cada uno al año.

En 2014 sólo visitaba una biblioteca un cuarto de la población, según el último Anuario de Estadísticas Culturales (Ministerio de Educación, Cultura y Deporte, 2016), mientras se aprecia que está bajando la media de horas abiertas al día. Además, se observa que las colecciones están envejeciendo debido, según Antonio María Ávila, director ejecutivo de FGEE, a la caída del gasto en adquisiciones: de 1,50 euros por habitante en 2009, a solo 0,56 céntimos en 2014 (Koch, 2017)

La tendencia se aprecia en otros indicadores. Por ejemplo, el promedio de préstamos por usuario en España (no se indica en la fuente si es sólo préstamos de ítems físicos o también digitales) sufrió una sensible caída en el período 2010-2013, con una mínima recuperación en 2014 y 2015, para volver a bajar en 2016, según los datos del Ministerio de Educación, Cultura y Deporte.

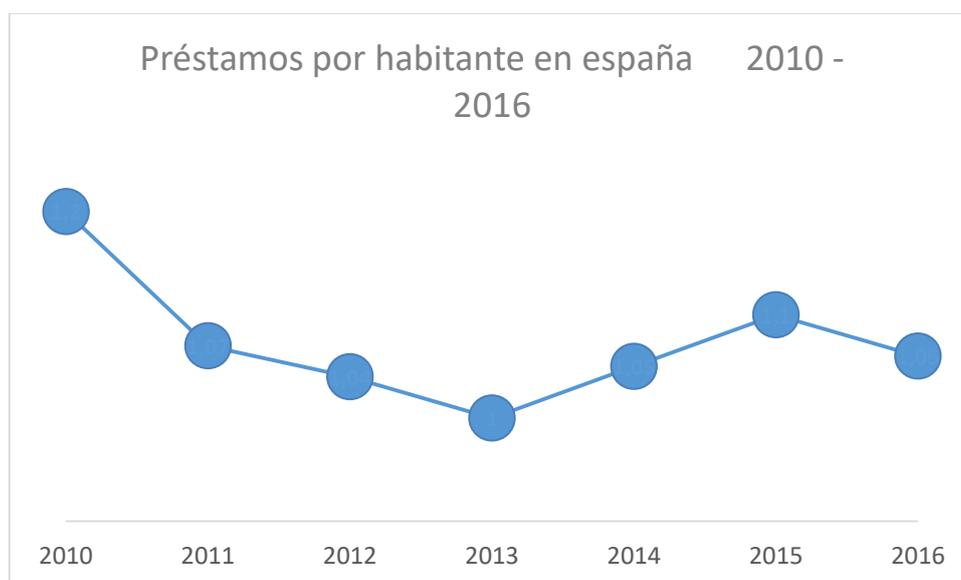


Fig. 2-2 Promedio de préstamos por usuario en toda España. Fuente: Ministerio de Educación, Cultura y Deporte, 2016 . Elaboración propia.

Si tomamos los datos por regiones, Andalucía, Aragón, o Castilla y León son de las que más varían en una tendencia que disminuye hasta 2014. Cataluña, siendo una de las regiones (junto con Castilla y León) que mejores datos tenían en 2010 (2,21% y 2,26% respectivamente)

desciende hasta el 0,07 durante 2012 y 2013, para volver a subir, aunque sin llegar a los valores pretéritos.

Tabla 2-1 Préstamos por usuario en porcentajes, en las distintas regiones de España (2010-2016). Fuente: Ministerio de Educación, Cultura y Deporte, 2016. Elaboración propia.

	2010	2011	2012	2013	2014	2015	2016
Andalucía	0,54	0,52	0,5	0,5	0,45	0,42	0,39
Aragón	1,79	1,73	1,58	1,48	1,35	1,29	1,32
Canarias	0,4	0,42	0,45	0,5	0,38	0,33	0,38
Cantabria	0,77	0,9	0,99	0,85	0,86	0,83	0,74
Castilla y León	2,26	2,22	2,08	2,06	2,02	1,92	1,85
Castilla-La Mancha	2,05	1,91	1,79	1,65	1,52	1,41	1,34
Cataluña	2,21	0,08	0,07	0,07	1,96	2,05	1,94
Ceuta	0,16	0,15	0,17	0	0,2	0,21	0,23
Comunidad de Madrid	1,29	1,3	1,26	1,2	1,09	1,08	1,03
Comunidad Foral de Navarra	1,89	1,82	1,79	1,68	1,49	1,44	1,35
Comunidad Valenciana	0,9	0,89	0,92	0,9	0,83	0,8	0,73
Extremadura	1,05	0,92	0,95	1,03	0,94	0,99	0,89
Galicia	0,78	0,82	0,88	0,9	0,86	0,84	0,8
Illes Balears	0,85	0,82	0,67	0,79	0,79	0,69	0,7
La Rioja	1,7	1,75	1,51	1,61	1,49	1,42	1,37

Melilla	0,13	0,12	0,12	0,11	0,12	0,09	0,05
País Vasco	1,5	1,49	1,48	1,41	1,33	1,44	1,41
Principado de Asturias	1,57	1,56	1,49	1,4	1,3	1,23	1,15
Región de Murcia	1,01	0,93	0,97	0,94	0,89	0,85	0,82
PROMEDIO	1,2	1,07	1,04	1	1,05	1,1	1,06

En términos globales, son 8.456.030 préstamos menos en toda España en 2015, con respecto a 2011 (Ministerio de Cultura y Deporte, 2016), pasando de los 60.176.521 préstamos de 2011 a 49.377.580 en 2016. Y el ritmo es vertiginoso. Cada año de esta serie, el número de préstamos disminuye en 2.330.000 aproximadamente, lo que nos lleva a visualizar un hipotético futuro en el que, de seguir esta tendencia, sin atender a otros factores y estirando al extremo, habrían desaparecido los préstamos de libros físicos (en papel) en España ¡en 2038!

Préstamos bibliotecarios en España (2011-2016)

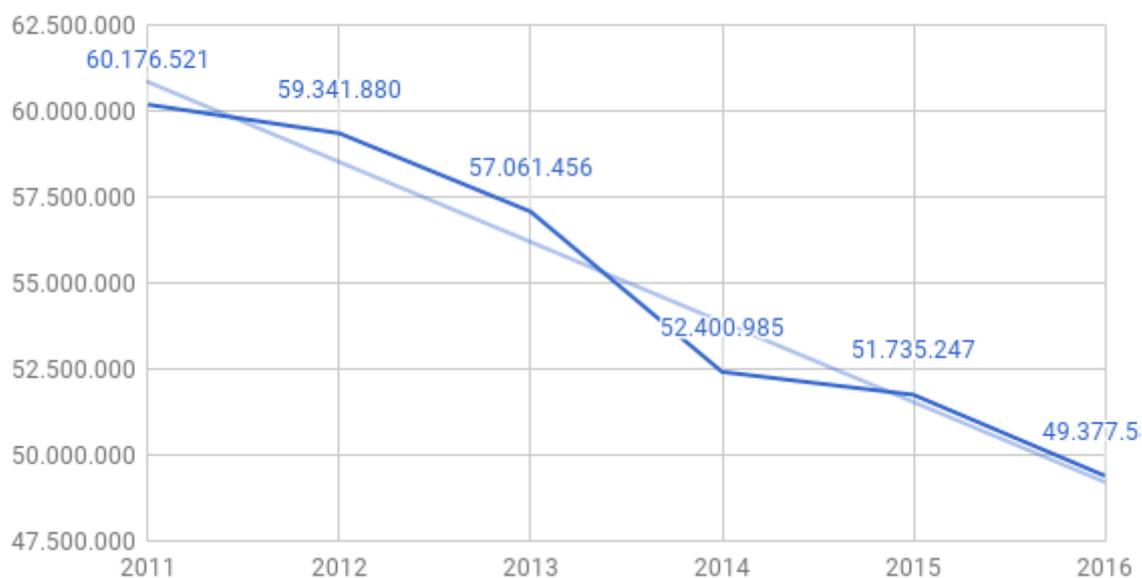


Fig. 2-3 Préstamos bibliotecarios en España (2011 - 2016). Fuente: Ministerio de Educación, Cultura y Deporte, 2016 . Elaboración propia.

Y esta tendencia no es privativa de España. Por ejemplo, según la encuesta elaborada por el Consejo Nacional para la Cultura y las Artes de México, en 2015 sólo el 15.9% acudía a las bibliotecas a conseguir libros (Quezada, 2016) Cabe preguntarse si eso quiere decir que se lee menos en el mundo, y particularmente en España, o si el rol de las bibliotecas en lo que a lectura de libros físicos va a cambiar, y si es así, en qué sentido.

A tenor de los datos que el Ministerio de Educación ha puesto a disposición del público por el momento, la información estadística sobre préstamos termina en 2016, por el momento. Mientras, los datos de préstamos de libros en formato electrónico (lo que engloba libros digitales y audiolibros) sólo están disponibles para 2015 (Ministerio de Educación, Cultura y Deporte, 2015), 2016 (Ministerio de Educación, Cultura y Deporte, 2016) y 2017 (Ministerio de Educación, Cultura y Deporte, 2017). Aun así, lo que parece apreciarse es que, mientras descienden los préstamos de libros “físicos” (lo que denominaríamos “en papel”) aumentan los préstamos de libros digitales y audiolibros en años posteriores.

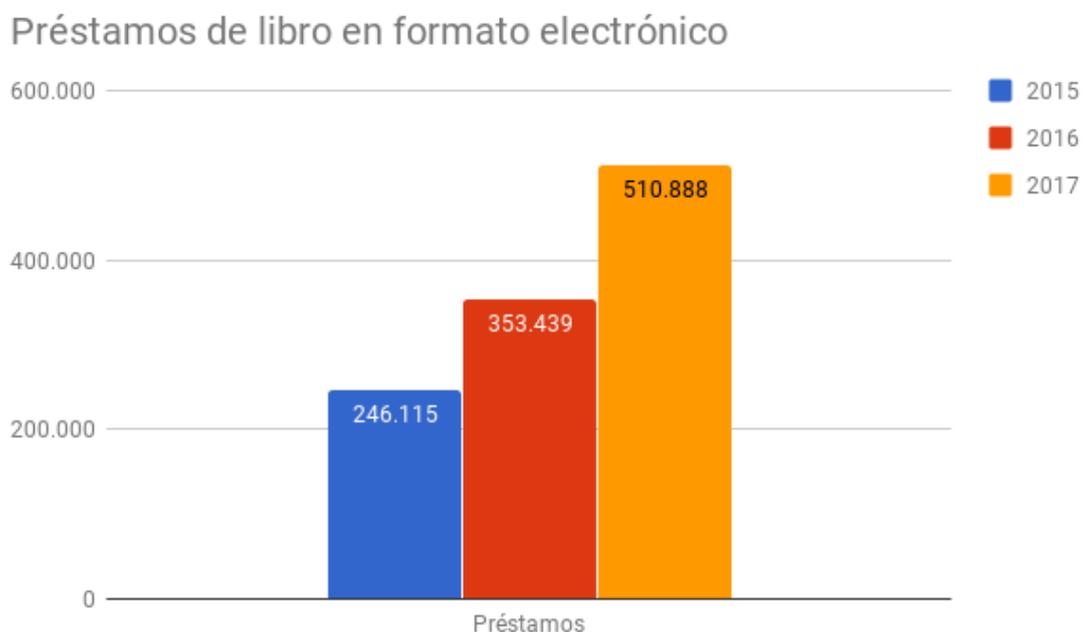


Fig. 2-4 Evolución de préstamos de libros digitales y audiolibros (2015-2017). Fuente: Ministerio de Educación, 2018 (elaboración propia)

La situación necesita de profundos análisis de los hábitos de lectura en nuestro país, para conocer de forma cuantitativa y cualitativa no sólo la evolución de esos hábitos, sino las motivaciones de los lectores y usuarios que hay detrás de esos hábitos. Es propuesta de esta

tesis que en un mundo tecnificado, las metodologías de investigación clásicas, como encuestas o entrevistas, puedan ser complementadas por nuevas técnicas de análisis, que ayuden no sólo a preguntar al usuario de forma consciente por sus intenciones de compra, adquisición y/o lectura, sino a descubrir otros aspectos de su conducta, de tal suerte que los sujetos de la investigación no sean conscientes de que están siendo investigados.

Distinto será cómo utilizar la información de inteligencia del mercado, o en este caso, del conocimiento del usuario / lector. La toma de decisiones dependerá del criterio del profesional, pues en opinión de Carrión para con las bibliotecas, “hay que buscar y prestar toda clase de colaboración, pero sin acciones de desembarco invasoras, sino con el ofrecimiento generoso y humilde de “aquí nos tienen para lo que gusten” y con la provisión de ideas y de incitaciones. Lo del hábito lector es otro cantar. Hay el peligro, si no se sabe bien qué es eso, de terminar en el hábito lector de “El código...”” (Jiménez Fernández, 2005, 32)

De ahí que el análisis del lector puede ser provechoso para las bibliotecas en muy distintas áreas. Por ejemplo, puede ser útil para ayudar en la clasificación e indización de los documentos y productos culturales que albergan, y así facilitar su posterior recuperación. A este respecto, diversas bibliotecas usan desde antiguo el *Manual de autoridades de la Biblioteca Nacional de España*, el cual en su introducción indica que “recoge la práctica que se ha venido siguiendo por esta institución en la normalización de sus puntos de acceso”.

A su vez, sigue, en lo que a políticas concretas y casos planteados, el *Manual de indización de encabezamientos de materia*, herramienta de apoyo al trabajo de los profesionales de la clasificación y la indización, a través del establecimiento de directrices que ayuden a realizar esas tareas de manera homogénea.

Este manual indica que, como herramienta que es, es en sí una obra abierta, por lo que puede seguir incorporando en el futuro “documentos e información de áreas temáticas en las que todavía no se han alcanzado acuerdos que permitan una uniformidad de criterios. Por ello, este manual será actualizado periódicamente, incorporándose nuevos contenidos y revisando los ya existentes “. (Biblioteca Nacional de España, 2013, p.7)

Por su parte, el *Manual de autoridades de la Biblioteca Nacional de España* indica las normas nacionales e internacionales y los criterios generales que usa para designar Registros de Autoridad derivados de la descripción y responsabilidad intelectual, y Registros de Autoridad de Materia. Respecto a los encabezamientos de materia, el Manual cita a la IFLA (International Federation of Library Associations and Institutions) como referencia máxima, de la que toma

los “Principios Subyacentes a los Lenguajes de Encabezamientos de Materia”, reflejados en el *Manual de autoridades de la Biblioteca Nacional de España* (Biblioteca Nacional de España, 2014, 5):

“Principios de construcción:

- Control terminológico (Principio de encabezamiento uniforme, Principio de sinonimia, Principio de homonimia, Principio de denominación).
- Orientación a través de la estructura paradigmática (Principio de semántica).
- Predictibilidad de representaciones (Principio de sintaxis, Principio de consistencia).
- Desarrollo dinámico y documentado (Principio de autoridad literaria o bibliográfica).
- Vocabulario orientado a la audiencia (Principio de autoridad del usuario).”

Nos quedamos con el último principio, el principio de construcción de Registros de Autoridad de Materia basado en el vocabulario orientado a la audiencia (Principio de autoridad del usuario). Este principio se invoca en más de una ocasión en el citado Manual, por ejemplo, cuando constata que “a menudo no es fácil encontrar un consenso entre los expertos sobre el término más apropiado para representar el o los conceptos. En general se debe preferir el uso más actual de un concepto.” (Biblioteca Nacional de España, 2014, 3)

Por tanto, si la normativa sobre principios de construcción de estos instrumentos de clasificación e indización está abierta a adiciones o cambios, y éstos pueden serlo (entre otros criterios) debidos a la construcción basada en vocabulario orientado a la audiencia, basándose en la autoridad del usuario, cabe pensar que sería muy útil para la continua construcción de autoridades, conocer el vocabulario utilizado por la audiencia, tanto cuando hace referencia a las autoridades conocidas como a las que haya por conocer.

También puede resultar útil en el ámbito del desarrollo de las colecciones bibliotecarias, uno de los grandes desafíos de cualquier institución bibliotecaria, así como en la continua evaluación de las mismas que, siguiendo a Lancaster, “debe ser dirigida principalmente a la identificación de puntos o limitaciones en un sistema existente y la formulación de mecanismos para incrementar su rendimiento” (Arriola y Butrón, 2008) siendo en general indistinguibles los objetivos de las tareas de desarrollo de una colección y su evaluación, pues la una retroalimenta a la otra (Rodríguez, Gómez-Díaz y Cordón-García, 2007).

Diversas son las metodologías posibles para desarrollar y evaluar una colección bibliotecaria, pero se detecta una evolución desde una metodología de observación directa de la colección, fijada en el ítem bibliotecario per se (Massísimo i Sánchez de Boado, 2002) para después

completar ese conocimiento del contexto de la colección con el estudio del uso de la colección (Massísimo i Sánchez de Boado, 2004).

Esas metodologías pueden variar dependiendo del tipo de biblioteca: en el caso de bibliotecas universitarias, Cuesta et. al (2014) investigan una muestra cualitativa para la toma de decisiones, o Aguado et. al., (2006) combinan indicadores cuantitativos y cualitativos. En otro tipo de bibliotecas como las públicas, priman los indicadores cuantitativos por ser prescriptivos, cuyos datos estadísticos son computados por el Instituto Nacional de Estadística (Barba Pérez et. al., 2006) y en algunas ocasiones sí se realizan análisis más cualitativos, como en el caso del estudio de las desideratas (Maroto-Galán, Niño-Herranz y Ruiz-Bautista, 2010).

La literatura científica abunda en estudios referidos a la medición de los servicios bibliotecarios o servicios de información, en especial a los universitarios, donde se evalúa la colección junto con otros servicios. En unos casos, se liga la evaluación bibliotecaria a la gestión de la calidad (Barba Pérez, 2009), utilizando las cartas de servicio como referente ante el cual contrastar el estado de la entidad.

Así, la evaluación de una biblioteca u otro servicio de información, ha visto aumentado el rango de indicadores y parámetros a medir y valorar en los últimos años. Por ejemplo, el modelo de evaluación de la Oficina de Referencia de la Biblioteca de la Universidad Pública de Navarra, define un conjunto de indicadores de rendimiento sobre aspectos tan variados como “la economía, penetración de mercado, eficiencia, eficacia, coste-eficacia, costes y productividad” (Puente Lanzarote, Campo Hilario y Ruiz de Luzuriaga Peña, 2001). En este sentido, tanto la penetración de mercado como la eficiencia y la eficacia, necesitan del análisis del usuario para medir eficazmente la mejora del servicio.

Otra de las áreas de mejora que el conocimiento del usuario ayudará a implementar, será la sede web de una biblioteca, que también puede ser utilizada para centralizar la promoción de la lectura y de los servicios de la entidad a través de Internet, por lo que también es susceptible de evaluación y mejora. Se identifican entonces las prácticas profesionales que sustentan esas iniciativas de promoción y marketing bibliotecario (Giraldo-Giraldo et. al., 2008; Marcos, 2012) o se listan las problemáticas que tienen las bibliotecas para darse a conocer ante el usuario potencial, como la “disminución de los presupuestos para la compra de bibliografía, el incremento del volumen de la información (...) el cambio cualitativo del perfil del usuario” (Gómez y Colombo, 1997) la percepción de los usuarios (Kronqvist-Berg, 2014) o el cambio en los procesos de interacción en el préstamo bibliotecario (Chiang-Martín, 2014).

Siendo como es la sede web una parte integrante de la presencia de la biblioteca en Internet, y por tanto, de la visibilidad de la entidad, es susceptible de ser partícipe del plan de marketing de la biblioteca. El marketing bibliotecario está adaptándose al marco de ese cambio de perfil del usuario, por lo que es imperativo entender mejor las dinámicas de uso de la información por parte de los usuarios, en tanto que ese uso puede serlo hoy día a través de la sede web de la institución, donde a la vez también está inserto el OPAC, el Catálogo en línea, a través del cual los usuarios pueden consultar la disponibilidad de los productos de información que buscan. Tanto es así, que las bibliotecas aprenden o deben aprender a diseñar un plan de marketing digital (González Fernández-Villavicencio, 2009) que puede ser útil para mejorar su propia imagen de marca o la de la institución a la que sirven, como en el caso de las universidades (González-Solar, 2016) o incluso influir en la posición de los rankings de las universidades, por lo que es necesaria una visión holística de esos servicios a la investigación, con la aplicación de técnicas de marketing específicas (Marcos Blázquez, 2012) para que se integren con otros aspectos de la experiencia del usuario en una biblioteca.

Ese plan de marketing digital de la biblioteca, que tiene un principio pero no un final, siguiendo un proceso iterativo de mejora constante, debe evaluarse de forma continua utilizando indicadores (González Fernández-Villavicencio, 2013) ligados a los canales en los que medir el ROI (Retorno de la Inversión o *Return of Investment*) de sus acciones, y que podrían verse optimizadas con un correcto y ético uso de herramientas de relación con los usuarios (CRM o *Customer Relationship Management*), para aumentar su fidelización y compromiso (*engagement*) con las marcas bibliotecarias, al conocer mejor sus necesidades, para así cumplimentarlas (Serrano-Cobos et. al. 2016).

Esas acciones tendrán una clara conexión con las expectativas de los usuarios cuando sean atraídos a la biblioteca, pudiendo ser múltiples y muy variadas. Por lo general, en la literatura al respecto se suele optar por acciones muy locales. Por ejemplo, Houghton-Jan (2007) menciona entre otros, la inclusión de la biblioteca en Wikipedia, en los calendarios de la comunidad local, la aparición de bibliotecarios en directorios de expertos o el envío de boletines.

También pueden ocurrir, en el mundo digital, que se hagan campañas de animación online a la lectura a través de weblogs (Ferrada-Cubillos y Jaña-Monsalve, 2007) o que se implemente la web social para generar comunidades online de lectores (Juárez-Urquijo, 2008), potenciando el cambio de percepción que sobre la información contenida en la biblioteca, a través del marketing relacional bibliotecario vía social media marketing (Kronqvist-Berg, 2014). También

se pueden generar enlaces y relaciones con otras instituciones relacionadas con sus lectores, para aumentar los caminos hacia la web bibliotecaria desde ellas (Welch, 2005) lo que entronca claramente con técnicas basadas en optimización en buscadores o *Search Engine Optimization*.

De hecho, según un informe de OCLC, *Perceptions of Libraries and Information Resources*, el 89% de los estudiantes americanos consultados, comienzan su proceso de investigación usando un motor de búsqueda en lugar del sitio web de la biblioteca, utilizada sólo por el 2% (Rushton, Kelehan y Strong, 2008). En otros casos se considera a los propios bibliotecarios como el objetivo de una campaña de marketing, para que ellos ejerzan una labor activa de influenciación y recomendación de ítems bibliotecarios a los usuarios finales (Chiang-Martín, 2014)

Las bibliotecas deben evaluar pues también el impacto de campañas completas de marketing (González Fernández-Villavicencio, 2014) especialmente el marketing basado en presencia en buscadores y el marketing en medios sociales, debido principalmente a su excelente relación coste-impacto. Y de forma agrupada, deberán evaluar las posibilidades que hoy día podrían aportar el análisis de cantidades de datos masivas de interacción de los usuarios, a través del estudio de los *logs* agrupados de las distintas bibliotecas, como de los datos de búsqueda en motores de búsqueda de Internet, lo que daría lugar a estudios de mercado de alto nivel no sólo cuantitativo, sino cualitativo (Serrano-Cobos, 2014)

2.2.EL ESTUDIO DE LOS HÁBITOS LECTORES

2.2.1. Qué es lectura, qué es lector

Antes de estudiar los hábitos de los lectores, cabría preguntarse qué es leer, qué es un lector, y si sólo leer un libro identifica a la “lectura” como práctica cultural. Porque según Jose Antonio Millán, “es un error asociar la lectura solo con las novelas. Si se tiene en cuenta periódicos, cómics o webs, el 92% de la población puede ser considerada “lectora” (Koch, 2017).

Sistemáticamente en los estudios analizados, se puede ver a entrevistados que se autodefinen como no lectores, cuando sí leen, desde periódicos a manuales, por lo que no podemos comparar fehacientemente estudios que computan sesgos distintos en los individuos encuestados (González, 2017) Si diversos sondeos de opinión muestran resultados significativamente distintos, si las propias encuestas de lectura necesitan una evaluación de su calidad (Pouliot, 2009) resultará científicamente difícil comparar resultados con los que diagnosticar los posibles problemas y sus causas.

Cabe por tanto la posibilidad de distinguir entre la lectura por el placer de la lectura y la lectura con otros fines, como así se comprueba en la encuesta de hábitos de lectura del PEW Research para Estados Unidos en 2012, donde un 26% de los adultos americanos consultados afirmaba leer motivados por el aprendizaje, frente al 39% que lo hacía por placer, evasión, entretenimiento o relajación.

Esa distinción lleva a investigar si difiere según el tipo de lector, por lo que es interesante separar lectores de libros digitales, de lectores de libro tradicional. Según la Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015 (Ministerio de Educación, Cultura y Deporte, 2015) el 94,9% de los españoles que leen al menos una vez al año lo hacen en soporte papel, y un 28,5% lee en soporte digital (obsérvese que suman más de 100%, por lo que hay una parte de esa población que lee indistintamente en ambos soportes)

A pesar de ello, la proporción de lectores en soporte digital crece, pues según el mismo estudio, en el último periodo 2014-2015 se está triplicando el número de lectores en soporte digital con respecto a 2011. Y según el estudio “Hábitos de lectura y compra de libros en España 2012” (Federación de Gremios de Editores de España, 2012) de entre los entrevistados que leían, el 58% leía en formato digital.

Lectura total vs lectura digital

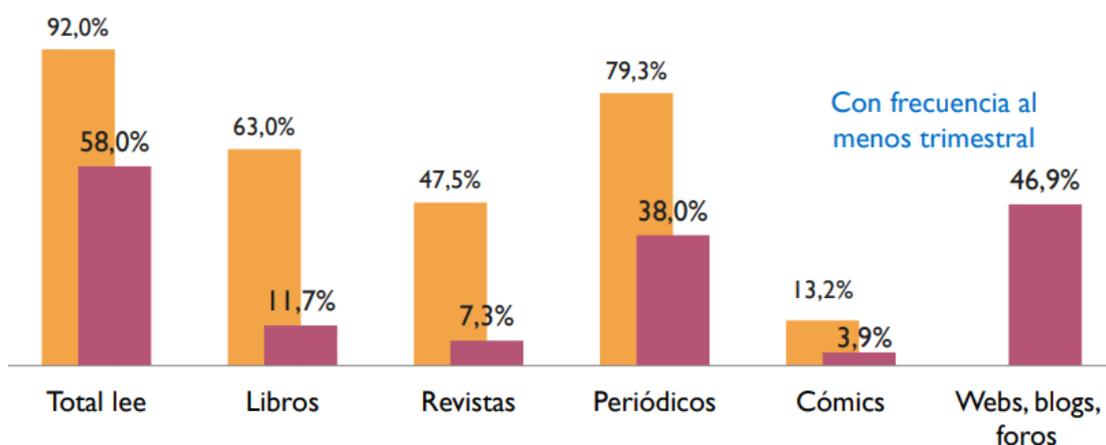


Fig. 2-5 Lectura total vs lectura digital. Federación de Gremios de Editores de España, 2012.

Diversos estudios con distintas muestras de habitantes vienen a ayudarnos a ver que los lectores entienden por lectura no sólo la de los libros. En el caso del estudio del Grupo de Investigación E-Lectra, de la Universidad de Salamanca, entre un segmento de lectores frecuentes que utilizan dispositivos digitales (Cordón-García, Gómez-Díaz y Alonso-Arévalo, 2013, p.189), los libros eran el tipo de contenido consumido por el 55% de los encuestados, siendo el resto cómics, novelas gráficas y artículos científicos. Otro estudio, el de “La Sociedad de la Información en España 2015” de la Fundación Telefónica, señalaba que el 78,6 % de la población española usa internet para leer noticias (Sáinz Peña, 2016).

Pero es lícito pensar que los lectores no separan sus intereses de manera excluyente. En el citado estudio de la Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015 (Ministerio de Educación, Cultura y Deporte, 2015) podemos comprobar que un 92,5% de los encuestados que han leído libros por motivos no relacionados con la profesión o estudios en un trimestre, consumen literatura de creación literaria, mientras que el 57,3% han leído otro tipo de literatura (científica, información y otros) por lo que hay cierto solapamiento (hay lectores que leen tanto un tipo de literatura como otra, pero lamentablemente no se refleja en el estudio qué porcentaje concreto elige ambas opciones).

3.14 Personas que han leído libros por motivos no relacionados con la profesión o estudios en un trimestre según los temas de los libros leídos

(En porcentaje del total vertical)

	Total han leído en un trimestre		Hombres	Mujeres	De 15 a 24 años	De 25 a 34 años	De 35 a 44 años	De 45 a 54 años	55 y más años
	Miles	%							
TOTAL (Miles)	17.744		7.442	10.302	2.392	2.857	3.913	3.598	4.982
%		100	100	100	100	100	100	100	100
CREACIÓN LITERARIA	16.407	92,5	90,8	93,7	92,4	92,2	93,1	92,5	92,2
Infantil, juvenil	1.188	6,7	4,8	8,0	7,2	5,8	12,7	7,0	2,0
Comics e historietas	473	2,7	4,6	1,2	5,6	3,4	3,0	2,4	0,8
Biografías o memorias	1.570	8,8	9,3	8,5	5,9	5,7	7,5	8,9	13,1
Teatro	293	1,7	1,6	1,7	3,8	1,0	1,3	1,2	1,6
Poesía	655	3,7	2,8	4,3	4,8	2,9	2,3	3,9	4,5
Novela	15.317	86,3	84,5	87,6	84,7	87,2	86,0	88,0	85,7
Novela clásica	4.380	24,7	20,6	27,6	19,2	19,4	25,2	26,4	28,7
Contemporánea	13.361	75,3	76,0	74,8	77,0	77,9	74,5	76,5	72,8
Histórica	5.477	30,9	34,7	28,1	19,2	23,1	28,3	34,0	40,7
Aventuras	3.544	20,0	22,5	18,1	27,3	21,2	20,3	18,6	16,5
Costumbrista	988	5,6	4,3	6,5	2,4	3,5	4,3	8,1	7,5
Sentimental/amor	2.287	12,9	3,3	19,8	18,6	17,3	11,9	11,6	9,3
Negra y policíaca	1.932	10,9	10,0	11,5	7,9	10,4	11,8	12,1	11,1
Ciencia ficción	1.746	9,8	13,1	7,5	16,8	12,9	10,8	8,9	4,7
Otros	3.493	19,7	19,5	19,8	19,0	21,0	19,9	22,2	17,2
DIVULGACIÓN, INFORMACIÓN Y LIBROS NO PROFESIONALES	10.176	57,3	63,1	53,2	54,8	57,4	55,1	59,0	59,1
Enciclopedias y diccionarios	481	2,7	3,4	2,2	3,5	2,3	2,3	2,6	2,9
Historia	3.799	21,4	25,3	18,6	15,4	16,7	17,5	23,4	28,6
Filosofía, psicología	1.261	7,1	6,5	7,5	6,8	7,5	7,1	7,6	6,7
Religión, teología	735	4,1	4,3	4,0	2,0	3,6	3,0	3,7	6,7
Política y economía	895	5,0	8,1	2,8	4,2	3,9	4,6	5,2	6,3
Ciencias sociales y humanidades	1.111	6,3	7,2	5,6	7,2	4,6	6,5	7,1	6,0
Ciencias médicas y naturales	813	4,6	4,1	4,9	5,2	4,8	4,5	4,2	4,4
Ciencias físicas, químicas, matemáticas e informática	594	3,3	5,3	1,9	5,9	3,8	3,6	3,1	1,8
Deportes y juegos	418	2,4	4,2	1,0	4,9	3,2	2,6	1,5	1,1
Artes plásticas y gráficas, fotografía	367	2,1	2,2	2,0	2,4	2,0	2,3	2,4	1,6
Cine, música, teatro	432	2,4	3,5	1,7	3,8	2,9	2,6	2,1	1,6
Toros	28	0,2	0,3	0,1	0,1	0,3	0,2	0,1	0,2
Otros	3.471	19,6	19,3	19,7	20,6	21,0	20,6	19,6	17,4
OTROS	242	1,4	1,4	1,3	2,0	1,5	1,4	1,3	1,0

Fig. 2-6 Personas que han leído libros por motivos no relacionados con la profesión o estudios en un trimestre según los temas de los libros leídos. Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015.

Es interesante contrastar estas cifras de lectura con otros estudios. En el caso del estudio de “Hábitos de lectura y compra de libros en España. Año 2004” realizado por Federación de Gremios de Editores de España, por encargo de la Federación de Gremios de Editores de España, se observaba que para la lectura en sentido clásico (lectura de libros) seguía reinando la novela como temática más leída en España, fluctuando del 71,5% en 2002 al 74,3% en 2004,

con crecimiento del ensayo, mientras que la lectura de libros de humanidades y ciencias sociales bajaba de 14,7% en 2002 al 11,0% en 2004.

Más de 10 años después, las tendencias eran similares. En 2017, (y aunque aquí la pregunta cambia) la mayor parte de los lectores encuestados en el “Barómetro de hábitos de Lectura y Compra de Libros en España 2017” (Federación de Gremios y Editores de España, 2017, p. 52) indican que la materia del último libro que han leído versa sobre “literatura” (74%) seguida de Humanidades y Ciencias Sociales (9,9%) Científico-técnico (3,0%) libros prácticos y enseñanza (6,4%) infantil y juvenil (2,6% y “otros” el 2,8%, con cierta tendencia a la baja en las dos primeras materias, como se puede ver en el gráfico 2.8.

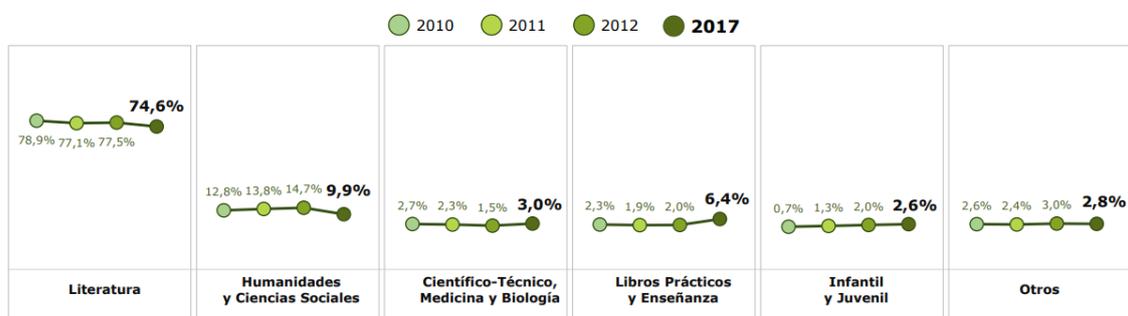


Fig. 2-7 ¿Cuál era la materia del último libro leído? Federación de Gremios y Editores de España, 2017.

2.2.2. Metodologías de investigación de los hábitos de lectura

El análisis del comportamiento lector, como “expresión social de la forma en que una persona representa y practica la lectura en el contexto de la cultura escrita que lo acoge” (Igarza y Monak, 2014, 22) puede enfocarse desde distintos puntos de vista, dimensiones, escala y contextos no necesariamente excluyentes. Por ejemplo y entre otros:

- Análisis cuantitativos: se encargan de cifrar los hábitos lectores mediante el cálculo (pasado o futuro) de indicadores cuantitativos, que sirvan por ejemplo para dimensionar el aumento de interés en un determinado tipo de literatura, autor, etc.
- Análisis cualitativos: dominado hasta la actualidad por técnicas como las encuestas y/o las entrevistas, que buscan ahondar en los porqués, en las motivaciones, tanto como en qué porcentaje de qué segmento de población se comporta de determinada manera. Lo habitual es ver cruces entre análisis cualitativos y cuantitativos.
- Análisis de la demanda: qué es lo que los usuarios quieren consumir o consultar sirve a escritores, editores, librerías y/o bibliotecarios para dimensionar sus esfuerzos futuros, basándose en lo que ha tenido éxito (o no lo ha tenido) entre determinado segmento lector.
- Análisis de la conducta: cómo se acerca el usuario / lector a la lectura, en qué momento, en qué términos, a través de qué canales (de marketing o de contacto).
- Análisis por formato: si hubo un momento histórico en el que no sea trivial preguntarse en qué formato lee el usuario es éste, puesto que la tecnología ha traído una creciente variedad de dispositivos de lectura e interacción: móviles, ordenadores, tabletas, televisores, libros electrónicos, videojuegos...
- Análisis geográfico: Segmentar el estudio geográficamente ayuda a precisar las diferencias contextuales locales.
- Análisis por país / idioma: en un planeta en el que los movimientos migratorios se han acelerado en todas las direcciones y en particular en España (tanto por entrada como por salida de migrantes, que también pueden ser lectores) el análisis por idioma, lenguaje, raza o país permitirá conocer mejor hábitos que no habían sido conocidos en nuestro país hasta hace bien poco, y que pueden no estar representados en los estudios clásicos.

- Análisis monodisciplinarios: aquellos que sólo emplean una técnica concreta, como la cuenta de libros en los carritos de las bibliotecas, para generar indicadores de consumo bibliotecario.
- Análisis multidisciplinares: aquellos que atacan el problema de investigación desde distintas áreas del conocimiento, como lo sería, ampliando el caso anterior, la entrevista a aquellos usuarios que depositan libros en los carritos de las bibliotecas, para indagar por qué los dejan y no los piden en préstamo o los leen en sala.

Existen manuales de métodos de estudio de hábitos lectores, como *Metodología común para explorar y medir el comportamiento lector* (Uribe et. al., 2011, ampliado al mundo digital en Igarza, 2014) refrendado por la UNESCO, o la “Guía metodológica para el estudio de las necesidades de formación e información de los usuarios o lectores” (Núñez, 1992) y en el mundo de las bibliotecas otros enfocados al análisis de los usuarios, que también aplica a este aspecto de la relación usuario - biblioteca (Sanz y Gómez, 1994; Gómez-Hernández, 1996; Villaseñor-Rodríguez, 2017).

Pero a pesar de ello, y como se apreciará en el siguiente apartado, existen multitud de estudios que utilizan, si no metodologías en sí distintas, sí preguntas diferentes, lo que por un lado nos brindará un abanico de dimensiones de análisis de gran complejidad sociológica, y por otro lado una gran incertidumbre en cuanto a la capacidad de comparar estudios de distinta autoría país y contexto temporal (Pouliot, 2009).

Las investigaciones al respecto muestran fundamentalmente dos vertientes: una cuantitativa, que descubre cuánto leen los que leen (Millán, 2017); y otra cualitativa, que cubre distintos aspectos de la experiencia de la lectura, desde las motivaciones por las que se llega a la lectura, hasta explorar qué, cuándo, dónde y cómo se lee (Ministerio de Educación, Cultura y Deporte, 2016).

En la literatura consultada domina abrumadoramente la metodología de investigación basada en encuesta y en entrevista (Federación de Gremios de Editores de España, 2004). Escasamente se observa otra metodología, como no sea la de los *focus groups* (Turner, 2016) en el contexto del uso de una biblioteca; y la etnografía, véase el ejemplo de proyecto Territorio Ebook sobre los hábitos de lectura (incluyendo la digital) de socios de bibliotecas públicas en España (Cordón-García, 2011).



Fig. 2-8 Prácticas de recomendación y obtención de lectura entre prestatarios de bibliotecas francesas. Fuente: RENARD, Hervé, 1995 (Cordón-García, 2006)

Se descubre aquí un primer problema metodológico, pues se observan múltiples definiciones para un mismo indicador. Probablemente, la primera pregunta recurrente vista en la literatura sobre análisis de hábitos lectores es “¿cuánto se lee?” La evolución es diferente si la pregunta es “¿cuántos lectores frecuentes hay?” Pero definir “lector frecuente” es en sí mismo un problema, puesto que en los distintos estudios localizados se observan diferencias de criterio que impiden una comparación de los resultados.

Así, el Ministerio de Cultura en el estudio "Equipamientos, Prácticas y Consumos Culturales de los Españoles" de 1991, señalaba como lectores habituales al 24,4% de los encuestados en aquella ocasión, definiendo “lectores habituales” como que aquellos que “leen todas las semanas un libro”, siendo así lectores diarios el 18,1 %, y no lectores el 41,7% de la población española adulta de la época (Perez, 1994) No se habla de otros medios ni de Internet, claro, puesto que estaba llegando todavía el CD-ROM, por lo que la lectura se identificaba todavía con un libro en papel.

La Federación de Gremios de Editores de España (2004) utiliza dos indicadores para revelar inconsistencias de los resultados, puesto que miden dos naturalezas distintas (objetiva y subjetiva):

- “1- Frecuencia de lectura: percepción que los entrevistados tienen de sí mismos como lectores.
- 2- Libros leídos al año: medida objetiva del número de libros que han leído enteros en un año.”

Según González (2017), la variable frecuencia de lectura ofrece la siguiente clasificación (sin distinguir entre tipo o dispositivo de lectura):

- “Lector ocasional: lee una vez al mes o al trimestre.
- Lector frecuente: lee todos los días o una o dos veces a la semana.”

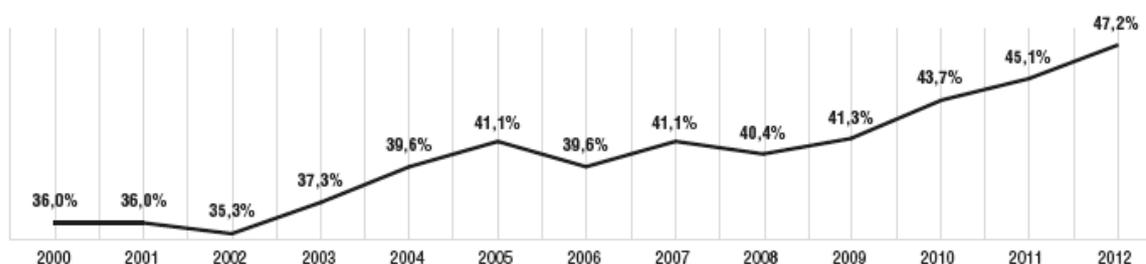


Fig. 2-9 Evolución del porcentaje de lectores frecuentes respecto a la población española. Fuente: Federación de Gremios de Editores de España, 2012.

Sin embargo, la “Encuesta de hábitos y prácticas culturales en España” (Ministerio de Educación, Cultura y Deporte, 2015) pregunta en su encuesta “¿ha leído al menos un libro durante el último año?” por lo que la definición de “lector frecuente” es radicalmente distinta. En este caso especifica al menos que se pregunta sobre el libro (sin separar entre dispositivos de lectura en la pregunta), aunque también refleja un incremento de 4,5 puntos en el período 2007 - 2015.

En el caso del CIS, la pregunta que se hace a los encuestados es sobre su percepción del grado de lectura en España (Centro de Investigaciones Sociológicas de España (CIS), 2016). Así, ante la pregunta “¿diría Ud. que en España la gente lee...?” los españoles opinan abrumadoramente

desde 1998 hasta 2016 que en España se lee poco (un promedio del 67,38% a lo largo de esos años, con poca variación interanual).

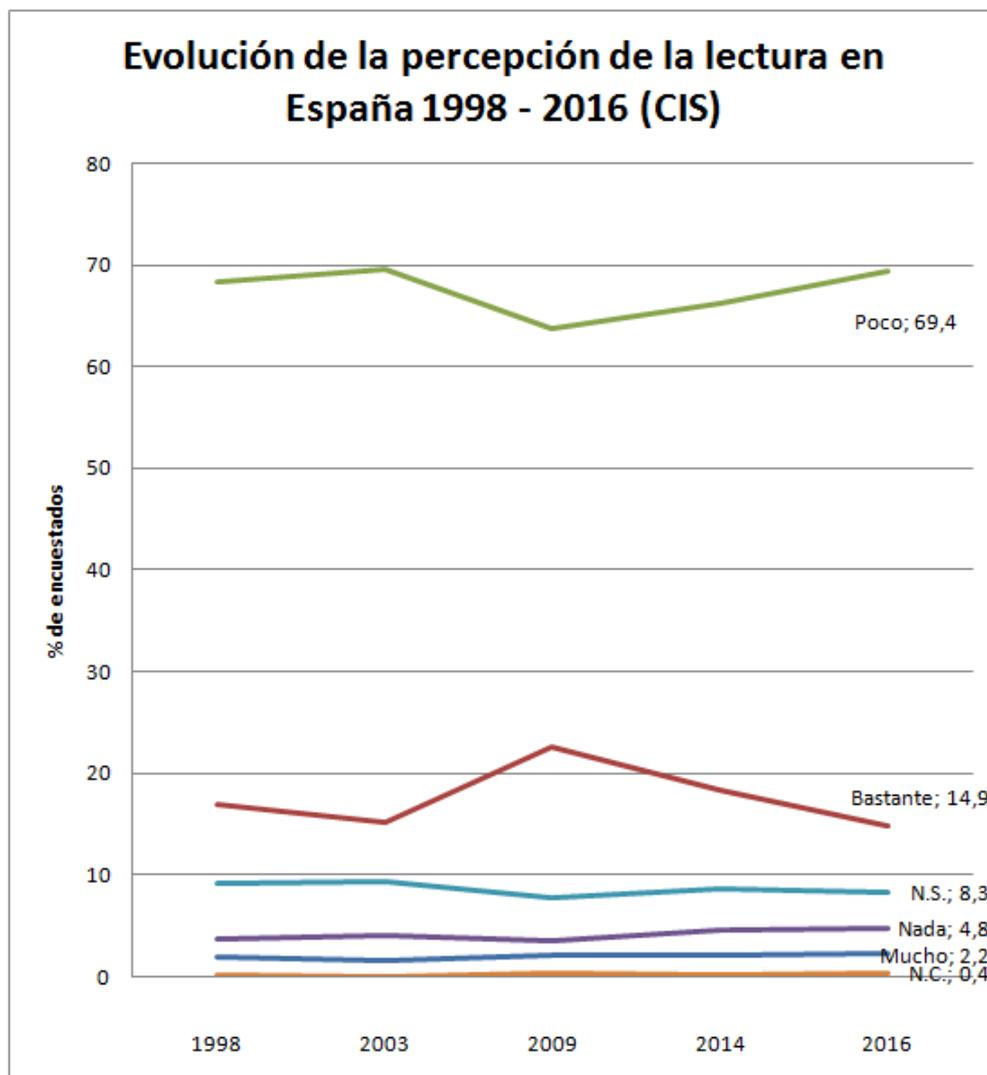


Fig. 2-10 Evolución de la percepción del grado de lectura en España 1998-2016. Fuente: Centro de Investigaciones Sociológicas de España (CIS), 2016. (Elaboración propia)

Fuera de eso, es prácticamente imposible obtener una serie temporal de las encuestas del CIS (1998, 2003, 2009, 2014, 2016) en las que se preguntó a los ciudadanos españoles sobre temas relacionados con la lectura, pues las preguntas cambian o desaparecen, por lo que no se puede obtener una visión histórica fidedigna de la evolución del comportamiento de los españoles a este respecto.

Por ejemplo, sólo en 1998 y 2003 se preguntó “... me gustaría que me dijera, en general, ¿qué tres tipos de actividades suele Ud. hacer durante su tiempo libre en un día laborable?” Se ha unificado aquí una tabla con los datos obtenidos, pero hay que hacer notar que en 2003 se añadió una posible contestación más, “navegar por internet”, que no existía en la entrevista de 1998, con lo que los resultados no son contrastables con los de 1998. En 2009 las preguntas cambiaron definitivamente.

Tabla 2-2 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.

	1998	2003
Ver la televisión	69.9	70.9
Escuchar música	26.5	25.2
Escuchar la radio (no programas de música)	17.7	17.8
Leer (prensa, libros, etc.)	34.5	33.1
Hacer trabajos manuales (bricolaje, punto, etc.)	13.8	10.8
Asistir a espectáculos (cine, teatros, conciertos, etc.)	4.2	3.2
Ver alguna exposición (pintura, fotografía, etc.)	1.2	1.0
Jugar con el ordenador en casa	4.0	3.3
Pasear, hacer ejercicio, deporte, gimnasia	36.2	35.1
Salir a 'tomar algo' con pareja/familia	11.4	11.8

Reunirse con amigos fuera de casa	19.1	15.5
Jugar con los hijos	8.7	9.4
Otras actividades	8.4	9.0
Navegar por Internet	X	5.6
No_sabe	0.4	0.2
No_contesta	0.6	0.6

No sólo en España evolucionan las preguntas que se hacen los estudios sobre la práctica de la lectura, y los indicadores utilizados son parecidos pero diferentes. Un caso más ilustra esta variedad de interpretaciones. En la *Encuesta Continua de Hogares (ECH)* del Departamento Administrativo Nacional de Estadística (DANE) de Colombia, se comprobó que, en 2005, la proporción de colombianos adultos que afirmaba leer de forma frecuente bajaba de 67,9% en 2000 a 65,9% en 2005 (Gamboa y Reina, 2006)

Ahora bien, en el estudio se usa la expresión “lectura habitual”, y no cuenta con una definición para los propios encuestadores en 2000 y 2005, por lo que la interpretación del término estaba sujeta a la subjetividad del encuestado, teniendo además la posibilidad de elegir como “lectura habitual” cualquier combinación de estas posibilidades: i) libros, incluidos textos escolares; ii) revistas; iii) diarios o periódicos; y iv) internet. Sin embargo y para el mismo país, no se encuentra referencia al concepto “lectura habitual” en el manual de Igarza y Monak(2014).

En México se analiza la propia *Encuesta Nacional de Lectura*, que recoge las diferencias entre libros, periódicos, revistas e historietas / comics, computadora e internet, para contextualizar mejor los segmentos de población lectora, pero que admite que todavía deja sin estudiar “algunos textos, actividades, o aspectos de la lectura escrita” (Zaid y Goldin, 2006) En el estudio de Gamboa y Reina (2006) para Colombia, se constata el aumento de lectura habitual pero en internet, que creció 144% entre los encuestados en el período 2000-2005.

Una vez más, no es fácil comparar resultados al usarse distintas métricas, pues durante el período 2013-2015, según la encuesta de Parametría, (con un margen de error de 3.5 puntos porcentuales y nivel de confianza de 95%) contrariamente a los datos de Colombia, se expone que aumentó el porcentaje de población mexicana (de 50 a casi 60%) que no ha leído ni siquiera un libro en los últimos 6 meses (Quezada, 2016)

2. ¿Qué tipo de libros lee y en qué formato?		
	Impreso	Digital
a. Libros infantiles	<input type="checkbox"/>	<input type="checkbox"/>
b. Libros juveniles	<input type="checkbox"/>	<input type="checkbox"/>
c. Textos educativos	<input type="checkbox"/>	<input type="checkbox"/>
d. Enciclopedias y diccionarios	<input type="checkbox"/>	<input type="checkbox"/>
e. Científicos/técnicos/profesionales	<input type="checkbox"/>	<input type="checkbox"/>
f. Historia, política y sociales	<input type="checkbox"/>	<input type="checkbox"/>
g. Operativos o de operación técnica	<input type="checkbox"/>	<input type="checkbox"/>
h. Literatura	<input type="checkbox"/>	<input type="checkbox"/>
i. Religión	<input type="checkbox"/>	<input type="checkbox"/>
j. Superación personal/autoayuda	<input type="checkbox"/>	<input type="checkbox"/>
k. Cocina	<input type="checkbox"/>	<input type="checkbox"/>
l. Otro	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 2-11 Ejemplo de pregunta en una encuesta sobre hábitos lectores. Fuente: Igarza y Monak Salinas, 2014.

El mismo estudio anteriormente mencionado de Gamboa y Reina (2006) denuncia el problema metodológico que supone constatar que aumentó el número de quienes afirmaron “leer habitualmente revistas (de 26,2% a 27,2%), diarios o periódicos (de 31,2% a 32%) e internet (de 4,9% a 11,9%)” pero con un cambio en las preguntas en esos años (2000-2005), haciéndose difícil extrapolar conclusiones, tanto sobre qué era lectura habitual o frecuente para los encuestados, como para saber si en su imaginario, en el concepto “leer” separaban la lectura del libro o la lectura en general sin distinguir entre dispositivos (papel y electrónico) o entre tipos de textos.

Ídem ocurre en las estadísticas bibliotecarias (CULTURABase, 2017), donde hasta 2010 se usa un tipo de fondo denominado “documentos electrónicos”, y a partir de esa fecha se utilizan tipos de fondo como “libros electrónicos” y “juegos electrónicos”, sin que se explique si es una división del tipo de fondo anterior o no, por lo que no es posible seguir una serie temporal más ambiciosa en el análisis de la presencia de literatura digital en las bibliotecas.

Otro tanto denuncia Gil, basándose en Igarza y Monak (2014), propugnando que, sin salir de la encuesta como metodología científica, se deben estandarizar y “unificar este tipo de estudios”, puesto que “cuando intentas comparar los datos de un país con los de otros es frecuente observar que a veces los ratios no miden lo mismo o las magnitudes son completamente divergentes” (Gil, 2015). Este punto se observa tanto al intentar comparar indicadores de distintos países (Alonso-Arévalo, Cordón-García y Gómez-Díaz, 2014) como al comparar distintos estudios para un mismo país (Quezada, 2016), siendo una pregunta no resuelta, como señala González cuando se pregunta “¿Cómo vamos a llegar a acuerdos sobre políticas de lectura cuando no todo el mundo habla de lo mismo?” González (2017, 81)

2.2.3. Hábitos lectores del no lector

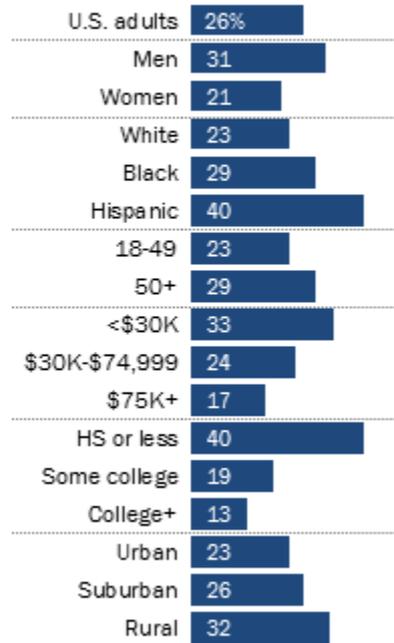
Tradicionalmente se ha estudiado a los lectores y sus hábitos, pero también será imperioso conocer qué hay de los no lectores y las causas de que no lean, para mejorar los servicios bibliotecarios y las estrategias de generación de contenido en las industrias culturales. En España, el Centro de Investigaciones Sociológicas (CIS) realiza un estudio sociológico a nivel nacional en torno a los hábitos de lectura en España. Según esta entidad, en 2015 “una de cada tres personas (35%) no lee 'nunca' o 'casi nunca’” (Dotras, 2015) siendo un 21% según otra encuesta, la de “Hábitos y Prácticas Culturales en España 2014 – 2015” (Ministerio de Educación, Cultura y Deporte, 2015)

Según la anterior encuesta, son más los hombres que no leen ni un libro (37,9%), siendo el 32,1% de mujeres que tampoco leen. La razón aludida en el 42% de los casos para no leer, es que simplemente no les interesa o no les gusta. En México, el 79.9% respondieron en 2015 que no leen por falta de tiempo y sólo el 14.6% porque no le gusta (Quezada, 2016), explorando los segmentos que conforman la población no lectora en el último año, que allá era en 2016 del 26% (Perrin, 2016)

En el caso americano (PEW Research, 2016), como se indica en la figura 2.12, también el 31% de hombres eran no lectores, siendo menor el porcentaje de mujeres no lectoras (21%) Así, el mapa de adultos que no leen en USA se forma en su mayoría por hombres, hispanos, aquellos con ingresos de menos de 30.000 dólares y los de menor formación (que sólo cuentan con diploma de secundaria).

Who hasn't read a book in the past 12 months

*% of U.S. adults who say they have
not read a book in any format in the
past 12 months*



Note: Whites and blacks include only non-Hispanics.

Source: Survey conducted March 7–April 4, 2016.

PEW RESEARCH CENTER

Fig. 2-12 Who doesn't read books in America? PEW Research, 2016.

2.2.4. El *best seller*

Es de destacar, por su importancia capital tanto en la estrategia del mercado editorial como en la de la biblioteca, el fenómeno del *best seller* que, para los análisis investigados hasta ahora, son objeto de estudio entre el grupo de lectores que leen lo que la opinión pública conoce como “libro” (es decir, la literatura consumida por el placer de leer)

Una de las preguntas que no es fácil encontrar en los estudios de los hábitos lectores, es por qué el lector lee lo que lee (Viñas, 2009) No es fácil desglosar, mucho menos en un sondeo de opinión, las verdaderas razones causales que expliquen si el fenómeno del *best seller* aparece o se provoca, es primero el huevo o la gallina. Por ejemplo, según Gómez-Escalonilla (2005) la precaria situación de las bibliotecas españolas lleva a los bibliotecarios a contestar a la mayor demanda (al *best seller*) que según afirma, viene provocada por el mercado editorial, el cual favorece la difusión de unos cuantos títulos "superventas", por lo que sugiere fomentar la diversidad editorial y el acceso a la lectura.

El análisis de distintos *best sellers*, muestra que no tiene por qué haber una correlación entre éxito editorial y calidad literaria (Alamo, 2009) Por tanto la pregunta persiste metodológicamente, siendo complejo detectar qué es una tendencia en lectura. Toda tendencia comparte un conjunto de rasgos universales, según el Future Today Institute (2016):

- Es impulsada por una necesidad humana básica.
- Es oportuna, pero persiste.
- Evolucionan a medida que emerge.
- Puede materializarse como una serie de puntos no conectados inicialmente, que comienzan apareciendo de forma marginal, y se mueven a la corriente principal.

Elegir la temática sobre la que escribir también es un factor relevante en el éxito potencial de una novela, por lo que conocer las temáticas que más gustan o disgustan a los lectores se convierte en un aspecto clave del problema. En España, el estudio de *Hábitos de Lectura y Compra de Libros en España* (Federación de Gremios de Editores de España, 2004) se lo preguntaba también, entrevistando a 10.412 lectores de novelas, mostrando que las temáticas que más gustaban en ese año eran las históricas (42,4%) las de intriga (35%) y las de aventuras (31,3%) siendo las románticas (33,2%), las de terror (23,3%) y las de ciencia ficción (19,6%) las

que menos gustan, donde resulta interesante observar que las románticas son de las obras que más polaridad presentan (o gustan mucho o no gustan nada).

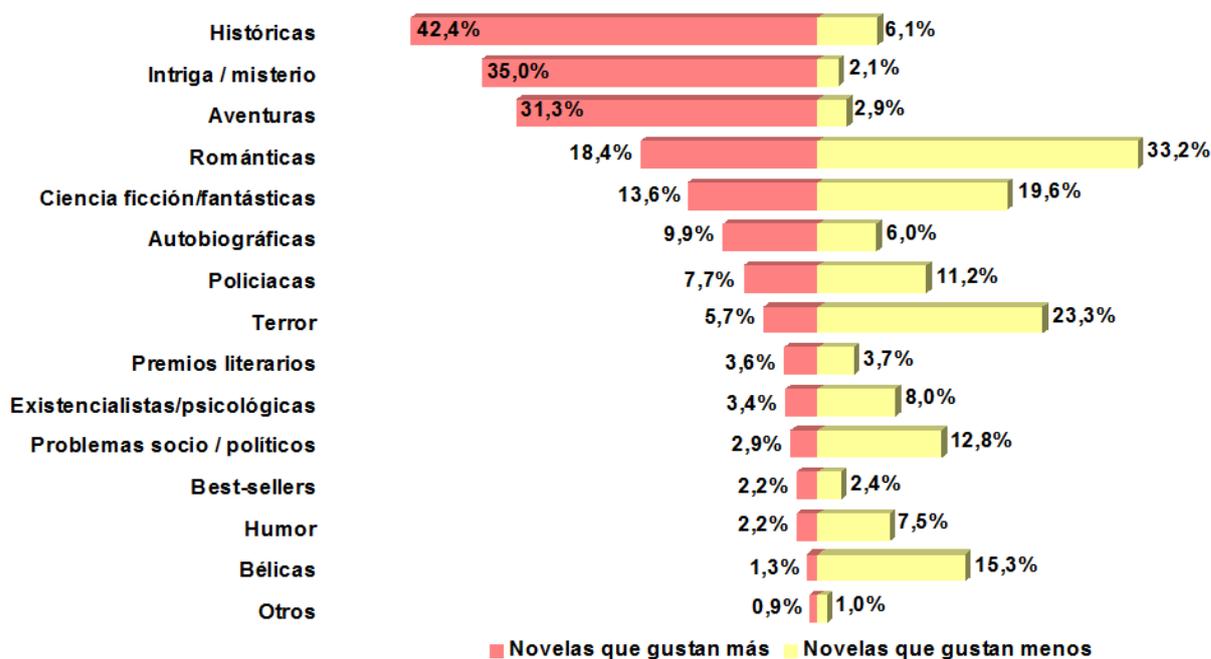


Fig. 2-13 Novelas que más y menos gustan. Año 2004. Fuente: Federación de Gremios de Editores de España.

Por qué una tendencia pasa de no estar conectada a ser parte de la corriente principal es sin duda, el quebradero de cabeza de editoriales y profesionales del marketing del libro. En lo tocante a este estudio, interesará detectar y observar esas tendencias, así como buscar similitudes prácticas entre ellas, que aporten más información sobre estos rasgos universales.

2.2.5. Tipos de lectores

Además de las encuestas generalistas, desarrolladas sobre universos de población amplios, como por ejemplo mayores de 14 años de España, en el *Barómetro de hábitos de Lectura y Compra de Libros en España 2017*, (Federación de Gremios y Editores de España, 2017), se observa que los hábitos de lectura también se han estudiado en ámbitos o segmentaciones concretos, detectando, segmentando y estudiando distintas tipologías de lectores, lo que lleva a dificultar la comparación entre conclusiones de distintos estudios.

Así algunos autores trabajan sobre estudiantes universitarios. Yubero y Larrañaga (2015) descubrían que el 7,9% de estudiantes universitarios de España y el 7,4% de Portugal no leen (libros) nunca (de forma voluntaria), lo que impide abordar una competencia genérica que es básica, porque “correlaciona en un alto grado con el nivel de consecución de otras competencias genéricas”. Otras conclusiones pertinentes son que:

- Las mujeres universitarias son más lectoras que los hombres, pero de éstos, los que leen, leen más habitualmente que en el caso de las mujeres.
- Cuanto más habitual es un lector, mayor es la variedad de tipologías literarias que lee.
- El puro gusto por la lectura es la motivación más común, seguida de la diversión, la evasión, y aprender.

Además, a los estudiantes españoles de España y Portugal, da igual su categoría, les gusta en general la novela, y en particular la novela romántica, la de intrigas y la de aventuras, las tipologías literarias que más gustan, de media.

	NL	FL	LO	LH
Ensayo	3,8	5,8	9,3	17,6
Poesía	8,0	8,9	10,4	18,5
Teatro	13,6	8,5	10,3	12,8
Novela	61,7	66,3	83,0	90,5
Histórica	12,1	16,9	26,6	41,1
Aventuras	35,1	38,7	39,0	34,2
Intriga	35,7	41,4	44,2	46,1
Romántica	42,8	38,0	44,2	36,9
Terror	13,0	12,5	9,5	14,3
Ciencia ficción	15,0	19,9	21,3	33,9

Nota: NL= No lector; FL= Falso lector; LO= Lector ocasional; LH= Lector habitual

Fig. 2-14 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.

En el caso de Yubero y Larrañaga (2015), una categoría o segmentación de lector interesante que se incluye en su estudio, caracterizada en una publicación anterior de los mismos autores Yubero y Larrañaga (2005), es la de “falsos lectores” (FL). El problema es la dificultad de la definición del público estudiado y la dificultad de encontrar estudios que trabajen con similares segmentaciones, pues estos “falsos lectores”, según estos autores, son aquellos lectores que no son ni habituales ni ocasionales, sino que “no leen durante su tiempo libre, aunque hayan leído más de 2 libros en el último año; o bien, aunque pueden leer durante su tiempo libre, no han leído más de 2 libros en el último año” (Yubero, 2009, 104). Este segmento de población no tiene un estilo de vida lector, pero llega a falsear las respuestas que da al entrevistador, debido a la “deseabilidad social” y utilizando el “autoengaño”, al mostrar una actitud favorable hacia su audiencia concreta (el entrevistador) por lo que busca responder con un dato cercano a lo deseable.

Los autores apuntan como razón de esta falta de integración de la lectura en su estilo de vida a un momento anterior en la educación del estudiante, pues “resulta contradictorio que entre los objetivos de la educación primaria se incluya el desarrollo del gusto por la lectura y que en las siguientes etapas educativas (obligatorias y no obligatorias) desaparezca este aspecto por

completo, quedando la lectura relegada exclusivamente a su valor instrumental y obligatorio” (Yubero y Larrañaga, 2015, 719).

En el caso de la Encuesta Nacional de Lectura de 2005 realizada por CONACULTA en México (Ramírez Leyva, 2008), revelaba otra clasificación distinta de los lectores, donde se identifica a los lectores frecuentes como “extensivos”, y a los que leen preferencialmente periódicos y de revistas como lectores “intensivos”, leyendo más libros al año los lectores extensivos, que sin embargo declaraban leer mayor variedad de temas.

Hay por tanto distintos estudios sobre tipologías de lectores concretas, por lo que es difícil generalizar conclusiones, siendo el contexto de estudio un factor clave para la utilidad del mismo. Por ejemplo, otra área de estudio de los hábitos de lectura por segmentación, es la de lectura en la edad temprana, durante la infancia, que es el momento en el que se afianza o destruye el interés por la lectura. En este sentido, Salazar y Ponce (1999) avanzaban en su investigación de niños de educación primaria de Perú, que “la escuela no acerca a los alumnos al libro y la lectura de forma adecuada”, pues “la mayoría de ellos asocia la lectura al aburrimiento y al castigo, y se entrega a esta actividad más por obligación que por voluntad propia”, siendo sólo un 19,9% de los niños encuestados, los que afirmaban “me gusta leer cuentos y novelas”.

Concluyen los autores que la escuela “familiariza al niño con el libro para cumplir con su objetivo de enseñar a leer para cubrir una necesidad”, lo que puede ser privativo de un país y un tiempo concretos, de la misma manera que las aplicaciones a políticas de lectura sean distintas en el contexto de los no lectores americano (Perrin, 2016) o el español en el de los usuarios de bibliotecas públicas (Ministerio de Educación, Cultura y Deporte, 2011).

2.2.6. Motivaciones para la lectura

Una encuesta realizada entre usuarios de biblioteca pública de Portugal, relacionaba las razones de lectura con la compra de libros, y revelaba que las mayores motivaciones para ambas prácticas venían dadas en su mayoría por las sugerencias de amigos, el conocimiento del tema, y el conocimiento del autor (Cordón-García, 2006).

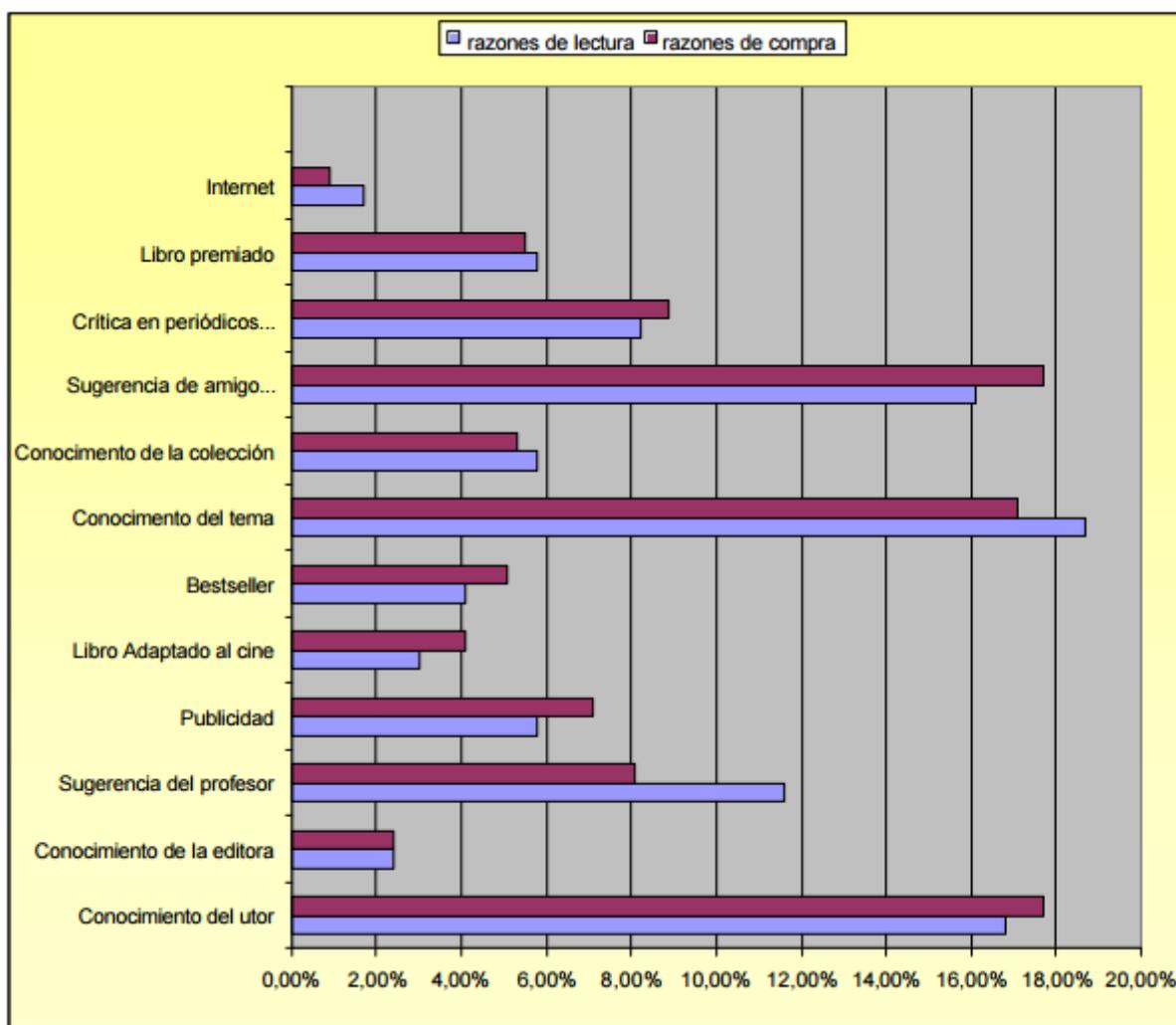


Fig. 2-15 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.

Ese conocimiento del autor puede venir por distintos motivos, entre otros el descubrimiento del autor en las bibliotecas. Una encuesta estadounidense publicada en *Library Journal*, citada

por Cerdón-García (2011) mostraba en 2010 que el 50% de los usuarios de las bibliotecas académicas americanas estudiadas, decían comprar libros escritos por un autor descubierto en la biblioteca.

En España, la encuesta de *Hábitos de Lectura y Compra de Libros en España* (Federación de Gremios de Editores de España, 2004) incidía en el crecimiento del consejo de amigos y conocidos (60,8%) como una de las mayores motivaciones de compra de libros. En este estudio Internet aparecía todavía tímidamente, con sólo un 2,3% en 2004, pero evidenciaba la bajada de la importancia de la biblioteca en el ciclo de vida de la lectura, bajando al 1,4% (es decir, que sin contradecir el estudio de *Library Journal*, el peso del descubrimiento de autores y libros era mayor en otras fuentes de referencias).

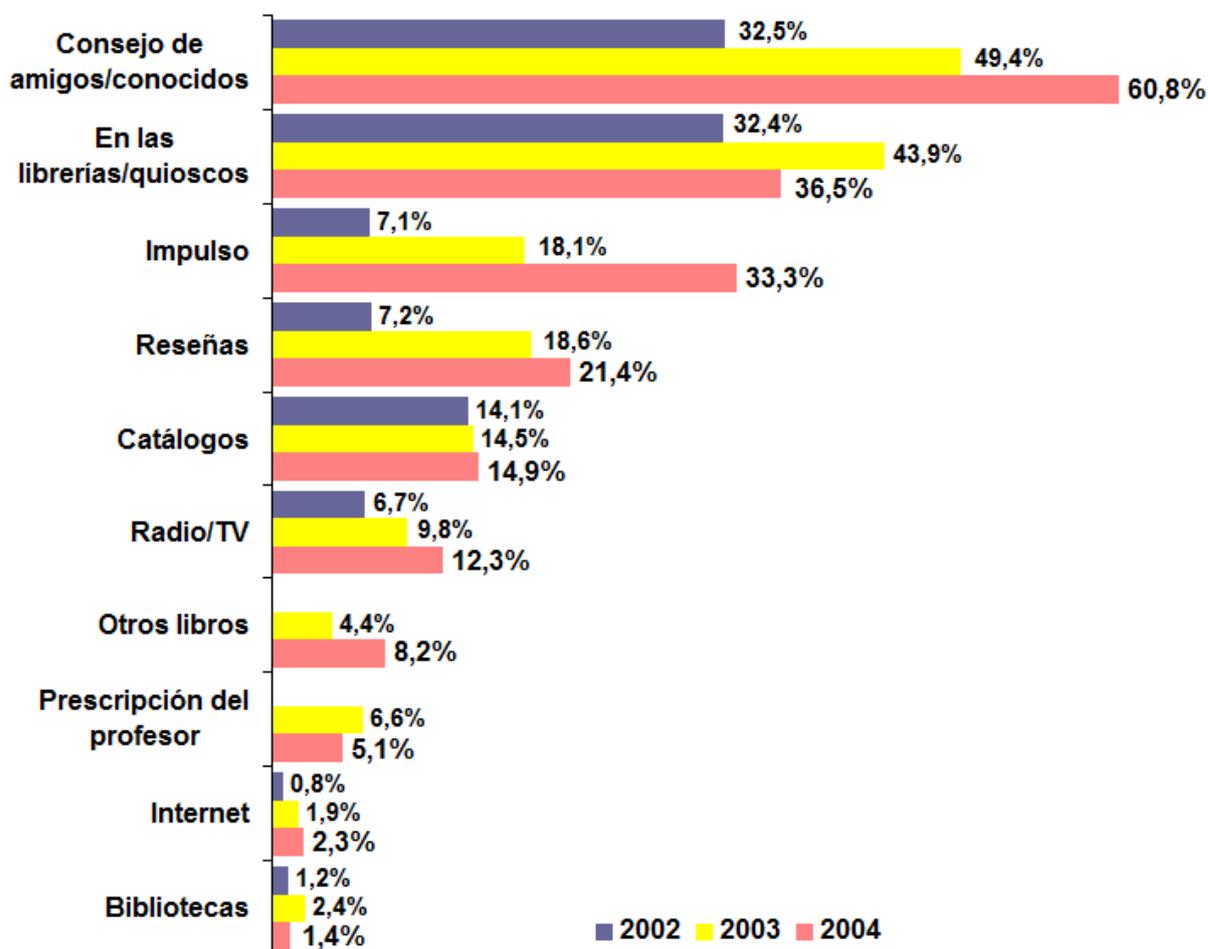


Fig. 2-16 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.

De nuevo, es complejo comparar encuestas cuando las preguntas son distintas o llevan a interpretación por parte del entrevistado. Por ejemplo, no se preguntaba lo mismo en las encuestas del CIS (Centro de Investigaciones Sociológicas de España, 2016) que en la realizada por Federación de Gremios de Editores de España en 2004, por lo que las respuestas pueden aportar conclusiones distintas.

Ante la pregunta “a la hora de elegir qué libro va a leer, ¿cuál es el criterio que tiene más en cuenta?” la mayor parte de la población española dice seguir su propio criterio (lo que no es lo mismo que la opción “impulso” de la encuesta de la Federación de Gremios de Editores de España de 2004) aunque con un cierta tendencia a bajar (variando del 69,5 al 60,7% de los encuestados) y donde sí se comprueba una tendencia al alza en confiar en la recomendación de familiares o amigos (del 18,5 al 26,5% en el caso del CIS), lo que se aprecia en ambos estudios.

Tabla 2-3 Criterio que tiene más en cuenta al elegir qué libro va a leer. Fuente: Centro de Investigaciones Sociológicas, 2016 (elaboración propia)

	1998	2003	2009	2014	2016
Su propio criterio	69,5	70,1	58,7	58,4	60,7
La recomendación de familiares o amigos/as	18,5	17,3	25,8	27,1	26,5
La orientación de profesores/as y/o especialistas	4	3,2	4,3	4	3,8
La información de medios de comunicación (crítica, publicidad)	6,6	7,1	8,3	5,8	3,5
La recomendación de librereros/as	1	-	-	1,4	1,7
Recomendaciones u opiniones de Internet	-	-	-	2,2	2,3

Otro	0	1,4	1,9	0,5	0,8
N.S.	-	-	0,1	0,2	0,4
N.C.	0,5	0,8	1,1	0,4	0,4

En el otro extremo, es interesante entender las motivaciones para la no lectura. Se descubren aquí motivaciones para no leer distintas para los lectores y para los no lectores. En el caso de los que sí son lectores, un análisis de la red social de GoodReads muestra las razones por las que sus usuarios (lectores más o menos frecuentes) abandonan una lectura (Cordón-García, 2013). Así, el 46.4% de los encuestados lo deja porque resulta aburrido, sin argumento el 8.5%, estúpido el 8.8%, o con una escritura “floja” el 18.8%. Resulta interesante descubrir que el 4.9% lo deja porque no le gusta el personaje principal (Goodreads, 2013).

2.2.7. De la lectura en papel a la lectura en libro electrónico

El estudio de los hábitos de lectura se complica con la llegada del libro electrónico y la lectura en diversos soportes, comenzando por la propia oferta editorial y los nuevos modelos de negocio que afectan a esos mismos hábitos. En 2011 el mercado editorial español era de 3.000 millones de euros, donde el 34,6% de las editoriales publicaron en soportes no papel (el 17% de los libros publicados en 2011, según el ISBN), alcanzando ventas por valor de 213 millones de euros, y aumentando en el año 2012 hasta convertirse en la cuarta parte de los libros registrados por la Agencia del ISBN.

Las editoriales comienzan a innovar, con modelos de negocio en la nube y sistemas de venta que rompen la cadena editorial habitual, lo que según la Confederación Española de Gremios y Asociaciones de Libreros (CEGAL) ha provocado el cierre de librerías en toda España, como en 2014, cuando se cerraron 912 y surgieron 226 (Manrique, 2015; Millán, 2017; Asociación de las Cámaras del Libro de España, 2015) Así, “el 8,5% de la venta de libros editados en formato digital fue directa desde la web de la editorial, el 80,2% a través de e-distribuidores, el 0,5% a través de librería y el 10,8% por otros canales” (Cordón-García et. al. 2014).

Esta tendencia en España contrasta con la descrita por Forman, Ghose y Goldfarb en 2008, según la cual, en aquel momento, cuando se abría una tienda de libros física, los consumidores (americanos) que estaban cerca de la zona, compraban menos libros a través del canal online. En el contexto español, al mercado editorial tradicional hay que sumar en la actualidad no sólo a las editoriales independientes, sino los nuevos modos de autoedición y autopublicación por parte de los autores, entre otras causas por la falta de transparencia en el reparto de las ganancias de los derechos de autores (Navarro, 2016). Esta tendencia hace que aumente la variedad de títulos publicados que escapan a los cálculos de las agrupaciones editoriales, y más difícil si cabe comparar el mercado del libro impreso con el digital. Tanto es así que Schavelzon (2015) afirma que no hay datos fidedignos, pues no hay forma de saber cuántos libros son leídos, sino solamente cuántos se venden, y eso sólo al circunscribirse a la lectura de libros.

En este sentido, a las cifras oficiales en España, se sumarían las de alguna plataforma de venta online, como Amazon, que certifica cómo la variedad de títulos vendidos ha ido aumentando con el tiempo (Brynjolfsson, Hu y Smith, 2011), siendo los libros de poca demanda una parte muy importante de las ventas (36'7% en 2008) cinco veces más que en el año 2000, efecto denominado Larga Cola o *Long tail*, debido a que sigue una distribución de tipo *power law* o

ley de potencias, en las que unos pocos ítems reciben la mayor atención, y muchos ítems reciben cada uno muy poca, pero que sin embargo, en multitud de casos, la suma de esos pocos resulta muy grande, o mayor que la de los pocos ítems que reciben mucha atención.

En palabras de Javier Celaya (2017) “las ventas de contenidos digitales no paran de crecer, representando en 2016 el 35% del total de las ventas en el Reino Unido; el problema para los editores es que los lectores están comprando los libros directamente a sus autores. En 2012 las ventas de libros autoeditados solo representaban el 12% del total de las unidades vendidas. En 2016 la autoedición representa ya el 22% de las ventas. Estos datos indican claramente que la cuota de mercado de las editoriales no para de empequeñecerse año tras año, al contrario que la de los contenidos autoeditados, que no para de crecer.”

Esta evolución al alza en el consumo de libros electrónicos se constata especialmente en el caso de las comunidades universitarias (Hernández Salazar, Nicholas y Rowland, 2009) pero también en el mercado global. Baye, Santos y Wildenbeest (2013) estudiaron la evolución de las búsquedas sobre lectores digitales (e-readers en el artículo) en el período 2004-2013 para el mercado estadounidense, con un espectacular crecimiento para términos “kindle” o “nook”, lo que sugiere que los americanos identificaron más el dispositivo por la marca que por su tipología.

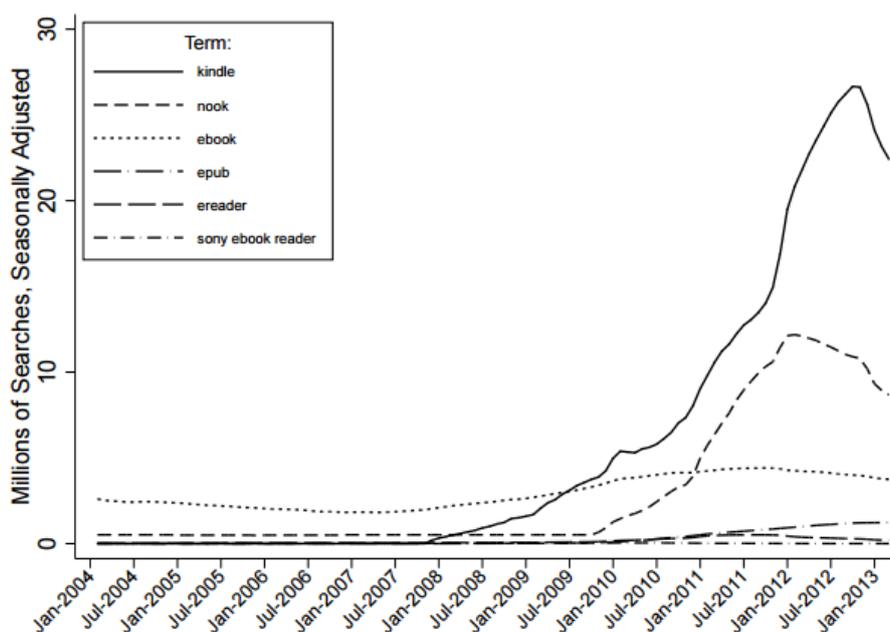


Fig. 2-17 Tipología lectora según el hábito lector (%) Fuente: Yubero y Larrañaga, 2015.

Pero sin duda al estudio de la lectura en el libro digital le falta recorrido histórico, puesto que llevamos cientos de años leyendo libros en soporte papel, y apenas unos pocos años de consumo digital, y ahora es frecuente todavía ver estudios sobre consumo de libro digital que trabajan con usuarios que están haciendo su propia transformación digital, pasando de un formato a otro, o conviviendo entre ambos (Cordón-García y Fernández, 2015).

En palabras de José Antonio Millán en la obra *La Lectura* (Basanta Reyes, 2010) “se refieren a personas forjadas en la lectura tradicional, y que llevan (en el mejor de los casos) dos lustros inmersos en la lectura digital.” Tanto es así, que en las estadísticas de *CULTURABase* (2017) del Ministerio de Cultura de España, el libro electrónico sólo se contabiliza como tal a partir de 2010, convirtiéndose en 2014 en el 5,93% de los fondos de las bibliotecas computadas en el estudio español, aumentado desde el 3,39% en 2010, así como los juegos (entendidos como un *medio* producto cultural de uso en las bibliotecas) aunque en menor medida, aumentando desde el 0,05% al 0,07% de los fondos bibliotecarios.

Tabla 2-4 Fondos por clase de fondos en Bibliotecas españolas. Fuente: *CULTURABase*, 201 (Elaboración propia)

	2010	2012	2014
TOTAL	183.221.933	205.454.248	216.609.833
Libros electrónicos	6.214.016	9.265.978	12.838.609
Juegos	91.284	134.738	145.856

Otro tanto ocurre en el CIS, que no incluye preguntas sobre el libro electrónico hasta 2014 (Centro de Investigaciones Sociológicas de España, 2014) cuando es tarde para comprender su evolución inicial, porque ante la pregunta “¿ha oído hablar o sabe Ud. qué es el libro electrónico o *e-book*? el 84,9% de los encuestados contesta que sí, frente al 15,1%, que afirma que es la primera noticia que tiene.

Eso sí, sólo el 15,3% de los que conocen lo que es un libro electrónico, han leído parcial o totalmente un libro en versión digital “en bastantes ocasiones”, mientras que han leído e-books “alguna vez” el 18,4% y “nunca” el 66,2%. Finalmente, el 33,3% de los conocedores de la lectura en soporte digital en 2014, opina que “la mayor parte de los libros serán electrónicos y habrá muy pocos libros impresos en papel”, mientras que el 42,0% piensa que “en el futuro ambos convivirán “.

2.2.8. Leer en internet

Si en el pasado no se contemplaba ni el libro electrónico en las encuestas, desde principio de siglo (Ministerio de Educación, Cultura y Deporte y Fundación Autor, 2003) se incorpora Internet a las mismas, para tener mayor contexto de la realidad de las prácticas de consumo de contenidos e información por parte del público.

Según Sáinz Peña (2015) en el año 2015 “un millón más de personas que el año anterior se conecta regularmente a Internet en nuestro país, alcanzando la cifra de 27,15 millones en la franja de edades comprendidas entre los 16 y los 74 años. Pero además lo hacen de forma más intensiva, ya que 22,2 millones de españoles (el 81,7 por ciento de los usuarios), se conectan todos los días a Internet. Es un millón y medio más que un año antes. Es decir, el 64,35 por ciento de los españoles se conectan todos los días a la Red; A destacar en este crecimiento, la franja de edad entre 55 y 64 años, que ha aumentado 6 puntos porcentuales revertiendo la brecha generacional que existía.”

Internet da pie a distintas prácticas lectoras, y también a distintos patrones de difusión de la lectura. En la era pre-internet, según autores como Hervé Renard, en su estudio entre prestatarios franceses, la viralidad se producía (y se produce) cuando un lector dejaba, prestaba o regalaba un libro, y la recomendación se discutía sobre gustos literarios en cafés y tertulias (Cordón-García, 2006)

Cuando usted ha leído un libro prestado que le gusta que hace después?	CONJUNTO DE PRESTATARIOS	PRESTATARIOS COMPRADORES	PRESTATARIOS O COMPRADORES
*Se lo aconsejo a un pariente o a un amigo	72,9	72,8	73
*Compro el libro para tenerlo	20,9	23,8	12,8
*Compro el libro para regalarlo	9,3	10,8	5,1
*Compro otro libro del mismo autor	13,1	14,8	8,3
*Compro otro libro del mismo genero	3,4	4,2	1,2
*Pido prestado un libro del mismo autor	69,8	69,8	69,8
*Pido prestado un libro del mismo genero	37	34,2	44,5
*Nada de esto	6,5	6	7,9

Fig. 2-18 Criterio que tiene más en cuenta al elegir qué libro va a leer. Fuente: Centro de Investigaciones Sociológicas, 2016 (elaboración propia)

Sin embargo, en la era de internet existen otras posibilidades, como:

- El fenómeno de los booktubers (García et.al., 2015)
- los likes en las redes sociales (Sánchez Ramos, 2016),
- los reviews en los marketplaces virtuales tipo Amazon, las conversaciones entre blogueros o las plataformas de encuentro virtual entre escritores y lectores, como Librarything o Goodreads (Alonso-Arévalo, Cordón-García, 2011),
- los efectos de algoritmos sociales de recomendación o la viralización vía plataformas *Peer To Peer* (Cordón-García, Gómez-Díaz y Alonso-Arévalo, 2011)

Así, según el Pew Research Center (2012) los propietarios estadounidenses de dispositivos de lectura electrónica son más propensos a pedir recomendaciones de libros de personas que conocían (81%) de librerías online, o de otros sitios web (56%).

Un problema de investigación es la separación del estudio de lectura entre el libro en papel y el digital, pues quienes usan dispositivos electrónicos también consumen libros en otros formatos (lector digital, papel, móvil, *tablet* e incluso audiolibros). Así, en 2015 el 83 por ciento

de los usuarios accede a Internet usando teléfonos inteligentes o *smartphones*, por encima del ordenador como dispositivo de acceso a la Red. El 69,6 por ciento para buscar información sobre bienes y servicios. (Sáinz, 2015) En otros países la tendencia es similar. El estudio de 2012 del Pew Research Center antes comentado, revela que los lectores digitales leían por entonces un promedio de 24 libros y los que sólo leen material impreso, 12.

2.2.9. Conclusiones parciales

Se ha visto a lo largo del apartado, cómo se ha abordado el análisis de los hábitos lectores tradicionalmente. Como disciplina, ha visto evolucionar tanto el público sobre el que se investigaba como los mismos conceptos de lectura y libro, y aún más en el caso de las bibliotecas, donde se han incorporado otros *media* o formatos de producción cultural (revistas, videojuegos, películas, internet...). Sin embargo, las metodologías generales han permanecido estables, destacando la encuesta y la entrevista como las técnicas mayoritarias. Esta dinámica cambiante en el objeto de estudio, ha desembocado en la proliferación de distintos problemas metodológicos, como:

- Dificultad de establecer definiciones de conceptos a estudiar, como lectura, libro, motivación, etc. (Pouliot, 2009; Koch, 2017)
- Diversidad de objetos de estudio (Federación de Gremios de Editores de España, 2004; *CULTURABase* 2017)
- Imposibilidad técnica de establecer el universo de estudio, en cuanto a productos culturales en sus diversos formatos (Schavelzon, 2015)
- Variedad en las segmentaciones de público al que investigar (Ramírez Leyva, 2008; Yubero, 2009; González, 2017; Ministerio de Educación, Cultura y Deporte, 2015)
- Variaciones en las preguntas a los encuestados a lo largo de los años, en ocasiones para un mismo estudio, como en CIS (1998 2003, 2009, 2014, 2016)
- Dificultad de aclarar en el encuestado el sentido de la pregunta.
- Diversidad en las preguntas entre estudios distintos (Gamboa y Reina, 2006; Gil, 2015; Quezada, 2016)
- Variedad en las motivaciones y contextos socioculturales que influyen en las mediciones (Pew Research Center, 2012; Cordon-García, Gómez-Díaz y Alonso-Arévalo, 2011; García et.al., 2015; Sáinz, 2015)
- Posibles intereses no científicos en el diseño de encuestas (Acevedo y Mercado Lafertte, 2013),
- Sesgos propios de las metodologías basadas en encuestas y entrevistas, incluso cuando los intereses son meramente científicos o técnicos (Pouliot, 2009; Yubero y Larrañaga, 2015).

La pregunta subsiguiente es si, en un contexto en el que la digitalización de la sociedad española es patente (Ministerio de Educación, Cultura y Deporte, 2015; Sainz, 2016), puede decirnos algo nuevo el análisis de los hábitos de búsqueda de los usuarios de Internet, algo que no digan ya las encuestas. Metodológicamente será interesante saber si puede mostrar este tipo de análisis, si los usuarios buscan lectura en un formato u otro, si puede ayudar a entender cómo se expresan los usuarios de Internet, cuando intentan encontrar qué leer, y si

puede servir esa información de utilidad para los distintos actores del sector del libro, y en particular a las bibliotecas.

En un mundo dominado por el consumo de información a través de Internet (Serrano-Cobos, 2016), hacen faltas metodologías y técnicas que trabajen con las posibilidades tecnológicas y las pautas de comportamiento del público en Internet. En el próximo apartado se tratará de conocer qué tipos de investigación se llevan a cabo para conocer esas pautas de comportamiento, en las que el consumo de la tecnología y los interfaces son claves en el uso y disfrute de productos culturales, de forma que se contextualice cómo el análisis de los hábitos de búsqueda en motores de búsqueda puede ser clave para aportar dimensiones de análisis no investigadas por las metodologías clásicas de estudio de hábitos lectores. Pues allá donde las encuestas nos dirán qué dicen los usuarios buscar, el análisis de logs de motores de búsqueda de Internet nos podrá decir qué y cómo buscan en realidad.

2.3. INFORMATION SEEKING

Como se especifica en la introducción y objetivos, esta investigación está orientada a aportar información que ayude a la toma de decisiones de distintos actores del sector del libro, especialmente en lo tocante al sector editorial y al sector bibliotecario. En el sector editorial, el cambio de formatos de uso de los productos culturales y de los canales de comunicación y recomendación, en particular Internet, provocan que este nuevo contexto deba ser estudiado con nuevos métodos. Asimismo, en el sector bibliotecario, esas mismas necesidades llevan a la inclusión de nuevas metodologías que complementen los estudios clásicos en lo referente a usuarios.

Pero mientras que en el mundo editorial no se ha observado ningún gran cambio metodológico, en las bibliotecas y servicios de información la incorporación de otras metodologías viene de lejos, aunque de forma gradual, pues se comienzan a utilizar otras técnicas de investigación y a contemplar distintos paradigmas, desde el momento en el que las colecciones empiezan a consultarse a través de catálogos a través de interfaces (OPACs) y con mayor motivo, a través de Internet.

Se verá en este apartado cómo ese interés de los investigadores en recuperación y gestión de información, en particular a través de interfaces y motores de búsqueda de Internet, ha generado una serie de líneas de investigación, de las cuales bebe esta tesis doctoral, que puede aportar una visión diferente y enriquecedora del conocimiento del usuario, para ambos sectores, el editorial y el bibliotecario (Given, Julien y Case, 2012).

Para ello se introducirá el estudio de los diferentes aspectos del OPAC que competen al usuario, como interfaz de consulta de catálogos en las bibliotecas. Su análisis y evaluación, primero localmente, después integrado en sedes web, llevará a incorporar la Analítica Digital para la medición de parámetros de uso. Y más allá, cuando los científicos sean conscientes de que el análisis de una sede web no se puede concebir sin completarlo con el análisis de su contexto dentro de Internet, conformado por sedes web que compiten por la atención del usuario, la Cibermetría llegará para ampliar la panoplia de recursos de investigación, con un enfoque más abierto y global.

Se dará cuenta también de la construcción del acervo teórico que han aportado distintos investigadores para solucionar los problemas de recuperación de información primero en bases de datos documentales y catálogos locales, luego en sedes web y más allá, en motores

de búsqueda de Internet, donde hoy día se libra la última batalla por conseguir atraer al usuario. Porque, en lo que respecta al canal Internet como forma de acercamiento del lector a la obra, si un usuario no encuentra un producto cultural, no puede utilizarlo.

Así, se hará una revisión de los modelos de conducta de búsqueda de información (concepto en inglés acuñado como *Information Seeking Behaviour*) en el marco del paradigma cognitivo de la Recuperación de Información. Dentro de éste, se hará hincapié en líneas de investigación en materia de análisis del comportamiento del usuario al utilizar motores de búsqueda, tanto buscadores internos de sedes web, como buscadores de Internet (p.ej. *Google*) dentro de la subdisciplina denominada *User Search Behaviour*. Esta última será la que ayudará a dar marco a la metodología utilizada en esta tesis doctoral.

2.3.1. Introducción

La necesidad de combinar nuevas metodologías a las clásicas encuestas y entrevistas de los estudios de usuarios en bibliotecas, viene dada por la transición tecnológica de las mismas. Los servicios de información lo son en cuanto respuesta a unas necesidades, necesidades que han cambiado con la tecnología. El *Manual de Biblioteconomía* de Gómez-Hernández (1996) indica que “la biblioteca ha de orientarse a los usuarios” y establece que “el punto de partida es, lógicamente, el estudio y conocimiento de los usuarios”, lo que apostilla Eric Lease Morgan (Serrano-Cobos, 2010, 212) que asegura de las bibliotecas que “*if they focus too much on books, then their future is less positive*” (lo que se podría traducir como “si se enfocan demasiado en los libros, entonces su futuro es menos positivo”).

Por su parte Núñez (1992) establece en una guía metodológica para el estudio de las necesidades, “un control total de usuarios, su categorización o segmentación y jerarquización o priorización; el estudio de necesidades de formación y de información de usuarios individuales o grupales, y su actualización sistemática” lo que evoluciona más tarde (Núñez, 2004) en una “metodología integral para la determinación y la satisfacción dinámica de las necesidades de formación e información en las organizaciones y comunidades (AMIGA)” que aplica luego a las bibliotecas de salud (Fernandez y Núñez, 2007)

	F U E N T E S			T É C N I C A S
	D O C U M E N T A L E S	N O D O C U M E N T A L E S		
		P E R S O N A L E S	A C T I V I D A D E S	
I) RELATIVAS A LOS PROBLEMAS, Y A LAS ACTIVIDADES QUE EL U.C. DEBE DESARROLLAR EN ELLOS: <ul style="list-style-type: none"> • TEMÁTICA • TIPO DE ACTIVIDAD (ESTRUCTURA DE PASOS) • CONDICIONES (TECNOLÓGICAS, ORGANIZATIVAS, MATERIALES, GEOGRÁFICAS, SOCIALES, ETC.) 	<ul style="list-style-type: none"> • DE ESTRATEGIA O POLÍTICAS (mundiales, regionales, nacionales, territoriales, de la organización, etc.) • DE PLANIFICACIÓN (planes o programas de trabajo, de investigación, de estudio, etc.) • ORGANIZATIVOS 	<ul style="list-style-type: none"> • DIRECTIVOS • EXPERTOS (nacionales, extranjeros, de la propia organización, etc.) • USUARIOS EN ESTUDIO 	<ul style="list-style-type: none"> • CONSEJOS DE DIRECCIÓN, TÉCNICOS, DE PRODUCCIÓN, CIENTÍFICOS, ETC. • REUNIONES (donde se toman decisiones sobre estas variables) • EVENTOS (Nacionales e internacionales, etc.) 	(SEGUN LA FUENTE): ANÁLISIS DOCUMENTAL OBSERVACIÓN DIRECTA Y PARTICIPANTE ENTREVISTA SOCIOMÉTRICAS (PARA EL TRABAJO EN GRUPOS)
II) RELATIVAS A LOS RECURSOS INFORMATIVOS DISPONIBLES: <ul style="list-style-type: none"> • FONDOS ACCESIBLES • FUENTES • TECNOLOGÍA DISPONIBLE • RECURSOS HUMANOS (CANTIDAD, PREPARACIÓN Y VALORES) • CANALES DE COMUNICACIÓN • RECURSOS FINANCIEROS • SERVICIOS EXISTENTES DISPONIBLES • OTRAS 	<ul style="list-style-type: none"> • DIRECTORIOS, CATALOGOS. • PUBLICACIONES PRIMARIAS, SECUNDARIAS O TERCIARIAS • INVENTARIOS DE MEDIOS BÁSICOS • TECNOLÓGICAS Y DE SISTEMA. • JURÍDICAS Y DE REGLAMENTO. • DE PRESUPUESTOS, COSTOS, GASTOS, SALDOS, ETC. • LISTA DE OFERTA DE SERVICIOS • REGISTROS DE PERSONAL • OTRAS 	<ul style="list-style-type: none"> • DIRECTIVOS (de información) • EXPERTOS (Nacionales o Extranjeros, en cualquier aspecto de los recursos y la actividad informacional) • USUARIOS EN ESTUDIO 	<ul style="list-style-type: none"> • CONSEJOS DE DIRECCIÓN, TÉCNICOS, DE PRODUCCIÓN, CIENTÍFICOS, ETC. (de la entidad de información) • REUNIONES (donde se toman decisiones sobre estas variables) • EVENTOS (Nacionales, internacionales, etc., de temas que influyen en la gestión de información, la comunicación, el conocimiento, la inteligencia o el aprendizaje) 	OTRAS, PARTICIPATIVAS, DE SOLUCIÓN DE PROBLEMAS Y DE DESARROLLO DE LA CREATIVIDAD INFORMÉTRICAS, PARA EL PROCESAMIENTO DE INFORMACIÓN OBTENIDA POR LAS TÉCNICAS ANTERIORES
III) RELATIVAS A LAS CARACTERÍSTICAS SOCIO-PSICOLÓGICAS DE LOS U.C.: <ul style="list-style-type: none"> • EXPERTICIDAD EN LA(S) TEMÁTICA(S) • EXPERTICIDAD EN EL TIPO DE ACTIVIDAD • DOMINIO DE IDIOMAS • HÁBITOS DE USO DE LA INFORMACIÓN • VELOCIDAD DE LECTURA O PROCESAMIENTO DE LA INFORMACIÓN (Aproximado) • TIEMPO QUE LE DEDICA A LA LECTURA Y PROCESAMIENTO DE LA INFORMACIÓN • FUNCIONES QUE DESEMPEÑA EN EL GRUPO • OTRAS 	<ul style="list-style-type: none"> • SOBRE EL USUARIO (Expediente laboral, científico, académico, etc., currículum vitae, Evaluaciones Técnicas, y otras) • GENERADOS POR EL USUARIO (artículos, ponencias, informes, obras plásticas, exámenes, cartas, etc.) • SOBRE RESULTADOS OBTENIDOS POR EL USUARIO (Premios, críticas, evaluaciones de su trabajo, etc.) 	<ul style="list-style-type: none"> • DIRECTIVOS (del usuario) • RECEPTORES DE LA ACTIVIDAD DEL USUARIO (pacientes, estudiantes, clientes, etc.) • COMPAÑEROS DE TRABAJO DEL USUARIO 	<ul style="list-style-type: none"> • CONSEJOS, REUNIONES O EVENTOS DONDE PARTICIPE EL USUARIO, DESEMPEÑANDO ALGÚN ROL. 	

Fig. 2-19 Tipos de variables, fuentes y técnicas en los Estudios de Necesidades (Núñez, 2004)

Autores como González-Teruel y Abad-García (2007) e incluso tardíos como Villaseñor-Rodríguez (2017) han estudiado la literatura sobre estudios de usuarios producida en español durante el período 1990–2004, descubriendo que se ha incrementado el número de publicaciones sobre la temática, predominando los estudios que se llevan a cabo en el contexto universitario, y que se tienen en cuenta aspectos relacionados con la tecnología disponible y el tipo de actividad de los usuarios.

En un principio, los estudios de usuarios de biblioteca, al igual que los de los hábitos lectores, al ser realizados en ocasiones por los propios bibliotecarios para autoevaluar sus propias colecciones, o por investigadores sin acceso a grandes cantidades de datos, suelen ser cualitativos, en forma de encuesta (Torre, 2009), y no tanto como observación rigurosa, sistemática y global de sus procesos de búsqueda (González Teruel, 2005, 2011) Pero sí se observan otras que se centran en el estudio de la búsqueda de la información y uso de la biblioteca por parte de sus usuarios. Así, Albornoz et. al. (2007) tuvieron como objetivo establecer los perfiles de comportamiento de los investigadores y docentes universitarios, identificando grupos a los que entrevistar.

	No.	%
One aspect		
User attitude to a service, system or source of information	1	1.1
Information needs	5	5.6
Actual requests made of a service	22	24.7
Use of a service, system or information source	24	27.0
Use of information	9	10.1
Satisfaction	2	2.2
More than one aspect		
Attitude to a service + User of a service, system or information source	1	1.1
Information needs + Use of a service, system or information source	3	3.4
Actual requests made of a service + Use of a service, system or information source	1	1.1
Use of a service, system or information source + Use of information	1	1.1
Use of a service, system or information source + Satisfaction	20	22.5
Total	89	100

Fig. 2-20 Aspectos del proceso de búsqueda de la información estudios en la literatura científica en español, período 1990–2004 (González-Teruel, 2007)

Bajando a los aspectos tecnológicos de la biblioteca, un aspecto a valorar en una biblioteca, cada vez más tecnificada, es el catálogo automatizado u OPAC (*Online public access catalog*). El OPAC como herramienta tecnológica de consulta e interacción con la colección bibliotecaria ha ido evolucionando, desde los primeros, desarrollados en la Universidad Estatal de Ohio en 1975 (Borgman, 1986), al OPAC extendido (Sellés-Carot y Serrano-Cobos, 2011)

Asimismo, en algunos casos, el OPAC ha ampliado sus horizontes hacia las herramientas de descubrimiento o *discovery tools*, donde habrá que tener en cuenta su integración con otros catálogos mediante el uso de datos enlazados (*Linked Data*) e interoperables (Sellés-Carot, Orduña y Serrano-Cobos, 2010)

Cuando un OPAC no sólo consulta la propia colección, sino que permite la búsqueda federada de información, la experiencia del usuario se encaja en un ente mayor que la propia colección de la biblioteca desde la que consulta, por lo que el cuadro de mando de esa evaluación puede necesitar un espectro mayor de indicadores basados en datos enlazados (Hallo, 2016)

La evaluación de OPACs considera distintos aspectos desde los principios del uso de estas herramientas de consulta del catálogo:

- La evaluación de la colección (Villén-Rueda, 2006; Selgas Gutiérrez, 2004).
- La evaluación de la recuperación de información (Tillotson, 1994, Hildreth, 1997).
- La evaluación de la interfaz de recuperación de información (Instone, 1997; Hansen, 1998).

- La evaluación de la interacción del usuario (Belkin, 1991; Hancock-Beaulieu, 1993).

La evaluación de la colección a través del OPAC permite conocer si ésta responde a las necesidades de los usuarios (si hay contenidos que respondan a la necesidad, si están disponibles o no, si existen en el formato que necesita el usuario o no). La evaluación de la recuperación de la información y la evaluación de la interfaz buscan optimizar la interacción con el OPAC, mediante el análisis de los usuarios que utilizan esos subsistemas o pueden ser más proclives a usarlos, estudiando sus conductas de búsqueda y navegación para dar sentido a la evaluación (Krug, 2000).

Y aquí entran en juego distintas disciplinas, que conjugan sus métodos para aportar luz nueva. Hay técnicas de investigación provenientes de la disciplina de *Human-Computer Interaction* o Interacción Persona-Ordenador que se pueden aplicar a este tipo de evaluaciones (Dix et. al., 1993; Floria, 2000, Shtykh, 2011; Olson et. al., 2014; Shneiderman et. al., 2016), como las técnicas de tipo observacional y experimental, estudios de laboratorio, etnográficos, por paneles, estudios de campo, *A/B testing*, o estudios de *logs*, (Dumais et al., 2014). En esta investigación se considerará el uso de estudios de tipo observacional basado en logs de búsqueda.

	Observational	Experimental
Lab Studies <i>Controlled interpretation of behavior with detailed instrumentation</i>	In-lab behavior observations	In-lab controlled tasks, comparison of systems
Field Studies <i>In the wild, ability to probe for detail</i>	Ethnography, case studies, panels (e.g., Nielsen)	Clinical trials and field tests
Log Studies <i>In the wild, little explicit feedback but lots of implicit signals</i>	Logs from a single system	A/B testing of alternative systems or algorithms

Fig. 2-21 Tipos de estudios en HCI. Fuente: Dumais et al., 2014.

El análisis de los logs de uso de OPACs de distintas empresas, tecnologías y versiones centra buena parte de las investigaciones en esta línea, donde podemos diferenciar desde los inicios de este tipo de estudios, distintos tipos de análisis (Boock, 2007):

- Estudio de patrones de uso de los usuarios (*user search patterns*) para confirmar si los usuarios encontraban muchos de los mismos problemas identificados en los estudios de comportamiento realizados sobre otros tipos de sistemas de recuperación bibliográfica (Borgman, 1986).
- Análisis de probabilidad de uso mediante cadenas de Markov, basado en los archivos logs (Tolle, 1983).
- Estudios de diseño de interacción de páginas de resultados de búsqueda centrado en el usuario según tests de usuarios o usability testing para iterar diseños de interacción con el fin de mejorar la experiencia del usuario (Blecic et. al. 1998).
- El testeo de nuevos algoritmos para mejorar la recuperación de la información, desde la búsqueda realizada en distintos idiomas (Bosca y Dini, 2009).

Millsap y Ferl (1993) identificaron indicadores que han seguido usándose durante los últimos 25 años, como el número de búsquedas, bases de datos buscadas, número de recuperaciones, comandos usados, problemas con los términos de búsqueda, y en general la efectividad de la búsqueda.

En español, tradicionalmente se han publicado más estudios sobre ficheros de autoridades de materia y las interfaces, desde el punto de vista de la usabilidad de la herramienta de consulta concreta (Maniega-Legarda, 2002) y una vez más, en el entorno universitario (Travieso Rodríguez, Alonso-Arévalo y Vivancos Sevilla, 2007), comparándose distintos programas del mercado (Játiva-Miralles, 2004), pero según Villén-Rueda (2006) son estudios demasiado escasos como para ayudar en la mejora de la indización y de la recuperación por materias de los mismos.

El análisis de los *logs* de un servidor también se ha usado en el ámbito de los museos (Peacock, 2002) permitiendo contabilizar cuantitativamente las respuestas a cuatro preguntas básicas de cualquier experiencia de usuario al utilizar sistemas de recuperación de información:

- (1) ¿Puedo encontrarlo?
- (2) ¿Funciona?
- (3) ¿Tiene lo que estoy buscando?
- (4) ¿Satisface mis necesidades?

A este respecto, cabe apuntar que será menester combinar la investigación del *log* de búsquedas del sistema de recuperación interno y de las expresiones de búsqueda que lleven

usuarios desde el exterior hasta la sede web bibliotecaria, con las realizadas en motores de búsqueda en Internet, como Google, puesto que el lenguaje natural puede variar (Serrano-Cobos, 2009; Delgado et. al., 2016).

Pero este análisis del OPAC, que hoy día suele formar parte de la sede web de la biblioteca, resulta inconcluso si no se integra con la misma, por lo que la familia de factores a considerar aumenta (Corda y Viñas, 2015) La evaluación, no sólo del OPAC, si no del sitio web, exigirá por un lado de un análisis de la usabilidad de la sede web bibliotecaria, por ejemplo en el caso del sitio web de la Biblioteca del Congreso de los Estados Unidos (Krug y Rosenfeld, 2004) tanto en la fase de prediseño como de postdiseño de la interacción de la misma; y por otro, de la cuantificación de la navegación de los usuarios que lleguen a la sede web bibliotecaria, para lo que será de gran ayuda la Analítica Digital.

2.3.2. Algunos modelos de *Information Seeking Behaviour*

La necesidad de información ha sido durante mucho tiempo un objeto de investigación dentro del área temática de la ciencia de la información, lo que incluye toda una familia de disciplinas relacionadas, incluyendo los estudios de usuarios. Los primeros análisis de este tipo ocurrieron en la década de 1940. Según Wilson (1994; 2000, 2010), el primer uso de "necesidad de información" ocurrió en la *International Conference on Scientific Information* (1959) en la que se hacía referencia a las necesidades de recuperación de información de los científicos (Cole, 2014) y cómo se comportaban ante esas necesidades. Ahí se enmarca el paradigma cognitivo en la Recuperación de Información (Ingwersen, 1992, 1996 o Kuhlthau, 1999, 2005), puesto que añade al paradigma físico o algorítmico, "la relación del usuario con el sistema de información" (Cruz Gil, 2014)

El "comportamiento informacional" (*Information Behaviour*) podría considerarse así el término general para una serie de disciplinas científicas anidadas, entre las que se incluye el subcampo del "comportamiento de descubrimiento de información" (*Information Seeking Behaviour*), que se ocuparía de estudiar la variedad de métodos que las personas emplean para descubrir y acceder a los recursos de información. Éste, a su vez, incluiría el "análisis de la conducta de búsqueda de información" (*Information Search Behaviour*, utilizándose *Behavior* en la forma americana) donde "búsqueda" (*search*) se definiría como una rama de la búsqueda de información que estudia la interacción con sistemas informáticos de información, en concreto sistemas de recuperación de información. (Wilson, 1999).

Aunque según González-Teruel et. al. (2015) en la disciplina hay un núcleo con pocos autores, que son quienes han desarrollado las teorías subyacentes, Fisher, Erdelez y McKechnie (2005) identifican 85 investigadores de 10 países como aquellos que aportan investigaciones de mayor calado sobre el problema de la conducta de búsqueda de la información, y en Google Scholar se podían localizar 4.490 publicaciones sobre la disciplina, hasta 2018 inclusive⁵. Puede considerarse una rama del estudio de la Interacción Persona-Ordenador (*Human Computer Interaction* o HCI), en sí muy compleja, donde las distintas tecnologías cambian constantemente, lo que puede incidir sobre la conducta de los usuarios, y las técnicas de

⁵ Fuente:

https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&authuser=1&q=%22information+seeking+behavior%22+or+%22information+seeking+behaviour%22&btnG=

investigación son múltiples, siendo difícil comparar estudios y modelos de *Information Seeking Behaviour* (Martzoukou, 2005)

Se reseñan aquí algunos de los modelos teóricos más conocidos entre los estudiosos de la materia (Choo, Detlor y Turnbull, 2000; Fisher, Erdelez y McKechnie, 2005; Tramullas y Garrido, 2011; Núñez y Zayas, 2016;) lo que permitirá contextualizar las aportaciones que se puedan hacer a partir del estado de la ciencia en esta área concreta desde distintas áreas de conocimiento.

2.3.2.1. Modelo de Wilson

Wilson se centra en estudiar el comportamiento humano de búsqueda de información, en lugar de enfocarse en el sistema de recuperación como tal, como algoritmo. El modelo de Wilson ha ido cambiando a lo largo del tiempo. En 1981, identificaba factores relacionados con las tareas asociadas a la búsqueda de información, así como las dificultades de su consecución exitosa. Entre esos, destacó los factores personales, fisiológicos y cognitivos (Wilson, 1981).

También identificó al sistema como un intermediario, no como un fin en sí mismo del proceso de recuperación de información, en el que los recursos de información recuperados por el sistema, son evaluados por el propio usuario. Más tarde, en 1999, integra otras áreas de investigación para crear un marco global en el que pone a la disciplina de la Conducta de Información (*Information Behavior*) como disciplina paraguas, y dentro incorpora disciplinas relacionadas pero separadas, como *Information-Seeking Behavior* e *Information Searching Behavior* (Wilson, 1999)

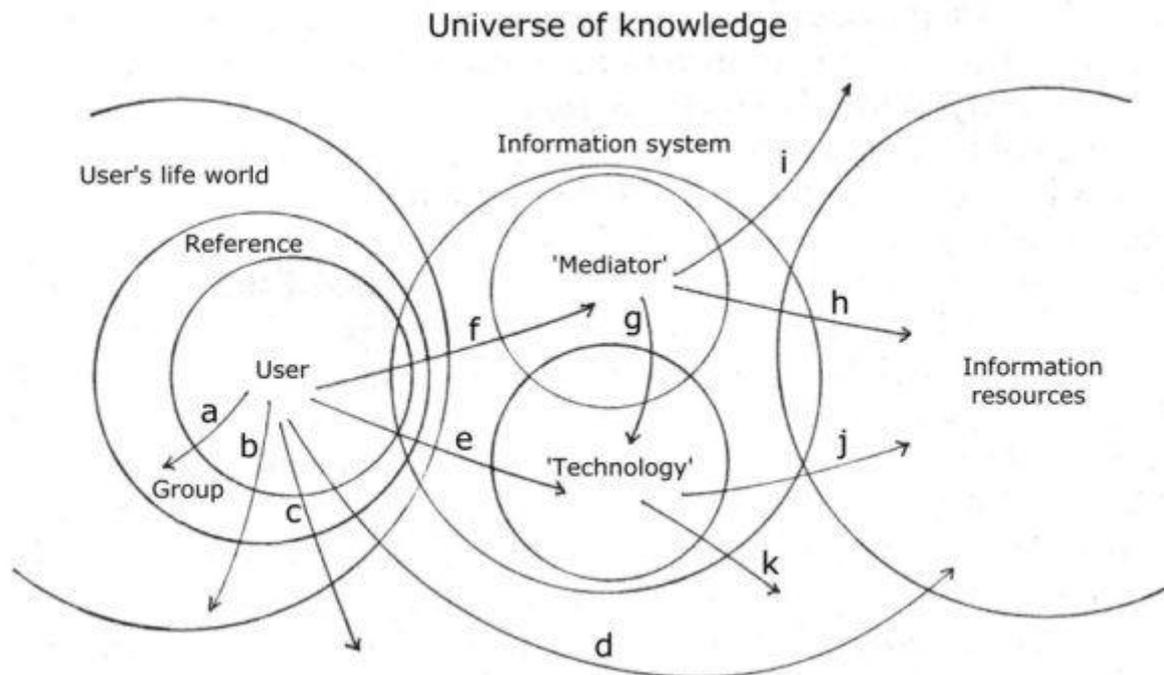


Fig. 2-22 Relaciones entre usuario, sistema de información y recursos de información. Fuente: Wilson, 1981.

2.3.2.2. Modelo de Belkin

Uno de los primeros autores en estudiar los retos ante los que se enfrentaban los usuarios en procesos de recuperación de información mediante el uso de términos de búsqueda en un motor de búsqueda fue Nicholas J. Belkin, quien en su artículo seminal *Anomalous states of knowledge as a basis for information retrieval* (Belkin, 1980) observó de que en ciertos casos, los usuarios de sistemas de búsqueda no podían formular sus necesidades de información con la precisión que necesitaba el sistema de recuperación de información, ni en su lenguaje de consulta específico.

Ante la falta de esos conocimientos específicos para interrogar al sistema, era más adecuado intentar describir el “estado anómalo de conocimiento” de un usuario que pedirle a éste que especificara su necesidad como una solicitud al motor de búsqueda, para llegar a lo que Belkin denominaba en 1980 y 1982 “coherencia”, tras resolver su Estado Anómalo de Conocimiento (*ASK* o *Anomalous State of Knowledge*) cuyo trabajo completo sobre este tema fue sintetizado por Achananuparp (2006).

La influencia de su estudio viene del interés por determinar por qué se realiza una búsqueda, lo que más adelante se materializará en el estudio de intenciones de búsqueda (*query intents*) como se verá más adelante, siendo parte integral de esta tesis doctoral.

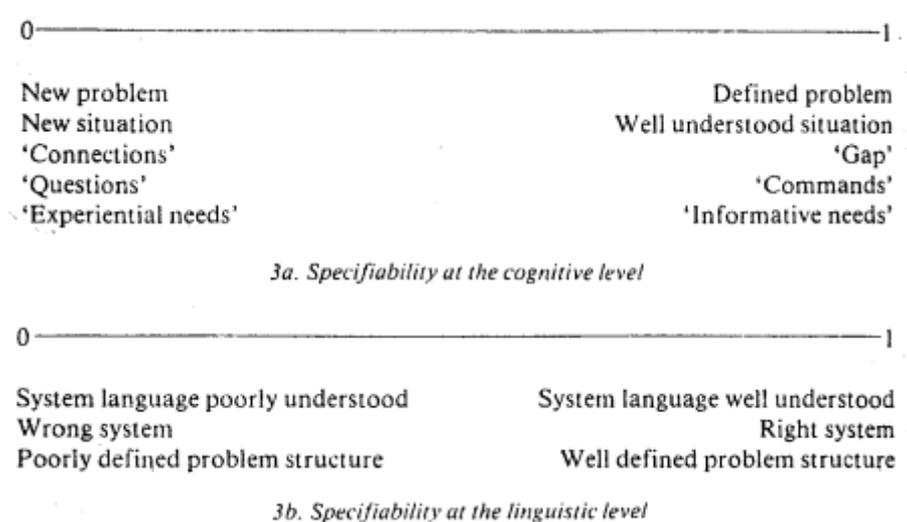


Fig. 2-23 Escalas y niveles de especificación de una necesidad de información. Fuente: Belkin, 1980.

17 años después, el problema persistía, como describen Shneiderman, Byrd y Croft (1997): *“Does the program interpret the query the way the user intended it, and -- even if it does -- does the user know that the program interprets it that way?”*, lo que se puede traducir por “¿el programa interpreta la consulta de la forma en que el usuario la usó, e incluso si lo hace, sabe el usuario que el programa lo interpreta de esa manera?”. Por tanto, el sistema de recuperación de información debía interpretar la intención del usuario, e indicarle al usuario que lo había interpretado de forma correcta.

2.3.2.3. Modelo Sensemaking

Sensemaking (también se puede encontrar como *sense-making*) relacionada con el concepto de conciencia situacional en psicología cognitiva, es en sí una metodología multidisciplinar, que aporta a las Ciencias de la Información (Information Science) y a la Ciencia de Computadores (*Computer Science*) en particular a la disciplina de Information Seeking para la primera, y al estudio de la Interacción Persona-Ordenador (*Human-Computer Interaction*) un método según el cual, los usuarios buscan dar sentido dentro de sus contextos a situaciones inciertas (Dervin, 1983).

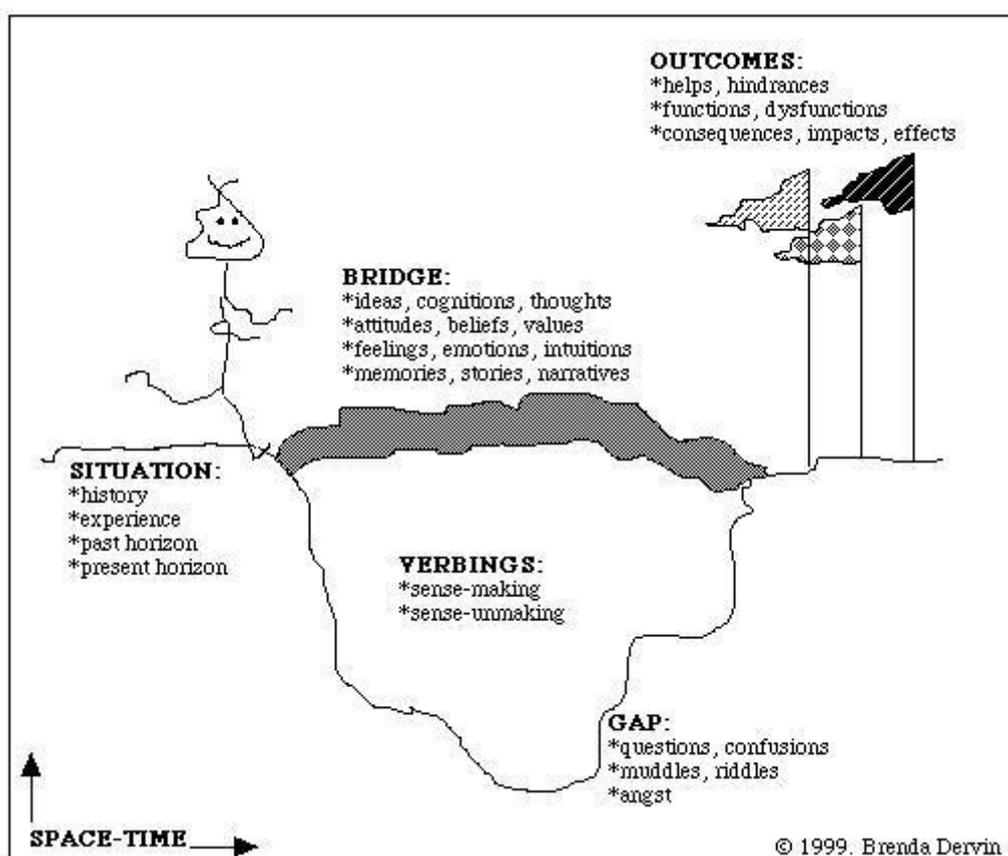


Fig. 2-24 Representación original del modelo Sensemaking. Fuente: Belkin, 1980.

Sensemaking busca detectar cómo el usuario interpreta la información para generar una toma de decisiones, cuando éste reconoce que su comprensión de cierta parte del mundo que le rodea, no es adecuada. Su enfoque viene de la Comunicación, disciplina de la que Dervin, la autora del modelo, proviene. Tiene aplicaciones en distintos entornos y

contextos, y admite distintas perspectivas, al servir como un proceso para representar el proceso de la toma de decisiones y la resolución de problemas (Cheuk y Dervin, 1999)

2.3.2.4. Modelo de Ellis

La aproximación de Bates al problema de la RI coincide con los trabajos de Ellis (1984, 1989a, 1989b, 1997, 2002) o Kuhlthau. En este sentido y según Ellis, los SRI o Sistemas de Recuperación de Información, necesitan aportar al usuario flexibilidad en el abanico de soluciones a utilizar, incluyendo este modelo en el que se va picando de una posible solución a otra hasta obtener un conjunto final de respuestas suficiente. Por tanto, según Ellis:

1. Cuantas más soluciones distintas de Recuperación de Información o RI (léase buscadores, enlaces organizados jerárquicamente o alfabéticamente, etc.) pueda aportar un sistema de información, mayor efectividad y eficiencia será posible.
2. Los usuarios ya usan estas opciones, por lo que los sistemas de información deberían adaptarse a formas familiares a los usuarios, de forma que éstos puedan interactuar con éxito y en el menor período de aprendizaje posible.

Tabla 2-5 Modelo de Ellis de Information Seeking Behaviours. Fuente: Choo, Detlor y Turnbull, 2000.

Scanning Modes	Information Need	Information Seeking	Information Use
Undirected Viewing	General areas of interest; specific need to be revealed	"Sweeping" Scan broadly a diversity of sources, taking advantage of what's easily accessible	"Browsing" Serendipitous discovery
Conditioned Viewing	Able to recognize topics of interest	"Discriminating" Browse in pre-selected sources on pre-specified topics of interest	"Learning" Increase knowledge about topics of interest
Informal Search	Able to formulate simple <i>queries</i>	"Satisfying" Search is focused on area or topic, but a good-enough search is	"Selecting" Increase knowledge on area within narrow

		satisfactory	boundaries
Formal Search	Able to specify targets in detail	"Optimizing" Systematic gathering of information about an entity, following some method or procedure	"Retrieving" Formal use of information for decision-, policy-making

2.3.2.5. Modelo *Berrypicking* de Marcia Bates

El denominado *Berrypicking model* de Marcia Bates de 1989 sobre conducta de búsqueda de información (*Information Seeking Behaviour*), está basado en el uso de bases de datos bibliográficas. Bates defendió anteriormente la necesidad de aumentar el campo de visión a la hora de evaluar un SRI (Sistema de Recuperación de Información), pasando del modelo clásico a otro más complejo, en el cual intervendría el ser humano y no sólo el algoritmo de recuperación de información (Bates, 1985).

En el modelo clásico de recuperación de información, al analizar los distintos SRI, se toma la pregunta, expresión de búsqueda o *query* del usuario como “la” unidad de análisis, una única concepción del problema. En la vida real, quiso demostrar que el proceso de búsqueda no era tan sencillo, y que había que introducir al usuario en la ecuación.

Sin embargo, antes fue Gerard Salton (1968), siendo uno de los pilares del paradigma objetivo o estadístico, quien había introducido anteriormente la idea del *feedback* o retroalimentación, como proceso iterativo en el que el usuario puede ayudar al sistema para que éste le aporte mejores resultados. En cada iteración, el usuario elegiría los documentos de su preferencia, hasta que la recuperación fuera óptima.

Según Bates, Salton seguía el modelo clásico, que se centra en los documentos y las búsquedas y no cuenta con lo que el usuario haya aprendido durante las diversas iteraciones, a lo que se debería añadir, en los modelos que tratan de solucionar el denominado *problem-solving process*, o “procesos de solución de problemas”, el contexto en el que se mueve el usuario (Gaslikova, 1999).

Pero en la vida real el usuario se mueve entre diversas fuentes: cada nueva pieza de información lo encamina hacia nuevas direcciones que seguir e incluso modifica su búsqueda, tanto la concepción mental que tenía de la pregunta como del lenguaje de búsqueda a usar, los términos de la búsqueda (*search terms*). Es lo que Bates denominó “búsqueda en evolución” (*evolving search*).

Por tanto, el problema de la Recuperación de Información incorporaba más elementos que un corpus final de documentos obtenidos en respuesta, desglosándose en una serie de elecciones de informaciones que modificaban la búsqueda y su solución, en constante evolución. Este modelo fue denominado “modelo de recogida de frutas del bosque” (*berrypicking model*)

denominado así por hacer alusión a la forma en que se pueden recoger unos arándanos (por piezas de información), uno a uno.

2.3.2.6. Modelo *Information Search Process* de Kuhlthau

Kuhlthau ha realizado diversos esfuerzos para modelizar el comportamiento de los usuarios de OPACs y otros sistemas de recuperación de información (Kuhlthau, 1989). Para ello estudió distintos tipos de usuarios en diferentes situaciones y en ocasiones también a lo largo del tiempo. Quería estudiar sobre todo las consecuencias que tiene la incertidumbre que experimenta el usuario sobre el proceso de búsqueda de la información. El resultado de sus investigaciones fue un modelo de conducta de búsqueda llamado ISP, *Information Search Process*, que constaba de diversos pasos:

- iniciación,
- selección,
- exploración,
- formulación,
- colección,
- y presentación de los resultados obtenidos.

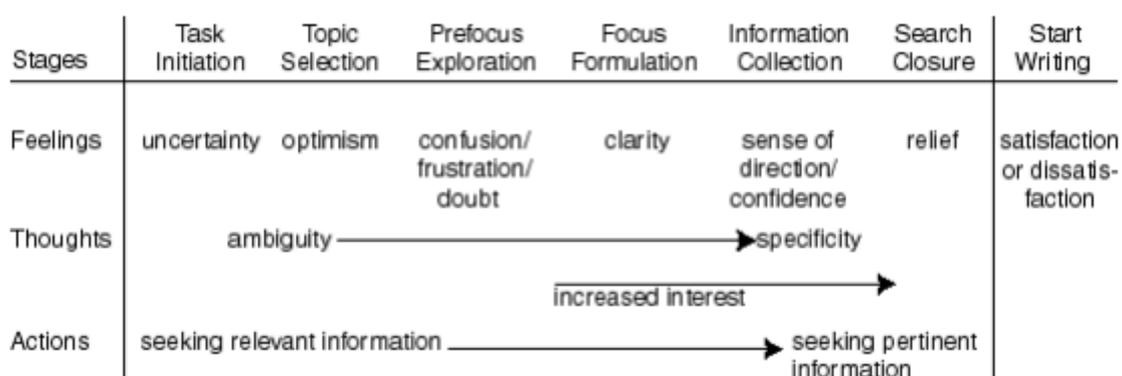


Fig. 2-25 Modelo de Kuhlthau, *Information Search Process*. Fuente: Kuhlthau, 1989

Al igual que Bates o Audunson (1999), Kuhlthau quería ir más allá de la perspectiva de concentrarse en el estudio del sistema de información y de la representación que de los textos hace el sistema para recuperar la información, para entender la recuperación de información desde el punto de vista de las necesidades de información de los usuarios, pues el modelo clásico de RI se basa en la certidumbre y en el orden, mientras que los problemas del usuario están caracterizados por la incertidumbre y la confusión (Kuhlthau, 1993).

Según este modelo, el ISP es la actividad de un usuario en su proceso de encontrar significado a partir de la información, para mejorar su estado de conocimiento en un problema particular o temática. “Proceso” es el elemento que diferencia al modelo clásico de éste, dado que el objetivo de la recuperación de información es resultado de una serie de encuentros con la información durante un período de tiempo, en lugar de un único momento en el que se obtiene un conjunto de resultados más o menos válidos.

Coincidiendo con otros autores como Belkin (1980), Chananuparp, 2006, o Kelly (2006), quienes habían incluido el paradigma cognitivo en sus investigaciones, Kuhlthau (1990, 1991) mostró la influencia que el “medioambiente” podía tener en la conducta del usuario, mediante factores como experiencias anteriores, conocimiento del tema, interés, información accesible, requerimientos del problema concreto, y tiempo disponible para la resolución del problema.

Más allá de que el modelo quede para estudiantes y no para otras tipologías de usuarios, el estudio de la experiencia afectiva en los procesos de búsqueda de información es importante no sólo desde el punto de vista de la evaluación del sistema de recuperación de información, sino más allá, del grado de utilización del sistema completo por parte del usuario. Un usuario insatisfecho es un usuario que, probablemente, no volverá a usar el sistema.

Asimismo, Kuhlthau (1988) comprobó en un estudio longitudinal cómo a lo largo del tiempo, el juicio de relevancia de un usuario puede cambiar en el tiempo, reflejando su conocimiento personal de una temática y su comprensión del problema, puesto que lo que era relevante tiempo ha, puede no serlo después, y viceversa. En su investigación, estudió a un conjunto de usuarios durante varios años para comprobar cómo cambiaba su conducta de búsqueda conforme mejoraba su conocimiento de los medios para realizar esas búsquedas.

La limitación de esta investigación es que no pudo ser realizada con un número estadísticamente relevante de sujetos de investigación, y estaba circunscrito al segmento de estudiantes de una temática concreta, por lo que debería ser estudiado más ampliamente, pero nos da una idea del cambio que se produce en el sujeto, donde se descubre que, dependiendo de su experiencia en un campo temático, el usuario selecciona un tipo de fuentes u otras. Especialmente llamativo (y por otra parte, lógico) es el caso de los clientes, que en un estado inicial no se tienen en cuenta como fuente de información, y pasados cinco años, el usuario los considera una fuente de información vital en su proceso de búsqueda de información.

2.3.2.7. Modelo de Marchionini

Quizá el modelo de *Information Seeking* más conocido es el formulado por Marchionini en 1995. Posteriormente citado hasta la saciedad, su gran virtud consiste en servir de modelo integrador para cualquier clase de búsqueda y para cualquier tipo de usuario.

El problema principal de la mayoría de estudios de usuarios de sistemas de información es que, en general, se realizan con pocos usuarios y en circunstancias muy concretas, con interfaces tan distintos unos de otros que es difícil extrapolar las conclusiones de forma que obtengamos un modelo general (Baeza-Yates y Ribeiro-Neto, 1999).

Para soslayar este problema, Marchionini realizó diversos experimentos en distintos interfaces, con distintos usuarios y en variadas situaciones. Como resultado, formuló el modelo más aceptado hoy día, un modelo que ha sido contrastado y testado en múltiples ocasiones por otros autores (Manglano et. al., 1998; González-Teruel, 2015). En general, el modelo parece mantenerse vigente, con algunas diferencias entre usuarios, dependiendo de sus características individuales o grupales.

El modelo de Marchionini agrupaba los pasos que los usuarios daban durante el proceso de búsqueda de información en:

1. **Reconocimiento y aceptación del problema de información:** El usuario admite que tiene un problema de información y que necesita usar un sistema de recuperación de información para solucionarlo.
2. **Definición del problema:** El usuario define mentalmente el problema, el tipo de información que necesita recuperar y qué soluciones prevé encontrar.
3. **Selección del sistema de búsqueda:** al enfrentarse al sistema de información, selecciona qué sistema de búsqueda prefiere utilizar. Es en este momento donde hay muchos usuarios que tienen los primeros problemas, bien por no saber utilizar el sistema de recuperación de información, bien por no conocer todas las posibilidades que tiene el sistema de información.
4. **Formulación de la pregunta o *query*:** en el caso de buscadores con distintas opciones, puede elegir entre búsqueda por lenguaje natural, utilizar comandos específicos u operadores booleanos. Una vez más, el desconocimiento o la costumbre de utilizar unos sistemas u otros puede ralentizar o dificultar la tarea.
5. **Ejecución de la pregunta en el sistema de recuperación de información:** uno de los estudios más interesantes se realiza sobre la base de lo que los usuarios buscan en forma de términos o palabras clave. El estudio del archivo digital de las expresiones de búsqueda insertadas en el formulario de búsqueda, denominado en inglés *log* del buscador, o *search log*, nos permite conocer qué palabras clave suelen usar los

usuarios, dependiendo del sistema, la interfaz, la materia (temática) que cubre, y el público que suele consultar el sistema. A este respecto se suceden dos tendencias: por un lado, se observa que un pequeño número de palabras clave son usadas en sus búsquedas por la mayoría de los usuarios del sistema, cumpliendo la ley 80/20 de Pareto o la ley de Zipf, y general, las leyes de potencias o *power laws*. Por otro lado, esa distribución en la repetición de términos de consulta sigue asimismo una forma denominada *long tail* o de cola larga, en la que unos pocos términos son repetidos muchas veces, y muchos términos son repetidos pocas veces (Brynjolfsson, Hu, y Simester, 2011).

6. **Examen de los resultados:** se realiza un proceso de *browsing* u hojear entre los resultados que ofrece el sistema. La disposición de los elementos, los resúmenes de los resultados, la relevancia de las respuestas, el grado de conocimiento de los usuarios y otros factores que afectan a los usuarios, influye en la elección de los resultados que van a ser consultados en profundidad.
7. **Extracción de la información:** Una vez elegidas y exploradas las respuestas, se extrae la información susceptible de contestar la pregunta formulada y, sobre ésta, solucionar el problema inicial.
8. **Iteración / Finalización:** dependiendo de la validez de las respuestas, el grado de conformidad del usuario y la naturaleza del problema, el usuario puede reformular su estrategia de búsqueda, e incluso cuestionar el propio problema inicial. Así, las iteraciones influirán en el comportamiento del usuario al reiniciar el proceso.

Una simplificación de este modelo lo observamos en Xie, Yu y Cen (2011), quienes dibujan el proceso típico de interacción con un motor de búsqueda, basando su modelo en el análisis de 750 millones de búsquedas realizadas en un buscador chino.

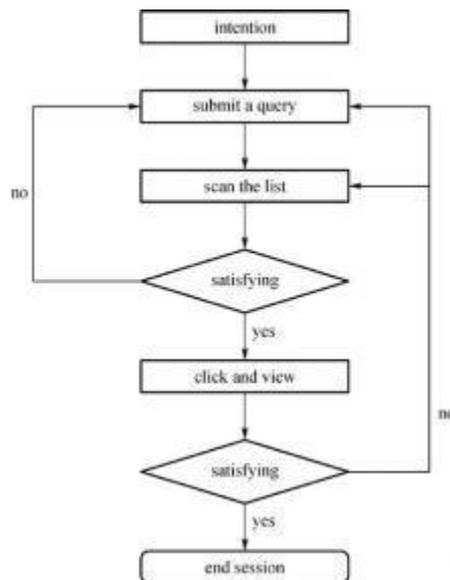


Fig. 2-26 Interacción entre usuario y motor de búsqueda. Fuente: Xie, Yu y Cen, 2011.

Los estudios de Mangano y Beaulieu (1997) y sobre todo Beaulieu junto con Robertson y Rasmussen en 1996, en su intento de refutar este modelo clásico de Marchionini de 1995, estudiaron distintos usuarios en distintas interfaces de recuperación de información para conocer conductas de búsqueda comunes, independientemente de la interfaz. Aportaron otras conclusiones, como la coincidencia con Kralisch y Berendt (2004) en afirmar que la mayoría de los usuarios estudiados mostraron una consistente preferencia por un tipo determinado de herramienta de recuperación de información o la preferencia por usar operadores booleanos frente a otras formas de uso de la búsqueda avanzada, como la elección de tipo de documento, fecha, lenguaje, etc.

Además, constataron que la herramienta más usada es siempre la que aparece por defecto, y que la reformulación de las búsquedas es una práctica muy común (del orden del 85-90% de los casos) aunque no sabemos si este hecho es independiente de la calidad del buscador o depende del nivel de credibilidad del mismo.

Del modelo de Marchionini, consignaron que el paso en el que más tiempo se pierde es en el 6, el examen de los resultados. El título es la parte de las respuestas (snippets) de la página de resultados (SERPs, Search Engine Result Pages, páginas de resultados de los buscadores) que más usa el usuario para evaluar cada resultado, seguido de los abstracts o resúmenes de cada respuesta. Otro factor de relevancia destacable que usaban los usuarios era conocer el documento previamente, o saber de él, por ser citado o haber leído sobre él.

2.3.3. De *Information Seeking Behaviour* a *User Search Behaviour*

A lo largo del apartado anterior, se ha visto cómo la Recuperación de Información hoy día es más que un algoritmo. En torno a esta disciplina, el concepto de “recuperación” (*retrieval*) se ve completado, por el de *Seeking*, término de difícil traducción (¿“búsqueda”? ¿“descubrimiento”? ¿“exploración”? ¿“indagación”?), en el sentido de que la Recuperación implica conocer el resultado de antemano, y el concepto de *Seeking* habla de incertidumbre, de descubrimiento personal del nuevo conocimiento adquirido (Saracevic y Kantor, 1988).

Information Seeking es por tanto un concepto más abierto y más centrado en el usuario (Kuhlthau, 1993), que va más allá de los documentos (objeto del modelo de RI tradicional) para intentar entender cómo interactúan los usuarios con el sistema de recuperación de información para encontrar las respuestas que necesitan.

Si esto es así, se han de tomar en consideración, dos factores necesariamente inseparables:

- El usuario, que se convierte en un factor crucial en el éxito de la correcta recuperación, tanto por los condicionantes sociales, educativos como de otro tipo del mismo;
- La interfaz, como vehículo del algoritmo de recuperación de información directa (buscador) como de otros elementos de RI (sea organización de enlaces jerárquica, alfabética o de otro tipo, para ayudar en el *browsing*)

Por tanto, de la panoplia de experiencias acumuladas en este apartado, se pueden extrapolar algunas ideas, tales como que los principios heurísticos de usabilidad enunciados por Molich y Nielsen (1990) deben probarse en cada interfaz; o que, ante tipologías de usuarios distintas, podemos presentar arquitecturas de información distintas, incluso dentro de la misma interfaz (Brusilovsky, 1996).

Este hecho lleva a confirmar que, cuantas más posibilidades de Recuperación de Información (jerárquica, facetada, alfabética, motor de búsqueda), mejor se puede preparar el sistema de información para responder a las distintas demandas, en distintos niveles, de los distintos usuarios, tipologías que no siempre se podrán prever.

Es interesante señalar el hecho de que diversas investigaciones han ofrecido similares conclusiones, incluso cuando la conclusión principal es que todo cambiaba dependiendo del diseño de la interfaz. En cualquier caso, se ha constatado que existen modelos comunes de

conductas de búsqueda, y que predominan dos formas de búsqueda básicas: directa (*searching*) y *browsing*.

Este hecho es significativo, en tanto que la Recuperación de Información tradicional no apreciaba el *browsing* como un elemento de estudio, concentrándose en el desarrollo de mejores algoritmos de recuperación mediante búsqueda directa (*searching*). Ahora se entiende que un sistema de información de suficiente entidad, debe comprender ambas aproximaciones al problema.

Estos estudios necesitan complementarse hoy día con las investigaciones realizadas directamente sobre la consulta de *search logs* (*User Search Behaviour*), en parte debido al cambio de conducta provocado por el efecto de Google en los hábitos de búsqueda de información de los usuarios de Internet. El estudio de la conducta de búsqueda de los usuarios, se revela como una disciplina determinante para comprender mejor cómo buscan y exploran los usuarios en Internet en las primeras décadas del siglo XXI, a través del análisis de universos de usuarios más amplios y variados, y estudios cuantitativos / cualitativos de amplio espectro.

Se indagará a continuación la relación entre los procesos de *browsing* y *searching* dentro de una conducta de búsqueda de información, a través de distintas investigaciones, para después centrar el enfoque en la conducta de búsqueda propiamente dicha en torno al *searching* (aquella en la que se introduce una expresión de búsqueda en un formulario, para recuperar contenidos o información, ya sea dentro de una sede web, ya sea en todo Internet).

2.3.4. *Browsing versus Searching*

A partir de la disciplina del análisis de conductas de búsqueda de información (*Information Seeking behaviour*), sobre todo en lo que representa para el diseño de sistemas de información que integran sistemas de recuperación de información con arquitectura de información de contenidos hiperenlazados, surge el estudio del binomio *browsing versus searching*.

Cualquier sistema de información en entorno web actual, como en multitud de portales bibliotecarios, integra ambos elementos de recuperación de información, y son los usuarios los que utilizan cada uno. La pregunta que subyace es si se pueden separar ambos estudios, o si se complementan en su tarea de ofrecer al usuario la información que necesita.

Para contestar a esta pregunta se pueden utilizar técnicas cualitativas, propias de la Interacción Persona-Ordenador o técnicas cuantitativas, para las cuales se utilizan herramientas de análisis de logs y etiquetas, hoy día integrables en torno a las posibilidades técnicas de aplicaciones como *Google Analytics*, que han dado pie a la proliferación de estudios y a acuñar una nueva disciplina, la Analítica Digital.

2.3.4.1. Nueva disciplina: Analítica Digital

La Analítica Digital (Jansen, 2009) es la disciplina cuantitativa que permite conocer los hábitos de los usuarios que visitan una sede web o una aplicación online. Antes conocida como Analítica Web, pero hoy ampliada a Digital por su aplicabilidad en distintos dispositivos, (como aplicaciones o apps) o entornos (como *Internet of Things* o IOT) es una disciplina que trabaja sobre distintas tecnologías, como los análisis de logs del servidor, los códigos insertos en html, o las cookies. Estudia los hábitos del uso e interacción de los usuarios en Internet que llegan a un sitio web a través de diversos canales y fuentes de datos (Digital Analytics Association, 2011).

En los últimos años la evolución de la analítica digital, que ha pasado técnicamente de utilizar logs, a combinar logs y tags para medir la interacción de los usuarios de sedes web y aplicaciones online, permite análisis cuantitativos en las sedes webs bibliotecarias (Jorquera-Vidal, 2011 y Fagan, 2014) que también son más cualitativos, donde no sólo se pueden usar indicadores de desempeño o *Key Performance Indicators* (KPIs) sino que se pueden relacionar y correlacionar variables, para deducir conclusiones cualitativas, e incluso evaluar el éxito de campañas de marketing bibliotecario en base a su ratio de conversión a descargas o a préstamos (Whang, 2007).

Los artículos sobre analítica digital para sedes web bibliotecarias comenzaron siendo basadas fundamentalmente en el estudio de casos (Black, 2009) pensados para un profesional que debe estar al día constantemente de nuevas tecnologías, y que suelen basarse en las funcionalidades de la herramienta *Google Analytics* (Yang y Perrin, 2014), utilizando esta herramienta fundamentalmente por su gratuidad. De hecho, la profundidad de estas publicaciones ha avanzado conforme han ido apareciendo más prestaciones en el servicio (Turner, 2010).

La Analítica Digital mide fundamentalmente interacciones dentro de un sitio web o una aplicación. Parte de esa interacción puede tomar efecto en forma de clics en enlaces y visitas a páginas internas del sitio, como las fichas de libros en librerías online o las sedes web, incluyendo los OPACs bibliotecarios (Vecchione et. al., 2016), lo que se denomina "*browsing*" de los usuarios, el "hojeo" que éste realiza entre los enlaces organizados mediante una arquitectura de información determinada dentro del sitio web, como en el caso de una sede web bibliotecaria (Barba et. al., 2013).

Pero también puede permitir el estudio del *search log*, el archivo histórico de las consultas, expresiones de búsqueda, palabras clave o frases de búsqueda (*queries, keywords, keyphrases* en la literatura científica en inglés) que se realizan directamente y en lenguaje documental o más comúnmente hoy día, en lenguaje natural, en el motor de búsqueda interno del sitio web. Esta interacción es denominada *searching*, en contraposición a la de *browsing*, en tanto que el usuario que busca mediante una consulta expresada mediante palabras en un texto de búsqueda, lo hace a través de un sistema de recuperación que permite “atajar” su navegación hacia los resultados de búsqueda más apropiados (más relevantes) para contestar su pregunta, su necesidad de información.

De hecho, el usuario normalmente hará *browsing* entre los resultados, clicando aquí y allá hasta encontrar la respuesta correcta, lo que puede ocurrir en el primer intento o dilatarse a través de diversas interacciones, en función de varios factores, como el algoritmo de búsqueda, el contenido a recuperar, el contexto del usuario, el dispositivo, etc.

Asimismo, la tecnología de Analítica Digital basada en *tags* o etiquetas (códigos insertos en el html), permite cruzar el análisis del *search log* (de los textos de consulta junto con el momento de la misma u otros datos de segmentación, como dirección IP del usuario, etc.) con el análisis de las respuestas clicadas y de la navegación tanto anterior como posterior al momento de la consulta (Tan, Shen y Zhai, 2006; Serrano-Cobos, 2009), por lo que se mezclan análisis de *searching* y *browsing* en la misma indagación de los hábitos de búsqueda del usuario (Rosenfeld, 2011).

2.3.4.2. Casos de estudio

Son muchos los autores que han trabajado distintos aspectos de la interacción de los usuarios dentro de un sitio, como se verá fundamentalmente mediante el estudio de casos concretos. Por ejemplo, Koch et. al (2004) estudiaron el porcentaje de actividad de los usuarios del portal renardus.com. Al utilizar el archivo log de visitas del portal, con unas 155.000 sesiones de usuarios, pudieron analizar un volumen de sujetos a estudiar más representativo que en los clásicos tests de usuarios utilizados en experimentos de observación (*focus group*, *thinking aloud*, etc.) en los que suelen participar pocos sujetos. Descubrieron que alrededor de un 80% de la actividad del portal se ejercía sobre las diversas formas de *browsing* disponibles en el mismo.

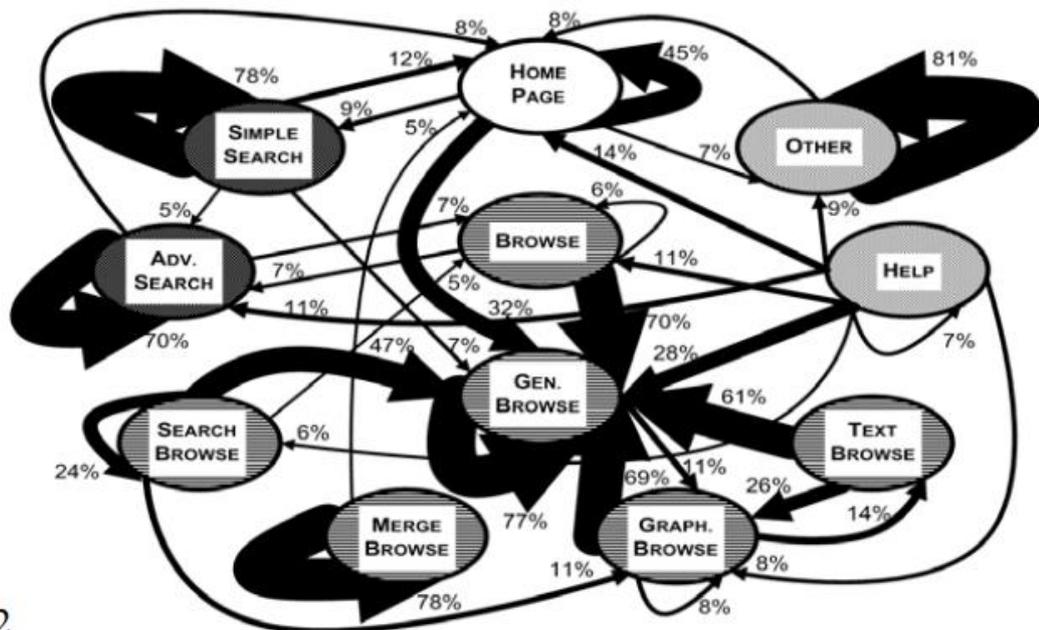


Fig. 2

Fig. 2-27 Diversas formas de browsing en Renardus.com. Fuente: Koch et. al. (2004)

A tenor de los autores, este hecho parece deberse al elevado número de entradas en el portal a través de enlaces internos y no de la home o portada, y a la disposición en pantalla de las herramientas de recuperación de información, que favorecía el *browsing* sobre la actividad de búsqueda mediante buscador (*searching* propiamente dicho), siendo en su mayoría secuencias seguidas de hasta 10 pasos en su actividad de *browsing*.

En la búsqueda exploratoria o *browsing*, el usuario puede conocer o no su necesidad de información con precisión. En ocasiones, carece de los conocimientos necesarios para formular consultas precisas y con el lenguaje de consulta apropiado para el software de consulta (Catledge y Pitkow, 1995), por lo que el análisis de la conducta exploratoria a través de los *logs* de uso del sistema, puede dar luz sobre las necesidades del usuario y la forma óptima de organizar o reorganizar la información en el sitio web. Wang et. al. (2009) estudiaron los patrones de conducta de los usuarios al comenzar su navegación a partir de la búsqueda por palabra clave, integrando esos patrones en un *topic map* sobre el que visualizar sus pasos.

Broder et al. (2005) estudiaron la integración entre el estudio del *browsing* o navegación guiada y el *searching*, y descubrieron tres escuelas de pensamiento:

1. La escuela centrada en la búsqueda, que considera el *browsing* como superfluo, gracias a que la búsqueda es tan buena, que los usuarios pueden satisfacer todas sus necesidades a través de los motores de búsqueda.
2. La escuela de navegación por taxonomía, afirma que los usuarios tienen dificultades para expresar las necesidades de información.
3. La escuela centrada en los metadatos, que aboga por el uso de metadatos para reducir los grandes conjuntos de resultados, muy utilizada en el comercio electrónico, donde se usa la búsqueda por facetas.

Otro aspecto del problema de la recuperación de información en el contexto de uso de un sitio web es el lingüístico. No es raro ver usuarios buscar información dentro de portales que muestran su contenido en diferentes idiomas, por lo que los Sistemas de Recuperación de Información (SRI) de los portales que deben ser diseñados pensando en que vayan a ser usados por usuarios de distintas nacionalidades y cuyo lenguaje materno es distinto al común del SRI.

Por su parte y como se ha visto anteriormente, Kralisch y Berendt estudiaron en 2004 la conducta de búsqueda de usuarios de habla nativa y no nativa de un portal sobre medicina, para determinar el impacto que los problemas lingüísticos derivados del desconocimiento lingüístico de la lengua del portal podían producir en su conducta.

Los resultados de su investigación apuntaron la conclusión de que el esfuerzo adicional requerido para usar las diferentes herramientas de recuperación de información tenía poco impacto en los usuarios no nativos, observándose similar conducta en unos y otros:

- En ambos casos, los enlaces organizados alfabéticamente fueron los más usados, y el buscador interno la opción menos usada.
- Asimismo, cuando se optaba por utilizar un sistema de recuperación de información concreto de los tres mencionados, frecuentemente se usaba de modo casi exclusivo, por lo que al usuario le costaba esfuerzo cambiar de sistema.
- Los contenidos organizados alfabéticamente se consideraron en ambos casos la alternativa por antonomasia cuando fallaba cualquiera de las otras opciones probadas en un primer momento de la navegación.

Sin embargo, la conducta de navegación de éstos sí era distinta de la de los usuarios nativos en los primeros momentos de la navegación. Los usuarios nativos de este portal médico utilizaban en mayor medida los contenidos organizados jerárquicamente. Los no nativos usaban en general más el buscador y los enlaces organizados alfabéticamente, lo que llevó a pensar a los autores que los nativos no eran grandes conocedores de la materia y los no nativos sí, idea que se corresponde con lo señalado por Allen en 1991, Vakkari et al., (2003), o Zhang, Anghelescu, y Yuan en 2005 en lo tocante a la relación entre los cambios en el grado de conocimiento de una temática y la conducta de búsqueda.

Angiolillo (2003) se quejaba de la poca atención que el estudio de los *searchlogs* suscitaba entre la comunidad científica. Por tanto, utilizó tests de usuarios y *focus groups* para contrastar las hipótesis a las que llegaba tras consultar el archivo de movimientos y palabras claves más buscadas de los usuarios en el buscador interno de un sitio web. En su estudio, extrajo de los archivos de su *search log* una interesante tendencia: la mayoría de los usuarios que buscaban, intentaban antes encontrar links que resultaran apropiados a su necesidad de información.

Si el usuario no los encontraba, entonces buscaba algo en el buscador interno. Además, aunque existiera un link adecuado, cuanto menor era la prominencia del link, es decir, cuanto más abajo en la página estuviera visible, más era buscado por los usuarios en la cajetilla del buscador.

Este punto parece confirmado por las observaciones hechas con la herramienta de Eyetools.com, que nos muestra, web a web estudiada, que el ojo humano tiene en general tendencia a buscar con la mirada (*seek*) primero lo situado arriba a la izquierda de la pantalla, luego va bajando con tendencia a terminar en el punto de la pantalla situado abajo a la derecha de la pantalla. Por otro lado, Angiolillo también constató en el mismo estudio que los usuarios noveles de ese sitio web, eran los que más usaban el buscador interno, al contrario

que los asiduos, que sabían adónde debían ir, y, por lo tanto, centraban su atención en seguir los links adecuados, a los que ya se habían habituado.

2.3.4.3. Del análisis de un caso al análisis de la Red: la Cibermetría

Como herramienta de análisis de los hábitos de búsqueda de los usuarios, la Analítica Digital necesita ampliar la variedad de fuentes de datos, para no circunscribirse únicamente al universo de los usuarios que utilizan esa aplicación o sede web (sea ésta un blog, una librería digital, un OPAC bibliotecario...) bien sean usuarios fidelizados, bien esporádicos. Porque al estar mediatizado por el contexto del caso concreto, será difícil tomar el conjunto de conclusiones obtenidas, como base estadísticamente válida para deducir conductas más generales y menos sesgadas de los lectores / y extrapolarlas a conjuntos más amplios de usuarios (de toda una ciudad, una comunidad autónoma, un país, un idioma...)

Pero la Analítica Digital dedica sus esfuerzos al análisis de un caso concreto, de una sede web particular, o de unas pocas, debido a que el potencial de su análisis se basa en el acceso a los datos de interacción de la sede web. En el momento en el que el investigador no dispone de esa calidad de información cuantitativa, necesita utilizar otras técnicas, que le permitan estudiar o comparar varias sedes webs, varias cuentas sociales, varias marcas o productos... Es ahí donde entra en juego otra disciplina emergente, la Cibermetría.

La Cibermetría es la disciplina, a su vez relacionada tanto con la Informetría como con las denominadas “Ciencias de la Web” o Web Sciences, que consiste según Shiri (1998) en la “medición cuantitativa, estudio y análisis de cualquier tipo de información y medio de información existente y que funcione en el ciberespacio, usando técnicas bibliométricas, cientiométricas e informétricas” Aunque su nomenclatura y diferencias con la bibliometría y la cientiometría estén sometidas a debate, la amplitud de elementos de estudios y técnicas han aumentado, por lo que su definición ha evolucionado hasta la de convertirse en “el estudio y caracterización del espacio red a partir del análisis de sus elementos constitutivos (especialmente en los aspectos relacionados con su creación, estructura, topología, difusión, interrelaciones, evolución, consumo e impacto) mediante técnicas cuantitativas de investigación social” (Orduña-Malea y Aguillo, 2014, p. 29).

La misma bibliometría comparte con los objetivos de esta tesis doctoral su aplicabilidad al conocimiento de los usuarios, puesto que su objetivo según Sanz Casado y Gómez Caridad (1994) es el “estudio de la producción de información, a partir de la utilización de métodos matemáticos (principalmente estadísticos)”, y al deber su origen a ese anhelo de eficacia

presupuestaria de las bibliotecas, utilizará los indicadores bibliométricos para el diagnóstico de las necesidades de los usuarios de nuestras bibliotecas (Carsen, 1991 y Pérez López, 2002).

En este sentido, en la Cibermetría se incorporan indicadores que provienen del uso y producción de información en Internet, que toman la forma de métricas alternativas o complementarias (*altmetrics* o *alternative metrics*) lo que en castellano se recoge en la disciplina de la altmetría (Torres-Salinas, Cabezas-Clavijo y Jiménez-Contreras, 2013) En estos últimos años están tomando mayor relevancia, puesto que pueden ser usados como complemento de la bibliometría (Glänzel, 2004) para explicar mejor las dinámicas de citación entre científicos (Costas, Zahedi y Wouters, 2015).

Estos indicadores permiten por ejemplo medir el impacto social de una investigación, no sólo el impacto entre pares (Alonso-Arévalo, Cordón-García y Maltrás-Barba, 2016) así como ayudar a los bibliotecarios a mejorar su selección de fondos usando (entre otros) indicadores basados en el uso y consumo de la información a través de Internet (Konkiel y Madjarevic, 2016) Más allá, el análisis longitudinal de las relaciones entre científicos y, por agrupación, de estructuras científicas nacionales e internacionales, al ser analizada a la luz de otros factores que la mera relación temática, en el contexto de un mundo hiperconectado en Internet, lo que estudia la Cibermetría y más allá, las Ciencias de la Web, ofrece hoy día una nueva dimensión de estudio que añade a la bibliometría, más capas de complejidad (Orduña-Malea et. al. 2010).

La Cibermetría trabaja con conjuntos de datos amplios en comparación con la Analítica Digital (la cual, recordemos, se circunscribe al estudio de un único sitio web o aplicación) extraídos de distintas fuentes de datos, para buscar conclusiones que se puedan aplicar a mayores conjuntos de objetos de estudio. Como ciencia multidisciplinar, no sólo permite, sino que alienta el uso e integración de distintas metodologías científicas para enriquecer sus hallazgos.

Así, distintas técnicas que estudian diversos aspectos y objetos de estudio de este espacio red. Por ejemplo, los análisis de enlaces permitirán descubrir las relaciones entre las sedes web de un sector o un país visualizadas a través de grafos, incluso su evolución temporal (Amitay et. al., 2004); los análisis de menciones las redes de liderazgo y conectividad entre usuarios sociales (Alonso-Arévalo et. al., 2016); los indicadores cuantitativos (como el volumen de páginas de un determinado dominio) aplicados al análisis de sedes web ayudarán incluso a generar rankings de entidades como universidades, marcas, hospitales, etc. (Orduña-Malea y Aguillo, 2014); los análisis de comportamiento en sedes web, a comparar y descubrir las mejores prácticas en diseño de interacción (Liu et. al., 2009).

Llevado al contexto de esta tesis doctoral, una de esas técnicas a integrar, derivada a su vez de la investigación en Recuperación de Información (*Information Retrieval*) y la Interacción Persona-Computadora (*Human Computer Interaction* o HCI) es el análisis del search log de motores de búsqueda de Internet, globales, no circunscritos a una única sede web (Angiolillo, 2003), que sean utilizados por miles o millones de personas. El análisis de los *logs* de consultas realizadas en alguno de estos motores de búsqueda ayudará a ampliar el contexto de investigación y sus descubrimientos, minimizando los sesgos de las búsquedas en una sede web concreta (Jansen, 2006).

Esta multidisciplinariedad posibilita la formulación de trabajos científicos basados en el análisis de las conductas de búsqueda (*user search behaviour*) como ámbito específico de la disciplina de *information seeking behaviour* (Nicholas, Huntington y Watkinson, 2005) junto con el análisis del posicionamiento en buscadores (SEO) y la altimetría, combinando así otros parámetros de estudio que aporten indicios de calidad y visibilidad de los autores y sus artículos, y las estructuras científicas en las que se enmarquen, o de los contenidos de las sedes web bibliotecarias o las editoriales con presencia en Internet.

El carácter dinámico de la tecnología, tanto de los algoritmos de búsqueda como de los interfaces de consulta, aportará para la disciplina del estudio de Conductas de Búsqueda del Usuario o *User Search Behaviour* (Dumais et. al., 2014) a través del estudio de *search logs* (*Search Log Analysis*) un reto a la hora de extraer conclusiones en el tiempo, especialmente de forma longitudinal (Jansen, Spink y Taksa, 2008). El estudio de Conductas de Búsqueda del Usuario vía análisis de archivos de expresiones de búsqueda en search logs de motores de búsqueda globales, supondrá para el sector de la información una aproximación complementaria a técnicas de investigación anteriores (Jansen y Spink, 2000). Al mismo tiempo, permitirá al científico desvelar la historia oculta de una consulta realizada por pocos o muchos usuarios a lo largo de los últimos años, en un contexto de uso creciente, como es el de los motores de búsqueda globales como Altavista (Silverstein et. al. 1999) o Google (Jansen y Spink, 2006) para localizar información de todo tipo, incluyendo aquella disponible en Internet sobre el libro, la lectura u otros formatos multimedia ofrecidos por las bibliotecas.

Se explorarán a continuación diversas técnicas de investigación, experiencias y variables empleadas por los investigadores para conocer mejor los procesos que competen a las búsquedas realizadas en motores de búsqueda, de forma que sirvan de base metodológica para esta tesis doctoral.

2.3.5. User Search Behaviour

2.3.5.1. Multiplicidad de variables de estudio en una búsqueda *viasearching*

Anteriormente se ha explicado cómo los usuarios pueden realizar la tarea de encontrar información a través del hojear de páginas (*browsing*) y de la búsqueda directa (*searching*). En este apartado se desglosarán investigaciones sobre los procesos de búsqueda directa de los usuarios, especialmente en motores de búsqueda generalistas (es decir, no aquellos utilizados dentro de una sede web, sino que buscan entre el contenido incluso de millones de sedes web, como es el caso de Google.com).

La importancia de este tipo de investigaciones hoy día es muy importante, tanto desde el punto de vista económico, por el negocio que mueve el Posicionamiento en Buscadores, como desde el punto de vista científico, por la puerta que abre a conocer patrones de comportamiento del ser humano, incluso en su globalidad. Por ejemplo, en 2011 se publicaba un estudio sobre la desaparición de dos millones de palabras en los últimos 200 años (Petersen et al, 2011). Es por tanto interesante conocer si en el ámbito de la búsqueda sucede lo mismo, entender cómo es el lenguaje de los usuarios de buscadores de Internet y su evolución, profundizando en otros aspectos relacionados, como por ejemplo cuándo se busca, en qué contexto, por qué se busca, a partir de en qué momento, o con qué motivación e intención (*Query Intent*), lo que se irá explorando a lo largo del apartado.

Existen diversas metodologías para acometer la investigación de las tareas de búsqueda en buscadores (Choo, Detlor y Turnbull, 1998), primando el análisis a través de archivos *log* de motores de búsqueda. Esta técnica permite en ocasiones sólo obtener información general de las expresiones de búsqueda de información y su momento de inserción en el formulario de búsqueda (Ginsberg, 2009). Pero en otras ocasiones, permite recopilar técnicamente y para cada usuario o para cada sesión de búsqueda, el conjunto de expresiones de búsqueda que se utilizan, así como los resultados de búsqueda del motor de búsqueda, junto con las respuestas que se hayan visitado (Jansen, 2006).

Nótese que, siguiendo a Van Gysel, Kanoulas y de Rijke (2016), se define una sesión de búsqueda como una secuencia de interacciones entre el usuario y el motor de búsqueda, donde se inserta una consulta consistente en uno o más términos, y el sistema devuelve una página de resultados que consta de un conjunto de resultados devueltos por el motor de

búsqueda (también conocido como SERP). Esas expresiones de búsqueda, antes generadas mediante combinaciones de comandos de búsqueda y operadores booleanos, están formadas por palabras y frases escritas, hoy día en lenguaje natural.

El potencial de investigación del comportamiento humano en su interacción con tecnología de alcance mundial, ha transformado tanto el desarrollo de los algoritmos de software como el marketing basado en buscadores, aunque se deba tener especial atención a mantener la privacidad de los usuarios (Cooper, 2008) y tenga como obvia limitación las posibilidades de los motores de búsqueda de dar acceso a los investigadores a sus *search logs* (Amitay y Broder, 2008), así como contemplar la posibilidad de combinar el análisis de *logs* con otros tipos de análisis, más cualitativos, como el estudio de campo de las interacciones los usuarios, o el seguimiento de paneles de usuarios (Grimes, Tang y Russell, 2007)

Todos estos factores impulsan investigaciones que son en sí poliédricas, con muchas variables de estudio interrelacionadas, difíciles de separar para buscar relaciones causales, por lo que será difícil encontrar investigaciones que aislen los factores y, en cierta medida, traten el proceso de la búsqueda como un desafío sistémico de la recuperación de información (Baeza-Yates, 2017; Norman, 2017).

Entre otras variables, se distinguirán investigaciones dedicadas a la influencia de la interfaz de búsqueda en la conducta del usuario y su evolución, así como al estudio de las interacciones de los usuarios durante el proceso de búsqueda, desembocando en el análisis longitudinal de las búsquedas y su evolución a través del tiempo.

Aparte, se buscará indagar qué aspectos cognitivos del usuario se han analizado en la literatura científica, con el fin de segmentar el comportamiento del usuario que busca. En particular, se descubrirán estudios sobre el idioma y lenguaje empleados en las expresiones de búsqueda, lo que llevará a la comunidad científica a preguntarse por las motivaciones de la búsqueda, o más bien, las intenciones de búsqueda (*query intents*) que las expresiones de búsqueda denotan a través del análisis de los textos de las mismas, lo que será clave para esta tesis doctoral, al analizar las intenciones de búsqueda en su relación con el sector de la lectura y las bibliotecas.

2.3.5.2. Interfaz de búsqueda

Toda búsqueda de información en un motor de búsqueda se realiza sobre una interfaz, en cuya optimización se revelará clave el trinomio usuario-diseño-algoritmo, en cuanto a que no será fácil separar los cambios en uno de los cambios en los otros (Hearst, 2009). Algunos autores intentarán separar variables, con limitado éxito, puesto que los resultados obtenidos lo son en un momento dado y con una interfaz dada. Por ejemplo, Scaife y Rogers (1996) utilizaron la presentación de la información, la interfaz, como variable de análisis.

En esa línea Jones, Cunningham y McNab (2000) de la Biblioteca Digital Nacional neozelandesa, trabajaron sobre la interfaz, constatando que los usuarios raramente cambian las opciones marcadas por defecto, corroborando lo indicado por Nielsen (1999), aunque reconociendo que esta conducta podría estar íntimamente relacionada con la disposición de las opciones de reconfiguración de la búsqueda. En el estudio se cambiaron las opciones por defecto, sin que se produjera un aumento de reconfiguraciones, lo que induce a pensar que el usuario simplemente acepta la configuración por defecto como óptima.

En su experimento trabajaron con un tipo de interfaz que permitía operadores booleanos, hoy un sistema de interacción más difícil de encontrar, y por tanto una investigación difícil o imposible de replicar. Observaron que el operador booleano más usado fue el de intersección (AND, Y) seguido del de unión (OR, O), y después por el de negación (NOT, NO). Dependiendo del tipo de interfaz que se le ofrece al usuario, la proporción de elecciones de operador variaba, aunque respetando ese orden de prioridad de elección de operador, siendo menos usados los operadores booleanos cuando el algoritmo de ranking de respuestas los incorpora, lo que ocurre en el buscador Google, por ejemplo, el cual incorporaba en sus inicios el operador NEAR por defecto (Serrano-Cobos, 2003).

Por otro lado, Jones, Cunningham y McNab (2000) coinciden con otros autores en que el contenido disponible, algoritmo de recuperación e interfaz de consulta pueden modificar en ciertos aspectos la conducta de búsqueda de los usuarios, por lo que cada sitio web tendrá peculiaridades únicas, y las conclusiones de cada estudio pueden no ser siempre extrapolables a otros contextos.

El análisis de la reformulación de búsquedas viene de la necesidad de mejorar en un objetivo práctico por parte de los diseñadores de interacción y de sistemas de recuperación de información, como lo es conseguir minimizar el fracaso en la resolución de las necesidades

informativas del usuario. Ese objetivo lleva a experimentar con diversas técnicas. Una de ellas, como se ha visto anteriormente, es la de conjugar las conductas de búsqueda y hojear (*searching* y *browsing*) pero en este caso, en la misma página de respuestas o SERP. Siguiendo la estela creada por Martin Belam (2003), que a su vez seguía a Morville y Rosenfeld (1998), Angiollo (2003) apuntaba que en Verizon se utilizaba la técnica de *Best Bets*, también denominada *Best Matches*, constatando otra aplicación práctica de la Ley de Zipf (1950) por la que se tomaban las búsquedas más repetitivas y se preparaban un conjunto de respuestas a mano que según los documentalistas correspondieran mejor a la *query* o búsqueda formulada.

Así, cuando un usuario buscaba algo, se le ofrecían dos juegos o subconjuntos de respuestas, las que ofrecía el algoritmo del buscador interno (basado en Verity) y las naturales. El subconjunto mejor aceptado por los usuarios fue el hecho a mano, los *Best Bets*.

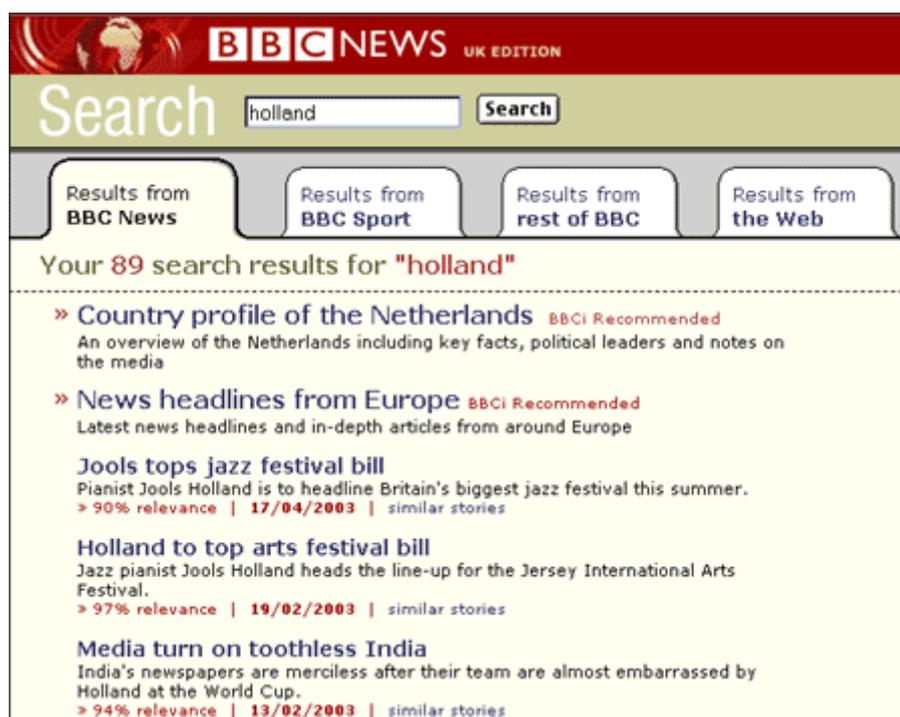


Fig. 2-28 Best bets en BBC. Fuente Martin Belam, 2003.

Estos intentos de eliminar o disminuir las visitas a enlaces que no respondan adecuadamente a las necesidades de información de los usuarios, viene dada no sólo por un deseo de mejora en el proceso, sino por una necesidad de menguar el problema de la excesiva cantidad de respuestas, lo que se agrava más cuando se trata de buscadores globales, como Google o

Altavista. Así, autores como Bradford y Marshall denotaban que, en las SERPs o páginas de respuestas de los buscadores, aparecía un elevado número de páginas duplicadas en sitios web diferentes o de contenidos idénticos en URLs distintas, alrededor de un 14% del total de páginas (Bradford y Marshall, 1999).

Estos duplicados ralentizaban la consulta de los listados de respuestas, al obligarlos a visitar sitios web que aparecían muy cercanas entre sí, como respuestas muy similares, descubriendo al visitarlas que eran el mismo contenido, lo que frustraba sus pesquisas, por lo que abogaban por el uso de URN (*Uniform Resource Locator*) para identificar unívocamente cada página.

Además, los mismos Bradford y Marshall estudiaron las estrategias de reformulación de búsquedas más comunes. Frecuentemente, el usuario abandonaba la búsqueda si la primera intentona no resulta satisfactoria. Una búsqueda podía diferir enormemente de la siguiente, por lo que la personalización de agentes de búsqueda parece problemática. Siguiendo a Zipf (Breslau, 1999), el 50% del tráfico en buscadores se realizaba a través del 10% de búsquedas, y estas búsquedas eran realizadas por un pequeño pero activo grupo de usuarios.

Estudiaron cómo mejorar la interfaz de inclusión de términos para añadir a la búsqueda mediante información explícita e implícita, observando qué documentos resultaban interesantes al usuario y cuánto, y además en qué momento, en qué iteración fue elegido el documento como relevante. Así, se obtuvo información de la relevancia de los términos a usar en la expansión de las búsquedas tanto por el propio contenido de los documentos relevantes en sí mismos, como por la conducta experimentada durante el proceso de búsqueda.

En este análisis de la interfaz, un aspecto a estudiar es el diseño de las páginas de respuestas del buscador o SERP (*Search Engine Response Pages*) Así, la aparición de resúmenes apropiados en las respuestas de las SERPs, resulta de importancia significativa para discriminar entre respuestas que son útiles o no, confirmando la evolución en este aspecto que se produjo entre la disposición textual que se observaba en los buscadores tipo Altavista hasta 1998 y la que se ha ido imponiendo a partir de la aparición de Google, que quizá pudo seguir los consejos de Nielsen al respecto (Nielsen, 1999). Nielsen recomendaba el uso de titulares en los resultados de páginas de respuestas ante una búsqueda (SERPs o *Search Engine Response Pages*), y que se utilizara negrita en las palabras que coincidieran con la frase de búsqueda o *query*, siguiendo a su vez el concepto de la esencia de información (*information scent*) dentro de la teoría de forrajeo de información, o *Information Foraging Theory* de Pirolli (Pirolli, 1995, Pirolli et.al. 2005).

Silverstein et. al. (1999) observaron que los usuarios miraban escasamente los primeros 10 resultados de la primera SERP o página de respuestas, y que rara vez modificaban los términos de su consulta. Pero esa conducta se aplica a un público (internacional) un buscador (Altavista) y un momento de la madurez de los usuarios de internet concretos (final de los noventa)

Prácticamente en el mismo año, Jones, Cunningham y McNab (2000) constataban en su caso que la mitad de consultas reformuladas consignadas en su estudio, se construían a partir del texto anterior, teniendo al menos un término en común, añadiendo o eliminando algún término. Una década después, en el estudio de Xie, Yu y Cen (2011) más del 56% de las búsquedas llevaban aparejadas posteriormente una reformulación o “*Query Refining*” de la búsqueda, donde el 82,3% de los usuarios chinos cambiaba totalmente la expresión de búsqueda, y sólo un pequeño porcentaje añadían términos (9%) o eliminaban términos (1,43%) para reformular sus búsquedas.

Term A	Term B	Count together ($O(AB)$)	ρ
cindy	crawford	118558	0.7098
persian	kitty	75716	0.6830
pamela	anderson	453467	0.6451
visual	basic	177971	0.6325
www	http	2355010	0.3862
buffy	slayer	12340	0.3989
slayer	vampire	13640	0.3088
buffy	vampire	12986	0.2766
lang=ko	domain=kr	1030416	0.7281
date=restricted	applet	1165565	0.7273
referred=yes	sessmodlen=4+	76257842	0.6388
referred=yes	sessmod=restart	46359290	0.5482
the	qwords=6+	2417838	0.1886

Fig. 2-29 Análisis de correlación de pares de términos en expresiones de búsqueda. Fuente: Silverstein et. al., 1999.

Como se ve, distintos estudios a lo largo de las últimas décadas muestran algunos elementos comunes, difíciles de integrar y comparar con exactitud, pero mostrando que, a pesar de los cambios de algoritmo o interfaz, hay conductas de los usuarios que permanecen estables, y que pueden ser empleadas en distintos escenarios y casos (Pirulli, 1995, Pirulli et.al. 2005). Es

por tanto interesante conocer otras investigaciones, que se centren en los usuarios, y no tan sólo en la interfaz.

2.3.5.3. Interacciones durante la búsqueda

No es fácil que los estudios de los usuarios, en sus procesos de búsqueda en buscador (searching) puedan separar sus investigaciones de los cambios de interfaz o algoritmo, y lo normal será ver casos como Ruthven et. al., (2003), que se han concentrado en la conducta de los usuarios durante el proceso de retroalimentación de los sistemas de recuperación (*Relevant Feedback Process*) y mejora de las cadenas de búsqueda (*query expansion*). Hay que leer a Wilson (1997) para encontrar estudios que identifican categorías de búsqueda y adquisición de información después de una encuesta de investigación que incluyó la búsqueda de información de salud.

- Atención pasiva: como escuchar la radio o ver programas de televisión, donde no puede haber búsqueda de información, pero donde la adquisición de información puede tener lugar;
- Búsqueda pasiva: parece una contradicción en términos, pero significa aquellas ocasiones en que un tipo de búsqueda (u otro comportamiento) resulta en la adquisición de información que resulta relevante para el individuo;
- Búsqueda activa: es el tipo de búsqueda más conocido en la literatura de la ciencia de la información, donde un individuo busca activamente información;
- Búsqueda en curso: donde la búsqueda activa ha establecido el marco básico de las ideas, creencias, y valores, pero donde la búsqueda continua ocasional se lleva a cabo para actualizar o ampliar el marco de la misma.

Ellis et al. (1989) en su integración de la búsqueda y el *browsing* u hojear de la información, habla de distintos *search moves*, también conocidos como *web moves* o solamente *moves* (Fidel, 1985) para entre los que destaca *Starting, Chaining, Browsing, Differentiating, Monitoring, Extracting* (traducible por "Iniciando, encadenando, navegando, diferenciando, supervisando, extrayendo") y despunta una clasificación de expresiones de búsqueda en función de sus interacciones posteriores al acto de explicitar su consulta en el motor de búsqueda.

Bates había introducido el concepto de *search moves* en su día (1979), partiendo de las búsquedas bibliográficas, como parte del estudio de las estrategias de búsqueda de información, en la que *search move* era la táctica de búsqueda o el movimiento realizado para ejecutar una búsqueda. Nombró veintinueve tácticas, definidas y discutidas en cuatro categorías: monitoreo, estructura de archivos, formulación de búsqueda y término.

Tabla 2-6 "Web Moves" según Ellis. Fuente: Choo, Detlor y Turnbull, 2000.

Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting	
Literature Search Moves (Ellis et al., 1989, 1993, 1997)	Identifying sources of interest	Following up references found in given material	Scanning tables of contents or headings	Assessing or restricting information according to their usefulness	Receiving regular reports or summaries from selected sources	Systematically working a source to identify material of interest
Anticipated Web Moves	Identifying Web sites/pages containing or pointing to information of interest	Following links on starting pages to other content-related sites	Scanning top-level pages: lists, headings, site maps	Selecting useful pages and sites by bookmarking, printing, copying and pasting, etc.; Choosing differentiated, pre-selected site	Receiving site updates using e.g. push, agents, or profiles; Revisiting 'favorite' sites	Systematically searches a local site to extract information of interest at that site

Fidel (1985) indica que los *moves*, movimientos o cambios en la reformulación de la consulta se realizan para resolver tres situaciones problemáticas:

- (1) cuando los conjuntos recuperados son demasiado grandes;
- (2) cuando son demasiado pequeños;
- (3) cuando los conjuntos recuperados están fuera del objetivo.

Al analizar noventa búsquedas obtuvo una lista de dieciocho movimientos operativos, o modificaciones de formulación de consultas, que mantenían invariable el significado de los componentes de consulta, y doce movimientos que cambiaban el significado de los componentes de consulta.

Table 1. Summary of moves and check list of options in online searching

	Operational moves	Conceptual moves
	<i>Moves to reduce the size of a set</i>	
Weight 1	Limit a descriptor to be a major descriptor.	Intersect 1
Weight 2	Intersect free-text set with a broader descriptor.	Narrow 1
Weight 3	Limit free-text terms to occur in a predetermined field.	Narrow 2
Weight 4	Require that free-text terms occur closer to one another in the searched text.	Intersect 2
Weight 5	Limit to documents of a certain form.	
Negate	Eliminate unwanted elements by using the AND NOT operator.	
Limit 1	Limit to documents written in a particular language.	
Limit 2	Limit to documents published, or indexed, in a particular period of time.	
Limit 3	Limit to documents retrieved from a specific portion of the database.	
Limit 4	Limit to sources that have, or do not have, a certain term in their titles.	
Cut	Submit only part of the retrieved answer set, arbitrarily selected.	
		Intersect a set with a set representing another query component. Intersect a descriptor set with a set created by more specific free-text terms. Qualify descriptors with role indicators. Intersect sets with role indicators.

Fig. 2-30 Moves operacionales y conceptuales. Fidel, 1985

<i>Moves to enlarge the size of a set</i>		
Add 1	Add synonyms and variant spellings.	Expand 1
Add 2	Add descriptors as free-text terms.	Expand 2
Add 3	Add descriptors assigned to relevant citations retrieved.	Expand 3
Add 4	Add terms from database's index that have a high number of postings.	Expand 4
Include	Group together a descriptor with all the descriptors that are its narrower terms.	Exclude
Cancel	Eliminate restrictions previously imposed.	Expand 5
<i>Moves to increase both precision and recall</i>		
Refine	Find a "better" descriptor.	Probe 1 Probe 2
		Construct an indexing-probe set. Use the difference among the number of postings for a search term in various databases to decide how to represent components in each database.

Fig. 2-31 Moves para aumentar e incrementar precisión. Fidel, 1985.

Por su parte, Choo, Detlor y Turnbull (1998) combinaron el uso de encuestas a usuarios y el seguimiento remoto de sus interacciones, para determinar las necesidades de información de los usuarios y su conducta de "Information Seeking". Los autores relacionaron las

“motivaciones” (estrategias y modos de ver respuestas y de buscar) con los “movimientos o *moves* (las tácticas implementadas por los usuarios para encontrar y utilizar la información) Anteriormente Shneiderman (1997) investigó sobre el tipo de tarea a realizar durante la búsqueda.

Después, de nuevo Choo, Detlor y Turnbull (2000) resumen estos incipientes esfuerzos por clasificar la forma en que los usuarios interactúan en relación a sus búsquedas, integrando el modelo de Ellis con el de Aguilar (1967) que hablaba de *Undirected Viewing*, *Conditioned Viewing*, *Informal Search* y *Formal Search*, como se puede apreciar en esta tabla:

Tabla 2-7 Tipos de Web Moves. Fuente: Choo, Detlor y Turnbull, 2000.

Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting	
Undirected Viewing	Identifying, selecting, starting pages and sites	Following links on initial pages				
Conditioned Viewing			<i>Browsing</i> entry pages, headings, site maps	Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	
Informal Search				Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	Using (local) search engines to extract information
Formal Search					Revisiting 'favorite' or bookmarked sites for new information	Using search engines to extract information

Estos estudios de las conductas de búsqueda de los usuarios son en su mayoría cualitativos y observacionales, y denotan que hay aspectos susceptibles de generalización, para las que se detectan clases distintas de movimientos, pero que necesitan incorporar otras variables. Esta conclusión lleva a descubrir otro tipo de estudios, que profundizan en la investigación de aspectos cognitivos y culturales, como posibles variables que intervengan en estos movimientos; por ende, en las conductas de los usuarios (Abadal y Codina, 2008).

2.3.5.4. Aspectos cognitivos del usuario

Una variable considerada entre los científicos del paradigma cognitivo en recuperación de información, ha sido precisamente la de la capacidad cognitiva del usuario en sus procesos de petición de información al motor de búsqueda, fundamentalmente en dos variantes: el grado de conocimiento de una temática (Navarro-Prieto, 1999), y el grado de experiencia en navegación a través de internet y de herramientas de búsqueda (Vakkari, 2000), y dentro de ambas, teniendo en cuenta en ocasiones la edad (Foss et. al. 2013; Foss y Druin, 2014). En la misma línea de considerar la capacidad cognitiva del segmento de usuarios a estudiar, otros autores se han decidido por explorar los aspectos culturales de los usuarios y su incidencia en sus conductas de búsqueda de información, como Sheppard y Sholtz (1999), Marcus (2000), Chau et. al. (2000).

Por su parte, Barber y Badre (1995, 4) lanzaron la hipótesis de que “los sitios en el idioma nativo del país de origen mostrarán marcadores culturales más específicos para esa cultura en particular, mientras que un sitio en un idioma no nativo estará más influenciado por marcadores culturales externos... Los marcadores culturales son elementos de diseño que se encuentran en las páginas web, y dichos elementos se convierten en marcadores culturales cuando demuestran ser altamente prevalentes dentro de un grupo cultural particular y menos prevalentes o ausentes en otros grupos”.

Borgman (1989) examinó las diferencias individuales que pueden modificar la conducta de los usuarios en material de recuperación de información, como características personales, aptitudes técnicas y orientación académica, y concluyó encontrando que estos factores estaban interrelacionados con su conducta. Biggs (1990) exploró la problemática del exceso de información recuperada en la experiencia de búsqueda (*Information Overload*) aunque limitada a búsquedas en OPACs.

Allen apuntó la idea de que el grado de conocimiento de una temática afectaba a la conducta de búsqueda de los usuarios (*Information Seeking*) en un catálogo de biblioteca online, así como a la efectividad de sus búsquedas en un sistema de información (*Search Effectiveness o Search Performance*). Los usuarios con mayor conocimiento usaban más expresiones de búsqueda hasta encontrar lo que buscaban en las SERPs. Allen señalaba que debería tenerse este factor en cuenta a la hora de diseñar sistemas de recuperación de información personalizados (Allen, 1991).

Marchionini et. al., en 1993, trabajó sobre la comparación en la conducta de búsqueda entre expertos en una determinada temática y expertos en técnicas de búsqueda. Su trabajo reveló que los expertos en una temática se fijaban más en el contenido de las respuestas y en cómo respondían a la pregunta o *query* formulada, mientras que los expertos en búsqueda de información se interesaban más por la formulación de la pregunta.

Anett Kralisch y Bettina Berendt (2004), a partir de autores como Rice, McCreddie, y Chang (2001), mostraron cómo la conducta de búsqueda de los usuarios podía estar determinada por circunstancias culturales, lo cual debía ser tenido en cuenta a la hora de diseñar las herramientas de recuperación de información internas de un sitio web. Propusieron así destacar unos servicios de búsqueda u otros, dependiendo de su nivel cultural.

Además, descubrieron que los usuarios con mayor nivel cultural demostraron tener mayor capacidad para absorber información contextual, y por lo tanto tendían más a utilizar los enlaces, organizados jerárquica o alfabéticamente, que aquellos usuarios con poca capacidad de absorción informativa, quienes rápidamente usaban el buscador.

Fidel y Efthimiadis (1999) observaron la conducta de búsqueda de información en internet de varios ingenieros en Boeing. En su investigación, constataron diversos patrones de búsqueda comunes:

- Todos los participantes en el experimento tomaron en consideración la relevancia y la confianza que ofrecía un sistema de información dado como los factores más importantes a la hora de decidir elegirlo para encontrar información sobre un tema concreto.
- A la hora de reformular las búsquedas, los ingenieros suelen desarrollar búsquedas más específicas que generales.
- Todos eligieron la facilidad de uso como el criterio más importante para decidirse por un método de búsqueda.

Cheuk (1998) siguiendo a Dervin (1983) trabajó según la metodología *Sense-Making Research* este tipo de patrones de conducta, con un estudio cualitativo sobre el proceso de búsqueda y uso de información de ocho auditores y ocho ingenieros en sus contextos laborales, con el fin de desarrollar empíricamente un modelo, denominado *Information Seeking and Use Process Model* (“modelo de proceso de búsqueda y uso de la información” o *ISU process*). El estudio utilizaba entrevistas cualitativas, con preguntas como:

- ¿Qué preguntas pasan por su mente en esta etapa particular del proyecto?
- ¿Qué estrategias usa para obtener respuestas a sus preguntas?
- ¿Por qué elige usar esta estrategia?
- ¿Qué problemas tiene para obtener respuestas?
- ¿De qué manera cada respuesta ayuda (o no ayuda) a continuar con sus tareas?
- ¿Cuál es su sensación en esta etapa?

Su estudio dio lugar a siete situaciones en el lugar de trabajo, pero estaba limitado a dos conjuntos de personas muy específicos:

Tabla 2-8 Seven Information Seeking and Using (ISU) Situations in the Workplace. Fuente Cheuk, 1998.

Task	Focus	Ideas	Ideas	Ideas	Ideas	Passing on
Initiating	Formulating	Assuming	Confirming	Rejecting	Finalising	Ideas
Situation	Situation	Situation	Situation	Situation	Situation	Situation

Asimismo, Cheuk con Dervin (1999) por su parte compararon 3 grupos distintos de usuarios: auditores, ingenieros y arquitectos mediante entrevistas. Confirmaron el modelo de Marchionini, pero a su vez, concordando con otros estudios anteriores, descubrieron diferencias de conducta dependiendo del tipo de usuario.

Sin embargo, Ellis y Haugan, en un estudio con ingenieros y científicos noruegos, habían identificado características similares: investigación, encadenamiento de búsquedas, monitorización de los resultados, *browsing* (hojeo), distinción de resultados relevantes, filtrado, extracción y finalización del proceso de búsqueda (Ellis y Haugan, 1997).

Vakkari y su equipo a su vez, (Vakkari, 2003) dirigieron un estudio en el que se investigó a 22 estudiantes de psicología mientras aumentaba su entendimiento de la materia y su experiencia en la interacción con buscadores. Su conclusión alude también al hecho de que el conocimiento de un dominio concreto del conocimiento tiene impacto sobre su conducta de búsqueda, aumentando el número de términos mientras aumentaba el grado de conocimiento de un tema dado (lo que coincide con lo indicado previamente por Allen) asumiendo que los usuarios tengan un conocimiento suficiente del sistema de recuperación usado.

Zhang, Anghelescu, y Yuan, trabajando en la línea marcada por Allen en 1991, constataron en 2005 que a medida que el nivel de conocimiento de un usuario sobre un determinado tema aumentaba, también aumentaba el número de búsquedas utilizadas para encontrar

información que se ajustara a sus necesidades, así como el número de términos utilizados en las consultas. Sin embargo, el nivel de efectividad (*Search Effectiveness*) en los resultados resultó el mismo para usuarios expertos en la temática y no expertos.

Por tanto, dedujeron que el nivel de conocimiento en una materia sí tiene un gran impacto en la conducta de búsqueda, no en la efectividad del buscador, por lo que la calidad de las respuestas debidas al algoritmo de recuperación sería independiente de la conducta del usuario, no así la reacción de los usuarios a su grado de relevancia y al interfaz de búsqueda.

Estas conclusiones coinciden con Wildemuth, quien investigó los efectos del conocimiento de un dominio en la formulación de las tácticas de búsqueda de los usuarios. Examinó las tácticas de búsqueda desarrolladas por 77 estudiantes de medicina en una misma base de datos durante un período de nueve meses y descubrió que las tácticas cambiaron a lo largo del tiempo conforme el grado de conocimiento de la materia sobre la que se buscaba información cambiaba (Wildemuth, 2004).

Ahora bien, otro aspecto interrelacionado con el dominio de una materia o la experiencia en el uso de sistemas de recuperación de información, es el lenguaje usado en esos procesos. La pregunta que se genera es si el idioma utilizado y el lenguaje aportan información en torno a marcadores culturales que puedan llevar a diferenciar o a generalizar conclusiones para toda la población humana (Petersen et. al., 2011). En ese aspecto, el volumen y la variedad de datos serán una constante en las investigaciones halladas, donde la investigación cuantitativa buscará trabajar con mayores cantidades de información, pero donde la investigación cualitativa tratará de identificar distintas características en el lenguaje y las expresiones de búsqueda utilizadas.

2.3.5.5. Idioma y lenguaje en las expresiones de búsqueda

Una de las investigaciones localizadas con mayor volumen de datos, es la publicada por Silverstein et. al. (1999) sobre 285 millones de sesiones de usuarios de Altavista, donde estudiaron diversas variables, como la duplicación de expresiones de búsqueda, las sesiones durante las cuales se producían diversas búsquedas, o la interacción de los términos que formaban esas expresiones de búsqueda. Entonces Silverstein y su equipo descubrieron que la mayoría de los usuarios utilizaban muy pocas palabras en sus expresiones de búsqueda (el 25.8% usaban sólo una palabra)

Otro estudio de gran volumen corrobora este patrón, en este caso el de Xie, Yu y Cen (2011) sobre 750 millones de expresiones de búsqueda, aquí realizadas en un buscador chino y en lenguaje chino. Lo que cambia en este estudio es el concepto de “término” en contraposición a “palabra”, porque en chino los “términos” son las partes contiguas de una expresión de búsqueda. Pero tanto midiendo términos como palabras dentro de una *query*, muestran una distribución de tipo “ley de potencias” (*power-law*).

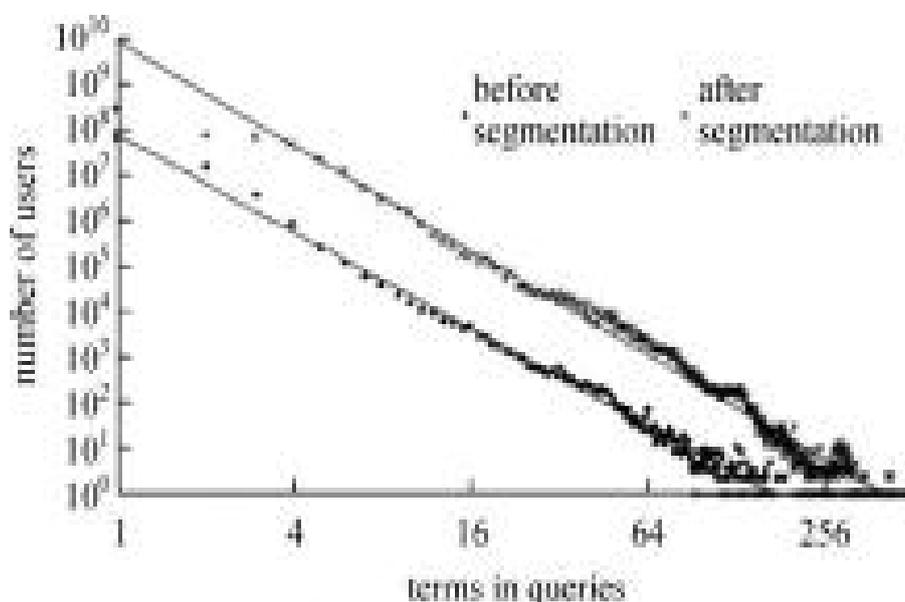


Fig. 2-32 Distribución de palabras y términos usados en búsquedas. Fuente: Xie, Yu y Cen, 2011.

Shneiderman, Byrd y Croft (1997) detectaron algunas variantes realizadas en las expresiones de búsqueda de los usuarios:

- Capitalización (mayúsculas y minúsculas)
- *Stemmed versions*: buscando "perro" encuentras palabras como "perrito" o "perruno".
- Partidas parciales: búsqueda de "biología" recupera "sociobiología" y "astrobiología".
- Variantes fonéticas, por ejemplo, a partir de N-grams u otros métodos, buscando "Johnson" encuentra "Jonson", "Jansen" y "Johnston".
- Sinónimos: buscando "cancer" encuentra "tumor maligno".
- Abreviaturas / siglas: buscando "Digital Equipment Corporation" encuentra "DEC".
- Términos más amplios o más precisos de un tesoro (en busca de "Nueva Inglaterra" encuentra "Vermont", "Maine", "Rhode Island", "New Hampshire", "Massachusetts" y "Connecticut" y viceversa).

Fujii (2008) analizó la distribución de términos de consulta que también aparecían en los textos de los enlaces que sirven de anclas, dado que los términos de palabras clave de tipo navegacional suelen contener esos *anchor texts* o textos de anclaje que llevan los enlaces hacia otras páginas. Pero hoy día coinciden menos con los términos de consulta, debido a que Google ha debilitado su potencial como factor de relevancia para el SEO al penalizar su uso excesivo, como práctica anti spam (Simmons, 2013).

Por su parte, en su estudio de 2008, Barr, Jones y Regelson se encontraron con que los nombres propios constituían el 40% de los términos de consulta estudiados en habla inglesa, y los sustantivos y nombres propios juntos constituían más del 70% de los términos de consulta.

Muramatsu y Pratt (2001) investigaron los modelos mentales de los usuarios de motores de búsqueda, en referencia a su uso de palabras vacías, el orden de las palabras dentro de una cadena de búsqueda (*Term Order Sensitivity*) o el uso de sufijos en expansiones de búsquedas, aunque básicamente lo que encontraron fue, de nuevo, la dificultad de éstos para interpretar los interfaces de los buscadores, y la interpretación que éstos hacían de las expresiones de búsqueda de los usuarios.

El estudio del lenguaje de los usuarios se ve mediatizado de nuevo por los cambios en la interfaz, o mejor dicho por los cambios en los algoritmos de búsqueda, que pasaron de buscar coincidencias o matching de relevancia entre palabras o frases usadas por los usuarios y las usadas por los contenidos a recuperar, a trabajar también con conceptos semánticos (Li y Xu, 2014; Pernas, 2017). Por eso, y de forma compatible con las investigaciones sobre lenguaje, aspectos cognitivos e interfaz, buena parte de las investigaciones han derivado en trabajar otro aspecto de las expresiones de búsqueda, el descubrimiento, identificación y caracterización de las intenciones de búsqueda de los usuarios o Query Intent (Simmons, 2013).

La aplicación última de este tipo de investigaciones es múltiple, pues sirve a los desarrolladores de sistemas de recuperación de información para interpretar las consultas más allá del puro texto y aparición de palabras coincidentes con la búsqueda (*matching*) y derivar hacia contenidos más relevantes, y a los diseñadores y creadores de contenido, para entender qué información crear y cómo presentarla para resultar más relevante a ojos del usuario que llega a él desde los motores de búsqueda.

2.3.5.6. Query Intent, Intención de Búsqueda

Nielsen argumentaba en 1997 que la mitad de los usuarios eran “*search-dominants*”, esto es, que utilizaban el buscador directamente para localizar la información que deseaban (Nielsen, 1997). Otros autores matizaban esta conducta, no como una estrategia única concreta versus el hojear o browsing, sino combinando ambas estrategias, en lo que se denominaría “*Exploratory Search*” o “búsqueda exploratoria” (Strohmaier et. al., 2007; Strohmaier et. al., 2008).

Marchionini definía el concepto de *Exploratory Search* (Marchionini, 2006) como el proceso de búsqueda que mezclaba *Querying* (búsqueda mediante palabras clave en un buscador) y estrategias de *Browsing* u “hojear”, tanto entre los resultados de esa búsqueda como dentro de los mismos, describiendo una serie de actividades dentro del proceso de búsqueda en el que apuntaba algunas motivaciones o intenciones de búsqueda, como navegación, transacción, verificación, respuesta a preguntas...

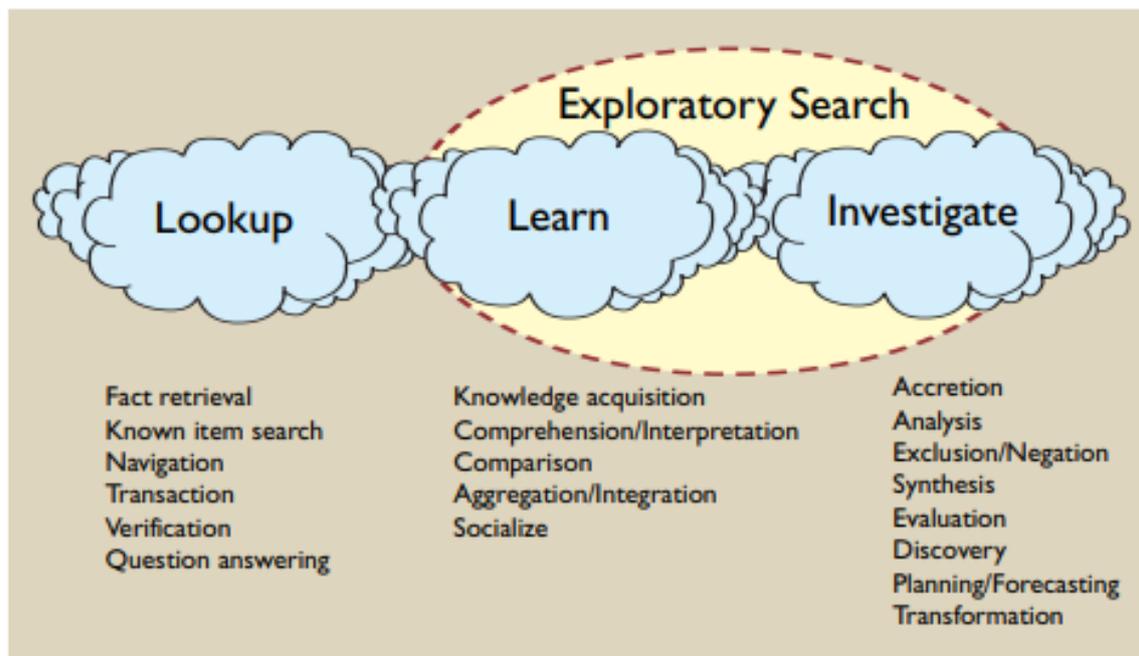


Fig. 2-33 Actividades de búsqueda. Fuente: Marchionini, 2006

Comprender así la intención que hay detrás de la consulta de un usuario en un motor de búsqueda, explicitada a través de una expresión de búsqueda (denominada *query* o *keyword* en la literatura al respecto) puede ser muy útil en multitud de tareas, como por ejemplo ayudar al motor de búsqueda a mejorar la recuperación de información, o ayudar a empresas e instituciones a entender cómo una población (o un segmento de ella) interpreta un dominio concreto a través de la formulación de sus inquietudes o necesidades de información (Richardson, 2005).

Lo más complejo en este sentido es representar, agrupar o clasificar las temáticas y las intenciones que hay detrás de las consultas o búsquedas. Después de los intentos previos, a partir de trabajos como los de Belkin, Shneiderman, Ellis y otros, serán finalmente trabajos sobre clasificaciones de la intención de las búsquedas (*User Intent*) como la de Broder (2002) o Rose y Levinson (2004) los que ganarían tracción entre la comunidad internacional, y que han derivado en multitud de estudios con aplicaciones distintas (Casado Sánchez, Lloret-Romero y Serrano-Cobos, 2011)

Andrei Broder (2002) en su *paper* seminal "*A taxonomy of web search*" afirmó que la Recuperación de Información clásica estaba "inherentemente predicada en los usuarios que buscan información", lo que denominamos "necesidad de información", concepto relacionado con el "estado anómalo de conocimiento" de Belkin (1980). Pero la necesidad detrás de una búsqueda en la web es a menudo no informativa. Así, Broder describe 3 categorías de intención de búsqueda en los usuarios:

1. **Navegacional.** La intención inmediata es llegar a un sitio web en particular.
2. **Informacional.** La intención es adquirir alguna información que supuestamente esté presente en uno o más contenidos web ("páginas web" en el original de Broder, aunque hoy día esa información puede no estar sólo en una página web per se).
3. **Transaccional.** La intención es realizar alguna actividad en Internet (comprar, alquilar, visitar... puede utilizar un verbo o no)

En su investigación inicial, realizó una encuesta y la comparó con la clasificación de 1.000 expresiones de búsqueda extraídas al azar del log de búsquedas de Altavista, motor de búsqueda mundial todavía activo en aquella época, dando como resultado que las búsquedas navegacionales eran compuestas por un 20% de las analizadas, por 48% de informacionales y un 30% de transaccionales, lo que tendría amplias consecuencias para el desarrollo del comercio electrónico posterior, donde la publicidad basada en buscadores se convirtió en una poderosa fuente de inversión publicitaria.

Type of query	User Survey	Query Log Analysis
Navigational	24.5%	20%
Informational	?? (estimated 39%)	48%
Transactional	> 22% (estimated 36%)	30%

Fig. 2-34 Clasificación de intencionalidad de búsqueda. Fuente: Andrei Broder, 2002.

Rose y Levinson (2004) ahondaron en esta clasificación primaria, desglosando las búsquedas informacionales en directas (y a su vez éstas en cerradas y abiertas) indirectas, consejos o sugerencias, localizaciones, y listas. Las transaccionales las renombraron como "de recurso", que según los autores serían expresiones de búsqueda en las que se buscaría obtener un recurso que no era información accesible vía página web, subclasificando las de esta tipología en búsquedas sobre "descargas", entretenimiento, interacción, y obtención.

SEARCH GOAL	DESCRIPTION	EXAMPLES
1. Navigational	My goal is to go to specific known website that I already have in mind. The only reason I'm searching is that it's more convenient than typing the URL, or perhaps I don't know the URL.	aloha airlines duke university hospital kelly blue book
2. Informational	My goal is to learn something by reading or viewing web pages	
2.1 Directed	I want to learn something in particular about my topic	
2.1.1 Closed	I want to get an answer to a question that has a single, unambiguous answer.	what is a supercharger 2004 election dates
2.1.2 Open	I want to get an answer to an open-ended question, or one with unconstrained depth.	baseball death and injury why are metals shiny
2.2 Undirected	I want to learn anything/everything about my topic. A query for topic X might be interpreted as "tell me about X."	color blindness jfk jr
2.3 Advice	I want to get advice, ideas, suggestions, or instructions.	help quitting smoking walking with weights
2.4 Locate	My goal is to find out whether/where some real world service or product can be obtained	pella windows phone card
2.5 List	My goal is to get a list of plausible suggested web sites (i.e. the search result list itself), each of which might be candidates for helping me achieve some underlying, unspecified goal	travel amsterdam universities florida newspapers
3. Resource	My goal is to obtain a resource (not information) available on web pages	
3.1 Download	My goal is to download a resource that must be on my computer or other device to be useful	kazaa lite name roms
3.2 Entertainment	My goal is to be entertained simply by viewing items available on the result page	xxx porno movie free live camera in l.a.
3.3 Interact	My goal is to interact with a resource using another program/service available on the web site I find	weather measure convertér
3.4 Obtain	My goal is to obtain a resource that does not require a computer to use. I may print it out, but I can also just look at it on the screen. I'm not obtaining it to learn some information, but because I want to use the resource itself.	free jack o lantern patterns ellis island lesson plans house document no. 587

Fig. 2-35 Clasificación de intencionalidad de búsqueda. Fuente: Rose y Levinson, 2004.

Investigaciones posteriores han tomado en consideración esta clasificación, siendo útil incluso para distintos idiomas (Wang, 2008) e incluso el propio Broder ha trabajado en diferentes aproximaciones, todo lo cual ha dado lugar a toda una subdisciplina (Gabrilovich, 2009) e incluso una competición, la Copa *KDD Cup on Web query classification* (Li, Zheng y Dai, 2005)

Baeza-Yates, Calderón y González-Caro (2006) realizaron una clasificación de User Goals o Query Intents algo distinta. En ella separaron las consultas de tipo "informativa", que coincidían con el original de Broder (2002), las de tipo "no informativa" (concepto que abarcaría a las búsquedas de tipo "resources" de Rose y Levinson (2004) y por tanto a las

transaccionales de Broder) pero además, incorporaron las búsquedas de tipo “ambiguo” (*Ambiguous Queries*) que definen como aquellas “cuyo objetivo o intención no puede inferirse directamente de la consulta (en algunos casos porque el usuario tiene un interés ambiguo)”.

Como se ve, las técnicas para clasificar expresiones de búsqueda son múltiples, aunque en los últimos años se centran en buscar métodos automáticos y escalables, como buscar los enlaces que se clican ante una expresión de búsqueda dada y generar grupos o *clusters* automáticamente (Beeferman y Berger, 2000) Baeza-Yates ha seguido investigando en agrupamiento de expresiones de búsqueda (*clustering*) a través del uso del modelo vectorial, integrando las selecciones del usuario y el contenido de los documentos seleccionados, con vistas a dos problemas: la mejora de los *rankings* de relevancia y la recomendación de consultas relacionadas (Baeza-Yates, Hurtado y Mendoza, 2007).

Asimismo, dos discípulas de Baeza-Yates atacaron el tema desde dos ángulos distintos: utilizando técnicas supervisadas de *machine learning* (González-Caro, 2011, González-Caro y Baeza-Yates, 2011) mediante correlaciones, y técnicas no supervisadas, estableciendo la última que “éste no es un problema unidimensional, sino que debería ser considerado como un problema de múltiples dimensiones, donde cada una de dichas dimensiones, o facetas, contribuye a clarificar y establecer cuál es la intención del usuario” (Calderón-Benavides, 2011, 6).

Li, Wang y Acero (2008) utilizan los clics en las respuestas, para clasificar la intención de búsqueda por medio del aprendizaje semi-supervisado, infiriendo la pertenencia de las frases de búsqueda a las clases de consultas, según su proximidad a aquellas etiquetadas, en grafos, mediante *Social Network Analysis*.

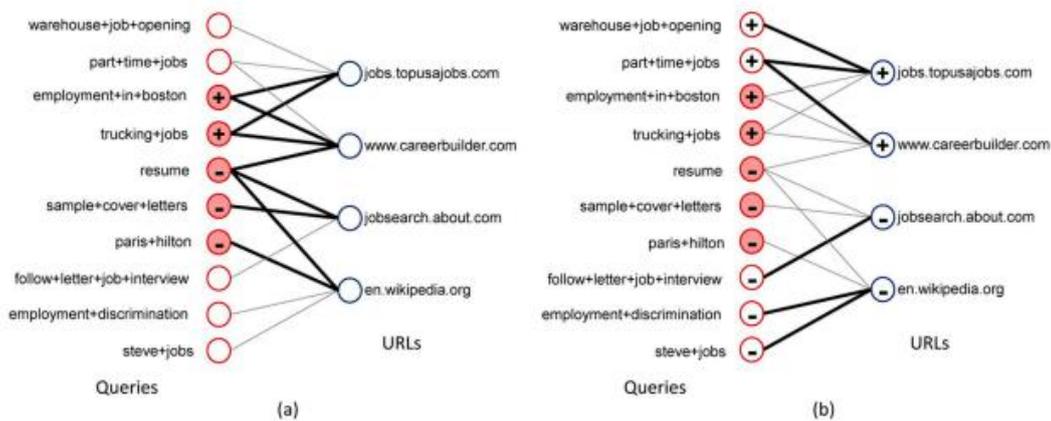


Fig. 2-36 Clasificación de intencionalidad de búsqueda mediante aprendizaje semi-supervisado. Fuente: Li, Wang y Acero, 2008.

Kong y Allan (2013) exploran grupos de términos semánticamente relacionados extraídos de los resultados de búsqueda para facetar los términos de búsqueda, enseñando a un sistema supervisado de la probabilidad de que un término sea candidato para cada faceta. Consideran que la inferencia exacta es intratable, por lo que capturan la probabilidad de que dos términos se agrupen en una faceta, y así sus dependencias.

Hu et.al. (2012) estudian, dentro de las búsquedas multifacéticas, la identificación de subtemas de consulta, denominados “subtemas por búsqueda” (*subtopics per search*) y “aclaración de subtítulos por palabra clave” (*subtopic clarification by keyword*). Al analizar los clics de los usuarios en las respuestas ante las búsquedas, determinaron que cuando un usuario hacía clic en varias URL en una consulta, las URL clicadas tendían a representar la misma intención o faceta (“subtema por búsqueda”), agrupando subtemas por agregación de palabras adicionales al expandir la consulta cuando el usuario aclara su intención de búsqueda, siendo estas palabras añadidas un indicativo de la faceta. A su vez, Bang aportó en 2007 un esquema de facetación de intenciones de búsqueda que constaba de cuatro facetas genéricas: ambigüedad, sensibilidad de autoridad, sensibilidad temporal y sensibilidad espacial (Bang, 2007).

Una de las técnicas de análisis más conocidas en los últimos años para determinar los patrones de uso de la información interactiva, y en concreto de los motores de búsqueda, es la técnica del *Eye-tracking*, técnica que captura el movimiento del ojo al mirar a la pantalla y también ha sido usada en conjunción con las clasificaciones de intención de búsqueda.

Un conocido estudio de Nielsen (2006) mostró cómo el patrón de interacción más común entre los usuarios analizados daba lugar a una visualización de los datos en forma de F (*F-Shape*) según el cual la mayoría de los usuarios ponen el foco más tiempo en los primeros resultados que aparecen en la zona superior de la SERP, esto es, arriba a la izquierda, sucediendo lo mismo columna de información a columna (como se puede ver en la imagen 2-32, en este caso en la columna que muestra los anuncios por palabra clave o *adwords*) independientemente de la intención de búsqueda del usuario.

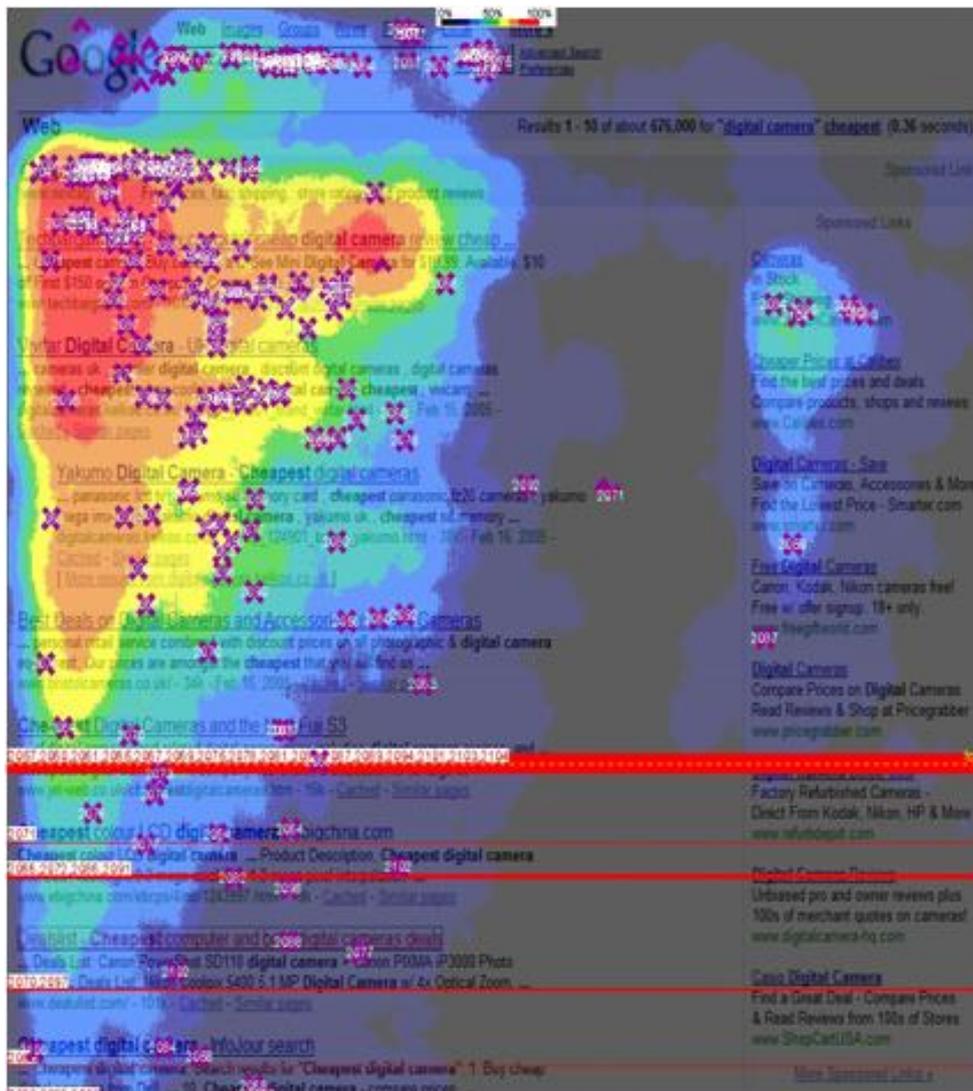


Fig. 2-37 Visualización del análisis con Eye-tracking de una página de respuestas (SERP) en Google. Fuente: Jakob Nielsen, 2006.

Esta técnica de estudio de la interacción se ha utilizado también atendiendo a su intención de búsqueda. Así van Gisbergen, van der Most y Aelen (2003) analizaron mediante eyetracking la forma que tenían los usuarios de mirar a la página de los resultados, pero en este caso, dependiendo de si las consultas eran informacionales o transaccionales.

Pero usando esta técnica también se han detectado diferencias en la interacción de los usuarios con los resultados de búsqueda, dependiendo de la intención de la misma. Van Gisbergen, Van der Most y Aelen, (2003) estudiaron el tiempo que los usuarios pasaban mirando los resultados de búsqueda para cada tipología de intencionalidad, lo que visualizaron en gráficos de tres dimensiones (ver imágenes 2-38 y 2-39), donde se puede apreciar que cuanto más tiempo mirando pasaban los usuarios estudiados, más alta es la montaña tridimensional del gráfico. En el caso de las búsquedas navegacionales, seguían claramente y de forma muy acusada el modelo *F-Shape*. Los usuarios centraban más su atención en los primeros resultados (arriba y a la izquierda, por cada columna de texto) lo que sugiere que los usuarios que buscan información en un motor de búsqueda de Internet, desean respuestas rápidas, sin detenerse a hojear entre las posibles respuestas.

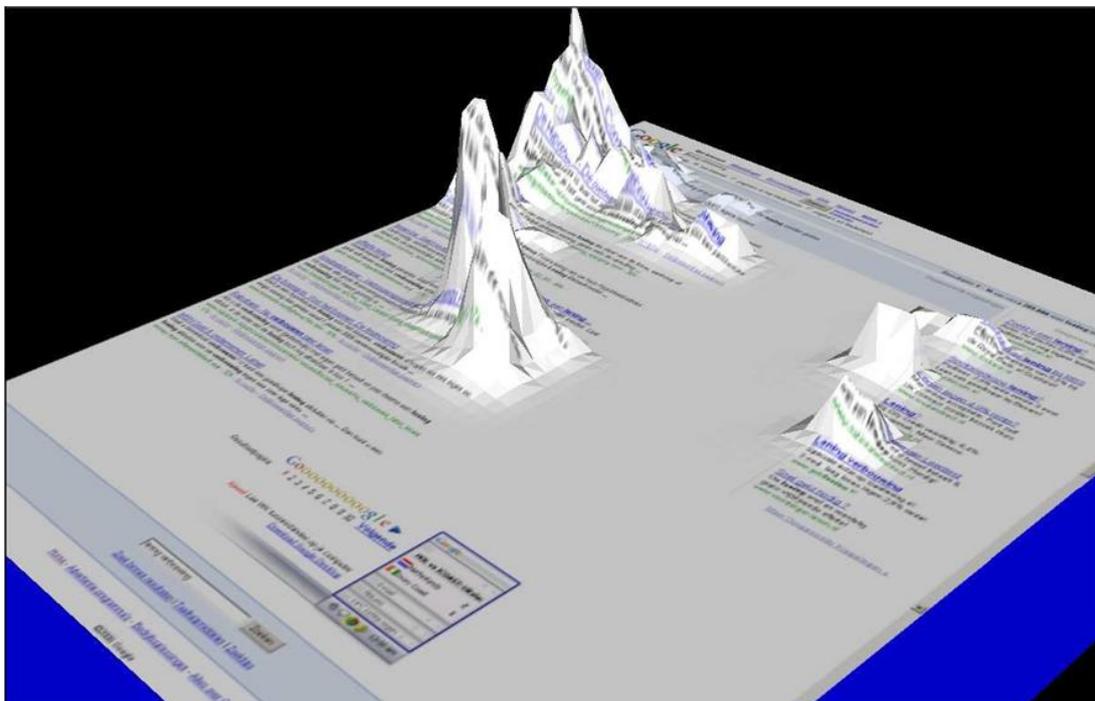


Fig. 2-38 Visualización de análisis con Eye-tracking de búsquedas informacionales . Fuente: Van Gisbergen, Van der Most y Aelen, 2003.

Pero en el caso de las búsquedas transaccionales, se observó que los usuarios, aunque también seguían el modelo *F-Shape*, repartían más la visualización de las respuestas, entre las que también estaban más abajo. Es decir, que no se fiaban tanto de las primeras respuestas, y comparaban más resultados para determinar que enlace iban a clicar.

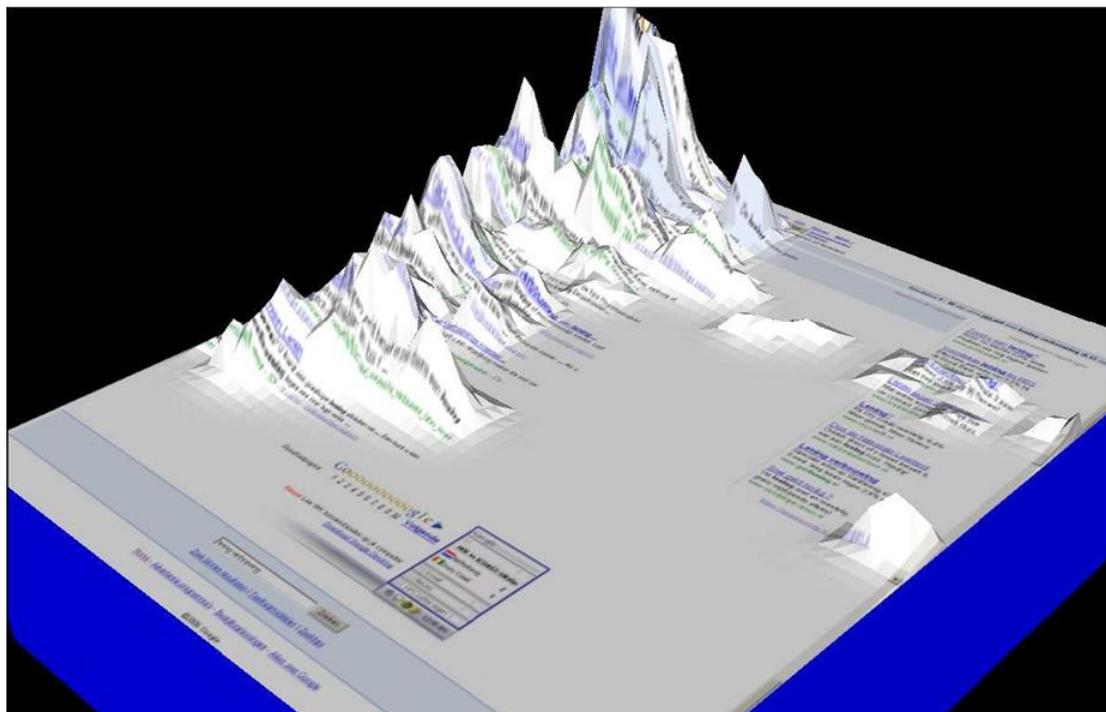


Fig. 2-39 Visualización de análisis con Eye-tracking de búsquedas transaccionales. Fuente: Van Gisbergen, Van der Most y Aelen, 2003.

Joshi et.al (2007) propusieron agrupar las expresiones de búsqueda semánticamente, para construir superconceptos, formados por familias semánticas basadas en el lenguaje natural de los usuarios, y así mejorar la recuperación de información por conceptos, en lugar de por palabras clave. Hu et.al. (2009) usaron los conceptos de la Wikipedia para clasificar las intenciones de búsqueda en torno a ellos. Wolk y Wöber (2008) utilizaron la técnica de Escalamiento Multidimensional o *Multidimensional Scaling* (MDS) para construir un mapa de las expresiones de búsqueda similares que generaban los usuarios del portal visiteuropeancities.info, para entender qué términos acompañaban de forma más común a las distintas ciudades turísticas buscadas. Ashkan y Clarke (2009) utilizaron las palabras que aparecen en las respuestas de las *SERP* para cada *query*, para extraer su intención de

búsquedas y clasificarlas en búsquedas navegacionales, informacionales y transaccionales, en particular, búsquedas comerciales.

En otras ocasiones, se utilizan técnicas de diferenciación gramatical o sintáctica (Olvera-Lobo y Robinson-García, 2009; Herdagdelen et. al., 2010) y vocabularios controlados de temáticas verticales, para comparar los términos del vocabulario con los del *search log* estudiado, identificando las coincidencias y agrupando en consecuencia. En el caso de Tsur et. al. (2016) localizaban así un tipo concreto de búsquedas informacionales, las denominadas *CQA-intent queries*, debido a que las respuestas a esas preguntas eran contestadas habitualmente en comunidades sociales (*Community Question Answering* o CQA), como foros o medios sociales especializados en preguntas-respuestas, aunque el método se combinó con la posición de ciertas palabras que denotaran una interrogación, como por ejemplo “*where can I sell mercury*” o “*how to get a boxer body*”.

Más allá, Pinter, Reichart y Szpektor (2016) realizaron una subagrupación de este tipo de búsquedas mediante análisis sintáctico de expresiones de búsqueda en inglés, detectando cuatro tipos de construcciones en las frases de tipo CQA: frases completas (*full sentence*), sentencias incompletas o rotas (*incomplete or broken sentence*), frases complejas (*complex phrase*) y frases sintácticamente desconectadas (*syntactically disconnected phrases*) como puede apreciarse en la imagen 2-35.

	Type	Example queries
1	Full sentence	how many bags of food does a horse eat; what does bold mean; my spleen hurts when i walk
2	Incomplete or broken sentence	muscle in leg is called; why page takes so long to load; how to find rate
3	Complex phrase	bed sheet that goes with pink and white room; inability to make eye contact
4	Syntactically disconnected phrases	modem internet off light; missing malaysia airplane psychic; resignation letter unhappy

Fig. 2-40 Ejemplos de distintas composiciones de expresiones de búsqueda de tipo CQA. Fuente: Pinter, Reichart y Szpektor, 2016.

Otra técnica de extracción de clasificación de la intención en las búsquedas es el Procesamiento del Lenguaje Natural (PLN) una disciplina que “relaciona directamente la informática con la lingüística. La misma persigue como objetivo, poder conseguir que el lenguaje coloquial (el lenguaje de uso cotidiano de todos nosotros) pueda ser utilizado como una entrada en un sistema informático” (Zamoszczyk et al., 2012).

Más allá de las clasificaciones de intenciones, otros científicos han investigado la interacción antes y después de las búsquedas. Más que la semántica en sí, modelizan las interacciones resultantes, pues responden a distintas necesidades de información mediante reyes bayesianas (Piwowarski, Dupret y Jones, 2009) Una variante de esta técnica es la inclusión de la distribución de los textos de los enlaces en el análisis de las interacciones en los resultados de búsqueda o *User-Click Behavior* (Lee, Liu y Cho, 2005).

Brenes y Gayo-Avello (2009) realizaron un análisis estratificado, agrupando expresiones de búsqueda y sesiones de acuerdo a su frecuencia de clicks. Liu et. al. (2009) utilizaron el análisis de las elecciones de respuestas de los usuarios en las páginas de respuesta de los buscadores ante búsquedas con intención informativa (*Click-Through Data Analysis*) para evaluar automáticamente la relevancia de un motor de búsqueda (en este caso chino).

Jansen, Booth y Spink (2007) descubrieron que más del 80% de las expresiones de búsqueda realizadas en Internet son informacionales, siendo aproximadamente el 10% navegacionales, transaccionales o ambiguas. Anteriormente, Jansen y Spink (2006) habían comprobado que los usuarios estaban viendo menos páginas de resultados que en los primeros años de la aparición de los motores de búsqueda en Internet, que los usuarios americanos utilizaban más operadores de consulta que los de Europa, y que los resultados cambiaba drásticamente de un motor de búsqueda a otro, lo que fue confirmado por Mahoui y Cunningham (2009). Esta segmentación de usuarios se ha estudiado también, dependiendo de las posibilidades técnicas de consulta de los *logs* de los buscadores. Por ejemplo, Efthimiadis (2008) estudió los hábitos de búsqueda de los griegos, analizando entre otros indicadores las frecuencias de búsqueda, distribuciones de términos, tipos de lenguaje de consulta usados, y patrones de longitud de las distintas búsquedas.

Es más, algunos como Alspaugh et. al (2014) realizan este tipo de estudios en herramientas que no son buscadores per se, como la herramienta de análisis de datos Splunk, descubriendo elementos de interacción similares al uso de motores de búsqueda, pero asociados a las posibilidades de filtrado de interacción (sinónimo de reformulación de búsqueda, si se llevara al dominio de los motores de búsqueda).

Uno de los problemas técnicos de este tipo de estudios es que, en sus esfuerzos por mejorar la experiencia del usuario, los algoritmos de los motores de búsqueda y sus interfaces se han ido adaptando, recogiendo la información estadística de la interacción de los usuarios y actuando ante ella (Amitay, 2006). Esta práctica conlleva cambios en la interacción forzados por la

herramienta de consulta (Bilder, 2006), incluso entre usuarios con gran experiencia en el uso de buscadores o con gran dominio de la temática sobre la que se busca (Jamali y Asadi, 2010) lo que presumiblemente seguirá sucediendo con la entrada de los interfaces de búsqueda por voz (González, 2016).

Debido a que esos factores cambian con el tiempo, algunas investigaciones atacan el problema de entender cómo busca el usuario midiendo su interacción a lo largo de períodos de tiempo más o menos extensos, a través de estudios longitudinales que trabajan con los datos que pueden recoger de los archivos logs de los buscadores, cuando éstos se lo permiten. Estos estudios tendrán limitaciones técnicas, como la anonimización de la identidad de los usuarios y sus sesiones, (Korolova et. al., 2009) pero permiten su análisis semántico (Reisinger y Pasca, 2007) y en ocasiones, permiten el uso de datos masivos, lo que permite una gran escalabilidad, dependiendo de la investigación (Wang, Berry y Yang, 2003).

2.3.6. Análisis longitudinal de la búsqueda

Como se ha visto, los motores de búsqueda responden a una necesidad concreta de información en un contexto determinado para cada usuario, definido por diversos parámetros, entre ellos el momento de la búsqueda, utilizando un dispositivo concreto y a través de una interfaz concreta (Fishkin, 2017), lo que obliga a ser precavidos en la generación de conclusiones generalizadas. El análisis longitudinal puede brindar otra dimensión a los análisis de los usuarios en este contexto, pues supone un estudio de series de observaciones obtenidas sobre los mismos objetos de investigación, a lo largo del tiempo (Belle et. al., 2004).

Este tipo de análisis puede hacerse tanto en sitios web o intranets, donde se podrán controlar mejor las distintas variables modificadoras de las conductas estudiadas (Stenmark y Jadaan, 2006) como en motores de búsqueda generalistas (Google News Lab, 2016) Sus descubrimientos, podrán afectar incluso al diseño de los algoritmos de recuperación de información, puesto que, al añadir la variable temporal, lo que es relevante hoy, puede no serlo mañana (Moulahi, Tamine-lechani y Yahia, 2016)

El aspecto más frecuente a investigar a lo largo del tiempo en el análisis de los search logs es el lenguaje usado por los usuarios (Kulkarni et. al., 2015). Las frases que los usuarios han utilizado para expresar su necesidad de información es algo que los search logs pueden ofrecer, en ocasiones gratuitamente. Para los investigadores, las limitaciones de usar el lenguaje natural empleado en esos search logs, vendrán dadas porque, a no ser que el estudio esté realizado por el dueño del archivo de consultas, normalmente los datos estarán anonimizados, y no se podrá seguir la evolución de las expresiones utilizadas por un usuario dado, sino la evolución del uso de una expresión de búsqueda por cientos, miles o millones de usuarios no identificados, a lo largo del tiempo analizado (Jones et. al., 2007).

Es decir, a menos que sean estudios más concretos y por tanto con menor representatividad estadística (Wu, Turpin y Zobel, 2008), el investigador que utiliza datos de fuentes como Google u otros buscadores generales, no sabrá si lo que está investigando (la expresión de búsqueda) ha sido utilizado por un mismo usuario muchas veces a lo largo de su historia como usuario del motor de búsqueda, o si la expresión es utilizada esporádicamente por muchas personas. En estos estudios encontraremos ambas formas de análisis, unos más cualitativos, en los que sí se puede medir la interacción de ciertos usuarios (Joho, Jatowt y Blanco, 2015) y otros en los que se sigue es la cantidad de veces que se busca una expresión para una

segmentación dada (país de origen de las búsquedas, lengua, marco temporal...). En algunos casos, también se puede encontrar información de interacción anonimizada (Beeferman y Berger, 2000) lo que permite combinar variables (lenguaje, tiempo, interacción).

Una vez delimitado este aspecto, el análisis longitudinal de expresiones de búsqueda permite aportar más información al estudio de las expresiones de búsqueda, al comparar tanto la evolución del lenguaje natural de búsqueda, analizado los términos utilizados, como en ocasiones la evolución paralela de los interfaces y algoritmos de recuperación de información, a partir del estudio de los resultados de búsqueda (Jones y Diaz, 2007) para cada término en distintos momentos (sea del día, mes, año, década...) incluso en tiempo real (Jansen, Campbell y Gregg, 2010), Por ejemplo, en 2002, Cothey detectó cómo los usuarios de un motor de búsqueda cambiaban su conducta (a lo largo de 5.431 días de estudio) conforme ganaban experiencia.

Uno de esos aspectos es la expansión de búsquedas, o Query Suggestions. Esto ocurre en ciertos buscadores, donde ante palabras con un significado ambiguo o polisémico, el motor de búsqueda ofrece alternativas, mediante un proceso de desambiguación semántica, para interpretar el sentido más apropiado de la expresión de búsqueda, en un contexto determinado, y así aportar las respuestas más relevantes. Song et. al. (2009) estimaron que aproximadamente el 16% de las expresiones de búsqueda en un buscador dado, son ambiguas. Por eso, los motores de búsqueda como Google usan fuentes de información semántica o recursos lingüísticos como diccionarios, tesauros y ontologías, como WordNet (Pinto, 2008) con los que realizar una expansión de consultas relacionadas con la que mejorar la interpretación del sentido de las mismas.

Esta línea de investigación lleva al descubrimiento de búsquedas relacionadas entre sí, por diversos motivos, no sólo semánticos. Shi y Yang (2006) utilizaron reglas de asociación basadas en el histórico de las expresiones usadas anteriormente, técnica que también usaron Xue (2002) et. al. para mejorar la relevancia de las búsquedas en un buscador interno (el motor de búsqueda que se utiliza dentro de un sitio web, o dentro de una intranet) Wang y Zhai (2008) buscaron la adición de términos basados en los patrones de contexto de las búsquedas (Context-Sensitive Term Addition Patterns) He et. al. (2009) se concentraron en comparar los patrones de búsquedas de los usuarios con los patrones históricos globales de búsquedas, para realizar recomendaciones predictivas de términos a los usuarios, en base a su semejanza con esos patrones globales, como también han investigado Zhang y Nasraoui (2008) o Naccarato et. al. (2018).

El análisis longitudinal de búsquedas se utiliza hoy día en muy diversos sectores económicos, con distintos usos. Por ejemplo, se usa para predecir cambios en la bolsa basados en las búsquedas lanzadas por los usuarios sobre finanzas (Ball, 2013), o más allá, para generar el *Google Inflation Search Index* (GISI) que emplea indicadores cibernéticos en contraposición o como contraste de estudios cualitativos (vía encuesta a expertos) de previsión de inflación económica, como el *Livingston Survey* o el *Michigan Survey* (Guzmán, 2011), o a técnicas basadas en la de recopilación automatizada de información o scraping desde contenidos digitales (Edelman, 2012), como por ejemplo Ormerod, (2013) para calcular el *Michigan Consumer Sentiment Index*, un índice psicológico usado en Economía.

Así, en el caso de Preis et. al. (2013) su estrategia para gestionar bolsa utilizando patrones de búsqueda en Google Trends, les produjo un beneficio de un 326% (línea azul en el siguiente gráfico), en comparación con los métodos tradicionales de gestión de activos de bolsa.

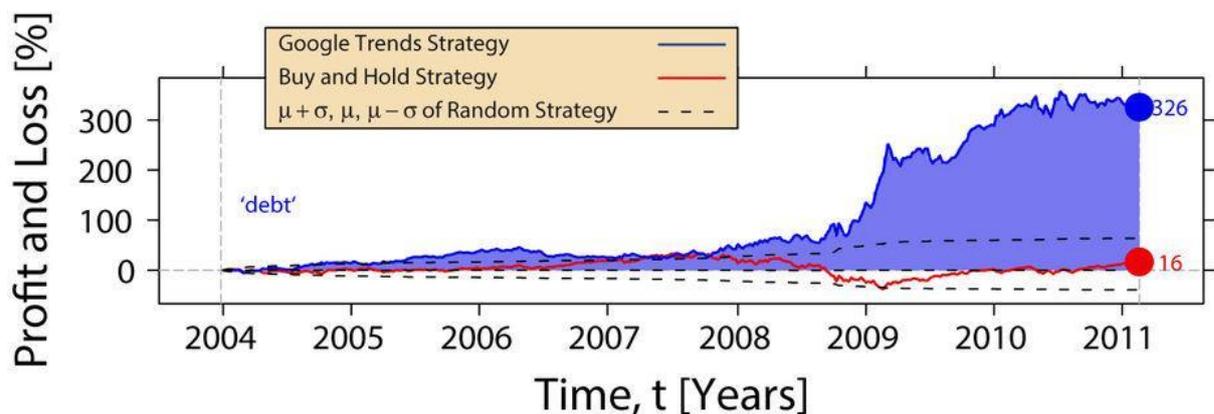


Fig. 2-41 Beneficios y pérdidas de estrategia de inversión en bolsa utilizando Google Trends. Fuente: Preis et. al. (2013)

También se ha utilizado en otros ámbitos, con fines predictivos, como el éxito de películas, ventas de videojuegos y ranking de canciones (Goel et. al., 2010) el interés por la comida o por ciertos ingredientes (Google News Lab, 2016) por la biodiversidad (Troumbis, 2017) o por el turismo (Pan, Litvin y O'Donnell, 2007), donde se refleja que los usuarios no siempre identifican correctamente lo que buscan, y expresan su necesidad a través de un concepto con mayor tasa de recuerdo, como por ejemplo preguntar por los guerreros de terracota, en lugar de preguntar por el mausoleo de Qin Shi Huang (Wu, 2017) La propia empresa Google utiliza

esa información para realizar visualizaciones de datos que ayuden a público e investigadores a descubrir patrones en los datos (Google News Lab, 2017)

Muchas de estas investigaciones se han dirigido al tema de la salud, preguntándose cómo podía utilizarse la medición del interés de los usuarios por ciertas informaciones relacionadas, para mejorar la información sanitaria (Schootman et. al. 2015), prevenir enfermedades (Carneiro, Mylonakis, 2009; Cavazos-Rehg et. al. 2015), o predecir su propagación (Johnson, Mehta, 2014). Quizá uno de los casos más sonados al respecto, es el análisis de expresiones de búsqueda relativas a la gripe mediante la herramienta *Google Flu Trends*, para predecir la expansión geográfica y temporal de la enfermedad (Google.com, 2008; Ginsberg et. al., 2009; Pervaiz et. al., 2012)

Google Flu Trends generó controversia e incluso respuesta en *Nature* (Butler, 2013) no en sí por su metodología en el momento de obtenerse las primeras conclusiones, sino por los efectos causales de los contenidos referidos a la enfermedad en años posteriores, ya que los informes y noticias en prensa aparecidos, pudieron haber desencadenado muchas búsquedas relacionadas con la gripe por personas que no estaban en realidad enfermas, lo que provocó que el sistema ofreciera estimaciones muy por encima de los datos del CDC estadounidense (*Center for Disease Control*). Casos como éste nos indican que los datos de *Google Trends*, aunque útiles para entender el lenguaje de búsqueda de los usuarios, no siempre servirán para predecir comportamientos; puesto que en esos casos, la conducta se verá afectada por factores externos, eventos concretos y/o impredecibles.

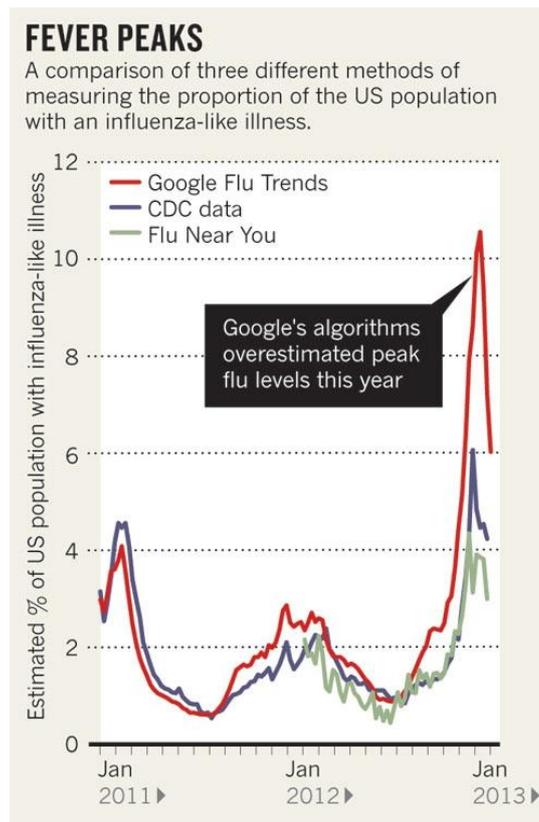


Fig. 2-42 Comparación de datos de Google Flu / CDC / Flu Near You. Fuente: Butler, 2013

Aunque no hay unicidad o interoperabilidad en los search logs de distintos motores de búsqueda a la hora de presentar su estructura de datos (Veríssimo, 2009), las características de estos datos (cuando se pueden explotar) en ocasiones permiten llegar a caracterizar aspectos muy concretos de las conductas de búsqueda de los usuarios, como las horas del día en que más busca una expresión de búsqueda (Beitzel et. al., 2007, Bridwell, 2011) mientras que otros investigadores se fijan en la regularidad de la frecuencia de búsqueda de las expresiones a lo largo de largos períodos de tiempo (Bauckhage y Manshaei, 2014).

Los estudios longitudinales pueden combinar también el análisis lingüístico de las expresiones de búsqueda con las respuestas clicadas en la página de respuestas del buscador, *SERP* o *Search Engine Response Pages* (Tseng, Yun-Fei y Hsu, 2006; Jin, Zhou y Mobasher, 2004), de forma que se pueda contrastar la calidad de las respuestas en función del *Clickthrough Ratio*, es decir el porcentaje de clics en los enlaces de esas respuestas para esas expresiones de búsqueda dadas, sobre el total de impresiones o visualizaciones, a lo largo de distintas sesiones de búsqueda (Tan, Shen y Zhai, 2006; Park, 2009). En otras ocasiones, se puede combinar la incidencia de las búsquedas con noticias u otros datasets (como la temperatura de

cada día, volúmenes de ventas o similares), para buscar correlaciones no obvias (Google Labs, 2011; Vanderkam, 2011), o con menciones de marcas en medios sociales (Mukherjee y Jansen, 2016).

Otros como Lau y Horvitz (1999) combinaron la longitud de las expresiones de búsqueda con la frecuencia de búsqueda y las intenciones de búsqueda, durante procesos de reformulación de las búsquedas o *Query Refinement*, cuando el usuario no está satisfecho con los resultados obtenidos y cambia algún aspecto de su expresión de búsqueda original, bien en todo o en parte. En su caso describieron esta tipología de reformulaciones de búsqueda:

- Nuevo (*New*): búsqueda de un tema no buscado previamente por este usuario dentro del alcance del conjunto de datos (veinticuatro horas).
- Generalización (*Generalization*): una consulta sobre el mismo tema que la consulta anterior, pero buscando información más general que la consulta anterior.
- Especialización (*Specialization*): Una consulta sobre el mismo tema que la consulta anterior, pero buscando información más específica que la consulta anterior.
- Reformulación (*Reformulation*): Una consulta sobre el mismo tema que puede ser visto como ni una generalización ni una especialización, sino como una reformulación de la consulta previa.
- Interrupción (*Interruption*): una consulta sobre un tema buscado anteriormente por un usuario que ha sido interrumpida por una búsqueda en otro tema (realizada por el mismo usuario, o por otro usuario que utiliza la misma sesión del mismo dispositivo).
- Solicitud de resultados adicionales (*Request for Additional Results*): Solicitud de otro conjunto de resultados en la misma consulta del servicio de búsqueda.
- Consultas en blanco (*Blank Queries*): páginas de respuestas que no contienen ninguna consulta. Estas páginas surgen cuando un usuario hace clic en el botón de búsqueda sin especificar ninguna consulta.”

En este apartado se ha visto cómo el análisis longitudinal de las expresiones de búsqueda puede servir a diversos propósitos. Por ejemplo, para intentar predecir el éxito del motor de búsqueda (Shah, Hendahewa y González-Ibáñez, 2016) es decir, mirando a futuro, o prediciendo los intereses de los usuarios (Choi y Varian, 2012) pero también mirando al pasado (Zhao, Bhowmick y Gruenwald, 2005) considerando el log de búsquedas como un archivo del pasado. Cabe preguntarse ahora, si para el tema que ocupa a esta investigación, hay investigaciones relacionadas con el estudio de los hábitos de lectura, desde el área del análisis de la conducta del usuario, o más allá, desde el estudio de la conducta de búsqueda del usuario, que se puedan aplicar.

2.3.7. *User search behaviour* en el sector de la lectura

Una vez se han desglosado distintas vertientes investigadoras que tratan la conducta de búsqueda del usuario, se ha explorado la posibilidad de observar investigaciones de esta rama de la ciencia que apliquen al caso de la búsqueda de libros en Internet. Así se descubren investigaciones que por un lado analizan los hábitos de búsqueda de lectura en motores de búsqueda generalistas, como Google o Bing, pero también en ocasiones y hasta donde cada tecnología o proveedor de datos lo permita, en buscadores de portales especializados (Baye, Santos y Wildenbeest, 2013) o librerías online como Amazon o Barnes & Noble.

La importancia del análisis de la conducta de búsqueda de lectura de los consumidores es crucial para los modelos de negocio digitales de editoriales y librerías online, pues conocer lo que buscan los lectores potenciales les permitirá crear contenido que tenga más probabilidades de ser encontrado a través de los motores de búsqueda (Arnold, 2006), y además, porque cada vez éstos buscan con mayor frecuencia a través de menos canales de información, de forma más centralizada, lo que modifica la política de precios de las editoriales y librerías, y transforma así todo el sector editorial y del libro, así como otros sectores (Badia, 2010).

Este efecto de transformación del sector, debido a la incidencia en este caso del entorno digital como factor de disminución de barreras entre el usuario y el producto final (Ministerio de Educación, Cultura y Deporte, Observatorio de la Lectura y el Libro, 2016), ya fue descrita por Stigler en su artículo seminal *The Economics of Information* (1961) antes de la era Internet. Stigler postulaba que cuanto mayor fuera el coste de búsqueda de un artículo (debido a la diversidad de fuentes en las que un usuario puede buscar para comparar precios) siendo otros factores iguales, mayor es la dispersión de los precios.

Su modelo ha sido criticado y ampliado (Baye et. al., 2006) y se han multiplicado los factores que deben aplicarse al análisis de precios hoy día. Pero es cierto que los motores de búsqueda generalistas como Google, o especializados como Amazon, centralizan esas búsquedas, por lo que el costo de búsqueda disminuye y también lo hace la dispersión de los precios. Brynjolfsson, Hu y Simester (2011) atestiguan ese efecto, reflejado en el concepto de *Long Tail* o Larga Cola de Internet, el cual no se debe únicamente al aumento de la selección de productos, sino que también puede deberse al menor coste de búsqueda en los principios del siglo XXI gracias a los motores de búsqueda anteriormente mencionados.

En ese interés por conocer los hábitos de búsqueda de lectura de los usuarios, se puede ver cómo en 2003 Chevalier y Goolsbee constataban que las grandes librerías online tenían una selección mucho mayor de títulos que incluso las mayores librerías físicas, que pueden tener hasta 150.000 títulos, mientras que Amazon y BN.com afirmaban tener millones de títulos disponibles, lo que aumentaba sus probabilidades de posicionamiento para más búsquedas de Larga Cola. Con el fin de obtener una mejor imagen de la naturaleza de los términos de búsqueda que más se usaban para acceder a esas librerías online (Amazon y Barnes & Noble) desde los motores de búsqueda, Baye, Santos y Wildenbeest (2013) agruparon más de 2.000 términos de búsqueda relacionados, y los agruparon en 5 categorías:

1. Nombre del sitio, lo que incluye el nombre o la dirección URL (también errores ortográficos) del sitio,
2. términos sobre productos que no fueran libros,
3. libro, lo que incluía libros físicos y libros electrónicos,
4. Lector electrónico, incluyendo búsquedas de hardware o software relacionado con el dispositivo, y
5. otros, lo que incluía nombres y direcciones URL de otras librerías online (incluyendo errores ortográficos).

Un hábito de búsqueda interesante detectado en su estudio, es el uso mayoritario de búsquedas navegacionales para buscar y navegar hacia librerías online con mucho *branding* o marca, como son para el mercado americano Amazon o Barnes & Noble, siendo para ambas librerías la búsqueda del nombre del sitio web la opción más utilizada.

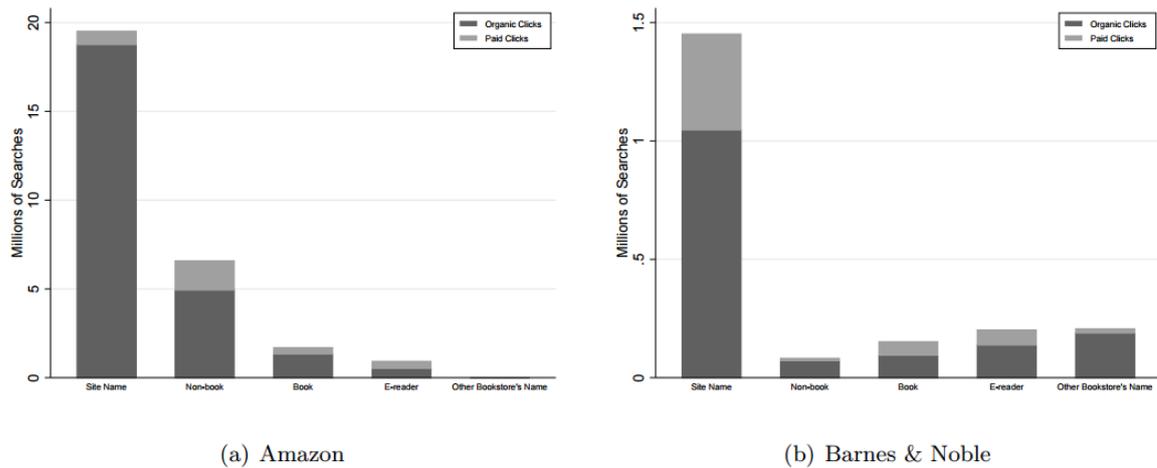


Fig. 2-43 Búsquedas navegacionales para Amazon y Barnes & Noble, con datos de comScore, 2012. Fuente: Baye, Santos y Wildenbeest, 2013.

Otro efecto interesante fue la constatación de la caída en búsquedas de títulos de libros en Google a lo largo de los meses, un patrón que debería ser tenido en cuenta por editores y responsables de webs de librerías y bibliotecas, tanto para el diseño de colecciones como para las acciones de marketing futuras basadas en la oportunidad o probabilidad de que la mejora del posicionamiento para esas búsquedas específicas sobre títulos concretos, atraigan más tráfico a sus sitios web.

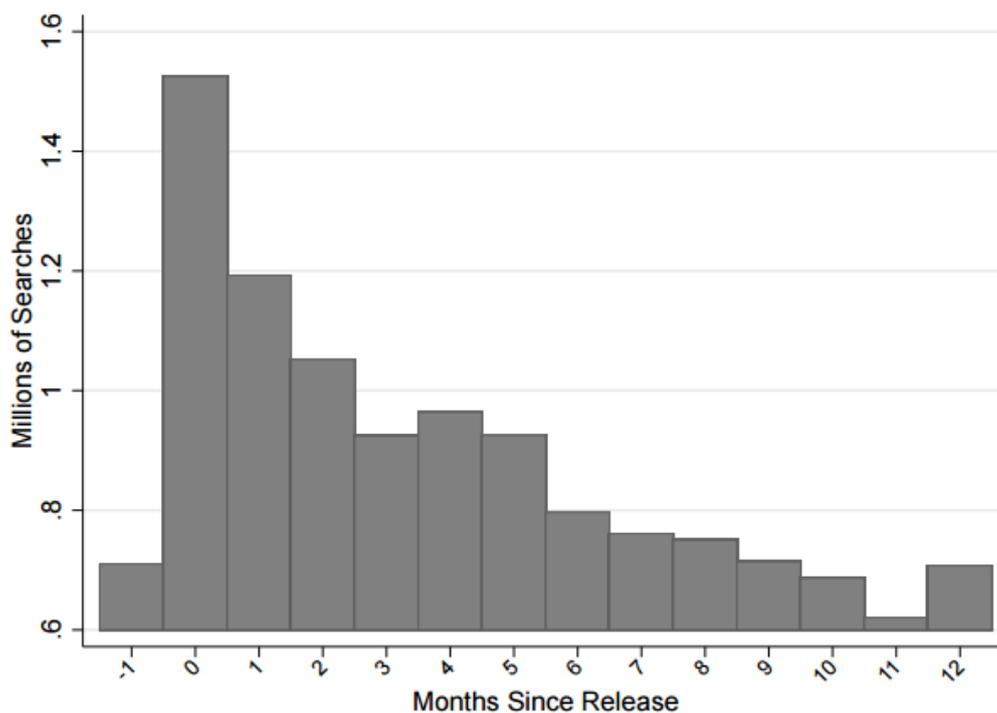


Fig. 2-44 Análisis longitudinal de búsquedas sobre títulos de libros en Google tras la salida del título. Fuente: Baye, Santos y Wildenbeest, 2013.

Las conclusiones de estas investigaciones deberían tener efectos prácticos para las librerías digitales, las sedes web bibliotecarias y otro tipo de creadores digitales, en su necesidad de mejorar el posicionamiento en buscadores de sus contenidos (ya sean fichas de libros, revistas, videojuegos, películas u otros productos culturales). En todos esos casos, la comprensión de las conductas de búsqueda de los usuarios puede ayudarles en varios aspectos, desde dinámicas de organización de información a la propia creación de los contenidos.

En particular, es vital en proyectos de optimización y posicionamiento en buscadores, los cuales comienzan entendiendo las necesidades de los compradores potenciales, para posteriormente realizar una auditoría de la sede web (Cueto-Álvarez-De-Sotomayor et. al., 2004) que, en función de cómo responde ésta a un conjunto de factores de mejora del ranking en buscadores, consensuados por la comunidad de *webmasters*, determinará una serie de acciones a seguir, en donde la caracterización del usuario permitirá segmentar el posicionamiento, siguiendo la personalización que hacen los motores de búsqueda actuales (Jansen et. al., 2009).

Como se ve, esta auditoría lo será a partir del conocimiento de las expresiones de búsqueda o consultas que realicen los usuarios que más interés captar a una sede web a través de los motores de búsqueda, así como las entidades semánticas que integran (Macias-Alegre, 2016), sus intenciones tras la búsqueda, sus motivaciones, su segmentación y su contexto de búsqueda. Por tanto, será útil entender al usuario (potencialmente interesado, lo que se denomina “usuario cualificado”, por contraposición al usuario al que se llega por casualidad y que tendrá menos probabilidades de estar interesado en el producto, servicio o contenido al que haya llegado) y sus hábitos de consulta (Serrano-Cobos et. al., 2016).

Las acciones a realizar en el posicionamiento en buscadores de una sede web podrán ser, tanto técnicas como estratégicas, tanto dentro del sitio web (lo que se denomina *SEO onsite*) como fuera del sitio web (*SEO offsite*) para obtener más y mejores enlaces entrantes y señales sociales de interés y *engagement* por parte de la audiencia más relacionada temáticamente o en virtud a otras segmentaciones de utilidad para la personalización y socialización del algoritmo del ranking del buscador. De esta forma, la sede se alineará con el objetivo último de relevancia y pertinencia del algoritmo del buscador, que pretende optimizar la asignación de relevancia para minimizar la tasa de rebote o *bounce rate* desde la página de respuestas del buscador (Serrano-Cobos, 2015).

Hay diversas modalidades de posicionamiento en buscadores (Enge, Spencer y Stricchiola, 2015), tanto en función de los intereses de quien posiciona (imagen de marca o *branding*, comercio electrónico, captación de *leads* o clientes potenciales) las técnicas (con tipologías como técnicas de “sombrero blanco” o *white hat*, “sombrero gris” o *grey hat*, “sombrero negro” o *black hat*, siendo *white* el estilo de posicionamiento en buscadores que más respeta las limitaciones de la política establecida por los motores de búsqueda, y *black* la que menos) o los contenidos a posicionar (textos, imágenes, vídeos...) que en ocasiones lleva a buscar el posicionamiento en buscadores especializados, como Google Scholar, Web of Science o Scopus en el caso del posicionamiento de contenidos científicos.

En este entorno del marketing académico o científico, que busca aumentar las probabilidades de visibilidad científica y así de citación bibliométrica y alométrica, se utilizan técnicas de posicionamiento en buscadores científicos como Google Scholar, práctica denominada por Beel, Gipp y Wilde (2010) ASEO (*Academic Search Engine Optimization*) para conseguir que los contenidos científicos propios aparezcan más arriba en los resultados de búsqueda de una expresión de búsqueda o consulta realizada en ese motor de búsqueda (Orduña-Malea et. al., 2010).

La importancia de las técnicas de ASEO ha aumentado en los últimos años al haberse comprobado un cambio en los hábitos de búsqueda de información científica entre los científicos, por lo que se puede mejorar la visibilidad científica optimizando el posicionamiento de publicaciones científicas en Google Scholar (Serrano-Cobos, 2016). En la práctica, entran en juego distintos factores para el posicionamiento en el algoritmo de Google Scholar que en el de Google para contenidos no científicos, desde la redacción de los contenidos, a la red de citas y enlaces a través de las sedes web de las distintas entidades científicas susceptibles de citar este tipo de publicaciones (Orduña-Malea, Serrano-Cobos y Lloret-Romero, 2009). Estas mallas de enlaces generarán incluso redes multinivel, en las que unas entidades más grandes (léanse grandes multinacionales o universidades) y otras más pequeñas (micro marcas en el caso de las empresas, o departamentos en los casos universitarios) pueden ser analizados para descubrir los enlaces o caminos más interesantes hacia los contenidos a posicionar (Orduña-Malea et. al., 2015).

Lo mismo puede decirse de sedes bibliotecarias, sedes de webs corporativas de editoriales, de librerías online, de webs personales de escritores, y de otros tipos de creadores de contenidos digitales, blogs, revistas online, portales, medios sociales y un largo etcétera. En todos estos casos, la importancia de conocer qué, cómo y cuándo buscan los usuarios y clientes potenciales no sólo es crucial para mejorar el posicionamiento de sedes web del sector, es su principio y final (Baeza-Yates, Hurtado y Mendoza, 2004).

2.3.8. Recapitulación

El interés en conocer los hábitos de búsqueda de información de los usuarios de Sistemas de Recuperación de Información, nace con el objetivo de mejorar los servicios de información, y aportar una dimensión distinta de análisis a la disciplina de la Recuperación de Información, desde el paradigma cognitivo (González-Teruel, 2016). Con el paso del tiempo, ese interés se ha visto ampliado desde otras ramas del conocimiento y con objetivos cada vez más interesados en el marketing y la aplicabilidad económica de esos análisis, tanto para quienes realizan el trabajo de filtro de información, como para quienes buscan mejorar su posicionamiento en Internet.

Ese interés cada vez mayor, depende en primera instancia de aplicar e integrar distintas técnicas de análisis de las conductas de los usuarios, tanto dentro de un sitio web (mediante Analítica Digital) como en un contexto mayor en el que se observen tendencias de mayor volumen y variedad (Cibernetría). La evolución de esta área de investigación ha visto la irrupción de distintas disciplinas científicas, que trabajan en muchas ocasiones de forma inextricablemente unidas, donde la multidisciplinariedad impide dividir a unos científicos de otros claramente (Spink y Cole, 2006). Así, se ha visto cómo desde la disciplina de *Information Seeking Behaviour*, al trabajar con interfaces, la Interacción Persona-Ordenador (*Human Computer Interaction* o HCI) diferencia procesos de búsqueda mediante hojear o *browsing*, y búsqueda directa mediante inserción de texto en formulario de búsqueda (*searching*) lo que ha dado lugar a la subdisciplina del estudio de la conducta de búsqueda del usuario o *User Search Behaviour* (Xie y Matusiak, 2016)

Esta variedad de perfiles de los investigadores y de ramas de la ciencia trabaja con distintas perspectivas, metodologías y objetos de investigación, en torno al objetivo común de entender cómo buscan los usuarios en el medio digital. Así, se han encontrado investigaciones de tipo cualitativo que intentan generalizar tipologías de acciones y modelos de comportamiento (Ellis, 1989). En otras ocasiones, se detectan estudios de caso, difíciles o imposibles de reproducir, pero que aportan valiosa información sobre el comportamiento del usuario en determinadas circunstancias (Vakkari, 2008). Además, se han observado investigaciones que trabajan en torno a archivos *logs* de los buscadores, a veces sobre grandes masas de datos, para investigar distintas variables de análisis, como la interfaz, el idioma, el lenguaje, la intencionalidad de las expresiones de búsqueda, o la evolución histórica de éstas mediante

técnicas tan distintas como la entrevista, la observación en laboratorio o mediante aprendizaje de máquina, en inglés *Machine Learning* (Greifeneder, 2014).

Las limitaciones de este tipo de estudios suelen ser técnicas: bien los archivos *logs* no permiten trabajar con datos personales (para estudiar la evolución de individuos) bien los proveedores de datos, no permiten lidiar con volúmenes de datos ideales. Por otro lado, se encuentran limitaciones metodológicas debido a la dificultad de comparar estudios, tanto de caso como otros más globalmente, especialmente al constatar cómo interfaz y algoritmos de recuperación cambian con el tiempo. Pero los problemas no deben impedir seguir estudiando estas dinámicas, especialmente cuando se ha constatado que gigantes de la industria como Google (Google Research, 2018) o Microsoft (Microsoft Research, 2018) están investigando en este campo continuamente, y aplicando su conocimiento de las conductas de búsqueda de los usuarios a mejorar sus propios Sistemas de Recuperación de Información.

2.4. Conclusiones generales del marco teórico

Como se ha visto en este apartado, las implicaciones del conocimiento de los hábitos lectores llegan a distintos actores del sector del libro con diferente intencionalidad: a las editoriales les interesa entender qué quieren (y cuánto) leer los lectores, para evaluar la cadena de valor editorial (Acevedo y Mercado Lafertte, 2013), con el fin último de mejorar su estrategia editorial; a las bibliotecas, y a la administración en general, para optimizar sus servicios públicos y poner en valor sus recursos. Se estudió así en este apartado la literatura disponible sobre el estudio de los hábitos de los lectores (y no lectores) en España, para conocer el estado de la cuestión en esta materia. Y todo en un escenario múltiple, en el que la tecnología trae la convivencia de un variado espectro de dispositivos de lectura, más allá del papel.

Las investigaciones del comportamiento lector se han cimentado fundamentalmente en investigaciones cuantitativas, derivadas del cómputo de compras o préstamos de productos culturales, y en cualitativas. Entre éstas se encuentran preferentemente encuestas y entrevistas, donde metodológicamente es difícil encontrar formas de comparar entrevistas salvo en las tendencias generales, puesto que se usan preguntas distintas en investigaciones diferentes, el sentido de éstas puede ser equívoco, así como los conceptos por los que preguntan, y ha aumentado la variedad de factores que pueden influir en las opiniones de los encuestados, debido fundamentalmente al avance de las nuevas tecnologías, tanto en formatos de uso y disfrute de los productos culturales, como los canales por los que llega la información al lector, posible lector, o usuario de una biblioteca.

Antes también, pero de forma creciente con Internet y la explosión en la disponibilidad de información que ha traído, conllevan un canal de información y recomendaciones que pueden no sólo influenciar en el objeto de nuestra investigación, sino aportar una serie de datos muy valiosos, lo que brinda la oportunidad de buscar nuevas formas de estudiar y analizar el comportamiento y las dinámicas de búsqueda de lectura y otros formatos de productos culturales. Ahí es donde se ha buscado indagar sobre las posibilidades metodológicas de otras disciplinas, como es la familia de disciplinas relacionadas con la del Análisis de la Conducta de Búsqueda o *Information Seeking Behaviour* (Kuhlthau, 1991).

Como se ha mencionado para el caso del análisis de los hábitos lectores, una de las dificultades inherentes a la investigación de las conductas de búsqueda, especialmente en lo indagado por

parte de los científicos del área de *User Search Behaviour*, es que es muy difícil comparar la conducta de unos usuarios en un motor de búsqueda y en otro, con interfaces distintos y soluciones de presentación de las respuestas distintas, que dan respuesta a usuarios de países distintos y con motivaciones particulares (Spink et. al, 2002).

La exploración de la investigación en la materia durante los últimos 20 años nos deja por tanto muchas publicaciones con resultados locales, donde conforme los proveedores de datos han permitido explotar sus archivos *logs* del buscador, se han visto más investigaciones, a pesar de las limitaciones técnicas, que dificultan que puedan resultar de utilidad horizontal (independiente de motor de búsqueda) y atemporal (independiente del año de análisis y por tanto de versión del algoritmo del buscador, diseño de interfaz, etc.) pero que pueden resultar de utilidad para conocer cómo buscan y han buscado información los usuarios de una determinada segmentación, a través del escrutinio de un gran volumen de datos.

La multidisciplinariedad de estos estudios añade capas de complejidad a la investigación Hartel et. al. (2012) identificándose diversas especialidades que vienen de ramas muy distintas, como las ciencias cognitivas, informática, ciencias sociales, marketing, inteligencia artificial, lingüística, o biblioteconomía y documentación (*Information Science*).

Teniendo en cuenta esta complejidad, y a tenor de lo inabarcable de su multidimensionalidad por parte de un único investigador, esta tesis doctoral se centra sólo en el estudio de las expresiones usadas en procesos de búsqueda de información sobre el libro, literatura y contenidos ofrecidos en las bibliotecas, de usuarios de Google en español y en España, a lo largo del período 2004 – 2016.

Los datos y conclusiones obtenidos en la tesis brindarán una información complementaria de utilidad para la toma de decisiones estratégicas de los actores implicados en el sector del libro (autores, editores y editoriales, bibliotecas, instituciones y empresas), con el fin de mejorar los indicadores de lectura en España. Nos arriesgamos, como decía Steinbeck, a que “por el grosor del polvo en los libros de una biblioteca pública (pueda) medirse la cultura de un pueblo” (Ortega Blake, 2013). O su equivalente digital hoy día.

3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

3.1.Introducción al análisis y discusión

En esta sección se analizan y discuten los resultados obtenidos tras la detección y clasificación de expresiones de búsqueda, y toma de datos en *Google Keyword Planner* y *Google Trends*. En total, el número de frases de búsqueda analizadas tras la eliminación de expresiones de búsqueda o frases de búsqueda no aptas para el estudio, asciende a 30.925, de las cuales, 7.250 generaron algún resultado en *Google Trends*, lo que supone un 23,4% del total de frases de búsqueda o expresiones de búsqueda recogidas a través de *Google Keyword Planner*.

Para entender el volumen de la muestra obtenida, se puede señalar que, para cada expresión de búsqueda de la muestra, se obtuvieron datos en *Google Trends*, de cada uno de los 12 meses de los años 2004 a 2016. Una vez eliminadas aquellas expresiones de las que *Google Trends* no ofreció datos, debido a que no superaban su *threshold* o límite mínimo (*Google Support*, 2018) de índice de búsqueda o *query index* mensual (Choi y Varian 2009, 2011) la muestra contabilizó un total de 1.131.000 valores de índice de búsqueda mensual para todo el conjunto de expresiones de búsqueda de las que se obtuvieron datos de *Google Trends*.

Se han realizado así dos aproximaciones de estudio:

- Estudios realizados sobre la base del conjunto total de las expresiones de búsqueda localizadas en *Google Keyword Planner*, donde no se trabaja con series temporales (es decir, con datos de *Google Trends*), sino sólo con las expresiones de búsqueda en sí. Se estudiará la validez de *Google Keyword Planner* como herramienta de extracción de datos para análisis del lenguaje de búsqueda de los usuarios. Se analizará la construcción de facetas (concepto explicado en el apartado 1.3.), en las que se han visto clasificadas las frases de búsqueda (cuya metodología de asignación de facetas fue descrita en el apartado 1.5.4.). Además, se comprobará si existe co-ocurrencia en la facetación, es decir, si una misma expresión puede clasificarse en varias facetas y/o en varios aspectos, y en qué medida.
- Estudios realizados exclusivamente a partir de la muestra de aquellas expresiones de búsqueda o frases de las que se pudieron obtener datos de *Google Trends*. En este apartado, se estudiará la validez de la herramienta para el estudio longitudinal de la

conducta de búsqueda de los usuarios en el contexto de la segmentación elegida (temporal, geográfica, lingüística y temática) También se estudiarán los patrones de conducta detectados tanto de forma global, de forma separada en las dos grandes subclasificaciones detectadas en su dimensión temporal (expresiones de búsqueda temporales y atemporales) como de forma más particular, en algunas familias de expresiones de búsqueda que han ofrecido datos que se salían de la tendencia global.

Se realiza en primer lugar una discusión sobre la metodología utilizada, tanto para la extracción de datos en *Google Keyword Planner* y luego en *Google Trends*, como de la metodología de selección de términos para las facetas y la técnica usada para desglosar las expresiones de búsqueda de las que se ha obtenido datos en *Google Trends*, en búsquedas temporales y atemporales.

Posteriormente, se discuten los resultados en sí, y los descubrimientos que su estudio han reportado. Se estudiará la distribución de expresiones de búsqueda en las facetas y aspectos, así como su co-ocurrencia. Respecto de la muestra de la que se tienen datos longitudinales, se explora la temporalidad y atemporalidad de la misma, y las tendencias que reporta la distribución de los índices de búsqueda a lo largo de la serie histórica, tanto de forma global, comparando el patrón ofrecido por expresiones temporales versus atemporales, como de forma segmentada para cada tipología de clasificación longitudinal por separado.

3.2. Discusión metodológica

En este apartado se investiga primero la metodología utilizada. Es importante entender la forma en que *Google Keyword Planner* ofrecía información relacionada a partir de unas palabras semilla, al menos en 2010 (cuando se realizó la toma de datos) lo que se ha explicado en el apartado 1.5.1. Pero una vez entendido, debe evaluarse su capacidad para aportar información relevante en forma de expresiones de búsqueda relacionadas para el contexto de esta investigación y, más allá, para futuras investigaciones sobre la materia.

Asimismo, se discute la idoneidad de la metodología utilizada para la clasificación de las expresiones de búsqueda en distintas facetas y aspectos, a la luz de otras técnicas de clasificación y metodologías de aplicación de clases, siempre adaptando las posibles mejoras y soluciones al problema de la asignación de clasificaciones, facetas y aspectos al ámbito que nos ocupa (en sus dimensiones temporal, geográfica, lingüística y temática)

Posteriormente se exploran las posibilidades que ofrece *Google Trends* para el análisis longitudinal de las conductas de búsqueda, las limitaciones y ventajas que aporta a este tipo de investigaciones, ampliando la experiencia particular de esta investigación y su contexto con otras investigaciones que han trabajado sobre la validez de la herramienta para la investigación cibernética.

Finalmente, se discutirá la técnica usada para desglosar las expresiones de búsqueda de las que se ha obtenido datos en *Google Trends*, en búsquedas temporales y atemporales, para explorar posibles mejoras al método empleado.

3.2.1. De *Google Keyword Planner* como herramienta para la obtención de expresiones de búsqueda relacionadas

Aunque en este estudio se ha preferido trabajar sobre las tendencias de búsqueda, y por tanto sobre datos extraídos de *Google Trends*, más que sobre el puro volumen de frecuencia de búsqueda de cada expresión (información ofrecida por *Google Keyword Planner*) es de interés estudiar cómo esta última herramienta ofrece las expresiones de búsqueda relacionadas, a la luz del dato de frecuencia de búsqueda que aporta, para comprobar su validez como herramienta de investigación cibernétrica. Además, mientras la validez de *Google Trends* como herramienta ha sido analizada en varios ejemplos en la literatura científica, como los recopilados por McCallum y Gwendolynn (2014), *Google Keywords Planner* no ha sido prácticamente estudiada entre los miembros de la Academia como herramienta de investigación cibernétrica, por lo que deberemos recurrir a las publicaciones de expertos en posicionamiento en buscadores, para encontrar algún análisis similar a éste (Fishkin, 2018)

Uno de los problemas o limitaciones metodológicas que plantea *Google Keyword Planner* para esta investigación, es que es opaco en su metodología de generación de recomendaciones de expresiones de búsqueda relacionadas con una expresión o frase semilla, por lo que hay ciertos datos que plantean dudas que impiden ir más allá a la hora de asegurar el volumen o la frecuencia de búsqueda de una expresión de búsqueda para un mes o la media mensual para un año, en comparación con otra u otras.

Por ejemplo, en ocasiones es muy difícil saber si el dato relativo al volumen o frecuencia de búsqueda de una frase pertenece realmente a esa frase, o a un conglomerado de frases más o menos parecidas. Véase el caso de “biblioteca de madrid”, cuyos valores de media mensual de búsquedas para el año en el que se hizo la consulta a la herramienta (2010), son idénticos para otras combinaciones muy relacionadas, como “biblioteca madrid”, “bibliotecas de madrid”, etc., lo que se puede apreciar en la tabla 3-1:

Tabla 3-1 Volumen de búsqueda caso "Biblioteca de Madrid"

Keyword	Media mensual de búsquedas en España

biblioteca de madrid	72.591
biblioteca madrid	72.591
bibliotecas de madrid	72.591
bibliotecas en madrid	72.591
bibliotecas madrid	72.591
madrid biblioteca	72.591

El problema es que, en otras ocasiones, palabras muy cercanas semántica o lingüísticamente, con escasa diferencia, ofrecen resultados completamente distintos, como es el caso de “libro” versus “libros” mostrado en la tabla 3-2, donde habiendo sólo una letra de diferencia, ofrece como resultado cientos de miles de búsquedas mensuales de distancia, lo que sugiere que el algoritmo de la herramienta sí diferencia ambas palabras en su cómputo.

Tabla 3-2 Volumen de búsqueda caso "libro" versus "libros"

Keyword	Media mensual de búsquedas en España
libro	2.530.909
libros	2.181.818

Una explicación plausible para este hecho sería que *Google* se fijara en la variabilidad de expresiones de búsqueda relacionadas, más allá de estas frases. Muy probablemente “libro” se combina con muy diferentes expresiones, al igual que “libros”, por lo que *Google* las considera semillas válidas para tratarlas como “entidades” independientes, al menos en lo que al cálculo de la frecuencia de búsqueda se refiere, aunque las utilice para relacionar documentos por co-ocurrencia de palabras en los textos (*Google.com*, 2011). Sin embargo, expresiones que generan poca variedad de frases de búsqueda, que se combinan con casi ninguna otra palabra

(como el conjunto de expresiones que hemos visto en torno a “bibliotecas” y “madrid”) son tratadas como una “entidad” común (Fishkin, 2018).

Otra posibilidad, siguiendo a Reyes, Majluf e Ibáñez (2018), es que Google analice las sesiones de búsqueda y los datos muestren tendencias divergentes, o combinaciones con otras palabras distintas en el caso de “libro” y “libros”. En el caso de sesiones de búsqueda en las que se reformule la búsqueda con pocos cambios en las palabras usadas, como podría ocurrir en el caso de “bibliotecas” / madrid”, lo que podría ocurrir, es que los usuarios busquen en la misma sesión distintas variantes de la frase, utilizando conectores distintos como “de” o “en”, o usando el singular, el plural, etc., para localizar la información. Finalmente, otra opción sería que Google utilizara información de los clics de los usuarios en las consultas, para así agrupar las variantes por resultados en los que los usuarios clican en los mismos resultados, y no, o no sólo, por co-ocurrencia léxica de palabras en el texto de las expresiones de búsqueda (Joachims, 2002; Google.com, 2012).

Pero al no tener más datos de la propia compañía es difícil, si no imposible, llegar a una conclusión clara. Por ese motivo no se intenta llegar a otras conclusiones que no sean la de tener en cuenta estos volúmenes de frecuencia como datos aproximados, redondeados, y en ninguna forma exactos, lo cual nos lleva a la conclusión de que la utilidad de la herramienta es parcial. *Google Keyword Planner* sirve para localizar expresiones de búsqueda relacionadas, sí, pero deberá ser utilizado con prevención, sin tener la seguridad de obtener datos reales ni exhaustivos.

3.2.2. De la clasificación facetada de expresiones

Al descomponer el conjunto de expresiones o frases analizadas en este estudio, se constata que está formado por 8.416 palabras distintas, que se combinan de distintas formas. A partir de ahí, se puede prestar atención a la pertenencia de las expresiones de búsqueda a una faceta o aspecto, a partir de cada palabra, y así saber cuántas frases necesitan incluir una o más de una palabra, dentro de la frase, para que se pueda clasificar esa expresión de búsqueda dentro de una faceta.

De forma global, se constata que el 80% de las frases de búsqueda necesita tan sólo de una palabra suelta para ser clasificada dentro de una faceta o aspecto. Por ejemplo, en la faceta *Ámbito*, la expresión de búsqueda "academia cervantes" contiene la palabra "academia", que es la que denota el ámbito que el usuario delimita. Otro ejemplo, en la frase "a mi manera autor" se usa la palabra "autor" para desambiguar la expresión de búsqueda, indicando así al motor de búsqueda que se busca el autor de la obra "a mi manera" y no sólo se buscan contenidos que contengan la frase "a mi manera", lo que sirve para incluirla en la faceta "Desambiguación".

Otros casos sirven de ejemplo para entender cómo una misma frase contiene palabras que sirven para clasificar la frase en varias tipologías de intencionalidad. Por ejemplo, una o varias palabras pueden servir para ubicar la expresión de búsqueda bajo una faceta con intención informativa, y otras para clasificarla bajo una intención distinta, como la tipología navegacional o la transaccional.

Es el caso de una frase como "wikipedia org wiki miguel cervantes", situación expresada gráficamente en la Figura 3.1. En ella se señala cómo "wikipedia", "org" son consideradas palabras que implican una intención de búsqueda por parte del usuario donde, en parte, se busca una entidad, o como en este caso, se especifica el sitio web ("Wikipedia" y "org", aunque el usuario se está refiriendo a "Wikipedia.org") y "wiki" denota un formato, mientras que también se especifica el contenido a recuperar, que es "miguel cervantes" (donde "miguel" es un "nombre personal", y "Cervantes" se ha tipificado como "autor"). Como se aprecia, y siguiendo a Jansen, Booth y Spink (2007), se mezcla la intención navegacional (el usuario quiere ir directamente al sitio web de "Wikipedia.org") con la informativa (busca la página informativa sobre Cervantes dentro de ese sitio web).

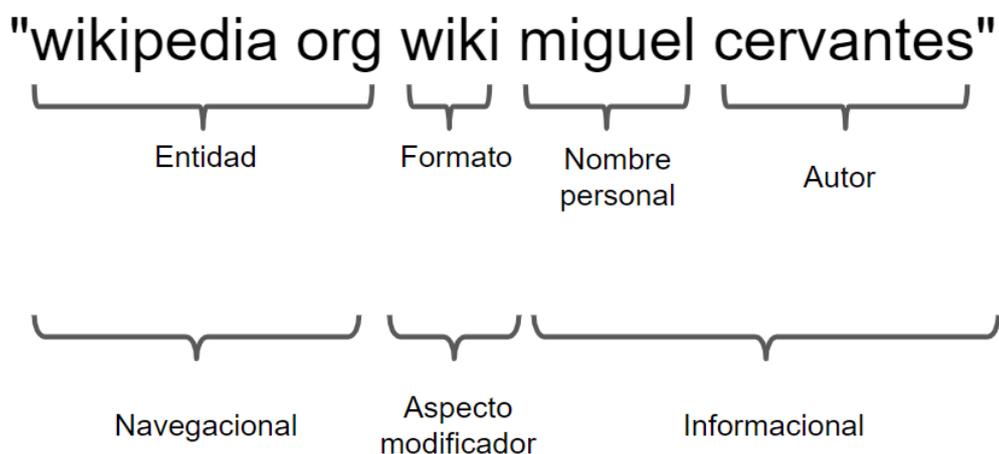


Fig. 3-1 Nota de Google Trends sobre cambios en asignación geográfica. Fuente: Google Trends (2018)

En pocos casos se encuentran tres palabras o más dentro la misma expresión de búsqueda que permitan clasificar una consulta bajo una misma faceta, como es el caso de "principales autores literarios hispanos y latinoamericanos", donde todas las palabras, excepto el conector "y" denotan formas de desambiguación por parte del usuario, y asimismo explicitan el uso del género (masculino en este caso) de la expresión de búsqueda. En el extremo se encuentran ejemplos como "*panic at the disco i write sins not tragedies traducción*", que usa 7 palabras que denotan la condición de Idioma.

Por tanto, los ejemplos sugieren que, aunque una única palabra puede servir para detectar la pertenencia de una frase a una faceta, lo que ocurre en multitud de casos, se puede encontrar casos en los que varias palabras, dentro de una misma frase, pueden servir para clasificarla a una misma faceta o aspecto. La cuestión es cómo detectar esa adscripción a cada faceta, primero palabra a palabra, después tomando la frase en su conjunto, o dividida en sintagmas formados por varias palabras.

En ocasiones es fácil encontrar elementos cuya clasificación es relativamente sencilla, lo que resulta útil para automatizar su identificación de forma escalable. Pero a su vez, cada elemento puede aportar distintas informaciones, que son útiles para entender la conducta de búsqueda de los usuarios, y su uso del lenguaje en un entorno digital. Véase en la Tabla 3-3, cómo de entre el conjunto de estas abreviaturas, hay elementos destacables. Por ejemplo, "8va", por "octava", "9no" por noveno, o "3o" y "4o" por "tercero" y "cuarto" respectivamente, son abreviaturas utilizadas por usuarios que, atendiendo al Diccionario panhispánico de dudas

(Real Academia Española, 2005) son un error. Pero este tipo de expresiones nos dan noticia de cómo evolución el lenguaje y su uso, y para los motores de búsqueda, identificar estas prácticas es clave para determinar la intención de búsqueda del usuario.

Tabla 3-3 Ejemplos de abreviaturas utilizadas en la muestra de expresiones de búsqueda.

Fuente: Google Keywords Planner. Elaboración propia.

Oscar	4º
1arte	4ta
1o	5ta
1º	6.1
2.0	6ta
2o	7ma
3d	7mo
3gp	7rm
3o	8va
4o	9no

El contexto de estudio (la lectura, el libro y otros y medios de consumo cultural en bibliotecas) complica la clasificación, o aun la adscripción de las expresiones recomendadas por *Google Keyword Planner*. En un análisis cualitativo, leyendo expresión a expresión, se pueden encontrar casos en los que es imposible, por el mero texto, deducir siquiera si la misma está relacionada con la temática en cuestión para el presente estudio. Tomemos como ejemplo “autores madrid”. En principio y dentro de la temática habida en el estudio y a tenor de las palabras semillas introducidas en la herramienta *Google Keyword Planner*, se inscribiría como una consulta que busca información o una lista de autores que sean de Madrid o residen en tal ciudad.

Existen otros casos, como “autores rusos”, “autores romanos”, “autores asturianos” y otras que van en esa línea, pero el caso del uso de la ciudad como especificación de la segmentación geográfica es particular, porque “autores” coincide con la marca de la empresa de autobuses “Autores” que realiza viajes por diversos lugares de la geografía española (hoy fusionada con otras y renombrada como “Avanzabus” (Avanzabus, 2018)

Esta polisemia de las palabras y por ende, del concepto al que hace referencia la frase de búsqueda, es un problema incluso para Google, cuyo interfaz presenta respuestas provenientes de ambos conceptos, de forma que sea el usuario el que desambigüe la consulta con sus clics en unas u otras respuestas (Baeza-Yates, Hurtado y Mendoza, 2005; Duan, Kiciman y Zhai, 2012; Park et. al., 2012; Gong et al., 2013; Preetha y Vimal Shankar, 2015; Dou et. al., 2015), por lo que podemos encontrar en esta y otras expresiones con el mismo problema, que una parte de ellos busque con una intencionalidad, o con otra.

The image shows a Google search interface with the query 'autores'. The search bar is at the top, and below it are navigation tabs: 'Todo', 'Imágenes', 'Noticias', 'Maps', 'Vídeos', 'Más', 'Configuración', and 'Herramientas'. A blue button labeled 'Iniciar sesión' is on the right. Below the search bar, it says 'Aproximadamente 158.000.000 resultados (0,33 segundos)'. The first result is an advertisement for 'Avanzabus.com | Viaja en Bus con Avanza | Acercamos personas' with a link to 'www.avanzabus.com/'. Below the ad are four service links: 'Tus billetes Online', 'Modificar Billetes Online', 'Atención al Cliente', and 'Bonos Madrid Salamanca'. To the right of these links is a 'Ver resultados relacionados' box for 'Autor (Función en un proyecto)'. Below the ad and service links are two organic search results: 'Bienvenido a Avanzabus.com - Avanza Autobuses' and 'Autores - Editorial Anagrama'.

Fig. 3-2 SERP o Página de Resultados de Google ante la expresión “autores”. Fuente: Google.es (2018)

Este hecho invita a trabajar en una mejora de la metodología de extracción de expresiones, que esté más dirigida al ámbito de estudio, cruzando más palabras semilla semánticamente relacionadas (Yin y Shah, 2010) para asegurar en la medida de lo posible la relación con el ámbito de estudio, y a una técnica de clasificación que permita delimitar expresiones ambiguas o imposibles de definir a priori, y así permitir análisis más ricos y más válidos internamente.

Por eso se han incluido en el anexo las expresiones que han aparecido en el proceso de extracción de datos, incluyendo aquellas que ya se desecharon en este estudio pero que *Google Keyword Planner* ofreció como relacionadas en 2010, para que puedan servir de ejemplo (tanto positivo en aquellos casos en los que se usen las expresiones sí admitidas como negativo en los casos de expresiones no admitidas) Estos ejemplos serán útiles para realizar futuros análisis que empleen técnicas derivadas de *clustering* o agrupamiento supervisado (Nettleton, Calderón-Benavides y Baeza-Yates, 2006; Parikh et. al., 2013) o semisupervisado (Beitzel et. al., 2005; Yu y Ren, 2012; Ren et. al., 2015) cuando complementen o se comparen con otras técnicas, como las de agrupamiento no supervisado (Tan y Peng, 2008; Mishra et. al., 2011)

Otra circunstancia sobre la que se debe discutir es la selección de facetas y aspectos que han sido determinados en el apartado 1.3.4 de la metodología. Como se explica en éste, el descenso del análisis al nivel de palabras sueltas, ha permitido llegar a un análisis muy granular, que permitía descubrir las facetas y aspectos comunes, que se iban repitiendo constantemente a lo largo del recorrido del listado de palabras, pero también identificar excepciones, que iban sumando a la lista de facetas y aspectos nuevos elementos clasificatorios, lo que obligó a revisar el listado varias veces, a la vez que enriquecía la variedad de posibilidades.

El descubrimiento de la riqueza de posibilidades de la facetación para la determinación de grupos y subgrupos de expresiones de búsqueda es uno de los factores que hacen a esta temática un campo de estudio tan atrayente, en el que se debería trabajar más en el futuro, más allá o en combinación con otras técnicas de clasificación de expresiones de búsqueda, como la clasificación semántica (Guo et. al., 2009; Limam et. al., 2010) la clasificación por longitud de frases o número de palabras (Bendersky y Croft, 2009) la frecuencia de aparición de combinaciones de palabras en una muestra temática (Downey, Dumais y Horvitz, 2007) o el análisis léxico de expresiones relacionadas previamente mediante análisis de series longitudinales de su volumen de búsqueda (Alfonseca, Ciaramita y Hall, 2009)

Ha de decirse que en el presente análisis se detectan ya posibilidades que apuntan direcciones muy interesantes de cara a ulteriores investigaciones. Por ejemplo, se generó un aspecto lingüístico “género” (donde “aspecto” no deja de ser otra forma de “faceta”) en el que se han integrado todas aquellas expresiones de búsqueda que incluyeran algún elemento que permitiera deducir que la búsqueda delimitaba o segmentaba uno o ambos sexos.

Es evidente que este aspecto podría haberse separado en más casos, permitiendo así desglosar las expresiones de búsqueda en aquellas que utilizan el género femenino (“autora”), las que usan el género masculino (“autor”), y aquellas que no lo determinan de forma concluyente, como algunas de las que usan también el plural (“autores”).

En este sentido, lo que sí se puede hacer y es una ventaja de la facetación como herramienta de recuperación de información, es cruzar facetas, o facetas y aspectos, que aportan dimensiones de análisis diferentes (Wu, Madhavan, Halevy, 2011). Por ejemplo, actualmente, y como sí existe “plural”, que es otro aspecto lingüístico, se puede recuperar un listado de expresiones de búsqueda que cumplan la condición de usar el género y esté en plural, con lo que se puede reducir una segmentación mayor en un subgrupo menor, aunque en este caso no se pueda deducir automáticamente que todas sean de un género, pues obviamente habrá plural de un género (“escritoras”), o indeterminado (“escritores”) lo que todavía es mejorable.

Por otro lado, en el curso del análisis de los resultados (apartado 3.3) se ha detectado la posibilidad de que existan familias semánticas y agrupaciones que vayan más allá de la relación textual. Por ejemplo, separando expresiones sobre bibliotecas universitarias y bibliotecas públicas y de otro tipo, dentro de la faceta biblioteca) O agrupando expresiones por su comportamiento o patrón longitudinal (Jun, Park y Yeom, 2014) entre otras cosas para poder realizar análisis predictivos que contemplen relaciones más adecuadas para la búsqueda de relaciones causales (Granka, 2009; Nyman, Ormerod y Tuckett, 2015), lo que en este estudio se ha hecho de manera meramente manual y cualitativa en una muestra intencionada, no probabilística.

Esa ampliación de técnicas de análisis enriquecería aún más la clasificación actual, evidenciando la necesidad de buscar más capas de estudio en un ámbito temático donde hay un amplio abanico de posibles relaciones causales, lo que invita a experimentar, en ulteriores investigaciones, con las posibilidades de combinar distintas técnicas de agrupamiento, utilizando como base o principio del que partir, la presente base de datos de expresiones de búsqueda, la clasificación ofrecida y los datos recogidos durante la serie temporal analizada.

Finalmente, cabe preguntarse por la idoneidad del criterio seguido en la investigación para clasificar unas expresiones de búsqueda en las facetas y aspectos determinados. Tómense tres ejemplos de las 30.925 expresiones de búsqueda recopiladas y clasificadas, para ejemplificar la dificultad de delimitar una política de clasificación que minimice los sesgos y aporte coherencia.

En primer lugar, se puede apreciar que la expresión de búsqueda “antonio machado biografia breve” puede ser clasificada bajo la faceta “autor”, por contener nombre y apellido de uno, “antonio machado”. También se ha delimitado que sea incluida en el aspecto modificador “tipo narrativo”, porque en la expresión se usa la palabra “biografia”, que determina una tipología o un género literario particular. Pero también se ha facetado la expresión, siendo contenida dentro del aspecto lingüístico “error”, debido a la falta de uso de la tilde en “biografia”, siendo la forma correcta en castellano “biografía”.

Como se podrá comprobar en el apartado 3.3.1, son miles aquellas expresiones en las que se encuentra algún elemento considerado como error. Buena parte de ellas lo son por esta razón, por la falta de uso de la tilde, lo que podría dar lugar a una subclasificación de errores en las que una de las categorías o subcategorías (subfacetas en este caso o aún subaspectos) lo fuera por este motivo.

Pero también cabe preguntarse si en el futuro deberán ser consideradas *stricto sensu* como erróneas (Kulkarni et, al., 2015), máxime cuando sistemas de recuperación como Google aceptan sin problemas el uso de palabras mal escritas y “educa” a los usuarios en esa redacción incorrecta, monitorizando sus acciones para comprobar cuándo es admisible el uso de errores sin que afecte a la recuperación de información (Amitay, 2006; Vallez y Pedraza-Jimenez, 2007; González Villa, 2016).

Otro ejemplo proviene del autor Dostoievski. Si se toma la expresión “apuntes del subsuelo dostoievski”, cabe preguntarse si “dostoievski”, está mal escrito o no. Un problema será determinar qué fuente seguir como primaria para delimitar este punto. La Biblioteca Nacional muestra como principal la forma “Dostoevskii, Fiodor Mijaïlovich” pero contiene diversas variantes, como “dostoievski”, “Dostoevskii” o “Dostoyevski” (Biblioteca Nacional de España, 2018)

Así, expresiones de búsqueda que contengan “fyodor dostoyevsky” podrían ser asignadas al aspecto de “Error” y también todas aquellas que usen la forma “dostoievski” que es la mayoritaria entre las utilizadas por los usuarios. Pero si el uso continuado de una forma es uno de los criterios utilizados por la propia Academia de la Lengua en la adopción de nuevas palabras (Real Academia Española, 2018) es lícito plantearse una adopción más laxa de la clasificación o aun la inclusión de una categoría, faceta o aspecto distinto, en el que se puedan incluir palabras candidatas a convertirse en el futuro en formas aceptadas o aceptables de expresar sobre un concepto o en este caso, un autor. Ese tipo de identificación, permitiría la

creación de relaciones entre las distintas formas de escritura de un concepto, (Marner, 1993; Lopes, 1995; Shiri, 2008) lo que ayudaría a controlar el vocabulario tal y como lo hace un tesoro o una lista de encabezamientos de materia (Ministerio de Educación, Cultura y Deporte, 2011)

Más allá, podrían generarse redes de palabras y conceptos relacionados (Miguel, Caprile, Jorquera-Vidal, 2008) u ontologías basadas en el uso de los usuarios de estas variantes idiomáticas (Nirenburg y Raskin, 2004; Godoy y Amandi, 2006; Bar-Ilan, Zhu y Levene, 2009; Aguilar-Lopez, Lopez-Arevalo y Sosa-Sosa, 2009; Kim y Ahn, 2012) donde una ontología sería una especificación explícita de una conceptualización (Gruber, 1993), en la que se genera una referencia común en un área determinada, en este caso para investigadores, pero también para permitir su computación, alrededor de un concepto, que puede tener relaciones semánticas pero también de otro tipo con otras conceptualizaciones o conceptos en el contexto de la expresión de una necesidad de información (Rodríguez et. al, 2017).

Aun así, se deberá tener en cuenta otro nivel de complejidad. El mismo ejemplo de “Dostoievski” nos sirve para identificar un ejemplo más de polisemia, puesto que este autor también es parte del título de un libro (Gide, 2016) o de muchos, lo que ocurre con muy diversos autores. En ocasiones el usuario especifica su necesidad desambiguando, facetando mediante lenguaje natural, y sin necesidad de entender el concepto de facetación, como en el ejemplo de la consulta “libro el idiota dostoievski”.

Pero cuando el usuario busca “Dostoievski”, cabe dudar de si busca un autor o un libro cuyo título sea éste. Muy probablemente se deberá dejar de usar la expresión “el usuario” como si todo usuario que utiliza una expresión, lo hace con un mismo sentido. Hablamos de “usuarios”, donde habrá una parte que tendrán una intención distinta de la otra, o de otras, indetectable por el análisis del texto simple, sin más contexto informativo. Ese hecho lleva a la necesidad de integrar con ésta, otras técnicas y disciplinas científicas, sobre todo las basadas en la interacción de los usuarios, su histórico de consultas y elecciones, así como distintos aspectos de su perfil contextual (dispositivo, hora o estación de la consulta, país desde el que busca, idioma, etc.) lo que diversos especialistas en posicionamiento en buscadores asumen que Google hace desde hace tiempo (Serrano-Cobos, 2015; Pernas Alonso, 2017).

Para ahondar en el caso de la problemática metodológica de clasificar expresiones de búsqueda en facetas minimizando los sesgos, si se toma la expresión “bibliografía dostoievski”, el problema será delimitar si se incluye a la consulta en el aspecto “Error “, debido a que se

escribe sin tilde “bibliografía”, o debido a que se usa “Dostoievski”, o a ambas. Este hecho lleva además a considerar la posibilidad de analizar en subsiguientes investigaciones, la fuerza de la pertenencia de una expresión de búsqueda a una faceta o aspecto en virtud, en este caso, de cuántas palabras pertenezcan a tal faceta o aspecto.

Otra variante de estudio sería calcular qué porcentaje de palabras de la expresión pertenecen más a una faceta o a otra. Por ejemplo, en la expresión “crimen y castigo Fedor Dostoievski” se vería que dos de las cinco palabras de las que consta la expresión pueden ser clasificadas tanto como “Error” como en “Autor” (“Fedor” y “Dostoievski”) y por tanto el peso de su pertenencia a “Error” es menor que el del anterior ejemplo “bibliografía Dostoievski”, donde ambas palabras que conforman la expresión han sido candidatas a ser consideradas como pertenecientes a “Error”. Nótese que no se ha considerado “y” como una palabra vacía, sin uso para el análisis, porque “y” puede servir para encontrar indicios de que una expresión forma parte de una frase.

Finalmente, en el ejemplo de la expresión “autor de la sirenita” el problema es decidir si “sirenita” hace pertenecer a la expresión a la faceta “Film” o no se usa dentro de la misma. Existe una película titulada “La Sirenita” y un libro titulado “La Sirenita”, por lo que la ventaja de la facetación ayuda a no necesitar elegir, pudiéndose clasificar bajo la faceta fílmica. Otra cosa es que la aparición de la palabra “autor” pueda sugerir que el usuario hace referencia a un libro.

Este caso, donde el lenguaje tiene tantos matices, ejemplifica problemas en ocasiones insalvables. Así se constata en el caso del uso de “síntesis”, “crítica”, “sinopsis” o “resumen” al lado de títulos de obras que han sido publicadas como libro, comic, película o aun como videojuego. Una exploración cualitativa, observando las respuestas que ofrece Google en primer lugar (Barry y Lardner, 2011; Lurie y Mustafaraj, 2018) antes distintas expresiones de búsqueda, revela que hay indicios para suponer que “síntesis” y “resumen” se utilizan más cuando se asocian a libros, pero “sinopsis” se asocia más a películas, mientras que “crítica” se presenta para diversos formatos.

Pero falta investigar más profundamente las expresiones que utilizan estas palabras, para delimitar tanto una posible adscripción como las otras, dejando en cualquier caso la posibilidad de que sean palabras ambiguas de por sí, y que pueden ser usadas por usuarios para diversas intenciones (buscar información sobre una película, un libro, etc., de un mismo personaje o saga)

Finalmente, hace falta ahondar más en las posibilidades de relación y clasificación de las expresiones detectadas o de otras por detectar, así como consensuar con otros expertos los criterios de relación, amén de investigar y aplicar otras técnicas de agrupamiento que permitan desarrollar clasificaciones más escalables, lo más libres posible de sesgos.

3.2.3. De *Google Trends* como fuente de datos cibernéticos de hábitos de búsqueda de lectura, libro y bibliotecas

Son miles las investigaciones realizadas hasta la fecha que versan sobre *Google Trends* que la utilizan como fuente de datos (Jun, Yoo y Choi, 2018) Una queja habitual por parte de los investigadores que utilizan esta herramienta para la extracción de datos cibernéticos es que, aunque pueda parecer que el dato que se muestra es el volumen de búsqueda de cada mes y año, los datos que *Google Trends* muestra y permite descargar, no son los datos de frecuencia de búsqueda (es decir, cuántas veces se busca una expresión de búsqueda en *Google Trends* para la segmentación aplicada, como sí hace, aunque redondeando, *Google Keyword Planner*) Lo que hace es mostrar un índice o *query index* (Choi y Varian 2009, 2011) para cada momento de la serie.

Este índice o *query index*, sí trabaja con las frecuencias de búsqueda, pero lo que hace es normalizar, al dividir el volumen de búsqueda de cada expresión entre el volumen total de búsquedas en Google de la segmentación determinada (que puede ser geográfica y por rango de fechas) Después, se toma como valor 100 el índice del momento con mayor volumen de búsquedas de toda la serie, y el resto de volúmenes de búsquedas se ordenan en consecuencia, al compararse con ese momento o índice de mayor valor de toda la serie. De esa forma, cuando se observa que una expresión de búsqueda ve decaer este índice a lo largo de los meses en una serie, es la popularidad en el uso de esa expresión la que decae, y no en número absoluto, siendo independiente de que haya podido crecer el número de usuarios en Google a lo largo de los años, y por tanto pudiera haber generado mayor número de búsquedas (Carrière-Swallow y Labbé (2010); Blanco, 2014)

Asimismo, utilizar el índice de búsqueda o *query index* de *Google Trends*, permite liberarse del problema de trabajar con una variable (el volumen o frecuencia de búsqueda por unidad de tiempo) que puede generar grandes diferencias entre expresiones de búsqueda distintas. Al trabajar con este índice se pueden igualar todas las expresiones de búsqueda, porque lo que indica es su relación con el momento de mayor frecuencia de búsqueda de la serie, y por tanto informa de los patrones de búsqueda, permitiendo que los datos sean equivalentes entre expresiones con volúmenes de búsqueda muy diferentes, lo que permite obtener información valiosa de aunar los datos históricos de docenas, cientos o miles de expresiones.

Por otro lado, cuando se estudia la tendencia histórica de la serie temporal que permite obtener *Google Trends*, debe hacerse notar que esta herramienta aplicó una corrección a su algoritmo de asignación geográfica en enero de 2011 (*Google Support*, 2018).

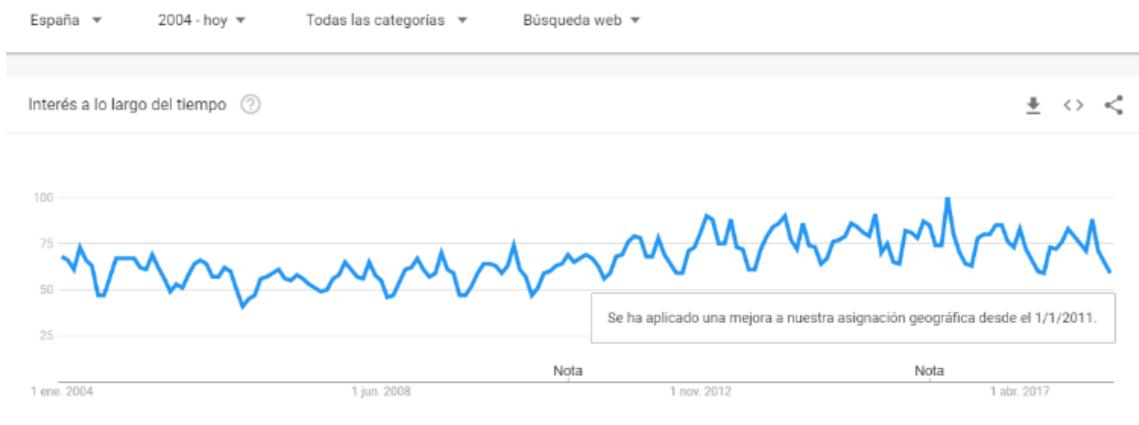


Fig. 3-3 Nota de Google Trends sobre cambios en asignación geográfica. Fuente: Google Trends (2018)

Asimismo, *Google Trends* también aplicó en enero de 2016, una mejora en el sistema de recogida de datos (*Google Support*, 2018) de la que Google no explicita tampoco en qué consistió, por lo que no es posible saber la naturaleza de estos cambios, lo que ha sido queja de diversos estudios al respecto (Nuti et. al., 2014; Lazer et. al., 2014; Cervellin, Comelli, y Lippi, 2017)

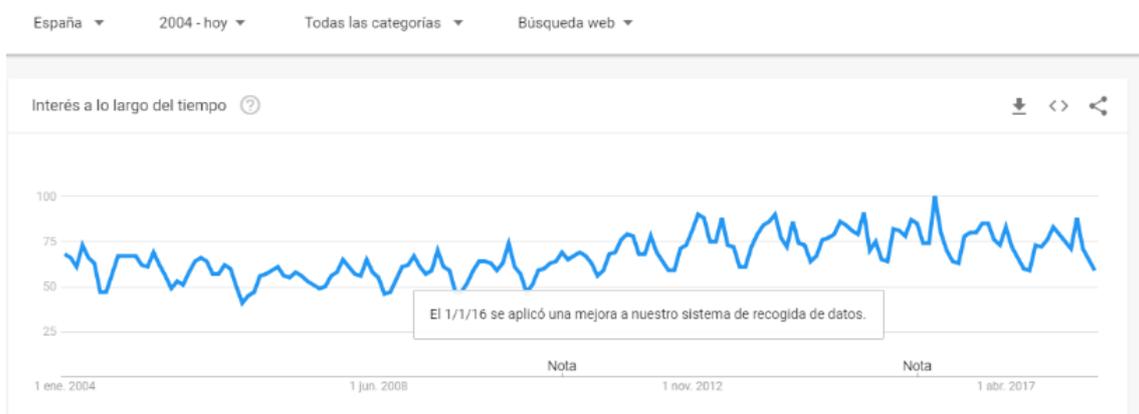


Fig. 3-4 Nota de Google Trends sobre cambios en sistema de recogida de datos. Fuente: Google Trends (2018)

Estos cambios sin embargo no afectan al estudio de tendencias en sí, puesto que la consulta a *Google Trends* se realizó en 2017, para cada frase de búsqueda, y como se ha visto en este apartado, el algoritmo de asignación de volumen de búsqueda para cada mes lo hace de forma relativa para toda la serie histórica, con lo que los datos recogidos ya están normalizados, algo ya visto en trabajos de otros autores (Alexandra Bota, 2014; Mccallum y Gwendolynn, 2014).

La validez o no del uso de los datos de *Google Trends* sigue sujeta a debate. Pero no es fácil determinar esta validez de forma generalizable, donde lo que se encuentran son investigaciones focalizadas en temáticas concretas (Shim et. al., 2001; Askitas, y Zimmermann, 2009; Vaughan y Romero-Frías, 2014).

Muchas de las investigaciones, incluso para una misma temática, no utilizan las mismas expresiones de búsqueda, utilizan pocas expresiones, o buscan coincidencias o similitudes en los volúmenes de los datos de *Google Trends* comparados con otros datos, cuando lo que ofrece *Google* no es volumen en sí, sino una tendencia. Así, Tran et. al. (2017) sugieren que la validez de los volúmenes de búsqueda de *Google Trends* para el pronóstico de comportamiento de las tasas nacionales de suicidio es baja, pero como decimos, se focaliza en volúmenes de búsqueda y no en tendencias.

El mismo autor identifica otros estudios de su misma temática que, sin embargo, otorgan una alta validez a los datos de *Google Trends*, (Gunn III y Lester, 2013; Ma-Kellams et. al., 2016) lo que también se detecta en estudios sobre otras temáticas como la opinión pública o la economía, que también avalan la validez de esta fuente de datos (Zhu et. al., 2012) Este mismo hecho es señalado por Preis et. al. (2013) quienes encuentran que los datos de *Google Trends* pueden servir como predictores de la evolución de la Bolsa, pero que su capacidad predictiva difiere con el término de búsqueda elegido.

Esta es una de las razones por las que en este estudio se han extraído datos de una gran cantidad de expresiones de búsqueda, como muestra de los hábitos de búsqueda de información sobre la temática analizada, de forma que fuera un conjunto amplio de las series de datos descubiertas el que nos diera una información, si no exacta, sí al menos indicativa de esas tendencias de búsqueda.

En segundo lugar, no debería intentar deducirse de ello volúmenes de búsqueda absolutos ni establecer relaciones causales aisladas, como ocurrió en el caso de Google Flu (Butler, 2013) donde debía observarse la influencia de los medios para entender la variación de popularidad de búsquedas relacionadas, influencia que también se ha comprobado en temáticas como los

ataques isquémicos (Abedi et. al., 2015) o en el caso de estudios sobre epidemiología, donde se detectó el efecto de las campañas de inmunización nacional en la estacionalización de las búsquedas sobre esta temática (Bakker, et. al., 2016). Por eso, se observan cada vez más investigaciones que utilizan Google Trends cruzando sus datos con los de otras fuentes (Blazquez y Domenech, 2018) para así tener en cuenta diversos factores con los que explicar las causas de las tendencias halladas.

La misma compañía Google ha puesto a disposición otra herramienta, Google Correlate (Google.com, 2011), para ayudar a los investigadores a correlacionar datos de otras fuentes con datos históricos sobre expresiones de búsqueda existentes en su log de búsquedas. Con respecto a este estudio, hará falta cruzar los datos aquí obtenidos con otros, de forma que permitan una mejor identificación de las posibles causas de determinados patrones de búsqueda, como la temporalidad o atemporalidad de las mismas.

3.2.4. De la técnica de clasificación longitudinal de expresiones de búsqueda en temporales y atemporales

Como se ha referido en el punto 1.5.7., una vez se consiguió reunir un corpus de datos longitudinales mediante *Google Trends*, de parte del conjunto total de expresiones de búsqueda localizadas mediante *Google Keyword Planner*, se procedió a separar la muestra obtenida en dos agrupaciones de frases de búsqueda.

Estas dos clases se denominaron temporales y atemporales, lo que permitió distinguir entre aquellas expresiones de búsqueda (temporales) cuyos datos mostraban una serie temporal en la que, tras un momento álgido en los valores de índice de búsqueda, se desarrollaba una sucesión más o menos rápida de valores a la baja donde primaba el valor cero de índices de búsqueda, y las atemporales, aquellas que incluso ofreciendo también un momento máximo de valor de índice de búsqueda en la serie, mantenían valores suficientes en la serie como para que no *Google Trends* no adjudicara valores cero en los meses posteriores.

De esta forma se pretendía separar las expresiones de búsqueda entre aquellas más estables, que podían crecer, decrecer o permanecer más o menos en la misma franja a lo largo de toda la serie (las expresiones atemporales) de las temporales, con un momento álgido y una rápida bajada e incluso desaparición, para entender las características subyacentes en ambas tipologías, que en principio identificaban conductas de búsqueda distintas.

El análisis cualitativo posterior, en el que se han visualizado las distribuciones obtenidas en ambas clases, ha dado como resultado el descubrimiento de aproximadamente un 5-10% de expresiones de búsqueda atemporales en las que se observaban tendencias que podían ser consideradas candidatas a ser expresiones temporales atendiendo a un criterio más amplio de asignación de la categoría “temporal”, observando patrones de comportamiento similares, aunque sus índices de búsqueda fueran suficientemente altos a lo largo de la serie como para no ser clasificadas como temporales.

Tomemos como ejemplo necesidades de información que se expresan al buscar sobre la Biblioteca Digital Mundial. Inaugurada en 2009 (Biblioteca Digital Mundial, 2009), ofrece expresiones de búsqueda tanto temporales como atemporales, cuando la tendencia es muy similar en ambos casos. En el ejemplo de la figura 3-5, generada a partir de los datos extraídos en *Google Trends* para la expresión de búsqueda atemporal “biblioteca digital mundial”, se puede determinar que muestra escasos valores en la muestra anteriores al momento de su

lanzamiento público (principios de 2009) Tras un momento álgido, con un crecimiento casi de cero a 100 en un mes, el nivel de los subsiguientes valores de índice de búsqueda quedan relegados a una secuencia bajísima, donde los valores no llegan a cero, pero quedan limitados a valores ínfimos.

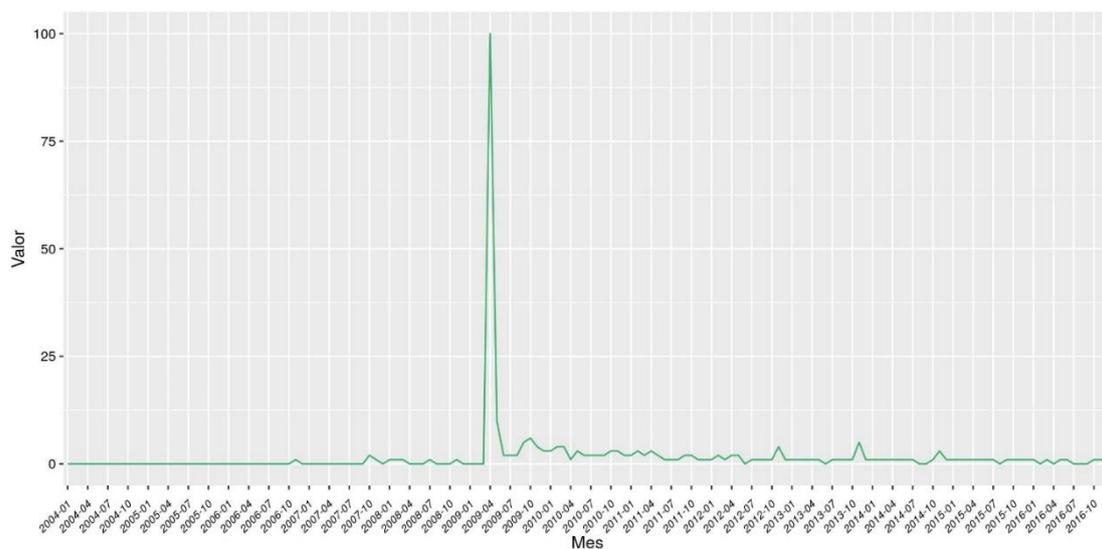


Fig. 3-5 Evolución de índices de búsqueda de la expresión atemporal “biblioteca digital mundial”. Elaboración propia, sobre datos de Google Trends.

Ahora tómesese alguna expresión de búsqueda clasificada como temporal, pero que lo sea sobre el mismo concepto antes mencionado como “biblioteca digital mundial unesco”, “biblioteca virtual unesco”, “biblioteca mundial virtual”, “biblioteca virtual mundial” o www.wdl.org biblioteca digital mundial. En este caso, como se puede apreciar en la figura 3-6, los datos de la frase “biblioteca digital mundial unesco” muestran una distribución muy similar, incluso con algunos valores más altos que los de la expresión clasificada como temporal.

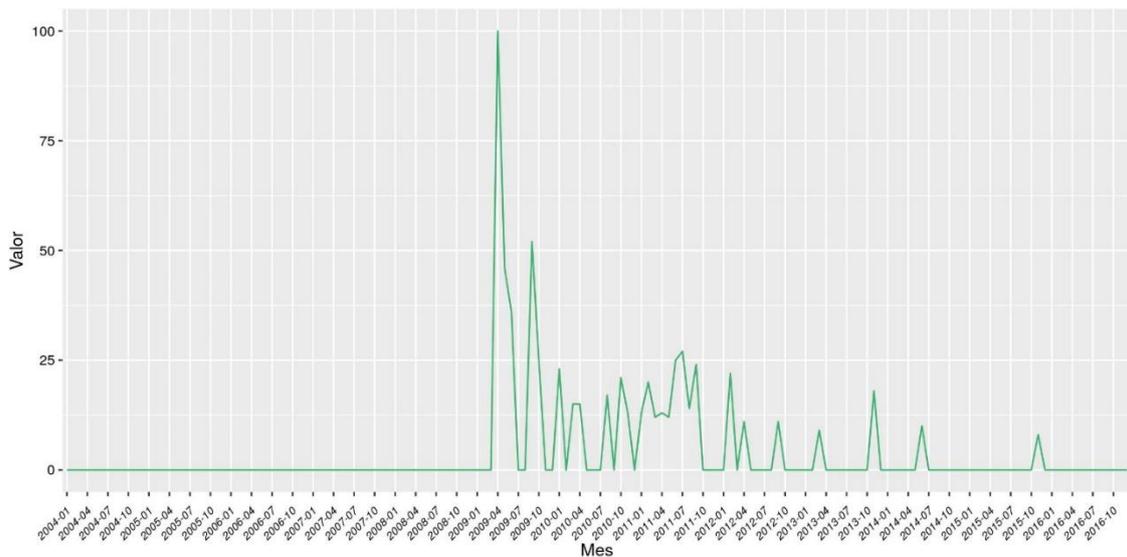


Fig. 3-6 Evolución de índices de búsqueda de la expresión atemporal “biblioteca digital mundial unesco”.
Elaboración propia, sobre datos de Google Trends.

La misma tipología de distribución se ha detectado en algunos casos concretos, lo que quizá obliga a reevaluar la fórmula utilizada para realizar la clasificación. Esta fórmula utiliza los valores cero como indicativo de la condición de temporalidad de una expresión de búsqueda, incluyendo los valores sueltos de ceros como al número de segmentos con valores seguidos de cero.

Pero un valor de índice de búsqueda cero en *Google Trends* no implica que en ese mes no haya habido ninguna búsqueda, sino que no ha llegado al límite mínimo o *threshold* que usa la herramienta para dar un valor diferente de cero, literalmente un “volumen de búsquedas bajo” según *Google Support* (2018). Por tanto, cabe preguntarse dónde está el límite, si sólo se utilizan los ceros como indicativo de la condición de temporalidad, y cómo decidir cuándo una expresión de búsqueda se puede clasificar como temporal o atemporal.

Tres parámetros podrían aportar más información para decidir sobre la temporalidad de una expresión: la diferencia de índice de búsqueda entre el valor más alto (100) y el resto de la serie, la existencia de estacionalidad, y la familia semántica a la que pertenezca la expresión.

La diferencia de índice de búsqueda entre el valor más alto de la serie (100) y el resto, cuando es muy acusada, es un buen indicativo de temporalidad. Pero también se observa esta diferencia en expresiones atemporales en las que, aunque hay mucha discrepancia entre el valor más alto y la media de valores de la serie, los valores del resto del histórico puede que

vayan bajando gradualmente y luego se mantengan estables, o tengan tendencias al crecimiento en otros momentos. Es la rapidez en la subida y bajada del patrón de búsquedas, el que nos da la clave (incluso intuitivamente al visualizar las gráficas) de que estamos ante una búsqueda de tipo temporal.

La existencia de estacionalidad ofrece pistas claras de que una búsqueda es atemporal, es decir, permanece en la serie, bien sea con una media de volumen medio a lo largo de la serie, bien bajo, con momentos de crecimiento y decrecimiento, normalmente en ciertos períodos de cada año. Este método nos permitiría identificar las claramente atemporales, pero no serviría como criterio exclusivo para segregar las demás como temporales.

Finalmente, se aprecia que la reunión de expresiones semánticamente relacionadas, que buscan mediante diferentes expresiones información alrededor del mismo concepto semántico (en el ejemplo, la Biblioteca Digital Mundial) permite detectar candidatos probables a pertenecer a la misma clasificación, ya sea temporal o atemporal, al generar similares patrones de conducta en los usuarios.

Aunque se seguirá realizando el análisis de las expresiones de búsqueda utilizando el cálculo definido en 1.5.7., la unión o combinación de otras estrategias de detección de la condición de temporalidad como las referidas, deberían ser estudiadas en el futuro, combinando criterios, de forma que se pueda mejorar la clasificación y posterior análisis comparativo.

3.3. Discusión de resultados

Una vez discutida la metodología y analizadas las herramientas de extracción de datos para el presente estudio, se plantea el examen de los resultados obtenidos durante el mismo. Si en el capítulo anterior se han descubierto posibilidades de mejora de la metodología, otro tanto será posible hacer al escudriñar los resultados, que servirán para aportar más información a la discusión de la metodología, por un lado, y por otro servir de apoyo a las investigaciones en materia de hábitos de lectura, demostrando cómo existe una dimensión no explorada en esta disciplina, (toda una subdisciplina en sí) la del estudio de los hábitos de búsqueda de lectura y otros medios de uso y consumo de productos culturales.

Durante esta discusión de los resultados, se hará referencia a dos grupos de datos: por un lado, se trabajará con el conjunto de expresiones de búsqueda identificadas gracias a *Google Keyword Planner*, y por otro, con la muestra menor de aquellas expresiones de búsqueda de las que *Google Trends* ha aportado datos de índice de búsqueda mensual durante el período 2004-2016, que a su vez forma parte del conjunto anterior de expresiones.

La muestra de expresiones de búsqueda obtenidas de *Google Keyword Planner* ha sido estudiada fundamentalmente a través de los textos, de las palabras y sintagmas que las componían, traduciéndose en un descubrimiento de la intención de búsqueda, facetas y aspectos que revelaban. Así, se escudriñará la distribución de las expresiones de búsqueda dentro de las facetas y aspectos, y la clasificación de una misma expresión en varias facetas o aspectos (lo que se denominará como “co-ocurrencia en la facetación”)

La muestra de expresiones subconjunto de la anterior, conformada por aquellas que han aportado resultados en *Google Trends* en forma de serie histórica, ha permitido el estudio de la temporalidad y atemporalidad de las expresiones de búsqueda. Se trabajará el análisis de la evolución longitudinal del conjunto tanto de forma conjunta o general, como comparando las expresiones clasificadas como temporales (las que aparecen y casi desaparecen en un breve lapso de tiempo en la serie longitudinal) versus las atemporales, que mantienen mayor presencia a lo largo del histórico de datos.

Su estudio, llegando al detalle de analizar cada tipología por separado, permitirá el hallazgo de diversos patrones de conducta de búsqueda de los usuarios descubiertos dentro del ámbito de la tesis. También descubrirá tendencias globales que pueden aportar información cualitativa y cuantitativa a los investigadores en el área del estudio de los hábitos de lectura, desde otra

perspectiva diferente de las clásicas, reportadas en el capítulo dedicado al marco teórico. Por último, alumbrará el planteamiento de varias cuestiones que podrían genera líneas abiertas de investigación futura.

3.3.1. Distribución de expresiones de búsqueda en las facetas

Comenzando por el análisis de las 30.925 frases de búsqueda, se puede observar en primer lugar cómo se distribuyen las mismas entre las distintas facetas obtenidas en 1.5.4. En la tabla 3-4 se puede observar el número TOTAL de expresiones de búsqueda o frases de búsqueda que han sido clasificadas en cada faceta y/o aspecto. Asimismo, y para cada una, se muestra MAX, que representa el número máximo de facetas que coinciden en una búsqueda clasificada con la faceta indicada en la fila. También se incorpora la MEDIA, que representa el número medio de facetas que coinciden en una frase de búsqueda clasificada con la faceta indicada en la fila, y MIN, el número mínimo de facetas que coinciden en una búsqueda clasificada con la faceta indicada en la fila.

La faceta en la que se han clasificado más expresiones de búsqueda ha sido “Error” (13.127 frases) y la faceta con menos palabras “Erótico”, con 58. El número medio de facetas que coinciden en una frase de búsqueda clasificada, entre todas las facetas, es de 3,49 facetas por expresión.

Si descendemos en el análisis a las palabras sueltas o términos que conforman las distintas expresiones de búsqueda estudiadas, el número de facetas máximo por el que ha sido clasificada una palabra es de 6, mientras que el mínimo ha sido 1, dando como resultado un promedio de 1.53 facetas por las que han sido clasificadas las palabras sueltas.

Estas cifras globales permiten entender que las frases de búsqueda no sólo pueden ser clasificadas dentro de categorías mutuamente excluyentes, como la clasificación de intenciones de búsqueda de Broder (2002), sino que pueden también ser clasificadas, aun dentro de la anterior clasificación, en torno a varias y diferentes facetas, siguiendo a González-Caro y Baeza-Yates (2011).

Tabla 3-4 Cifras globales facetas y aspectos. Elaboración propia.

FACETA	MIN	MEDIA	MAX	TOTAL
Abreviatura	0	0,29	11	2.308
Ámbito	0	4,70	11	8.593
Autor	0	1,57	10	3.872
Biblioteca	1	2,96	9	1.477
Ciencia	1	4,72	11	1.393

Deporte	0	2,94	9	99
Desambiguación	0	4,47	11	8.006
Entidad	0	1,21	10	3.730
Error	0	1,88	11	13.127
Film	2	5,21	11	62
Formato	0	4,42	11	3.950
Género	0	4,39	11	9.845
Idioma	0	1,16	11	7.334
Infantil	0	4,78	11	2.595
Lugar	0	2,80	10	2.095
Nombre personal	0	3,42	11	5.110
Personaje	0	4,78	11	2.777
Plural	0	3,58	11	10.879
Pregunta/frase	0	4,12	11	1.596
Tiempo	0	3,92	10	1.044
Tipo narrativo	0	3,79	11	11.866
Título	0	1,31	11	11.500
Topic Ficción	0	4,18	10	1.750
Topic No Ficción	0	4,02	11	8.043
Transaccional	0	3,77	10	3.237
Videojuego	2	6,01	10	91
Erótico	0	3,80	9	58

Al ordenar la tabla 3-4 por el número de expresiones de búsqueda que han sido clasificadas en esa faceta o aspecto (columna “total”) podemos apreciar en el diagrama de barras de la figura 3-7, cómo las facetas y aspectos que contienen más expresiones de búsqueda son “error”, “tipo narrativo” “título” y “plural”, lo que permite descubrir cuáles son las tipologías de expresiones de búsqueda para las que más variantes de frases de búsqueda se usan por parte de los usuarios que buscaron información sobre lectura en España y en español en el período analizado, al margen de su volumen o frecuencia de búsqueda.

Es decir, en este caso no se utiliza como variable de análisis el volumen de búsqueda mensual, no se ordena según cuánto se buscan las frases así clasificadas, sino según cuántos términos distintos, cuántas frases de búsqueda diferentes, se han clasificado en cada faceta o aspecto, lo que indica la variedad de expresiones de búsqueda que pertenecen a una clase y, por tanto, que ciertas facetas aportan mayor variabilidad expresiva.

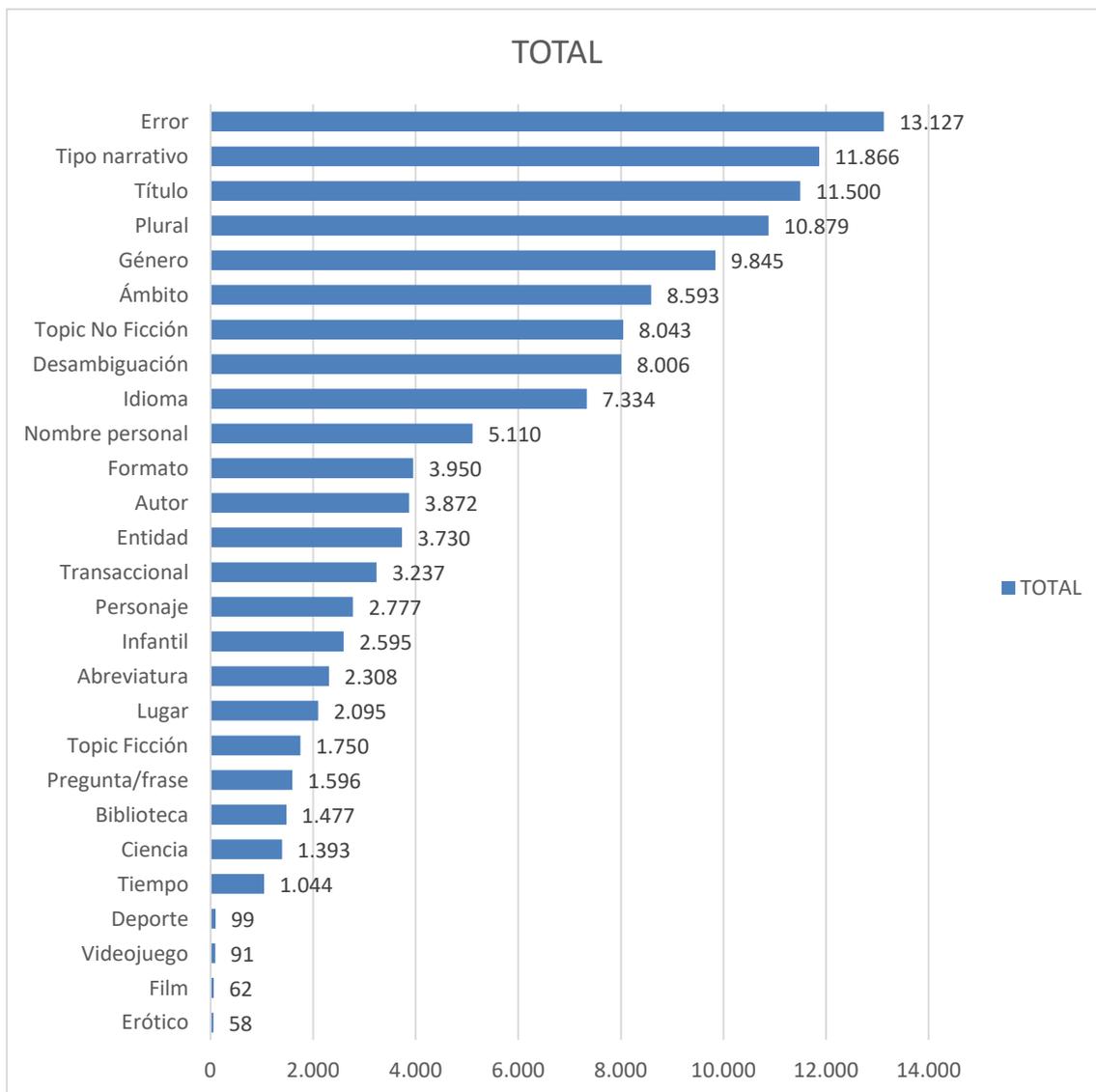


Fig. 3-7 Facetas ordenadas por número de expresiones de búsqueda clasificadas en esa faceta. Elaboración propia.

Estos datos llevan a deducir que, por la mayor variedad de frases pertenecientes a estas tipologías, los usuarios de Google que buscan en España utilizan una mayor variedad de expresiones de consulta al buscar lectura, en buena medida mencionando de una forma u otra el tipo narrativo de las posibles obras a recuperar, cometiendo errores ortográficos, en casi un 99% debido a la falta de tildes. Asimismo, se observa que, por ejemplo, hay un gran número de frases de búsqueda distintas que usan el plural (en concreto en 10.879 expresiones, el 35% del total de expresiones analizadas en este apartado), lo que apunta a que los usuarios buscan listas de ítems en las respuestas que esperan por parte de los motores de búsqueda.

Este punto debería comprobarse en una futura investigación con un número mayor de consultas, pero sirva como ejemplo, si se hace una búsqueda de este tipo de expresiones en Google, y como se puede apreciar en la imagen 3-8, que se observa cómo el motor de búsqueda selecciona (en 2018) para sus páginas de resultados (SERPs) y con respecto a ese tipo de expresiones de búsqueda, listados de ítems, bien sea autores, temáticas, obras, etc., coincidiendo con autores como Dou et. al., (2015).



Fig. 3-8 Página de resultados de la búsqueda perteneciente a la faceta Plural "poesias del dia del padre cortas" realizada en octubre de 2018. Elaboración propia.

Si se realizara tal análisis a mayor escala y se comprobara esta posibilidad para este tipo de consultas, permitiría a los profesionales de SEO, webmasters y creadores de contenidos digitales, entender mejor qué tipo de contenidos está primando o posicionando el motor de búsqueda, lo que podría servir para construir una metodología de detección de mejores prácticas (best practices) basada en la que se hace actualmente para detectar *features* o funcionalidades especiales en las páginas de respuestas de Google (Moz.com, 2018) con el propósito de redactar tanto contenidos digitales mejor relacionados y enfocados a cumplir con las expectativas de los usuarios, como a ayudar en la redefinición de arquitecturas de información de sedes web.

3.3.2. Co-ocurrencia en la facetación

El potencial de la facetación se explicita cuando ayuda a la recuperación de información, en este caso de expresiones de búsqueda que representen una determinada intención de búsqueda. La clasificación multifacetada de expresiones de búsqueda ayuda a explotar sus posibilidades, al permitir descubrir una mayor especificidad en la localización de expresiones de búsqueda no sólo muy relacionadas, sino que expresan mejor la intención de búsqueda que la clasificación monojerárquica (Kurland, Lewenstein y Porat, 2007; Belkin et. al., 2009; Croft et. al., 2010; Dou, et. al., 2011).

Por ejemplo, la expresión de búsqueda “librería general zaragoza horario” se puede englobar en tres facetas o dimensiones (query dimensions las denominarían Dou et. al., 2011). “librería” indica que se puede clasificar como búsqueda con intención de encontrar información sobre una entidad. “Zaragoza” implica que se especifica un “lugar”, donde el usuario puede pensar que debe indicar la ciudad para identificar una librería concreta de la marca “librería general”. Por último, el uso de la palabra “horario”, tiene una dimensión temporal, por lo que se puede clasificar dentro del aspecto “tiempo”.

Ahora bien, al clasificar de esta forma miles y miles de frases de búsqueda, permitimos a un sistema de recuperación de información sobre expresiones de búsqueda, recuperar segmentando, expresiones que indiquen mejor una intención de búsqueda muy específica, al permitir combinar facetas. En el ejemplo que nos ocupa, se podría localizar esta expresión de búsqueda segmentando a través de distintas combinaciones de facetas y aspectos:

- Entidad
- Lugar
- Tiempo
- Entidad + Lugar
- Entidad + Tiempo
- Entidad + Lugar + Tiempo
- Lugar + Tiempo

Por tanto, un aspecto a considerar en el análisis de la clasificación obtenida, es el grado de co-ocurrencia entre facetas y aspectos, donde co-ocurrencia implica la aparición de un mismo

ítem (en este caso expresiones de búsqueda) en pares de documentos, clases o en este caso facetas (Vilares, 2009; Irazú Hernández et.al., 2012)

Hay otras técnicas para localizar expresiones de búsqueda relacionadas, como el análisis de clics o selecciones de los usuarios de respuestas ante distintas expresiones de búsqueda en las páginas de respuesta de buscadores (Wang et. al., 2014) o analizando sesiones de búsqueda en las que co-ocurren los términos (Huang, Chien y Oyang, 2003), observar si existe co-ocurrencia de términos de búsqueda en distintas facetas y aspectos, o en qué casos hay un mayor número de expresiones que co-ocurren. Pero la existencia misma de co-ocurrencia en la facetación de estas frases nos permitirá identificar, siguiendo a (Saha Roy, 2013), que las consultas explicitadas mediante expresiones de búsqueda no son solo una forma de texto que usa secuencias aleatorias de palabras, sino que tienen propiedades distintas, propias de esta forma de expresión de las necesidades humanas de información (Ferrer I Cancho y Sole, 2001)

Si analizamos en su conjunto las facetas y aspectos obtenidos, observamos en la figura 3-9 la relación que hay entre pares faceta- aspecto, donde el mayor grado de relación viene dado por el porcentaje de frases clasificadas en una faceta (filas) que también se han clasificado como un aspecto (columnas). Así, se observa que se ha encontrado mayor coincidencia en pares faceta-aspecto como Lugar-Error, Biblioteca-Error o Deporte-Plural, lo que nos da información sobre qué facetas y aspectos están más relacionados, dónde aparecen más expresiones que podrían ser clasificadas tanto de una como de otra forma.

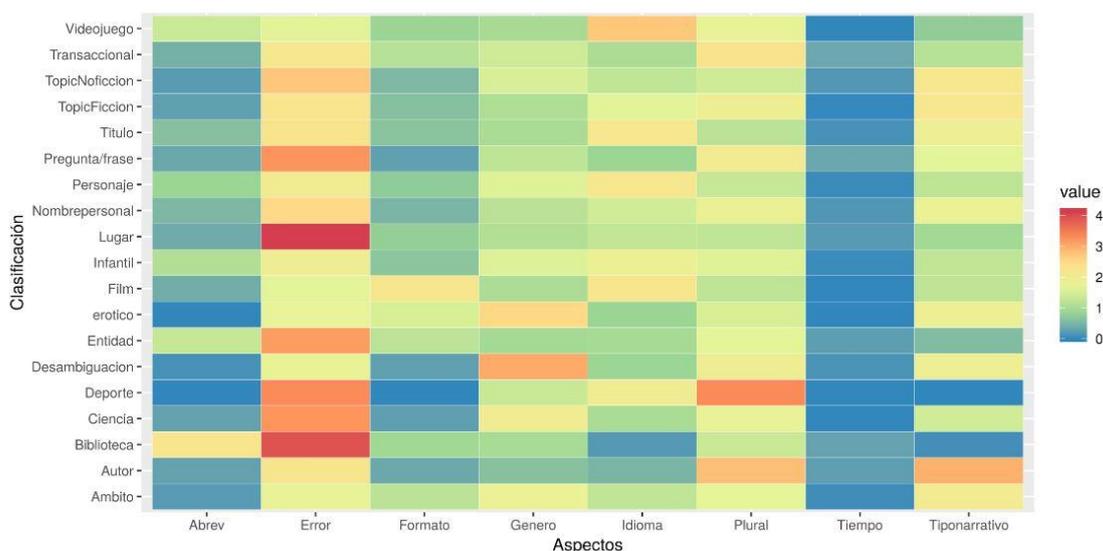


Fig. 3-9 Grado de coincidencia en la facetación o co-ocurrencia en la facetación entre aspectos y facetas. Elaboración propia.

Esta visualización se puede trasladar a la tabla 3-5, que expresa los valores en porcentajes de co-ocurrencia, y que asimismo permite tener una imagen más clara del volumen de aparición de frases de búsqueda que integran elementos o palabras que pueden hacer considerar la expresión de búsqueda como perteneciente a los pares de facetas-aspectos estudiados.

Tabla 3-5 Porcentajes de Co-ocurrencia en facetación (en %)

	Abrev	Error	Formato	Genero	Idioma	Plural	Tiempo	Tipo narrativo
Ámbito	2,09	17,65	11,82	18,34	12,40	16,50	0,60	20,60
Autor	2,92	21,95	3,68	6,22	4,95	28,17	2,54	29,57
Biblioteca	22,83	39,40	8,97	9,78	1,90	13,32	2,99	0,82
Ciencia	2,89	32,37	2,60	20,23	10,12	17,63	0,00	14,16
Deporte	0,00	33,33	0,00	13,33	20,00	33,33	0,00	0,00
Desambiguación	1,03	17,49	2,65	30,22	8,61	19,60	1,26	19,15
Entidad	12,87	31,52	12,00	9,54	9,63	16,29	2,45	5,69
Erótico	0,00	17,02	14,89	25,53	8,51	14,89	0,00	19,15
Film	4,08	16,33	22,45	10,20	22,45	12,24	0,00	12,24
Infantil	10,87	19,62	6,76	15,50	18,81	15,72	0,37	12,34
Lugar	3,89	41,35	7,64	10,95	12,54	12,39	2,02	9,22
Nombre personal	5,28	25,52	4,99	11,70	14,20	18,51	1,58	18,23
Personaje	8,48	20,59	7,20	15,73	22,19	13,23	0,43	12,16
Pregunta/frase	3,41	32,44	2,68	12,20	8,54	20,98	3,41	16,34
Título	6,10	23,09	6,56	10,00	22,23	11,86	1,04	19,11
Topic Ficción	2,69	22,90	6,06	10,61	16,16	19,19	0,17	22,22
Topic Noficción	2,12	27,42	5,30	14,89	12,29	13,98	1,73	22,27
Transaccional	4,49	21,41	11,40	14,02	10,29	23,48	3,52	11,40
Videojuego	13,58	16,05	8,64	9,88	27,16	17,28	0,00	7,41

Como se puede observar, si se buscan las facetas y aspectos más relacionados a tenor de la co-ocurrencia de frases en los pares de clasificaciones estudiadas, en general “Error” es el aspecto que más co-ocurre con buena parte de las facetas descubiertas (como un 32,37% con “Ciencia”, 33,33% con “Deporte”, o 32,44% con “Pregunta / frase”, por ejemplo). “Biblioteca” es la faceta más relacionada con otro aspecto, en este caso “Error” (39,40%) lo que llevaría a deducir que las búsquedas relacionadas con las bibliotecas ingresan en el buscador con multitud de errores (según el criterio usado para delimitar los errores, donde se observa una gran falta de uso de tildes, por ejemplo. “Biblioteca” también co-ocurre un 22,83% con el aspecto de “abreviatura”. Este dato ayuda a suponer que hay muchas expresiones que denotan una intención de búsqueda sobre entidades bibliotecarias que utilizan abreviaturas, lo

que es cierto en parte. Cuando se estudia el listado de expresiones de búsqueda que cumplen ambas condiciones, se localizan dos grupos de frases dentro de la faceta “Biblioteca”: aquellas que hacen mención de bibliotecas universitarias, y aquellas que hacen mención de otro tipo de bibliotecas, como públicas y otras.

En las búsquedas sobre bibliotecas universitarias, los usuarios suelen utilizar abreviaturas de las universidades para especificar a qué universidad pertenece la biblioteca que buscan localizar, lo que explica la alta relación entre esta faceta (“Biblioteca”) y este aspecto (“Abreviatura”). Se puede comprobar en diversos ejemplos, como “biblioteca usal”, donde “usal” significa “Universidad de Salamanca” o en “biblioteca upm”, donde upm significa “Universidad Politécnica de Madrid”.

Cuanto mayor es la relación encontrada en facetas donde se han recogido mayor número de expresiones de búsqueda, mayor es la representatividad estadística de la muestra y, por tanto, más interesante es para la posible generalización de conclusiones. En este sentido, hay que destacar la alta relación que hay en el conjunto de las expresiones de búsqueda clasificadas como “Desambiguación” (que tiene asignadas varios miles de expresiones) con el aspecto de “género”, con un 30,22% de co-ocurrencia.

En el extremo opuesto, es interesante denotar que las expresiones clasificadas bajo “Deporte” o “Erótico” no han sido clasificadas también como compuestas con alguna “Abreviatura” (0,00% de las expresiones), o que el aspecto “Tiempo” es el que muestra menos co-ocurrencia con el resto de facetas, mostrando valores entre 0.00% (para “Erótico”, “Film” o “Videojuego”) hasta 3,52% con la faceta “Transaccional”. Pero se debería tomar en menor consideración casos como “Film”, “Videojuego”, “Erótico” o “Deporte”, por la poca aparición de expresiones de búsqueda distintas, lo que podría llevar a generalizaciones equívocas.

Al ser exploratoria, esta investigación se podría intensificar en el futuro, realizando análisis de co-ocurrencias entre pares de facetas, entre pares de aspectos, y entre combinaciones de más facetas y aspectos, buscando aquellos conjuntos de facetas y/o aspectos que muestran mayor coincidencia de aparición de las frases. De esta forma, no sólo se trabajaría buscando coincidencias o co-ocurrencias entre pares, sino entre tripletas de facetas, o entre N conjuntos de facetas. Las expresiones de búsqueda con mayor co-ocurrencia en 3, 4 o más facetas a la vez, implicarían una mayor expresividad en el lenguaje utilizado, y una mayor intención por parte del usuario de ser específico en su lenguaje natural (Huang et. al., 2010; Gupta y Bendersky, 2015). Por otro lado, deberían complementarse con técnicas de clasificación que

no sólo partan de co-ocurrencias entre facetas, sino que se complementen del estudio de las relaciones semánticas entre los términos (Clough et.al., 2009; Veilumuthu y Ramachandran, 2009; Sloan, Yang, Wang, 2015)

3.3.3. Temporalidad y atemporalidad en facetas y aspectos

En este apartado, se analiza la clasificación obtenida en expresiones de búsqueda temporales y atemporales, atendiendo al número de expresiones de búsqueda clasificadas como otra tipología. Al extraer el porcentaje de expresiones de búsqueda clasificada como temporal o atemporal para cada aspecto y faceta, es posible comparar el grado de temporalidad o atemporalidad de cada faceta, es decir, será posible discriminar si una faceta tiene una tendencia mayor a comprender en general, más expresiones temporales (que se inician en un momento de la serie con un pico álgido en el volumen de búsqueda y un descenso abrupto en la serie temporal) o atemporales (con algo de frecuencia de búsqueda en toda o casi toda la serie)

Al observar el gráfico 3-10, donde las columnas más altas muestran mayor número de expresiones de búsqueda, queda patente el hecho de que lo común es observar frases de búsqueda de tipo atemporal, es decir, sostenibles en el tiempo y con una vida larga. En todos los casos, se aprecia que hay más expresiones de búsqueda atemporales que temporales, en ocasiones hasta 8 o 9 veces más, como en el caso del aspecto lingüístico “Plural”.

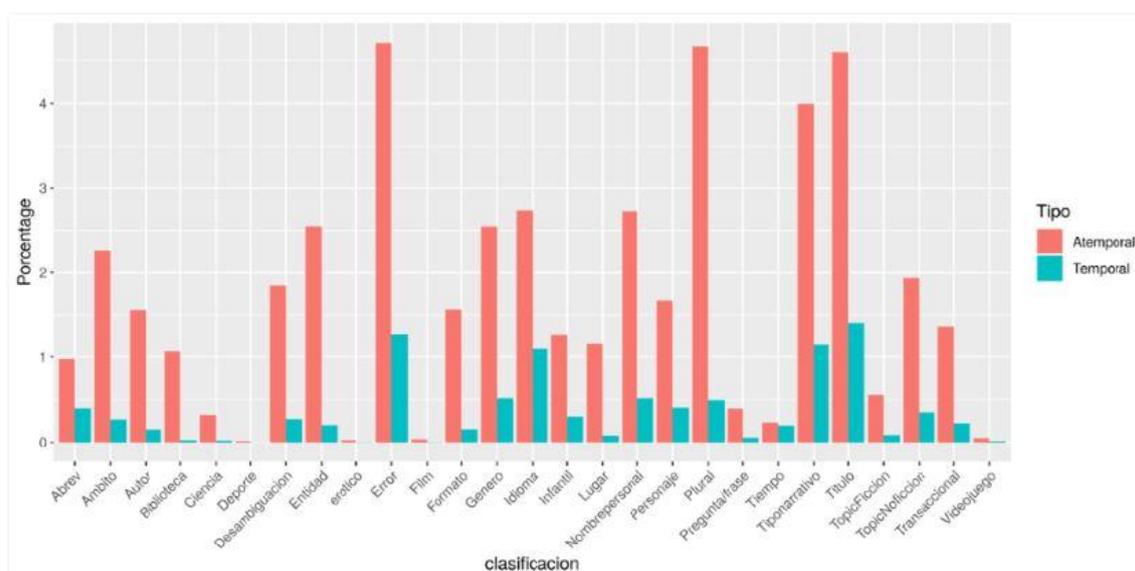


Fig. 3-10 Comparativa de temporalidad o atemporalidad en las facetas

Este hecho nos indica que ese tipo de frases de búsqueda, en los que se usa el plural, son no sólo muy comunes para el ámbito de los hábitos de búsqueda de lectura, libro y bibliotecas (recuérdese que miles de expresiones de las encontradas en el estudio utilizan el plural, siendo una de las facetas con mayor representación en la muestra) sino que, además, siendo atemporales, implican una gran estabilidad a lo largo de la serie analizada. Otros casos extremos, donde hay muchas más expresiones consideradas atemporales, serían el aspecto modificador “Tipo narrativo”, el aspecto lingüístico “Error”, o la faceta informacional “Título” y también en ambos casos, con una alta representatividad en la muestra global.

Por tanto, podría deducirse que, en el ámbito de la lectura, el libro y las bibliotecas, los usuarios tienden a utilizar expresiones de búsqueda que se mantienen más o menos estables en el tiempo (aunque muestren temporalidad, con momentos de mayor y menor frecuencia de búsqueda a lo largo de cada año de la serie) y a repetir similares tipologías de expresiones, utilizando en numerosas ocasiones el plural, buscando por títulos y / o por tipos narrativos, pero con profusión de errores.

Los casos donde se encuentran menores diferencias serían las facetas “Deporte”, “Erótico”, “Film”, “Tiempo” y “Videojuego”, aunque al coincidir con aquellas facetas y aspectos con menor número de expresiones localizadas, hay que tomar sus datos con mayor cautela, por la dificultad estadística de tomar sus resultados como válidos para conclusiones generalizables.

3.3.4. Índices de búsqueda en la serie histórica

En los gráficos 3-11 a 3-13, se puede apreciar el resultado de analizar los valores de frecuencia de cada mes, tomados del conjunto de expresiones de búsqueda de los que se ha podido obtener datos en *Google Trends* para componer la serie temporal de 2004 a 2016. Es decir, se ha tomado como variable de análisis los valores de frecuencia o volumen de búsqueda que aporta *Google Trends* para cada mes de la serie.

Así, en el gráfico 3-11 es posible estimar (para cada tipología o faceta y aspecto y tomando en su conjunto tanto expresiones de búsqueda temporales como atemporales), qué porcentaje de veces se encontraban valores de frecuencia de búsqueda de 0, 1-10%, 11-20%, etc., a lo largo de la serie histórica, siendo 100 el valor máximo de toda la serie, y que lógicamente sólo aparecerá una vez en la misma. Apreciaremos de este modo cuántos meses de la serie temporal muestran una frecuencia de 0, 1-10%, etc., lo que nos dirá cuánto predomina en cada faceta y aspecto cada rango de valores de volumen de búsqueda.

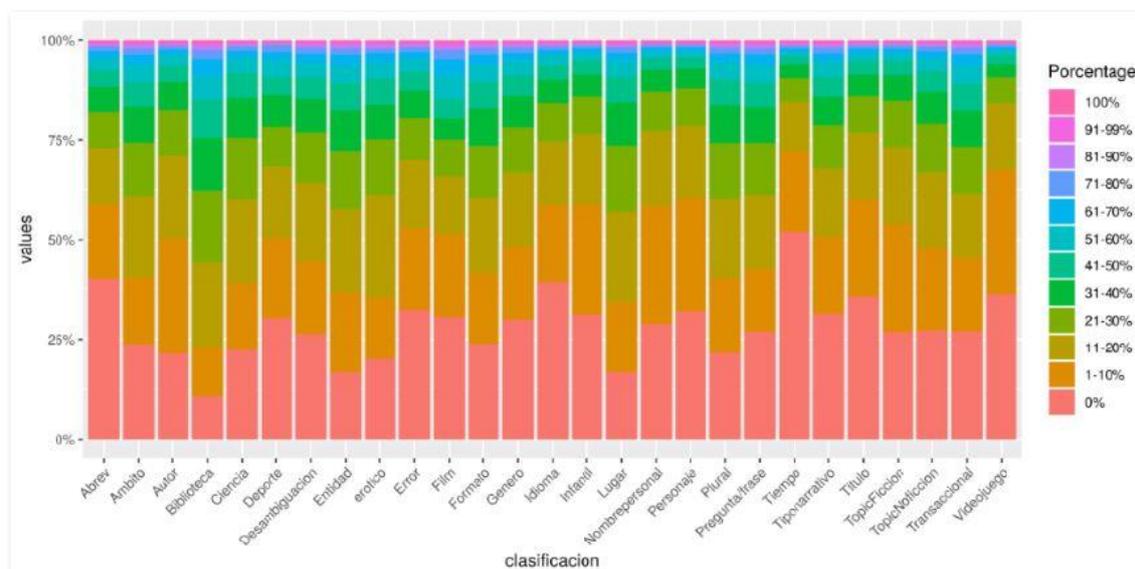


Fig. 3-11 Análisis global de facetas por frecuencias en la serie histórica

Se identifican así patrones de búsqueda a nivel histórico, donde destacan las expresiones de búsqueda analizadas que tienen un componente relacionado con el aspecto de “tiempo”, pues tienen un alto porcentaje (más del 50%) de meses donde la tendencia marcaba cero. Estos datos tienen sentido, si se piensa que buena parte de estas consultas hacen mención a momentos particulares para especificar su consulta, lo que nos da una indicación clara de que

estamos ante búsquedas temporales, que tienen una vida más o menos corta debido a que aluden a momentos concretos, y el conjunto de meses analizados aquí abarca más de una década.

A “tiempo” le siguen otras facetas como “abreviatura”, “idioma”, “título”, o “videojuego” que muestran un gran porcentaje de meses (entre el 25 y el 40% de toda la serie) en los que el valor de tendencia de búsqueda fue cero o cercano a cero, lo que indica que también en estos casos, durante buena parte de la serie se han usado poco este tipo de expresiones de búsqueda.

Pero hay que advertir, estas expresiones de búsqueda lo son en tanto a su relación con el libro y la lectura. Es decir, cuando se habla de expresiones de búsqueda sobre “*Film*”, el conjunto de frases surge de su posible relación (según *Google Keyword Planner*) con las temáticas usadas como semilla (el libro, la lectura...) y no deben tomarse como temáticas propias (“*Film*” o “Películas” sin más) puesto que esa consideración podría dar lugar a generalizaciones inexactas, y debería localizarse un conjunto diferente de frases de búsqueda, que podrían conllevar otras tendencias.

Al desglosar el conjunto de expresiones de búsqueda de los que se han obtenido datos a través de Google Trends, entre frases de búsqueda temporales (las que aparecen en algún momento de la serie temporales y tienden a desaparecer a lo largo de la misma) y atemporales (aquellas que muestran datos de altas y bajas en la frecuencia de búsqueda mensual, pero muestran algún tipo de actividad a lo largo de toda o buena parte de la muestra estudiada) es posible encontrar diferencias interesantes al respecto, que ayudarán a identificar y caracterizar ambas tipologías.

Si estudiamos por separado, para las mismas facetas, sólo aquellas frases de búsqueda clasificadas como “temporales”, aquellas que se inician en algún momento de la serie y suelen tener una vida más o menos corta en la serie histórica, se observa (gráfico 3-12) un gran número de meses en lo que el valor de la frecuencia de búsqueda es cercano a cero, ocupando casi hasta el 75% de los valores analizados en todas las facetas y aspectos para toda la serie.

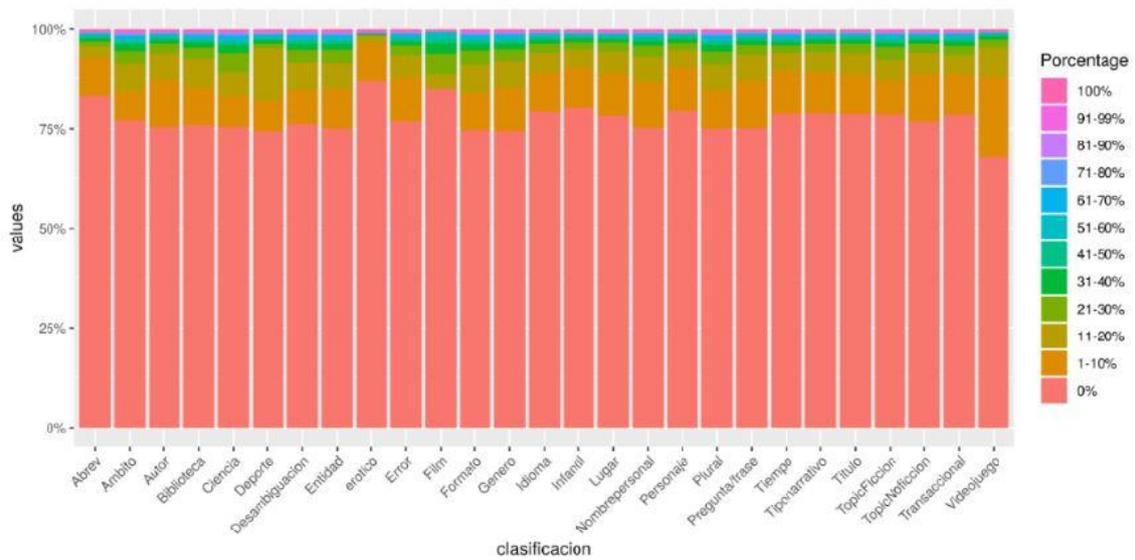


Fig. 3-12 Análisis de facetas por frecuencias en la serie histórica - Expresiones de búsqueda temporales

Además, y siendo consistente con el cálculo realizado para clasificar una expresión de búsqueda como temporal, en el resto de valores obtenidos se puede apreciar que entre un 10-15% de los valores restantes, para toda la serie, pertenece a frecuencias de búsqueda que están entre el 1-10%, y un pequeño porcentaje, en torno al 5-10%, de valores que están entre 21-40%, lo que indica que excepto picos de volumen de búsqueda muy concretos en la serie, en el resto de la misma suele haber poca o nula frecuencia de búsqueda.

En el caso de las expresiones de búsqueda de tipo atemporal, que comportan una cierta estacionalidad (con volúmenes altos y bajos a lo largo de la serie, pero sin mostrar claramente un principio y un final) se aprecia (figura 3-13) una menor incidencia de valores cercanos al cero, y un mayor número de valores intermedios.

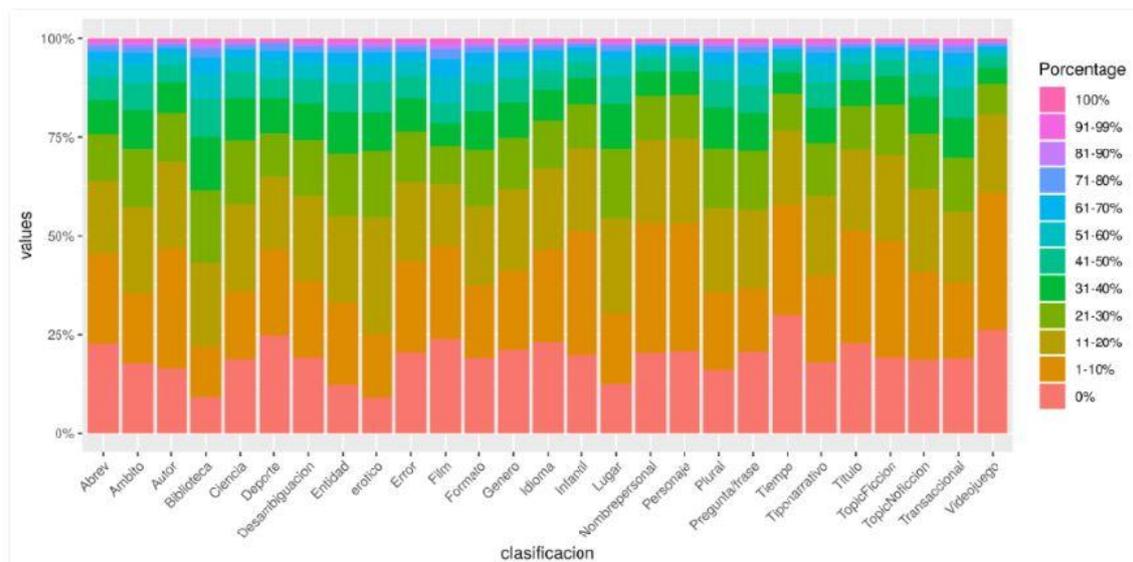


Fig. 3-13 Análisis de facetas por frecuencias en la serie histórica - Expresiones de búsqueda atemporales

Entre otras, cabe destacar facetas como “Biblioteca”, “Film” y “Transaccional”, que muestran que en torno a un 15% de los valores de frecuencias de las series asociadas a las expresiones de búsqueda clasificadas como tales, tienen los mayores porcentajes de valores entre 51 y 80% (colores azules) lo que indica que hay mayor cantidad de valores medio-altos en la serie que en las otras facetas, lo que no implica necesariamente que se busque mucho, sino que comparado con su momento de mayor búsqueda, no hay tantas diferencias a lo largo de la serie como en los casos en que se observa una mayor cantidad de valores cercanos a 0%.

Por el contrario, se observan otros casos en los que hay muy poca proporción de valores altos (en azules) y sí mayor proporción de valores de frecuencia cercanos al cero (en rojos), incluso en esta tipología de expresiones atemporales. Por ejemplo, “autor”, “infantil”, “lugar”, “nombre personal”, “tiempo” o “topic ficción”, tienen una bajísima incidencia de valores medio altos, donde entre el 30 a 50 por ciento, lo ocupan valores entre 0 y el 10%, lo que podría señalar que, incluso para expresiones de búsqueda clasificadas como temporales, es decir, con cierta continuidad en la serie, es sensible a los picos o momentos álgidos en la búsqueda, bien sea por la incidencia de eventos, estrenos de películas, muertes o aniversarios del autor, etc.

3.3.5. Análisis de evolución general - temporales versus atemporales

Como se ha visto en el apartado 3.3.3, se ha tomado como muestra en esta parte del análisis, el conglomerado de los datos de las series históricas de miles de expresiones de búsqueda de las que *Google Trends* ha aportado datos, y no el de unas pocas unidades o decenas. Se refuerza así la utilidad de trabajar con tendencias, con una amplitud y variabilidad de expresiones del lenguaje natural de los usuarios que reflejen mejor las tendencias globales de los hábitos de búsqueda en torno a la lectura para el período estudiado, en España.

De este modo, se minimiza el impacto que expresiones de búsqueda concretas puedan tener en la muestra, y se busca entender cuál es la evolución de la popularidad general de las frases de búsqueda detectadas. No obstante, también se desglosarán análisis específicos para consultas temporales y atemporales, para identificar familias de expresiones que muestren patrones distintos del global.

Para visualizar esta tendencia, siguiendo otras investigaciones como Marchal et. al., 2014; Bakker, et. al., (2016) o Zhang et. al., 2016, pero trabajando con miles de expresiones de búsqueda en lugar de con docenas, se han tomado los datos de todas las series temporales de todas las expresiones de búsqueda analizadas. A partir de ahí, se han obtenido los cuartiles, y la mediana, de cada mes del conjunto de datos, con los que se trazado un diagrama “de caja y bigotes” (*boxplots* o *box and whiskers*).

En este tipo de gráficos, se indican mediante barras verticales los cuartiles mínimos y máximos (siendo el mínimo marcado por el borde inferior, y el máximo por el borde superior) para explorar el rango de variabilidad de los datos para cada mes, y se usa como medida de tendencia central la mediana (la raya del medio) El color aumenta desde verde (valores de mediana cercanos a cero) a rojo intenso, lo que nos da una idea del valor de índice de búsqueda común para cada mes del conjunto de los datos de todas las expresiones analizadas para la serie temporal estudiada (2004-2016).

Nótese en la figura 3-14, donde se han unificado los datos de expresiones de búsqueda tanto temporales como atemporales, que en los primeros meses / años de la serie (2004-2005), se observa que la mediana es muy baja para esos primeros años, mientras que muestra una alta variabilidad en los índices máximo y mínimo (representada por los bigotes)

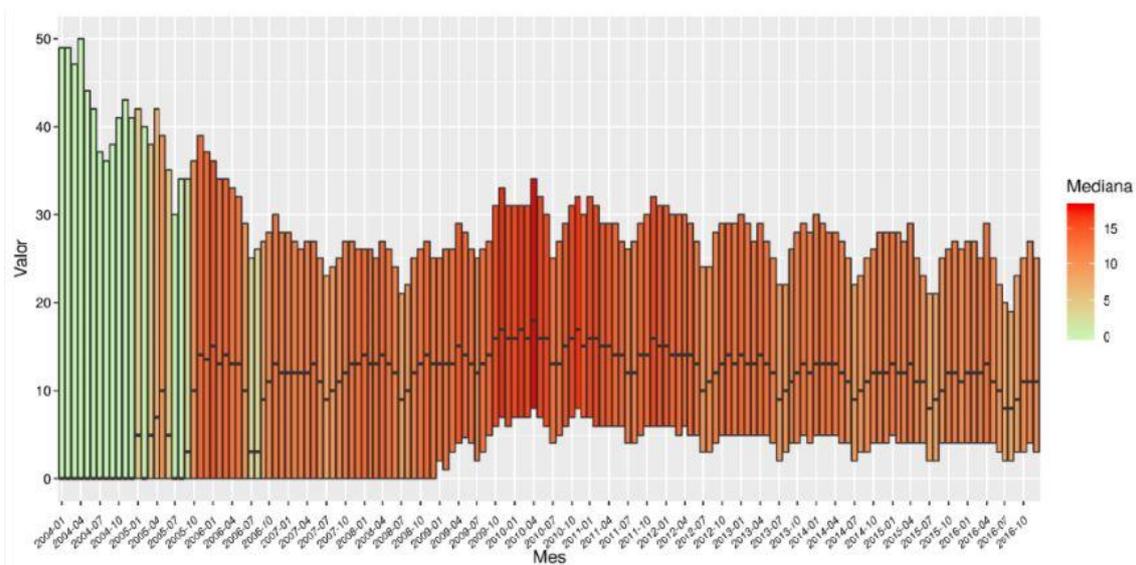


Fig. 3-14 Serie histórica global (todas las facetas, temporales y atemporales) Elaboración propia.

Ese efecto en el primer bloque o época de datos es debido al “ruido” que generan las frases de búsqueda temporales, que son aquellas que han comenzado en un momento dado de la serie temporal, y en general no en esos primeros años, aportando por tanto al cómputo global, para esos primeros años, una mediana cero. La presencia de datos de multitud de búsquedas temporales, que ya contaban con búsquedas realizadas en esos primeros años, hace que los “bigotes” lleguen a ser muy altos, lo que indica que en diversas expresiones de búsqueda sí había un alto índice de búsqueda para esos primeros años.

A partir de ahí, se identifica una época de estabilidad general, e incluso de crecimiento de popularidad entre las búsquedas de esta temática, concretamente durante el período de octubre de 2009 a abril de 2010. Desde entonces, y de forma continuada, se observa una tendencia marcada a la baja, lo que aportaría un indicio de que en general, en Google y en España, se utilizaron progresivamente en menor medida las expresiones de búsqueda detectadas se las que se obtuvieron datos de *Google Trends*, sobre lectura, libro y bibliotecas, durante el período 2010 - 2016.

Se ha de advertir que la muestra utilizada ha trabajado con miles de expresiones relacionadas descubiertas en 2010, por lo que puede que otras expresiones de búsqueda sean detectables en futuros estudios, y puedan aportar datos de un abanico mayor de nuevas formas de expresar las necesidades de información de los internautas que utilizan el buscador, en torno a

esta temática. Por tanto, este análisis lo es tomando en cuenta el lenguaje natural utilizado por los usuarios de Google en España en la época cercana anterior a 2010 (previsiblemente 2008-2010, a tener de las fechas utilizadas en la mayoría de las consultas) en el contexto del interfaz utilizado en Google en esos años y su evolución en años posteriores.

En cuanto a la composición de la visualización de la información de la gráfica 3-14, estos valores de mediana tan bajos al principio de la serie, se pueden expresar gráficamente de otro modo. Si se le da la vuelta a los datos de la serie, y lo que se estudia es la estacionalidad de las mismas se puede dilucidar en qué meses de cada año se busca con mayor o menor intensidad sobre esta temática.

En la figura 3-15 se aprecia claramente que año tras año, se busca menos sobre lectura en los meses de julio y agosto, viendo aumentar esa frecuencia de búsqueda global en los meses de Abril, debido probablemente a la Feria del Libro (Manrique Sabogal, 2015) y Noviembre, mes donde comienzan las compras navideñas (Torres Rodríguez, 2016; Deloitte, 2017).

Pero ese patrón anual se rompe en cierta medida especialmente en el año 2004, al inicio de la serie, donde la presencia en el análisis de expresiones de búsqueda temporales, que no han comenzado todavía, hace que los datos de índice de búsqueda sean mucho más bajos o inexistentes. Donde sí hay datos, muestran tendencia similar, como en el caso de los meses de verano e invierno.

Además, un aporte añadido de este análisis de la serie, es que permite visualizar el patrón de comportamiento estacional del conjunto de la muestra analizada y, por tanto, ayuda a identificar que la misma es en efecto y hasta cierto punto, homogénea, con tendencias anuales de subida y bajada en la popularidad similares en multitud de expresiones de la muestra.

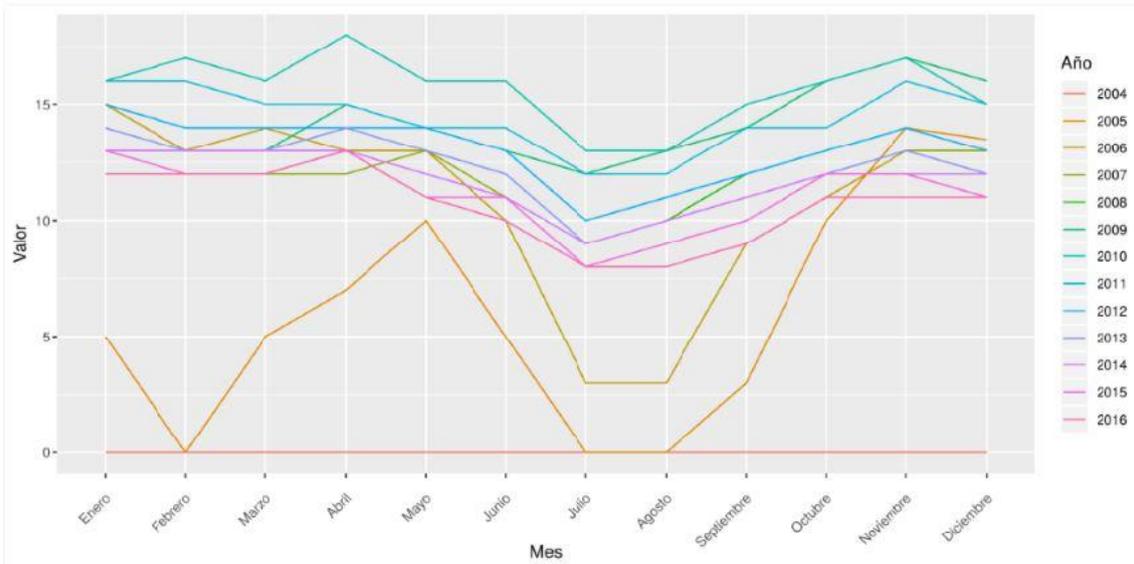


Fig. 3-15 Estacionalidad en toda la serie histórica (todas las facetas, temporales y atemporales) Elaboración propia.

El descubrimiento del posible influjo de las búsquedas de tiempo temporal en la tendencia global, necesita separar las búsquedas temporales de las atemporales, y estudiarlas por separado, para determinar el efecto detectado. Al visualizar la evolución de expresiones de búsqueda o frases de búsqueda atemporales en la figura 3-16, se observa una tendencia similar a la representada en el gráfico 3-14.

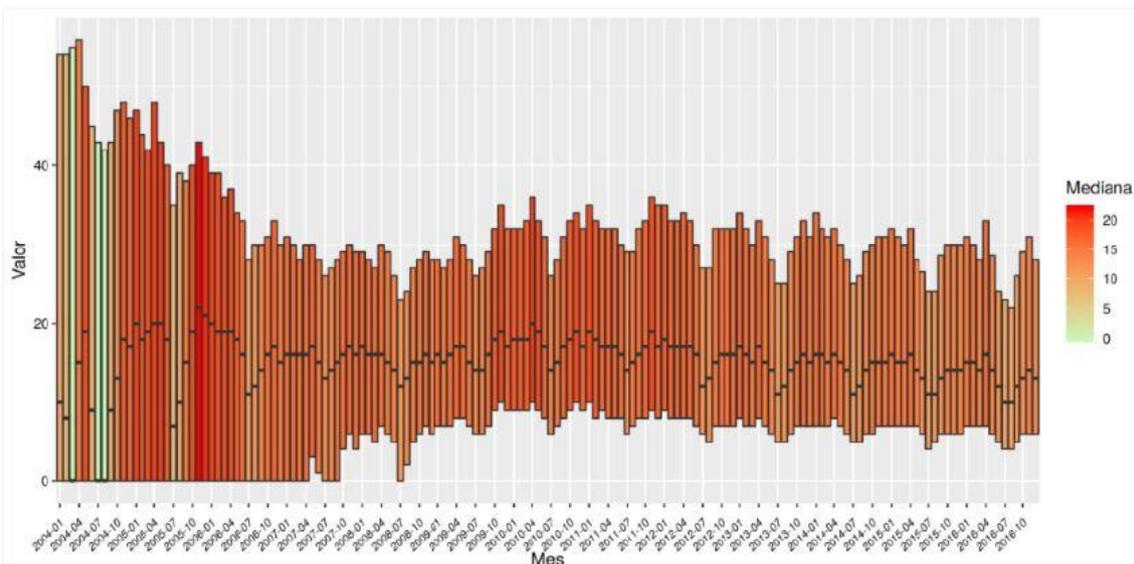


Fig. 3-16 Serie histórica (todas las facetas, atemporales) Elaboración propia.

Aquí, al eliminar las expresiones de búsqueda temporales, cuya baja mediana antes influía en la gráfica de la serie durante los primeros meses / años, se aprecian mayores valores de mediana en esos primeros años, con una alta variabilidad en el índice de búsqueda (valores muy altos, pero también muy bajos), como se puede identificar en esa primera parte izquierda del gráfico 3-16, donde los valores altos son los mayores de toda la serie temporal. Como se ha explicado para el conjunto total de datos (gráfico 3-14) a partir de 2010, la tendencia es claramente a la baja, con una estacionalidad global consistente en los meses mencionados para el gráfico 3-15.

Si ahora, al segregar los datos de las expresiones de búsqueda determinadas como temporales, se podrá visualizar su evolución general, en busca de diferencias apreciables en la conducta de búsqueda en los usuarios que buscan desde España a lo largo del período analizado con las frases de búsqueda clasificadas como atemporales.

En este caso, explicitado en el gráfico 3-17, se muestra claramente un nacimiento de las expresiones de búsqueda temporales localizadas para este estudio a lo largo de los años 2009 y 2010, lo que coincide con el momento en que se realizó la localización, identificación y extracción de expresiones de búsqueda, en 2010.

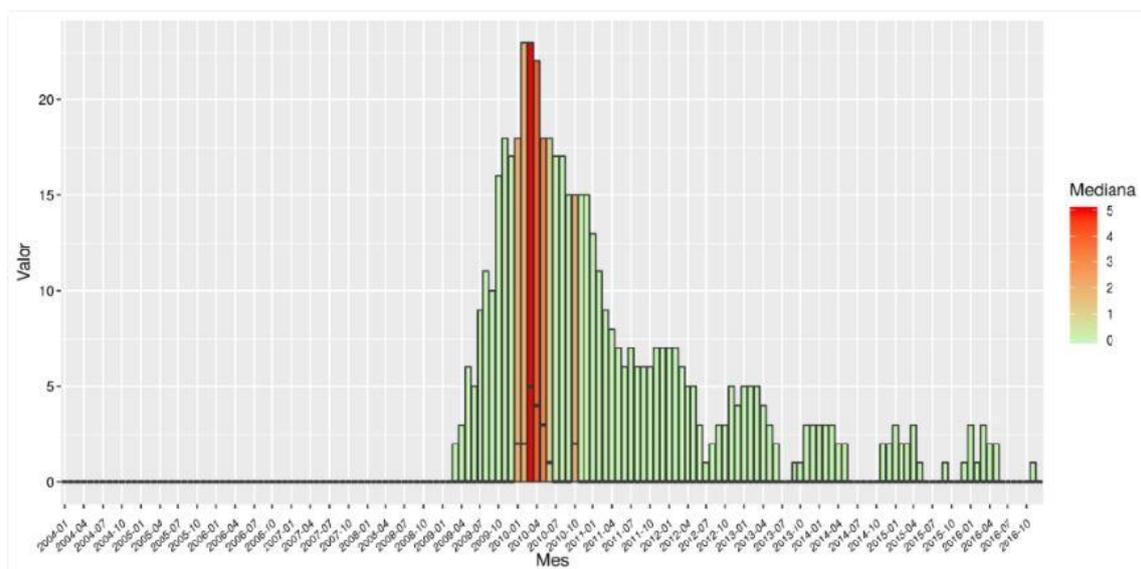


Fig. 3-17 Serie histórica (todas las facetas, temporales) Elaboración propia.

Este hecho, y el hecho de que no parece haber en la muestra cantidades significativas de expresiones que empiecen en años anteriores, resulta relevante para futuros estudios basados en *Google Keyword Planner* como fuente de datos. Al ser ésta la herramienta utilizada para localizar expresiones de búsqueda relacionadas con las palabras semilla, debe hacerse notar que estos datos aportan indicios de que la herramienta puede primar la novedad a la hora de ofrecer expresiones de búsqueda relacionadas, así como búsquedas más estables, o con temporalidad más estable.

Esta posibilidad implica que *Google Keyword Planner* no aportaría apenas expresiones más antiguas en las que la popularidad baje abruptamente o simplemente desaparezca con el tiempo, lo que tiene sentido si se piensa que la herramienta está pensada para ofrecer a los anunciantes de campañas de publicidad en buscadores. Si el sistema recomienda expresiones a los anunciantes, que han tenido o van a tener rápidamente una bajada en la popularidad de búsqueda, estos anunciantes no van a obtener tráfico de usuarios cualificados (relacionados temáticamente y por su intención de búsqueda) por lo que se sentirán frustrados y, eventualmente, podrían dejar de confiar en la herramienta para realizar sus campañas publicitarias.

Por tanto, es lógico pensar que el sistema de recomendación elimine del listado las expresiones de las que ya tiene histórico y de las que sepa que están decreciendo rápidamente, primando las más estables o con crecimiento al alza. De las que no tiene histórico todavía, como en el caso de las “recién nacidas” poco antes de 2010, no puede generar una estimación, apareciendo entre las expresiones de búsqueda relacionadas durante ese año.

A efectos prácticos, y para futuras investigaciones en la modalidad de expresiones de búsqueda temporales, parece recomendable asegurarse de realizar la búsqueda de este tipo de expresiones cada año o a lo sumo cada dos años, de forma que el sistema de recomendación incluya esas expresiones, y no se vean relegadas del estudio en un futuro lejano, lo que puede suponer un sesgo para este tipo de investigaciones, al opacar la detección de las expresiones temporales más antiguas.

3.3.6. Análisis específico de evolución de expresiones temporales

Atendiendo al grupo de expresiones de búsqueda localizadas de tipo temporal, en general nacidas en torno a 2009-2010, se observa un efecto de “pico” o “hype” con una gran subida en la popularidad de búsqueda, durante el momento inicial de este tipo de búsquedas, con un rango de intensidad de crecimiento que variará entre las distintas expresiones de búsqueda, seguido de una tendencia decreciente a lo largo del tiempo más o menos rápida, donde en general (ver gráfico anterior, 3-17) la mediana llega a 0 ya antes de enero de 2011, y donde incluso los mayores valores de índice de búsqueda comienzan a desaparecer en julio de 2013.

Las razones por lo que ocurre pueden ser variadas, lo que en ocasiones puede llevar a analizar expresión a expresión. Pero existen algunos elementos comunes, empezando por el hecho, ya explicado en este mismo apartado, de que suelen comenzar cerca de la fecha de extracción de datos de *Google Keyword Planner*. Otro factor detectado común a este tipo de búsquedas, es la aparición de elementos que expresan el aspecto de “Tiempo”. Tomemos el ejemplo de la frase “premio planeta 2010 ganador” cuya serie se expresa en el gráfico 3-18. Esta frase denota expresamente que se busca información relacionada con un evento, con una localización temporal muy concreta (año 2010).

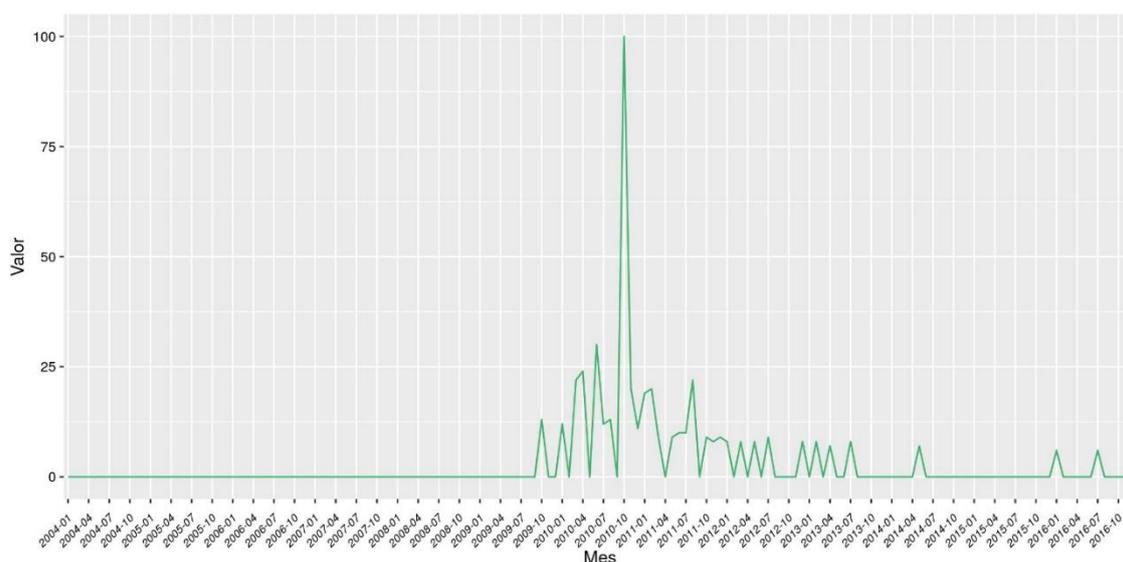


Fig. 3-18 Serie histórica para la expresión de búsqueda “premio planeta 2010 ganador”. Elaboración propia.

En este tipo de explicitación del momento temporal del cual se requiere información, el usuario indica su intención de recuperar información datada en esas fechas. Una vez ha

pasado el evento, conforme pasa el tiempo, se observa un menor interés en ese evento y momento concretos. De forma consistente, se observa asimismo que la mayoría de las expresiones relacionadas por *Google Keywords Planner* en el momento de la consulta que se integran en esta categoría, utilizan años pasados muy recientes con respecto al momento de la extracción de datos, fundamentalmente 2010 y 2009 (“feria del libro 2010”, “libros mas vendidos 2009”, por ejemplo) Más allá, se encuentran en menor medida expresiones de años pretéritos, como 2008 (“premio planeta 2008”) y escasamente algún caso en el que se busque información de 2006 (“premio nobel literatura 2006”) es decir, cuatro años antes del momento de realizar la consulta en la herramienta de *Google Keyword Planner*.

Otras expresiones de búsqueda son temporales, con un índice de búsqueda alto durante un momento muy concreto de la serie, y una rápida caída posterior, por otro tipo de razones plausibles. Por ejemplo, en frases relacionadas con el comic, es habitual encontrar la utilización del número del comic como especificación de la necesidad de información. Así tenemos ejemplos como “naruto manga 300”, “naruto manga 407”, “naruto manga 409”, y otros.

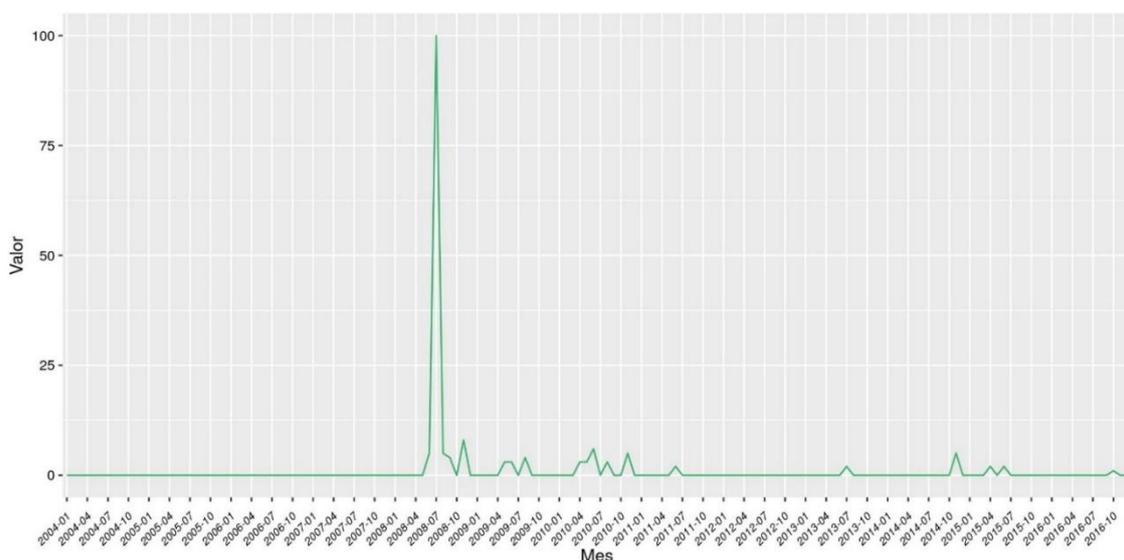


Fig. 3-19 Serie histórica para la expresión de búsqueda “naruto manga 407”. Elaboración propia.

En estos casos, la tendencia general es muy similar, donde se observa un momento único en la serie, de elevadísimo índice de búsqueda, seguido rápidamente de escasos momentos puntuales de interés en utilizar la expresión, menores cuanto más avanza el tiempo en el

futuro. En algunos casos, y como se puede apreciar en el ejemplo de la consulta “naruto manga 300”, la popularidad del ítem a buscar parece ser de mayor interés que en otros, lo que podría ser un indicativo de qué números de comic han tenido más éxito que otros, incluso pasados años después de la aparición del mismo, pero ésa es una posibilidad que debería analizarse con mayor profundidad y mayor cantidad de datos específicos para la temática del comic.

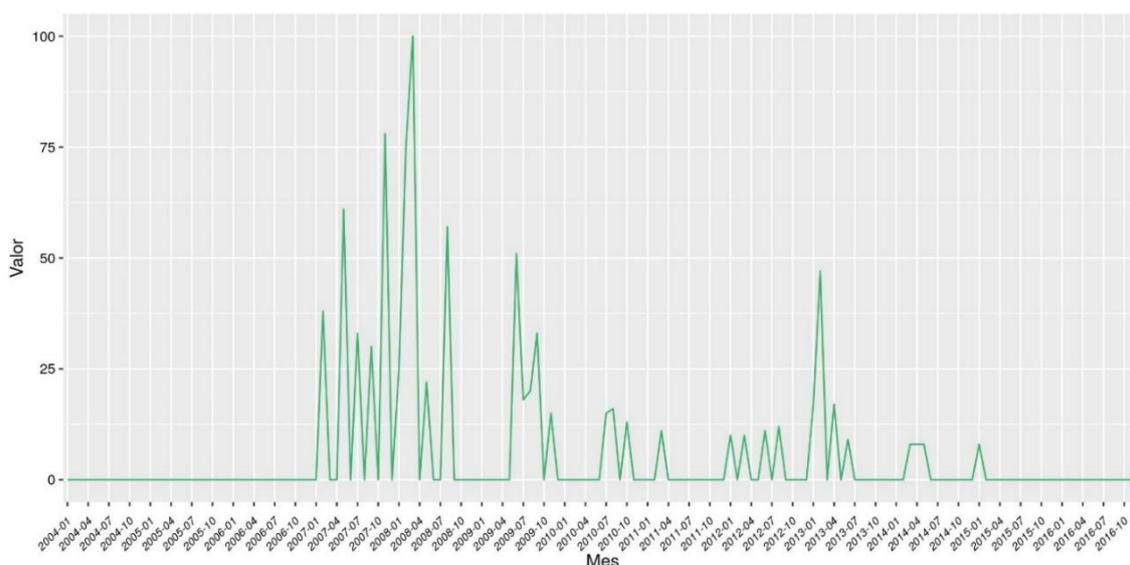


Fig. 3-20 Serie histórica para la expresión de búsqueda “naruto manga 300”. Elaboración propia.

Otras expresiones de búsqueda muestran una tendencia similar, a veces sin especificar un momento histórico concreto. Por ejemplo, varias expresiones distintas sobre el escritor Mario Varga Llosa generan gráficas similares (imagen 5- 16) donde “mario vargas llosa 2010” es la primera serie de la izquierda – arriba, siendo “mario vargas llosa biografia breve” representada a su dercha, y debajo, de izquierda a derecha, estarían las series temporales de “mario vargas llosa biografia resumida” y de “mario vargas llosa obras mas importantes” respectivamente.

Como se puede evidenciar, hay una gran similitud en las series, incluso cuando no se explicita un momento concreto. La razón más probable de esta conducta es el hecho de que el escritor recibiera el premio nobel en el año 2010 (Marcos, 2010) pero es significativo el hecho de que se descubren expresiones similares en cuanto a la tendencia detectada, incluso cuando muestran elementos que las permiten clasificarlas dentro de facetas y aspectos distintos,

como son “Tiempo”, “Tipo narrativo” y uso de “Nombre Personal”, más concretamente “Autor”.

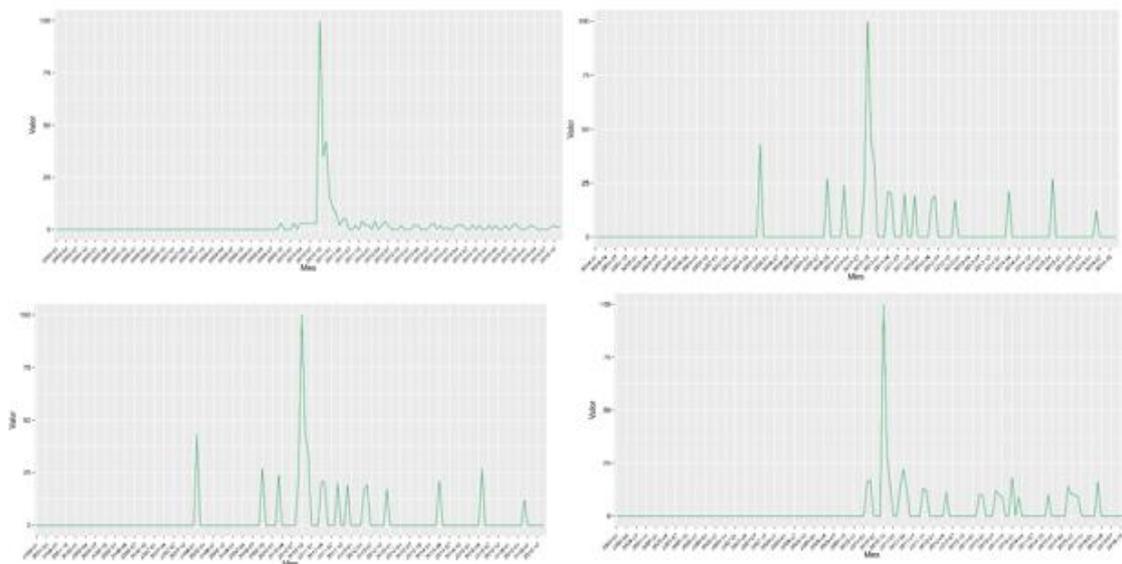


Fig. 3-21 Series históricas para cuatro expresiones de búsqueda relacionadas con Mario Vargas Llosa. Elaboración propia.

La cobertura mediática explica a priori muchas de las tendencias halladas como temporales. Así, diversas expresiones surgen tras la aparición de la Biblioteca Digital Mundial en 2009 (Biblioteca Digital Mundial, 2009). Es interesante observar cómo ante un mismo elemento, un mismo escritor, una misma entidad, un mismo evento, los usuarios utilizan distintas variantes para llegar a la información que buscan. En este caso destacan variantes como “biblioteca digital mundial unesco”, “biblioteca virtual unesco”, “biblioteca mundial virtual”, “biblioteca virtual mundial” e incluso “www.wdl.org biblioteca digital mundial”.

Esta práctica de los usuarios, por la que se descubren expresiones de búsqueda que podríamos denominar “sinónimas” en cuanto a su patrón temporal y/o a su intención de búsqueda en torno a la misma entidad semántica, ayuda a entender que la creación de contenidos debería tener en cuenta estas variantes para llegar a que el contenido (sea una noticia o una sede web, por ejemplo) tengan más probabilidades de ser localizadas por los motores de búsqueda de Internet.

En otras ocasiones, la explicación es aún más obvia. Por ejemplo, en el estudio aparecen múltiples consultas que contienen mención de la sede web “Megaupload”, portal en el que se

alojaban multitud de obras culturales de todo tipo, desde libros electrónicos a películas. El portal fue cerrado en 2012 por problemas con los derechos de los autores (ABC, 2012), y en la gráfica 3-22 se puede visualizar cómo los usuarios dejan de utilizar su mención para explicitar su intención, a la par navegacional y transaccional, ya que buscan ir directamente a la sede web, y ahí descargar algún tipo de contenido. En este caso, la intención transaccional no se denota expresamente en la expresión utilizada en la gráfica 3-22 (“audiolibros megaupload”) pero sí se observa en otras que usan la partícula “megaupload”, como “descargar harry potter y el misterio del principe megaupload”, por ejemplo.

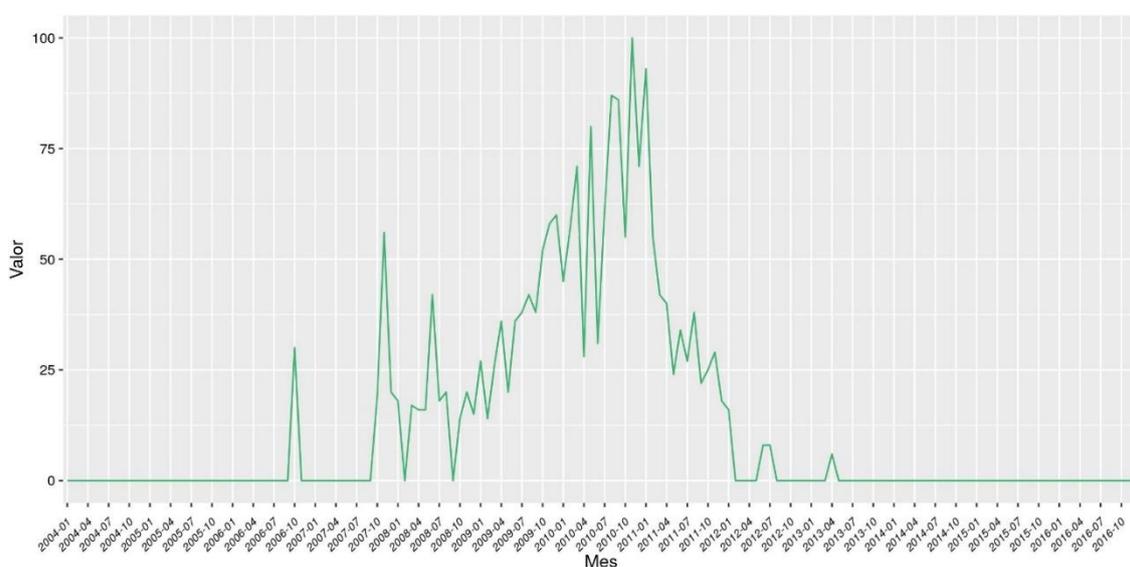


Fig. 3-22 Serie histórica para la expresión “audiolibros megaupload”. Elaboración propia.

Otro tanto ocurre con los lanzamientos de obras, especialmente *best sellers* o superventas. Por poner un ejemplo detectado, obsérvese el patrón histórico de consultas de la expresión “libro eclipse online”, mostrado en la gráfica 3-23, análogo al de los patrones de su familia de expresiones relacionadas (“resumen libro eclipse”, “audiolibro eclipse”, “critica eclipse”, “leer eclipse online”) todas ellas hacen referencia al libro “Eclipse” de la saga Crepúsculo, publicada en España algo más tarde de su publicación en inglés en 2007 (Meyer, 2009) Nótese cómo la popularidad de estas búsquedas nace abruptamente tras la publicación de la novela, y luego vuelve a crecer ante la inminencia del estreno de la película (IMDB, 2010), para después decrecer rápidamente.

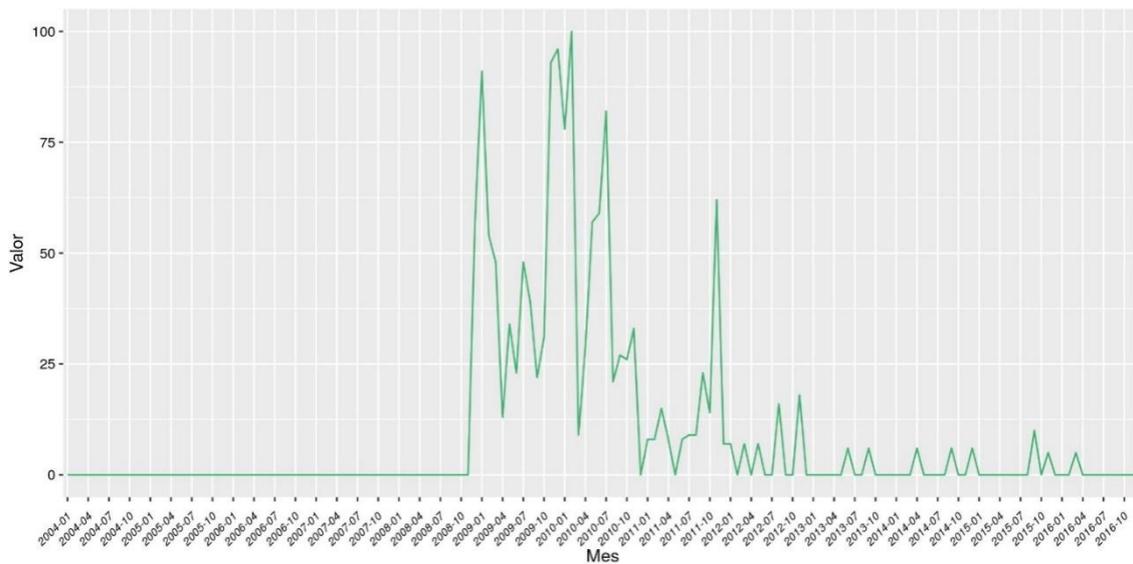


Fig. 3-23 Serie histórica para la expresión “libro eclipse online”. Elaboración propia.

El influjo y relación de los estrenos de películas basadas en obras literarias tiene diversos casos en la muestra analizada, pero quizá el que aporta más variedad al conjunto de expresiones de búsqueda descubiertas en este estudio sea el del fenómeno “Harry Potter”. El personaje y las obras escritas por su autora generan una gran diversidad de expresiones, que ayudan a detectar distintos aspectos de la conducta de los usuarios.

Algunas expresiones de búsqueda sí denotan de diversas formas el formato de su necesidad de búsqueda, aunque esté relacionada con la misma obra. Así, frases como “harry potter y el misterio del príncipe estreno” o “harry potter y el príncipe mestizo megavideo” expresan de forma indirecta su intención de búsqueda del archivo fílmico, bien sea por el uso de la palabra “estreno”, asociado con el estreno de una película y no con la publicación de una obra literaria, bien por el uso del portal “megavideo”, a la sazón similar al antes mencionado “megaupload”.

En otras ocasiones la definición de la necesidad de información es más expresa, como ocurre en “pack libros harry potter”, donde sí se explicita la necesidad referente al formato “libros” lo que ayuda a desambiguar, pero también puede ser más indirecta, como en “harry potter y el misterio del príncipe vagos” donde se hace referencia a la palabra “vagos”. Semánticamente por sí no indicaría nada, pero una búsqueda en Internet identifica ese “vagos” en ese contexto, con el portal “El rincón del vago”, nacido en 1998 (20minutos, 2008) en el que se alojan online multitud de documentos de todo tipo, pero no películas.

No sólo se encuentran explicitaciones directas o indirectas de la intención transaccional y el formato de la transacción (descargar o bajar algún archivo digital), también se pueden encontrar casos en los que la expresión como tal no permite deducir la intención concreta del usuario únicamente a través del análisis del texto usado. Véase (gráfica 3-24) en “harry potter y el misterio del principe on line”, “bajar harry potter y el misterio del príncipe” o “harry potter y el misterio del principe download”, cómo la expresión nos deja la duda de si el usuario se decanta por buscar archivos digitales de libros o de películas.

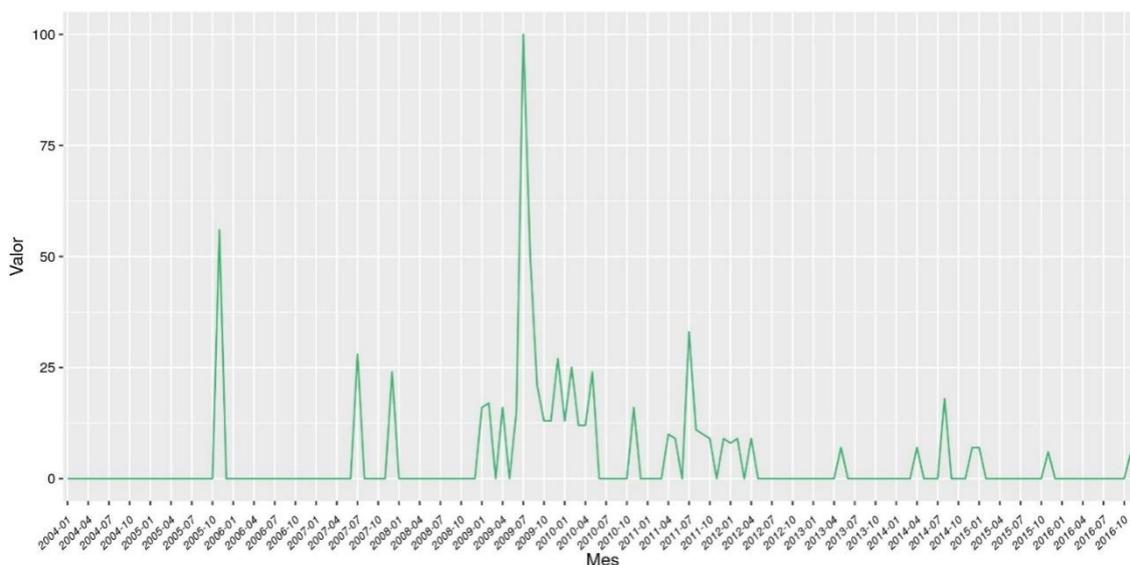


Fig. 3-24 Serie histórica para la expresión “harry potter y el misterio del principe on line”. Elaboración propia.

En estos casos, se considerará necesaria la intervención de otras técnicas para identificar la verdadera intención y necesidad de los usuarios, donde en unas ocasiones se decantará por un formato, y en otras en otro. Es decir, estamos en consultas cuyo lenguaje natural es compartido por usuarios con intenciones distintas, que dejan a discreción del motor de búsqueda ayudarles a determinar la misma, bien sea por su histórico de búsquedas (Jiang, He y Allan, 2014) bien sea por las elecciones que realicen de los resultados expuestos por los motores de búsqueda (Chan, Leung yLee, 2004; LunRadlinski, Szummer yCraswell, 2010) donde el escenario más complejo de computar será aquel en el que los usuarios no clican en ninguna respuesta ni reformulan su necesidad de información (Stamou y Efthimiadis, 2009)

Finalmente, no hay que olvidar que estas novelas también han dado lugar a productos culturales distintos, como videojuegos, lo que se manifiesta en expresiones como “harry potter y el misterio del principe ps2”, “lego harry potter psp Megaupload”, “lego harry potter pc

megaupload”, “lego harry potter psp taringa”, donde “psp”, “ps2” o “pc” identifican dispositivos en los que el usuario puede jugar a videojuegos relacionados con la obra literaria.

En conjunto, se observa gran variedad de expresiones de búsqueda relacionadas con la obra, con más de 70 entre las expresiones temporales con datos en *Google Trends*, y casi 800 distintas en total en el conjunto de frases identificadas en *Google Keyword Planner*, incluyendo aquellas de las que hay datos en google Trends y aquellas de las que no hay.

En este fenómeno relacionado con Harry Potter, se constata una gran variedad de contenidos que se muestran en diferentes medios de comunicación y consumo, lo que se denomina en la literatura como *transmedia*, que es aquella estrategia de generación de contenidos en la que se busca ampliar la narrativa mediante el desarrollo de distintas experiencias relacionadas con los personajes de una obra (Millán, 2017)

Esta ampliación puede ser tanto en distintos formatos narrativos (comic, película, cuentos cortos...) como dispositivos (consolas de videojuegos, libros digitales, audiolibros...) Tal variedad expresiva aporta indicios que permitirían preguntarse si a mayor generación de productos culturales distintos, mayor es la variedad de expresiones de búsqueda localizadas, lo que debería ser investigado en profundidad y en una gran variedad de casos.

3.3.7. Análisis específico de evolución de expresiones atemporales

La mayor parte de las expresiones de búsqueda de las que Google Trends ha ofrecido datos pertenece a la categoría de búsquedas atemporales. Atendiendo al análisis realizado en el apartado 3.3.5, incluso separando del análisis las consultas temporales, diríase que la tendencia hallada tanto en las de tipo temporal como en las de tipo atemporal, a la baja en los últimos años analizados, es común. Pero no se deben perder de vista diferencias encontradas en algunos subgrupos de expresiones, que nos aportarán información útil para entender la variedad de terminología y asimismo, la variedad de conductas, que se incluyen en el presente estudio. Esa variedad que surge incluso dentro de una misma temática, lleva a sugerir que se deberían explorar distintas técnicas de agrupamiento y extracción de aspectos (Paşca et. al., 2010) que ayuden a discriminar conductas comunes entre los casos localizados.

Tomemos el caso de las bibliotecas. Dentro de esta faceta se incluyen expresiones de búsqueda sobre diversos tipos de bibliotecas, mostrando una tendencia general más bien constante y periódica, con clara estacionalidad, en la que mediana sube y baja anualmente, sobre todo en los veranos, y una ligera tendencia global a la baja en los últimos años, más relevante si la comparamos con los datos de los primeros años de la agrupación, una década antes.

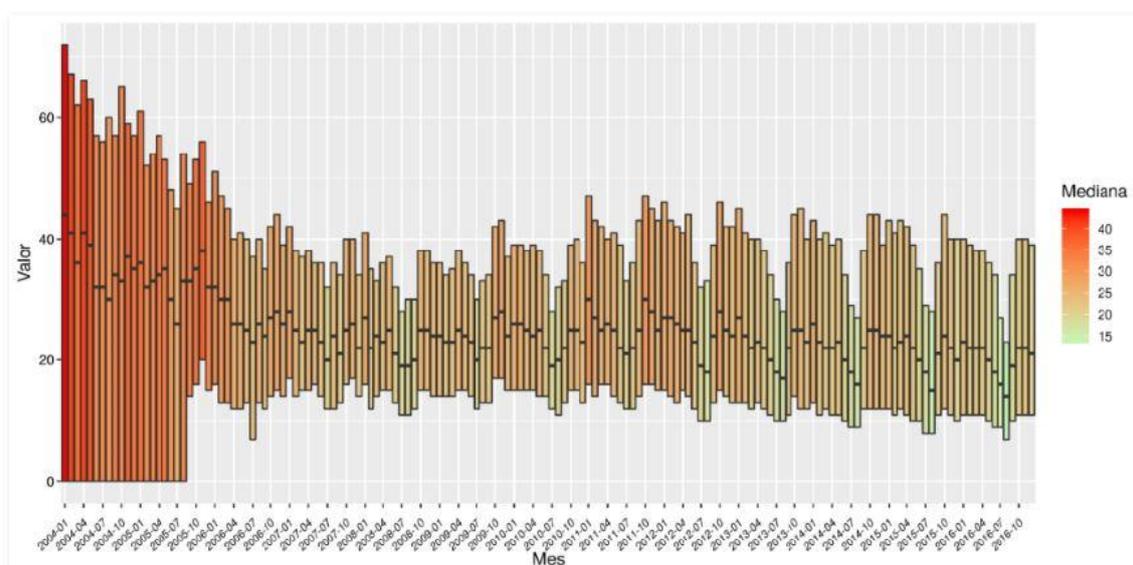


Fig. 3-25 Faceta Biblioteca- Evolución de queries atemporales

Pero este grupo de expresiones muestra al menos dos grandes grupos de bibliotecas, las consultas sobre bibliotecas públicas, y las pertenecientes a bibliotecas universitarias. Si nos fijamos en consultas relativas a bibliotecas públicas, la tendencia a la baja se confirma en multitud de casos, repartidos por toda la geografía española, como “biblioteca cordoba”, “biblioteca oviedo”, “bibliotecas navarra”, “bibliotecas valladolid”, “bibliotecas sevilla”, “biblioteca publica cadiz”, “biblioteca publica de zamora” y otras.

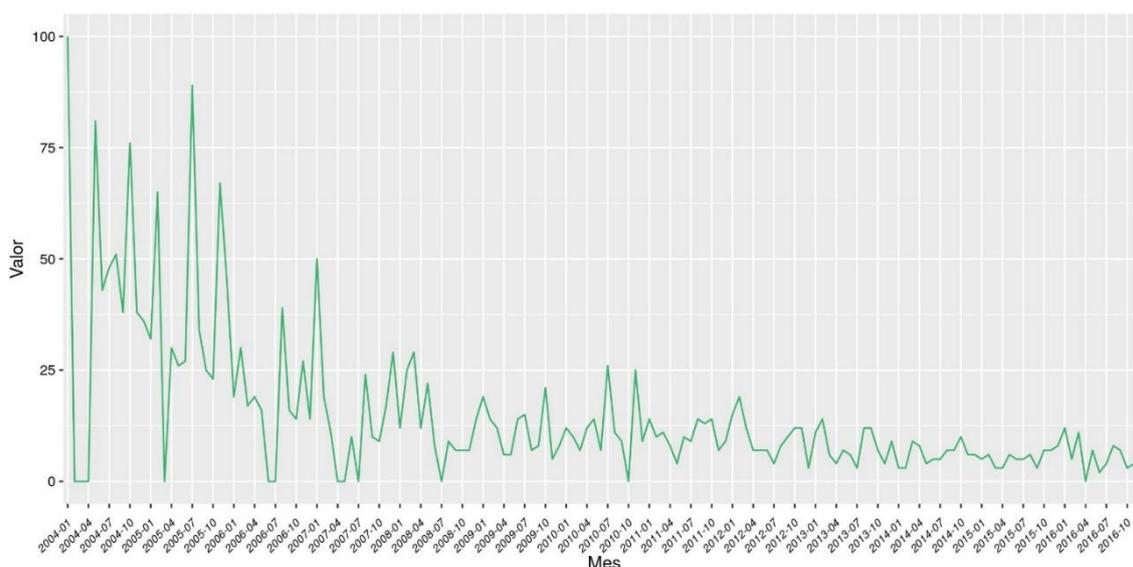


Fig. 3-26 Evolución de la expresión “biblioteca publica de zamora”. Elaboración propia.

Diversas son las razones que se pueden aducir para esta conducta. Lo cierto es que coinciden con lo expuesto en el apartado 2.1.2 de necesidades del sector de Bibliotecas, los datos de cierre de bibliotecas (Ministerio de Educación Cultura y Deporte, 2014) y de bajada de préstamos físicos (Ministerio de Educación, Cultura y Deporte, 2016) Siendo como son las expresiones de búsqueda un reflejo de la popularidad de los conceptos buscados, se puede determinar que entre la población internauta que utiliza Google en España y en español, hay una menor popularidad de expresiones de búsqueda sobre bibliotecas públicas. Y esta tendencia ocurre en otros tipos de bibliotecas, como por ejemplos el subconjunto de expresiones de búsqueda referentes a las bibliotecas virtuales o digitales. Así, “biblioteca virtual”, “virtual biblioteca” (la misma expresión, pero con los términos cambiados) “bibliotecavirtual” o “biblioteca virtual cervantes, comparten esta disposición en los datos.

En el caso de las bibliotecas universitarias la tendencia parece la contraria. Expresiones como “biblioteca ugr”, “biblioteca um”, “biblioteca ucv”, “biblioteca upm”, “biblioteca usal”, muestran series con un aumento gradual del índice de búsqueda o una estabilidad en el mismo. Obsérvese el uso constante aquí de la partícula biblioteca más una abreviatura de la universidad (gráfica 3-27), así como la clara estacionalidad marcada por el inicio del curso universitario, con su momento más bajo anual en verano.

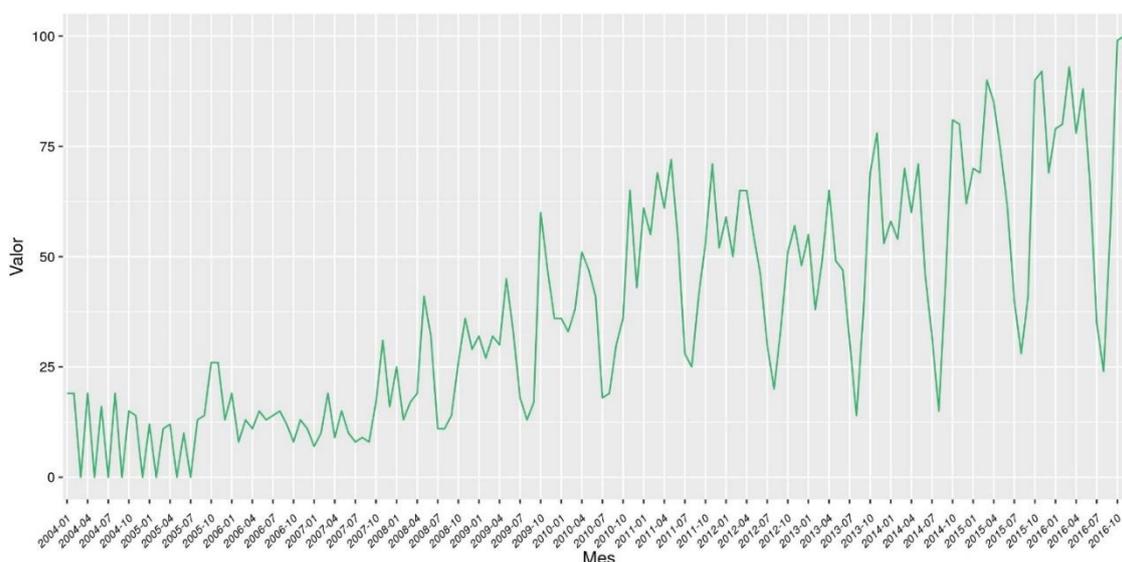


Fig. 3-27 Evolución de la expresión “biblioteca ugr”. Elaboración propia.

Junto con esa tendencia, es interesante observar que lo aumenta es ese uso de las expresiones de búsqueda que utilizan abreviaturas en la mención de la universidad, lo que indicaría una conducta de búsqueda que seguiría el principio del mínimo esfuerzo (Zipf, 1950; Kotler et. al., 2012) en la que el usuario que usa este tipo de expresiones, utiliza la abreviatura en lugar del nombre de la universidad desarrollado. Nótese, como se puede apreciar en la figura 3-28, que el empleo de la denominación completa de la universidad decrece conforme pasan los años, evidenciado en casos como “biblioteca universidad de sevilla”, “biblioteca universidad complutense” o “biblioteca universidad de valencia”.

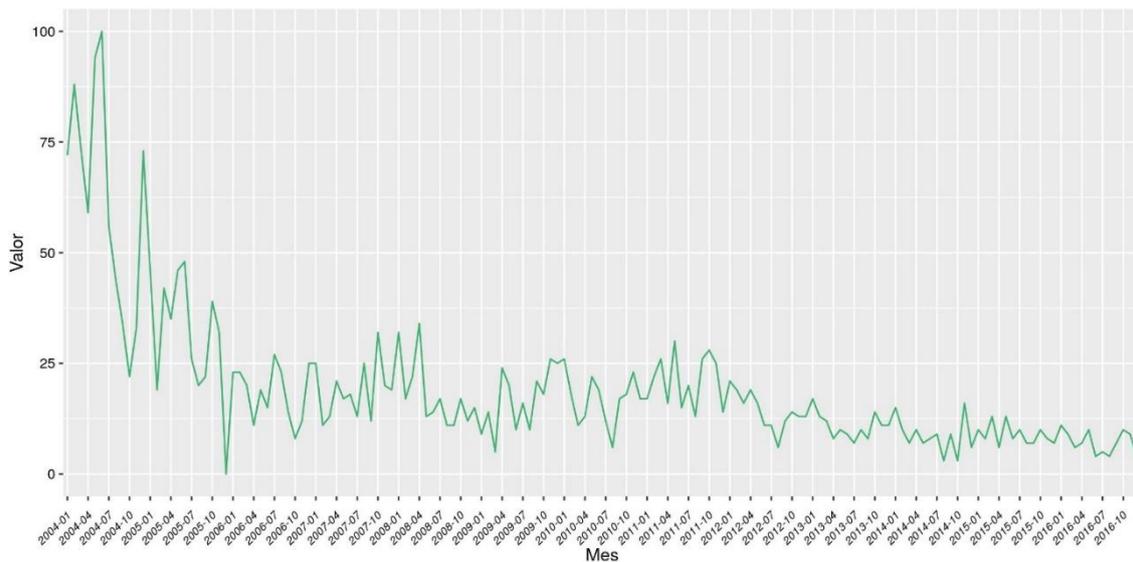


Fig. 3-28 Evolución de la expresión “biblioteca universidad de valencia”. Elaboración propia.

Hay familias de expresiones en torno a un concepto que muestran patrones de búsqueda en los que, como sucedía en el caso de ciertas búsquedas temporales, es fácil intuir relaciones causales debidas a eventos concretos en la serie histórica que, gracias a los medios, llegan a los usuarios, provocando un aumento de la búsqueda sobre el mismo. Es el caso de búsquedas como las relativas a Cervantes (“wikipedia miguel de cervantes”) donde aumenta el índice de búsqueda durante el año de la conmemoración de su muerte, en 2016 (Real Academia de la Lengua, 2016), y de forma análoga multitud de frases de solicitud de información sobre el escritor Gabriel García Márquez.

Diversas expresiones de búsqueda relacionadas con el célebre autor muestran una tendencia similar, que apuntan a una disminución en el índice de búsqueda paulatino o estable durante unos cuantos años, para volver a resurgir momentáneamente, en concreto 2014, cuando fallece y su óbito aparece en los medios. Es el caso de frases como “frases celebres gabriel garcia marquez”, “biografia de gabriel garcia marquez”, “cuentos gabriel garcia marquez”, “citas gabriel garcia marquez”, “gabriel garcia marquez obras” y otras. El mismo patrón se encuentra en el fallecimiento en 2014 de otro escritor, Juan Gelman, explicitado en la expresión de búsqueda “juan gelman poemas”, o en la muerte en 2009 de Francisco Ayala (cuando crece la búsqueda sobre “ayala escritor”)

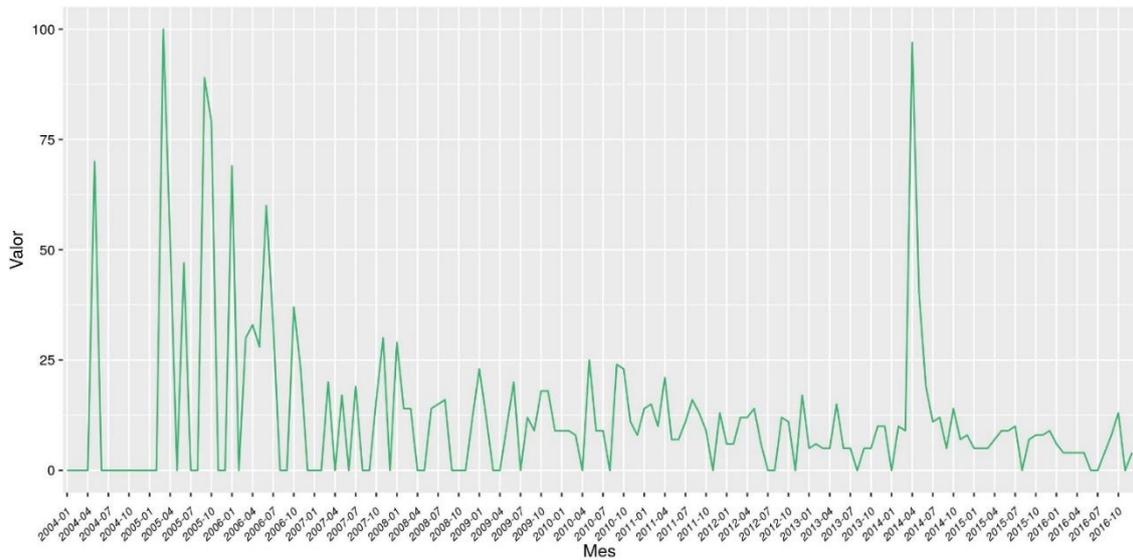


Fig. 3-29 Evolución de la expresión “frases celebres gabriel garcia marquez”. Elaboración propia.

En otras familias de expresiones de búsqueda la relación no es tan obvia, o al menos no tan inmediata, por lo que se pueden explorar diversos factores. Es el caso del comic como formato y como tipología narrativa, cuya popularidad en las búsquedas resurge poco a poco, a partir de final de 2013. Cabría preguntarse si ese resurgimiento se debe al avance de las películas basadas en cómics de editoriales como Marvel o DC. Ya se habían realizado películas anteriormente sobre este tipo de personajes, y no se puede obviar que este aumento en el índice de búsqueda no coincide completamente con el inicio del “Universo cinematográfico de Marvel” (en inglés, *Marvel Cinematic Universe* o MCU) que supone un esfuerzo organizado de interconectar películas sobre distintos personajes a través de distintas películas, y distintos formatos (Marvel, 2018)

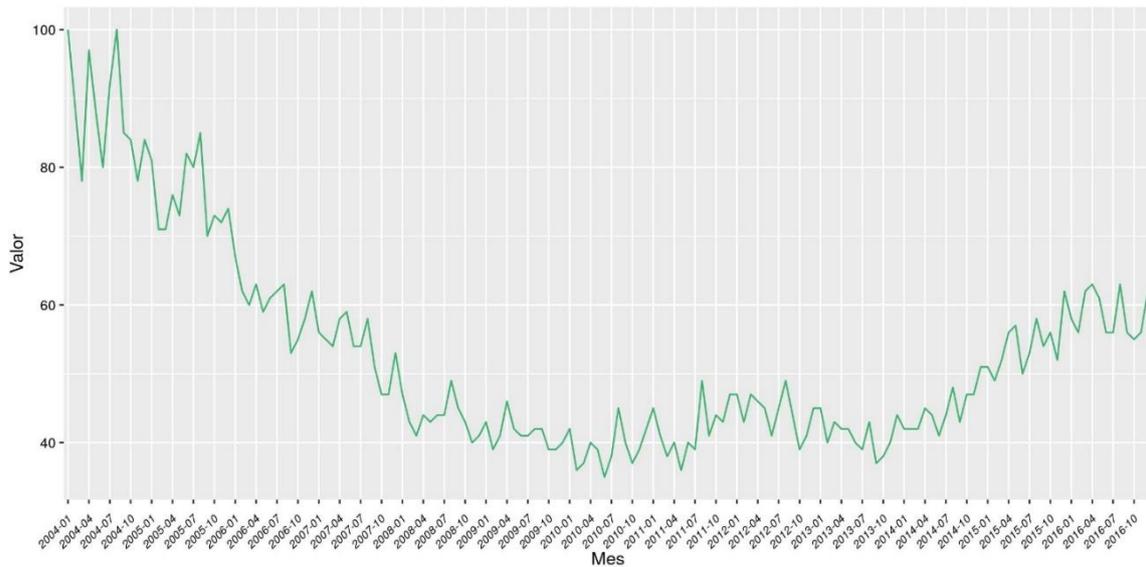


Fig. 3-30 Evolución de la expresión “comics”. Elaboración propia.

Muy probablemente esta tendencia para una expresión tan genérica como “comics” no debe interpretarse como generalizable para todo tipo de cómics. Si localizamos frases de búsqueda sobre comics más concretos, los datos pueden cambiar ampliamente con respecto de esta tendencia alcista. Así, “lobo comic” expresa una necesidad de información sobre un personaje concreto de los cómics, con una curva decreciente en la primera mitad del período estudiado, y una mayor estabilidad desde 2009 hasta el final del período, lo que coincide con los datos de ventas de su editorial, DC (Higuero y Paredes, 2018).

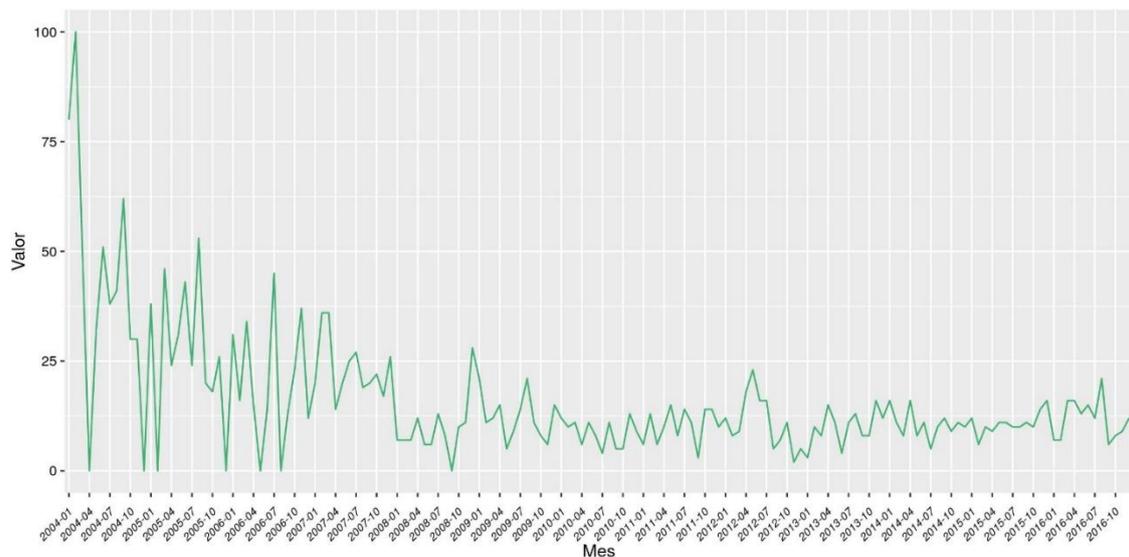


Fig. 3-31 Evolución de la expresión “lobo comic”. Elaboración propia.

Pero sí ocurre con otros cómics que se aprecia una coincidencia más concreta con la aparición de películas o programas de televisión. Por ejemplo, se observa un alza en búsquedas como “flash comic”, un personaje de la editorial DC, con el estreno en España de la primera temporada de la serie de televisión, en 2015 (Sánchez, 2015) Lo mismo puede decirse de otros cómics que utilizan el *transmedia* como estrategia editorial, véase el caso de los de la saga *Star Wars*. En expresiones como “star wars comics”, “comic star wars”, o “personajes star wars”, por ejemplo, hay una estabilidad en los datos del índice de búsqueda, salvo en momentos concretos que coinciden con el estreno de algunas películas de la saga, como 2005, 2012 (reestreno de películas en 3D), y especialmente en 2016, coincidiendo con el estreno de la película “Rogue One: Una historia de Star Wars”, dentro del universo de películas creadas alrededor de la saga.

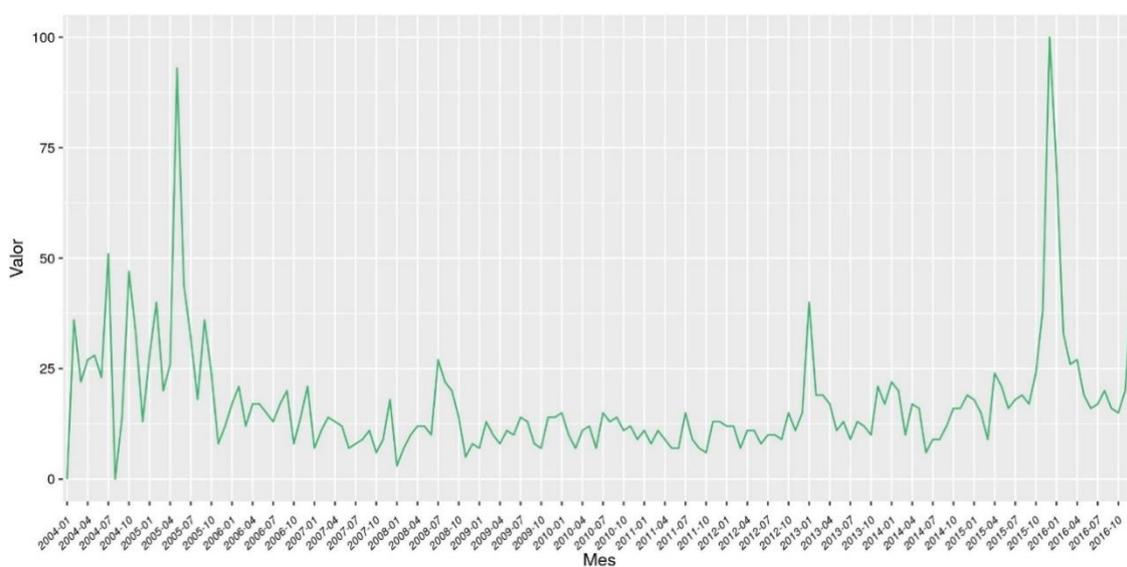


Fig. 3-32 Evolución de la expresión “star wars comics”. Elaboración propia.

Esta relación entre medios, especialmente entre cine y televisión con la literatura no se ve sólo circunscrita a los cómics. En 2015 hay un crecimiento del índice de búsqueda de la expresión “las puertas de anubis” novela de ciencia-ficción relacionada con la serie de "El Ministerio del Tiempo", de Televisión Española (Blánquez, 2016)

Pero no siempre es así. Ocurre con expresiones como “libros eróticos”, “novela erotica” o novelas eróticas”, cuya alza en el índice coincide en 2012, después de que se traduzca al castellano "50 sombras de grey". Sin embargo, no se aprecia un aumento excesivo del índice de búsqueda cuando aparece la primera película de la saga en 2015, acaso un momentáneo repunte, de lo que se deduce que cada caso debe ser tratado con cuidado para localizar las relaciones causales concretas de ciertas conductas.

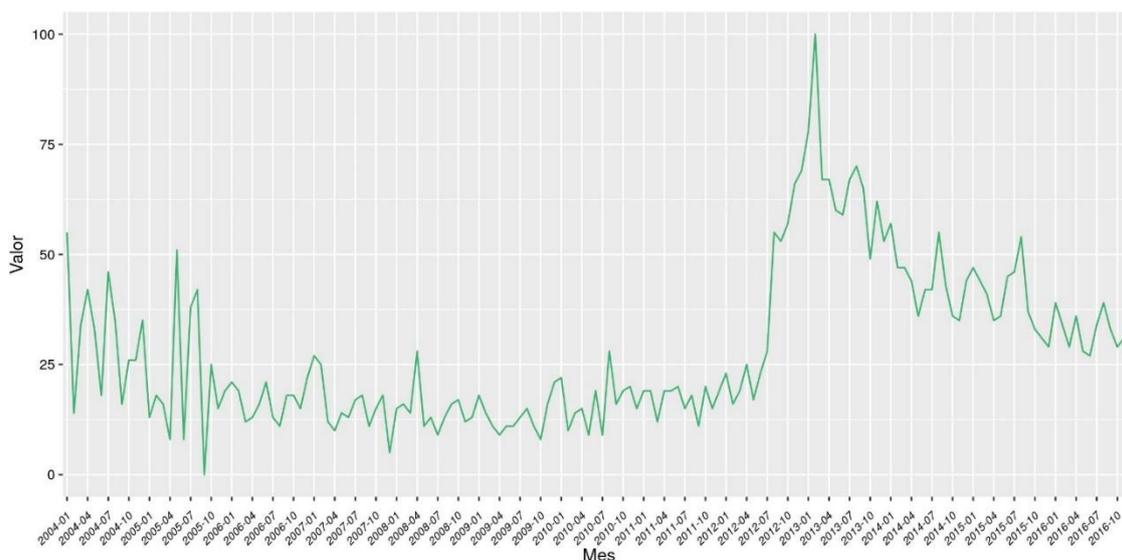


Fig. 3-33 Evolución de la expresión “libros eróticos”. Elaboración propia.

Pero de la misma forma que se identificaban búsquedas temporales en torno al personaje de Harry Potter, también hay expresiones de búsqueda que, sin llegar a ser clasificadas como temporales, también muestran un crecimiento en su índice de búsqueda cuando aparecen las películas. Lo que ocurrirá en muchos casos es que esa relación no se explicita en el texto de las expresiones de búsqueda, como en “harry potter 7”, “harry potter 6”, “harry potter descargar”, “harry potter reliquias” u otras.

El fenómeno de este tipo de obras merecerá ser estudiado en el futuro, como el de otros best sellers, para ahondar en las conductas de búsqueda de los usuarios, especialmente en el caso de aquellos que pasan de ser superventas a convertirse en clásicos. Uno de estos casos se detecta en torno a la obra El Principito, de Antoine de Saint-Exupéry, publicada en 1943 (Iberlibro, 2016) donde expresiones de búsqueda como “frases principito”, “pdf el principito” o “libro el principito”, no sólo no bajan en la línea temporal, sino que o se mantienen o crecen de forma continuada.

La influencia de la película podría centrarse a partir de 2015, momento en que se estrena, con mención a ella en expresiones de búsqueda como “el principito la película”, pero como se puede apreciar en la figura 3-34, aunque puede ser un factor que refuerce la evolución al alza de este tipo de consultas, venía observando un crecimiento desde aproximadamente 2009, momento que coincide con una reedición ilustrada (Saint-Exupéry, 2009)

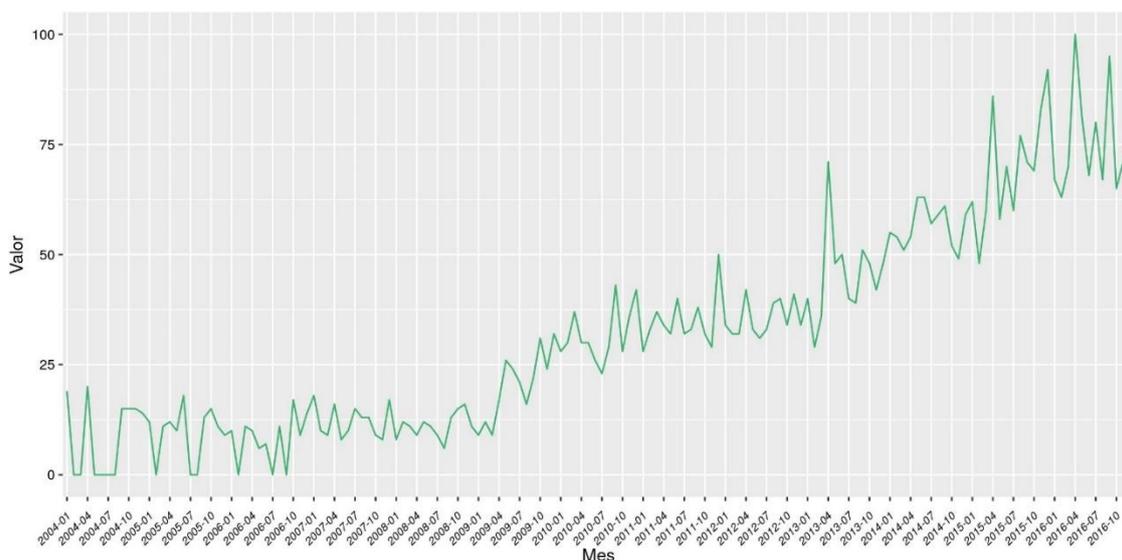


Fig. 3-34 Evolución de la expresión “frases principito”. Elaboración propia.

Otro tipo de lectura se mantiene estable o crece por diversos motivos. Al detectar expresiones de búsqueda que muestran esta pauta, surge un subgrupo de frases relacionadas con temáticas escolares o formativas, que utilizan una variedad de frases. Expresiones como “lengua y literatura Anaya”, “libros de texto baratos”, “libros de texto primaria”, “libros uned”, “lengua castellana y literatura”, “lengua y literatura bachillerato” pero también “poesías infantiles” o “cuentos infantiles” muestran una continuidad temporal o un crecimiento, especialmente acusado con el comienzo anual de las clases, en torno a septiembre-octubre.

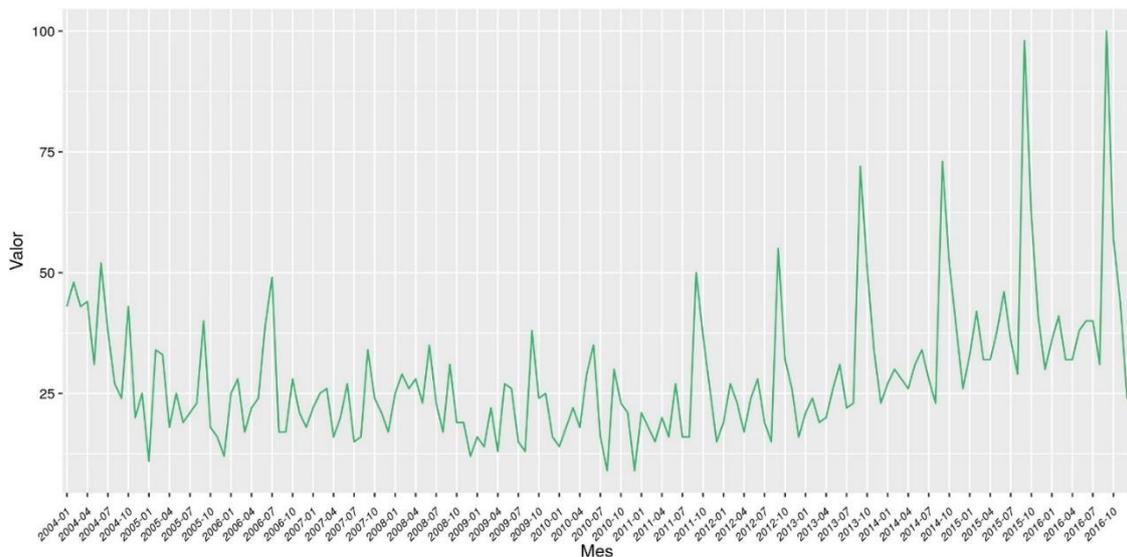


Fig. 3-35 Evolución de la expresión “lengua castellana y literatura”. Elaboración propia.

En particular, es interesante observar cómo hay expresiones que denotan una necesidad muy concreta cuando se relaciona la misma con el momento de inicio del curso. Casos como “celestina resumen” o “don quijote resumen” se mantienen o suben con marcada estacionalidad, lo que puede explicar al menos parte de las búsquedas hechas al respecto. Además, el aumento en el índice de búsqueda de expresiones como “verbo leer” y otras, suponen un indicio de que, con el aumento de la adopción de las nuevas tecnologías, las consultas pueden estar volviéndose más específicas, usándose más frecuentemente en los motores de búsqueda, lo que podría ser estudiado en mayor profundidad.

Ese incremento en el uso de expresiones de búsqueda más específicas se aprecia, no sólo en la literatura al respecto, sino en algunas de las frases detectadas. Así, y aunque como se ha explicado en el punto 3.3.5, gran parte de las expresiones de búsqueda que ofrecen datos de *Google Trends*, ofrecen una tendencia bajista en cuanto a su popularidad, con expresiones generalistas como “leer libros”, existen otras más específicas que muestran tendencias estables o al alza, como es el caso de búsquedas concretas sobre distintos tipos de novelas. Así se aprecia en los datos de “novela romantica descargar”, “novelas romanticas pdf”, o “libros novela romantica”; pero también en otras tipologías de novelas, como “ciencia ficcion libros”, “libros de ciencia ficción” o “novelas policiacas”.

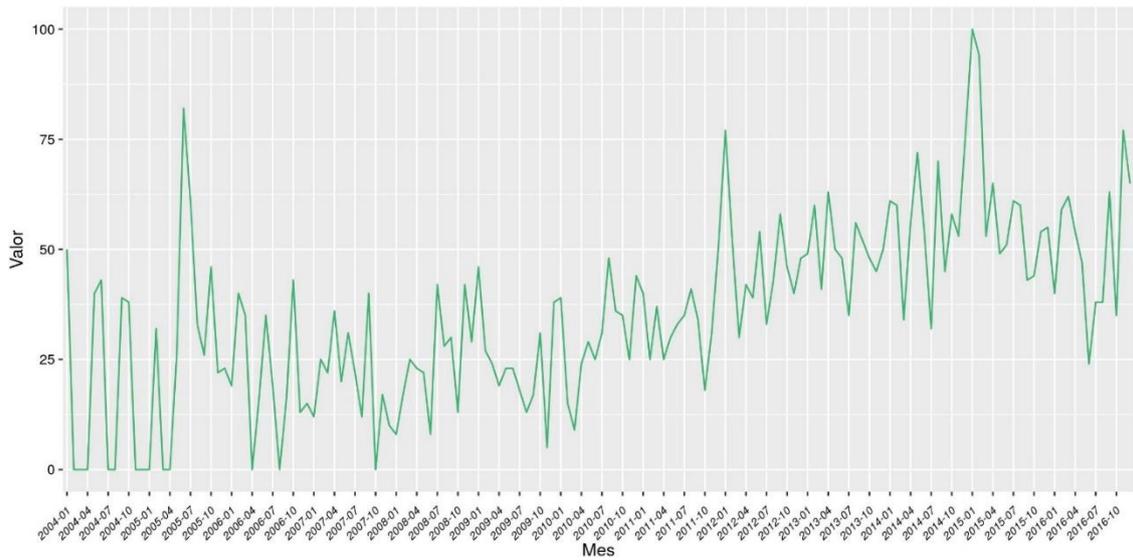


Fig. 3-36 Evolución de la expresión “libros novela romantica”. Elaboración propia.

Existen expresiones que muestran un crecimiento durante una etapa, pero que vuelven a decrecer con el tiempo. Es el caso de “libros electrónicos”, cuyo índice de búsqueda resurge en 2011-2012 y vuelve a bajar, tendencia en la que coinciden expresiones como “descarga libros electrónicos”, “descargar libros electrónicos”, “descargar libros electronicos gratis”, “bajar libros electrónicos” o “comprar libro electrónico”. Y ni el cambio de denominación para similares conceptos semánticos nos apartan de esta tendencia, con expresiones como “ebooks libros gratis” o “kindle” siguiendo pautas similares.

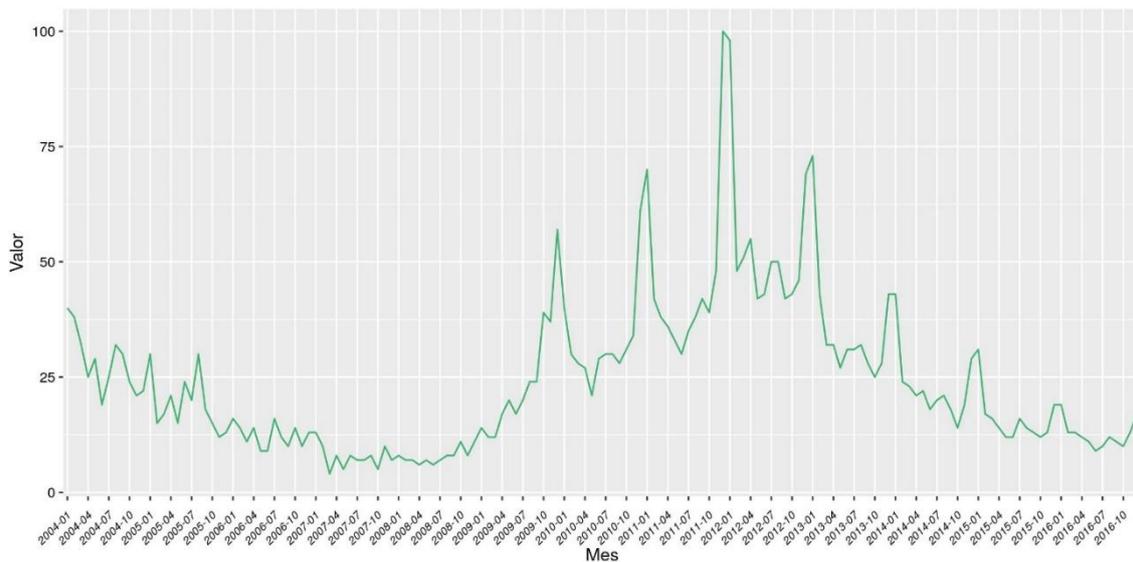


Fig. 3-37 Evolución de la expresión “libros electrónicos”. Elaboración propia.

La sospecha que produce la comprobación del aumento de la popularidad de expresiones más específicas, induce a pensar que no es que se busque menos sobre (en este caso) “libros electrónicos”, de forma general, sino sobre títulos concretos o géneros específicos concretos, como atestiguan expresiones de tendencia creciente como “novelas romanticas pdf “, “harry potter y la orden del fenix pdf” o “harry potter y la piedra filosofal pdf”, lo que necesitará de estudios también más específicos.

Finalmente, y en la línea de la tendencia global establecida en 3.3.5., sí que los datos revelan agrupaciones de expresiones con pautas descendentes. Tal es el caso de las expresiones relacionadas con las editoriales, es decir, búsquedas navegacionales que utilizan generalmente la combinación de la palabra “editorial” más la mención de una entidad o marca editorial, para formar su frase de búsqueda.

De este modo, los datos históricos de casos como “planeta editorial “, “editorial alfaguara”, “destino editorial”, “gredos editorial” y otras, atestiguan el descenso de la popularidad de búsquedas referidas a la marca de la editorial como método de búsqueda de información sobre las propias editoriales, o sobre las obras publicadas por las mismas.

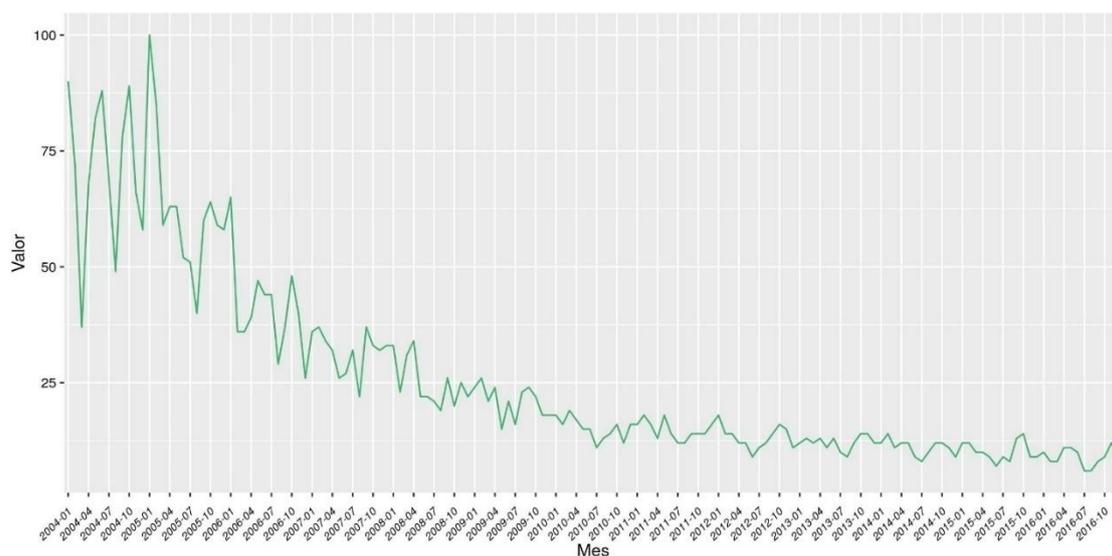


Fig. 3-38 Evolución de la expresión “planeta editorial”. Elaboración propia.

La aportación general de estos descubrimientos es que se debe bajar aún más en el desglose y agrupamiento de las expresiones de búsqueda con mayor detenimiento, cruzando los datos longitudinales con otros datos, como los que aporten información sobre la influencia de los medios y las estrategias de marketing y/o de generación de contenidos *transmedia*, para ahondar en las relaciones causales que competen a cada agrupación, donde las conductas de búsqueda pueden deberse a múltiples factores, raramente segregables unos de otros.

3.4. En resumen

En el apartado 3.2, se ha discutido la metodología utilizada, tomando muestras intencionadas no probabilísticas. Se han localizado diversos casos que ejemplifican las limitaciones y dificultades inherentes al estudio. Se han descubierto diversas posibilidades de mejora de la metodología, para implementar análisis multidisciplinares de las conductas de búsqueda a través de las expresiones de búsqueda utilizadas por los usuarios para el contexto aplicado, en cuanto a segmentación temporal, geográfica, lingüística y temática.

Así se han identificado diversas dimensiones de análisis o factores que han de tenerse en cuenta para optimizar la metodología. Además, se han probado las herramientas de extracción de datos empleadas (*Google Keywords Planner* y *Google Trends*) para establecer su posible validez en el contexto de la investigación, aportando datos y ejemplos de discusión.

Por otra parte, en el apartado 3.3. se ha pasado revista a los datos en sí, tanto de la muestra de expresiones localizadas con *Google Keywords Planner*, como de los datos longitudinales extraídos de *Google Trends*. De este modo se ha considerado el lenguaje natural de los usuarios en este contexto temático, aportando evidencias de una evolución en el lenguaje digital, que contrasta con las normas ortográficas y gramaticales clásicas.

Durante este proceso se han hallado indicios de una conducta general en la búsqueda de información relacionada con el ámbito de estudio, al utilizar el conjunto de datos históricos en su totalidad. Pero también se ha constatado que se deben separar tendencias diferenciadas según diversas segmentaciones (bien por familias semánticas o de expresiones relacionadas, bien por los factores intervinientes detectados, bien por la coincidencia de patrón en los datos de las series) que complican el establecimiento de generalizaciones, y que llevan a enriquecer la disciplina del análisis de hábitos de búsqueda.

A continuación, se unirán todos estos descubrimientos en el capítulo de conclusiones e investigaciones futuras, de forma que sirva como resumen global de la tesis y de los múltiples descubrimientos hallados, con el deseo de que sirvan como germen de nuevas investigaciones al respecto.

4. CONCLUSIONES E INVESTIGACIÓN FUTURA

4.1. Conclusiones y limitaciones por objetivo

4.2.

En este estudio se presentaba como objetivo principal determinar si los procesos y expresiones de búsqueda de información usados por los usuarios en motores de búsqueda, pueden considerarse como indicadores válidos para el análisis y estudio de los hábitos de lectura y posible interés en otros contenidos ofrecidos por las bibliotecas en España (como videojuegos o películas). Este análisis se había explicitado en los siguientes objetivos específicos:

- **Crear un corpus de expresiones o frases de búsqueda buscadas en Google, relacionadas con los hábitos de lectura y posible interés en otros servicios y contenidos ofrecidos por las bibliotecas en España, para que puedan realizarse distintos estudios de demanda online de información relacionada.**

Mediante una metodología que ha empleado como fuente de datos la herramienta *Google Keywords Planner*, que contiene millones de expresiones de búsqueda de todos los usuarios de *Google*, se han localizado y extraído manualmente más de treinta mil expresiones de búsqueda distintas sobre la búsqueda de información en *Google* sobre temáticas relacionadas con la lectura, el libro y las bibliotecas en España, a las que hay que sumar más de tres mil expresiones de búsqueda de diversa índole, que aunque compartían coincidencias textuales (como el caso de “autores”, mencionado en el apartado 3.2.2.) no tenían relación semántica con el ámbito de estudio.

Asimismo, se ha intentado obtener de *Google Trends*, herramienta de análisis longitudinal de la demanda de información de los usuarios en el buscador, datos históricos de todas las expresiones de búsqueda señaladas por *Google Keywords Planner* como relacionadas con las temáticas analizadas. De este proceso se ha podido obtener un subconjunto de expresiones (23,4% del total de frases) que ofrecieron datos de su histórico de búsquedas en *Google Trends* a lo largo de la serie temporal 2004-2016, debido a que sólo en esos casos se superaba el nivel mínimo de índices de búsqueda mensuales que aplica la herramienta para ofrecer datos.

Se ha constatado la utilidad limitada de *Google Keywords Planner* para localizar frases de búsqueda relacionadas, cuando lo que se busca es determinar la variedad del lenguaje de

consulta sobre una materia amplia. Para ello, se realizó un proceso de eliminación de expresiones de búsqueda no concernientes al ámbito de estudio, y de clasificación posterior de las sí concernientes, basándose en la adscripción de las expresiones a ciertas clases. Este proceso permitió descubrir ciertas características de recomendación de expresiones de búsqueda ante palabras semilla, inherentes al algoritmo de recomendación que *Google Keyword Planner* utilizaba en 2010. En algunos casos, como se ve en el apartado 3.2.1, también permitió exponer distintas posibilidades de funcionamiento de la misma, puesto que la herramienta (de software propietario) es opaca en cuanto a los algoritmos concretos de recomendación de expresiones que utiliza.

Estas características provocaron la aparición de numerosos casos de términos de búsqueda que no eran relevantes para la temática elegida para el estudio, y que los datos de *Google Keyword Planner* debieran ser tomados como referencias inexactas, debido a su alto nivel de redondeo. Aun así, se guardaron multitud de expresiones que, no teniendo relación temática, podían ayudar en futuras investigaciones para entrenar sistemas de agrupamiento de expresiones que utilizaran los ejemplos (tanto positivos como negativos) para mejorar el conjunto de frases detectadas.

Además, la metodología empleada no consiguió aportar un corpus todo lo exhaustivo que cabría desear. Aunque el conjunto de expresiones descubiertas pueda parecer grande (más de treinta mil), es pequeño si se compara con la producción editorial actual. Sólo en la Biblioteca Nacional se contienen más de 4 millones de registros de autoridad (Biblioteca Nacional de España, 2018) sin contar con que se ha constatado que, según la herramienta *Google Keyword Planner*, los usuarios españoles pueden buscar información sobre el ámbito a investigar en otros idiomas. Por tanto, el grupo de expresiones localizadas es claramente susceptible de aumentar con adiciones de otras búsquedas relacionadas con otras palabras semilla, que amplíen el corpus detectado, y sólo puede ser tomado como una muestra.

Se constató también la necesidad de utilizar otras técnicas de limpieza semántica que ayuden a revisar aún más la actual base de datos creada, determinando así las frases de búsqueda que realmente tengan una relación con la temática a investigar. Tales circunstancias, han llevado a proponer la necesidad de afinar más en el futuro el sistema de extracción y de tratamiento de los datos que salen de *Google Keyword Planner*.

Por otro lado, se examinó en el apartado 3.2.3., la posible utilidad de *Google Trends* para obtener datos históricos que representaran la demanda de información de los usuarios de una

determinada segmentación (en este caso España durante los años 2004 a 2016). Pero también se identificaron problemas, como la limitación de extracción de datos por parte de *Google Trends*, que sólo muestra el índice de búsqueda mensual a largo de la serie en aquellos casos en los que haya un mínimo de popularidad de búsqueda o *threshold*.

Esta circunstancia ha impedido obtener datos de todo el *dataset* o conjunto de datos identificado mediante *Google Keyword Planner*. De haberlo conseguido, se habría tenido una información aún más granular, al tener datos de expresiones de búsqueda más específicas. Lo que se ha conseguido ofrece datos históricos de aquellas expresiones que, al haber sobrepasado el mínimo exigido por la herramienta, tienen el mayor número de búsquedas mensuales de todo el conjunto global de expresiones localizadas mediante *Google Keyword Planner* en este estudio. Pero de haberse conseguido más datos de expresiones con menor volumen de búsqueda, podría haberse identificado un conjunto de conductas de búsqueda para subtemáticas más concretas, que han desaparecido en este segundo ejercicio, aun teniendo un cierto volumen de búsqueda.

Por contra, como se ha podido recuperar la información de aquellas expresiones de búsqueda suficientemente populares como para haber generado datos en *Google Trends*, se ha contado con una alta representatividad de lo que más se busca en España en ese período y sobre el conjunto de las expresiones descubiertas para la temática analizada.

Se procedió así a integrar la información obtenida de ambas plataformas, como se describe en 1.5.6. Pero no fue posible integrar los valores de las variables de frecuencia de búsqueda de *Google Keyword Planner* con los de *Google Trends*, al utilizar ambas herramientas distintas técnicas de medición y representación de esa frecuencia. En *Google Keywords Planner* se usa como variable de medición la frecuencia de búsqueda mensual que va de cero a miles, mientras que en *Google Trends* se utiliza un índice de búsqueda o *query index*, ponderado de toda la serie histórica segmentada, que va de cero a 100. Por tanto, se integró la información en base a las expresiones de búsqueda, ubicando las expresiones obtenidas en *Google Keywords Planner*, aportando la información histórica de *Google Trends* cuando fue posible, cuando sus valores de índices de búsqueda estuvieran por encima del mínimo exigido por la herramienta para mostrar datos históricos.

Entonces, para crear la base de datos desarrollada, al no ser posible su uso integrado con una única variable que mostrara el volumen o frecuencia de búsqueda de cada mes a lo largo de toda la serie, se utilizó *Google Keyword Planner* como sistema de identificación de expresiones

de búsqueda, y *Google Trends* como sistema de recogida de datos con suficiente volumen mensual de búsquedas. Este hecho, aunque ha limitado en cierta medida la calidad de los datos, ha permitido generar un corpus de expresiones de búsqueda sobre el ámbito expuesto, del que se espera que sirva como base para ser ampliado y mejorado en el futuro.

- **Identificar subtipos de intenciones de búsqueda propias del sector de la lectura, distintas a las meramente temáticas, como modificadores y características del lenguaje, que sirvan para completar las clasificaciones clásicas de intención de búsqueda (navegacional, informacional, transaccional), desde una dimensión complementaria de análisis.**

Partiendo de la clasificación genérica de intenciones de búsqueda para cualquier ámbito y sector temático, divididas en expresiones de tipo navegacional, informacional y transaccional (Broder, 2002) se ha buscado identificar otras formas de clasificación de las expresiones de búsqueda del ámbito de estudio, con el fin de escudriñar elementos que pudieran ser considerados privativos de este ámbito.

En primer lugar, se consultó la literatura al respecto del marco teórico relacionado con el estudio de los hábitos de lectura (apartado 2.2.), para detectar intereses comunes entre los investigadores y entidades relacionadas con los hábitos de lectura, así como para detectar posibles aportaciones del análisis de la conducta de búsqueda al análisis de hábitos de lectura.

Asimismo, se exploraron las posibilidades de análisis de la recuperación de información, en particular desde el paradigma cognitivo, en cuanto a su estudio del usuario a través de la disciplina del análisis de las conductas de búsqueda o *information seeking* (apartado 2.3) que pudieran ser aplicables al estudio. Ese análisis llevó a la identificación de una estrategia de clasificación de expresiones de búsqueda específica, que utiliza la facetación para determinar las posibles adscripciones de una expresión de búsqueda a varias clases de intención de búsqueda, y/o a otros aspectos identificables en el texto de la expresión (Baeza-Yates, Calderón y González-Caro, 2006).

Se procedió así a identificar facetas, aspectos modificadores y aspectos lingüísticos en las palabras que formaban las expresiones de búsqueda. Además, se llevó a cabo un segundo proceso de clasificación utilizando una nueva dimensión, la temporal. Utilizando las series históricas en aquellas expresiones de búsqueda de las que *Google Trends* había ofrecido

información estadística, se clasificó a estas expresiones en temporales y atemporales, habiendo sido a su vez ya clasificadas mediante facetas y aspectos.

La investigación posterior ha enfocado su interés en la identificación y clasificación de las palabras clave en torno a clasificaciones mutuamente excluyentes (Kathuria, 2010) o en última instancia a facetas (Jansen y Booth, 2010) y a otros aspectos detectados en las expresiones (Figuroa, 2015) En este sentido, la identificación de facetas y aspectos en el ámbito del análisis de los hábitos de búsqueda sobre la temática del estudio, ha permitido enriquecer y granular el estudio de esas expresiones de búsqueda, más allá de su adscripción genérica a una única tipología de clasificación de intención de búsqueda.

Así, se han localizado 27 categorizaciones, desglosadas en 19 facetas distintas, tres “aspectos modificadores”, y cinco “aspectos lingüísticos”, atendiendo a las diferentes perspectivas desde las que se han realizado la clasificación. Por un lado, se detectaron facetas que podían ser agrupadas dentro de las clasificaciones de búsqueda clásicas, y por otro lado se determinó que había facetas que en realidad eran horizontales, no ayudaban a englobar una expresión de búsqueda dentro de las clases de intenciones, sino que aportaban una información cualitativa distinta (por lo que se las denominó aspectos “modificadores” y “lingüísticos”).

Así, se detectaron aspectos que modificaban la expresión, ofreciendo una especificación del “formato” de lo buscado, o se hacía mención del momento o época temporal explicitada en la consulta (“tiempo”), o se brindaba información sobre el “tipo narrativo”, donde entró desde una tipología narrativa a escuelas literarias. Estos aspectos permiten una segmentación útil para el investigador que busca separar expresiones de búsqueda que mencionen por ejemplo al formato “pdf”, sin importar si son navegacionales, informacionales o transaccionales. Más allá, permiten separar ese conjunto en una o varias de las clases de intencionalidad, lo cual es una de los grandes avances de la facetación con respecto a la clasificación monojerárquica (Belkin et. al., 2009; Croft et. al., 2010; Dou, et. al., 2011).

Otro tanto sucedió con los aspectos denominados lingüísticos (“Abreviatura”, “Error”, “Género”, “Idioma”, “Plural”) Estos aspectos han aportado una dimensión de análisis también independiente de la clase o clases de intención a la que pueda pertenecer una expresión de búsqueda, centrada en las características lingüísticas de la frase de consulta.

De ellas cabe decirse que la categoría “Error” ha generado multitud de dudas al realizar la clasificación, como se describe en el apartado 3.2.2., siendo en ocasiones imposible delimitar la adscripción de una expresión de búsqueda basándose solamente en el texto (lo que ha

sucedido en diversos casos) El aspecto “Plural” es uno de los más factibles de detectar de forma automatizable, debido a la estabilidad de sus formas de expresión en español (“s”, o “es”), junto con “Abreviatura” y “Género”. El estudio de expresiones con aspecto “Género” ayudó a perfilar posibilidades de mejoras futuras, en concreto la subclasificación por los distintos géneros y aquellos que no pudieran ser determinados directamente, como los plurales genéricos (“autores”, por ejemplo) Aquí, la detección de expresiones que utilizaban el inglés y otros lenguajes resultó una sorpresa, pero no pareció deberse a un error de *Google Keyword Planner* en la detección del idioma. En unos casos las partículas en otro idioma se utilizaban directamente (por ejemplo, en “brontë sisters”), mientras que, en otras, formaba parte de una expresión en español (“cervantes meets shakespeare resumen”, por ejemplo) Este hecho abre nuevos frentes en el análisis del uso del lenguaje por parte de los usuarios de motores de búsqueda, en un mundo cada vez más globalizado.

De entre las facetas como tales, y dentro del tipo de intención de búsqueda navegacional, se identificaron tres facetas, denominadas “Entidad”, “Biblioteca” y “Lugar”. “Entidad” enmarcaba nombres de marcas, entidades administrativas o sitios web. “Biblioteca” era determinada por entidades que a su vez fueran bibliotecas, y “Lugar” hacía referencia a la mención de localizaciones geográficas. Atendían a la intención de búsqueda navegacional, porque hacía referencia a un intento por parte de los usuarios de localizar la ubicación física de una entidad o biblioteca, o su sitio web. Los estudios anteriores raramente han buscado detectar un tipo de intención de búsqueda tan concreto como “lugar” o “entidad” (Jones et. al., 2008; Jun, Yoo, Choi, 2018) y mucho menos en el ámbito de la “Biblioteca”, donde salvo error u omisión, es inédito.

También se ha identificado la faceta “Transaccional”, porque no se han descubierto clases o facetas “hijas” como tales de forma inequívoca. Lo que sí se ha apreciado es el uso de expresiones con una clara explicitación de la intención, por parte del usuario, de realizar una transacción o una acción (por ejemplo, “audiolibros gratis para descargar” o “comprar libros baratos”), mediante el uso de verbos. También se detectó otra forma más indirecta de pertenencia a la faceta “Transaccional”, indicada por la aparición de sedes web dedicadas a este tipo de acciones, como por ejemplo “harry potter y el principe mestizo torrent”, donde “torrent” indicaba una tecnología para compartir archivos entre usuarios de Internet.

Dentro del tipo de intención de búsqueda informacional, se detectó el mayor número de facetas distintas (“Ámbito”, “Autor”, “Ciencia”, “Deporte”, “Desambiguación”, “Erótico”, “Film”, “Infantil”, “Nombre personal”, “Personaje”, “Pregunta / frase”, “Título”, “Topic ficción”,

“Topic No ficción”, “Videojuego”) La aparición de facetas específicamente relacionadas con la temática a investigar, como “Autor”, “Personaje”, “Título”, “Topic ficción” o “Topic No ficción” supone no sólo una aproximación diferente frente a los estudios anteriores sobre clasificación de intenciones de búsqueda, sino una aportación útil para los estudios sobre hábitos de lectura. Estas facetas pueden contribuir con información sobre cómo se expresan los usuarios que buscan información relacionada con autores, con personajes de productos culturales, o cómo mencionan el título (en todo o en parte) para especificar su necesidad de búsqueda.

En cualquier caso, la propia condición de la facetación (que fomenta que una misma expresión pueda pertenecer a varias facetas, con distinto peso o fuerza de clasificación) así como el lenguaje de recuperación usado, en muchas ocasiones parco en palabras y capacidad de desambiguación, hacen difícil o imposible ser estrictos en la clasificación de unas expresiones como pertenecientes a una temática exclusivamente propia del sector de la lectura. En consecuencia, hace falta bien separar expresiones ambiguas de otras inequívocamente relacionadas con la temática, bien integrar más elementos de decisión para poder realizar una mejor clasificación, como el análisis de sesiones y clics en páginas de respuesta de buscadores o SERPs (*Search Engine Response Pages*) o el uso de relaciones semánticas (Clough et.al., 2009; Veilumuthu y Ramachandran, 2009; Sloan, Yang, Wang, 2015)

En multitud de ocasiones se aprecia una imposibilidad de desambiguación con los formatos de disfrute de una obra (como en el caso de “Film”, donde la búsqueda del título de una obra o de un personaje no implica que se esté buscando un libro de lectura, sino que puede ser una película, una serie, un comic o un videojuego). Esto es debido a que, como se ha explicado con diversos ejemplos en el apartado 3.2.2., la técnica de detección de facetas empleada (a través de un análisis *bottom-up* basada en el análisis de palabras sueltas, identificando las facetas a las que puede pertenecer una palabra, para luego clasificar esas facetas en torno a las clases principales clásicas de intención de búsqueda) ha resultado ser una técnica muy útil para identificar facetas concretas, pero no es suficiente para clasificar todas las posibilidades sin incurrir en sesgos o imprecisiones. En el mejor de los casos, se ha de reconocer que no es posible delimitar sólo mediante el texto de las consultas, si una parte de los usuarios pueden tener una intención u otra, incluso utilizando la misma expresión, exactamente el mismo texto.

Por tanto, por un lado, debe decirse que se han detectado facetas y aspectos que pueden ser compartidos por otros ámbitos temáticos; aspectos que aportan dimensiones de análisis horizontales, que se pueden cruzar a través de la facetación para localizar subconjuntos de expresiones de búsqueda más concretos; y facetas realmente propias del sector de la lectura,

que pueden aportar información cuantitativa y cualitativa a los investigadores del área del estudio de los hábitos de lectura. Por otro, que hay facetas que podrían seguir siendo desglosadas en otras más específicas, y así realizar estudios más granulares. Y en última instancia, que el texto no es siempre suficiente para determinar la intención de búsqueda, por lo que la técnica de clasificación de las expresiones debe ser mejorada mediante la integración de otras técnicas, para lo que habría que ejecutar nuevas investigaciones multidisciplinares.

- **Identificar relaciones entre los tipos de búsqueda y los elementos del lenguaje utilizados en la muestra.**

En el apartado 3.3.2., se exploró la posible relación entre las facetas y los aspectos descubiertos en las expresiones de búsqueda obtenidas de Google Keywords Planner, entendiendo aquí relación como “una forma de agrupamiento entre elementos” (Camarero, 2002, 380). Esta relación tomó forma como grado de co-ocurrencia en la facetación del conjunto de frases de búsqueda estudiadas, donde co-ocurrencia expresaba en este contexto la aparición de un mismo ítem (en este caso una expresión de búsqueda) en pares de facetas - aspectos (Vilares, 2009; Irazú Hernández et. al., 2012).

Este análisis permitió conocer en qué medida los usuarios utilizaban elementos lingüísticos o palabras que pueden encajar en distintas facetas. En general se ha visto cómo “Error” era el aspecto que más co-ocurría con buena parte de las facetas descubiertas (como un 32,37% con “Ciencia”, 33,33% con “Deporte”, o 32,44% con “Pregunta / frase”, por ejemplo). Se constató el uso intensivo de las abreviaturas en las búsquedas relacionadas con las bibliotecas (22,83%,) pero en particular las universitarias, y una alta tasa de coincidencia con "Error" (39,40%). Asimismo, se descubrió un alto uso del género, especialmente en las expresiones clasificadas dentro de la faceta “desambiguación” (30,22%).

Sin embargo, otras facetas mostraron poca o ninguna co-ocurrencia faceta – aspecto, como “Tiempo”, lo que supone un indicio de que esas clases de consultas tienen una menor incidencia de problemas ortográficos. Y al explorar los aspectos, se apreció que el aspecto menos utilizado en las consultas, sin importar su faceta, era el de “Tiempo”, mostrando valores entre 0.00% (para “Erótico”, “Film” o “Videojuego”) hasta 3,52% con la faceta “Transaccional”.

- **Conocer la evolución de la intención de lectura y posible interés en otros servicios y contenidos ofrecidos por las bibliotecas a través de la demanda de información en Google en España, durante el período 2004-2016.**

En los apartados 3.3.3 a 3.3.7, se analizaron y discutieron los resultados tras obtener datos históricos de una parte de la muestra de expresiones de búsqueda detectadas en *Google Keyword Planner*. Las expresiones de búsqueda se separaron en dos categorías en virtud a las características de sus series temporales, localizando frases temporales (identificadas por un inicio abrupto en el volumen de búsqueda, y una vida más o menos corta en la serie) y atemporales (palabras clave que se buscan a lo largo de una mayor parte de la serie o de toda ella, y sin incidencia de valores cero o secuencias de valores cero de índice de búsqueda) A partir de ahí, se pudo analizar la muestra tanto en su conjunto, como por separado, comparando sus características.

La visualización de los datos en forma de gráficos de caja y bigotes para los conjuntos de datos, y de líneas temporales individuales para ejemplos escogidos en una muestra intencionada (García, 1999), ha permitido mostrar la tendencia central del índice de búsqueda para cada mes de la serie (expresada con la mediana), junto con la dispersión en cuanto a esa variable, encontrando grandes diferencias entre la tendencia general, y las tendencias ofrecidas por determinadas agrupaciones de frases de búsqueda.

En referencia a la tendencia general, basada en la agrupación de medianas de las expresiones de búsqueda analizadas, se observó una disminución del valor de índice de búsqueda mensual en *Google*, para el rango temporal elegido, de frases de búsqueda referidas a la lectura, el libro y las bibliotecas, en especial desde 2010 hacia adelante. Es decir, que se significa una disminución de la popularidad de búsqueda para el conjunto de expresiones estudiadas, aquellas descubiertas en el año 2010 en el ámbito analizado, tanto aquellas que aparecieron tanto antes (al menos desde 2004) como aquellas que nacen en algún momento de la serie anterior a 2010, hasta el momento final de la toma de datos, en 2016.

Nótese que posteriormente a 2010, podrían haber nacido otras expresiones de búsqueda que hayan pervivido hasta nuestros días, pero este estudio no las puede contemplar, al haber realizado la extracción de expresiones relacionadas en 2010. Por tanto, estos datos deberían contrastarse en el futuro, para detectar nuevas incorporaciones al corpus, que puedan apoyar, desmentir o modificar en parte estas conclusiones.

Si bien el conjunto de datos analizado sólo puede ser considerado una muestra del total de los datos que podría aportar el estudio de la total producción de la industria del libro actual y pasada, los datos refrendan investigaciones mencionadas anteriormente, con datos de distintas fuentes y mediante distintas técnicas, desde el aumento de editoriales inactivas (Ministerio de Educación, Cultura y Deporte, Observatorio de la Lectura y el Libro, 2016) la caída en el número de librerías (Millán, 2017) el descenso de bibliotecas públicas (Ministerio de Educación Cultura y Deporte, 2014) o el de préstamos de documentos y productos culturales físicos en bibliotecas (Ministerio de Educación, Cultura y Deporte, 2016)

Por otra parte, se comprobó en el apartado 3.3.5. que el conjunto de series temporales, en su totalidad, tendía a mostrar un mayor índice de búsqueda en torno a los meses de abril-mayo, probablemente debido a las ferias del libro (Ministerio de Cultura, 2011) También obtenía buenos datos en noviembre, quizá debido en unos casos a las tareas escolares y en otros a la cercanía de la Navidad, aunque ambas posibilidades deberían comprobarse con un análisis más específico y el cruce con otros datos (noticias relacionadas, ventas...). Los valores más bajos fueron localizados en los meses de verano, julio y agosto, cuando el final del curso académico deja paso a las vacaciones, Pero esta hipótesis necesitaría de una investigación particular sobre expresiones de búsqueda referidas exclusivamente al entorno escolar, que las hay en el corpus desarrollado, lo que puede ser germen de ulteriores estudios.

Al segregar las series temporales en dos clases distintas (temporales y atemporales) y analizarlas por separado, las tendencias generales en ambos casos también resultaron ser a la baja, siendo en el caso de las temporales las que mostraban un ascenso y un descenso en los valores de índice de búsqueda más abruptos, en general alrededor de 2010. Las temporales fueron las que mostraron mayor uso de partículas que hacían referencia al tiempo en su lenguaje natural (“premio planeta 2009 ganador”, por ejemplo, véase 3.3.6.) mientras que en otras ocasiones fue debido a eventos concretos (muerte de un escritor, premio a un autor...).

Se constató que este hecho era debido a que las expresiones de búsqueda fueron recopiladas en 2010 y, por tanto, el uso temporal con respecto al pasado no puede hacerse en la serie a partir de 2010. Así, será recomendable realizar nuevas tomas de datos y extracciones de expresiones de búsqueda en años venideros, para realizar un mejor seguimiento de la evolución de este tipo de expresiones, y descubrir nuevas variantes y/o apariciones de expresiones relacionadas, que muestren cómo cambia el lenguaje.

Pero donde se encontraron más diferencias con el patrón común, fue al encontrar agrupaciones de expresiones más granulares, que atendían a distintas necesidades más específicas, donde en ocasiones se revelaron las relaciones causales más probables. Aquí sí se probó de utilidad el trabajo con series temporales de diversas expresiones muy relacionadas temáticamente, que aportaban distintas mediciones del mismo suceso, lo que ha ayudado a contrastar la incidencia de sucesos similares, conclusión explicada en el siguiente punto.

- **Identificar distintos subgrupos de expresiones que muestren patrones longitudinales específicos, y observar las diferencias con respecto a la tendencia general, en la evolución de la demanda de información durante el período y segmentación señalados.**

A pesar de la tendencia general a la baja en los datos históricos del conjunto de expresiones, tendencia ejemplificada en familias de palabras como las relativas a las editoriales, se exploraron las distintas series, para buscar excepciones a la regla. Como resultado del análisis del muestreo intencional en las series históricas, se descubrieron algunas prácticas específicas que se salían de la tendencia general. Así, como se explica en los apartados 3.3.6 y 3.3.7., se descubrieron elementos que aportaron indicios causales de momentos en la serie con un alto índice de búsqueda, como:

- Defunciones (“Juan Gelman”, “Francisco Ayala” o “Gabriel Garcia Marquez”, por ejemplo)
- Noticias (caso “Megaupload”)
- Premios (varios ejemplos referidos directa o indirectamente a “Mario Vargas Llosa” en 2010)
- Aparición de nuevos números (ejemplo “Naruto”)
- Aperturas (“Biblioteca Digital Mundial”)
- Eventos (conmemoración de la muerte de Cervantes en 2016)
- Salidas de nuevas ediciones (“El Principito”)
- Estrenos de películas u otros media (saga de libros, películas y videojuegos de “Harry Potter”, por ejemplo, además de DC, Marvel o Star Wars)

Previsiblemente y coincidiendo con otros casos de otras temáticas (Butler, 2013), la aparición de informaciones en los medios de comunicación sobre tales actos incidía en la popularidad de familias completas de expresiones de búsqueda relacionadas, y en el nacimiento de otras, aunque sería de utilidad investigar en profundidad este fenómeno, para buscar otras posibles variables intervinientes, como el efecto de las estrategias de generación de contenidos *transmedia*.

En el apartado de bibliotecas, fue clave la utilización del análisis de las series temporales para localizar dos subgrupos de expresiones, con tendencias opuestas. Mientras las expresiones relativas a bibliotecas públicas o de otro mostraron patrones a la baja en la mayoría de casos, las derivadas del interés sobre bibliotecas universitarias subieron a lo largo de la serie, pero sólo las que utilizaban abreviaturas de la marca de cada universidad, lo que induce a pensar en cambios en la forma de expresión del lenguaje de los usuarios, y en que no se deben tomar en cuenta un número bajo de series temporales de expresiones de búsqueda concretas para la toma de decisiones o la generalización, puesto que dos expresiones sobre el mismo concepto, pueden mostrar tendencias opuestas, llevando a falsas conclusiones.

En algunas expresiones de búsqueda resultaba difícil desambiguar la intención de búsqueda solamente mediante el texto (caso de búsquedas sobre Harry Potter), por lo que se comprobó apropiada la utilización de la segmentación temporal / atemporal para identificar frases de búsqueda más relacionadas con la versión fílmica o con la versión para lectura, especialmente en el caso de expresiones que no tienen por qué estar relacionadas semánticamente de forma directa (caso de “El Ministerio del Tiempo”). Pero este punto debería ser estudiado más detalladamente en el futuro, sistematizando y segmentando la captura de datos hacia objetos de estudio más específicos.

Finalmente, se constató la estacionalidad de ciertas variedades de expresiones, en concreto aquellas que tenían que ver con el temario del curso escolar, principalmente en septiembre-octubre. Ese hecho puede llevar a investigar específicamente el lenguaje de búsqueda de los alumnos (o de los padres) para entender qué necesidades tienen, y ayudar a los profesores a prevenirlas.

Uniendo estos casos diferenciados (expresiones influenciadas por eventos, bibliotecas universitarias, expresiones con estacionalidad...), ha de concluirse que el análisis longitudinal debe explorarse como otra vía para localizar subgrupos de expresiones de búsqueda, ya tengan relación directa semántica o no. El estudio de las series temporales se ha demostrado válido para detectar fenómenos y objetos de estudio más específicos, que permitan una reducción de las variables a considerar en su análisis. Además, también permitirá detectar indicios de los factores pertinentes que puedan concurrir en la explicación de los fenómenos que acontecen en la serie histórica.

4.3. Conclusiones generales

En este estudio se ha buscado abordar una dimensión nunca antes visitada en la disciplina del estudio de los hábitos de lectura. Fruto de la evolución tecnológica, el usuario, el lector, busca información sobre sus necesidades, y gracias a la tecnología, es más factible hoy día conocer esas necesidades.

La inexistencia de definiciones comunes para un mismo indicador en la literatura estudiada en el capítulo 2, ofrecía además una puerta a buscar nuevas metodologías. Teniendo en cuenta la denuncia de Gil, basándose en Igarza y Monak (2014) por la que se debían estandarizar y “unificar este tipo de estudios” (Gil, 2015), para poder comparar los datos de un país con los de otros, cabía la posibilidad de usar los datos de una herramienta globalmente utilizada, y con un mismo criterio para todo el planeta.

Esa circunstancia hacía que fuera necesario explorar la posibilidad de examinar otras formas de conocer cualitativa y cuantitativamente la conducta reciente y futura de los lectores y no lectores para el sector editorial, de los usuarios y no usuarios para el sector bibliotecario. En particular, la conducta de búsqueda de lo que se ha denominado genéricamente en esta investigación exploratoria “el usuario”.

A lo largo del estudio se ha determinado que los procesos y expresiones de búsqueda de información usados por los usuarios en motores de búsqueda, en concreto en Google, en España y en español, durante el período 2004-2016, aportan una información semántica de una granularidad nunca vista antes en los estudios de los hábitos de lectura, toda vez que en buena parte de las técnicas empleadas regularmente en esta disciplina se utilizan datos, encuestas y entrevistas cualitativas, que no investigan la dimensión aportada por esta investigación.

Igualmente, en el apartado 3.2.2. se han descubierto ejemplos de expresiones de nuevo uso (por ejemplo, nuevas abreviaturas), al analizar el lenguaje natural en la búsqueda de información a través de motores de búsqueda, que no han sido todavía aceptados por la Academia (Real Academia Española, 2005). Esta circunstancia invita a pensar que puede resultar una herramienta útil para los investigadores de la evolución del lenguaje, en particular el relacionado con el ámbito del estudio.

No sólo eso, se ha demostrado la existencia de formas particulares de lenguaje natural que utilizan la facetación de forma habitual como lenguaje de consulta. Aquí, la técnica de clasificación utilizada para identificar y clasificar las expresiones de búsqueda por su intención de búsqueda se ha aplicado con relativo éxito. Por un lado, se han desvelado diversas facetas, aspectos modificadores y lingüísticos, en ocasiones específicos del ámbito de estudio. Ese descubrimiento es en sí un aporte valioso para la disciplina del estudio de hábitos de lectura.

No obstante, el mismo análisis y discusión de la metodología denota que algunas de estas facetas podrían ser desglosadas en subfacetas más específicas, lo que haría futuros análisis más granulares y específicos. Además, tanto el método de extracción de datos como la clasificación manual, distan mucho de ser concluyentes, y mucho menos escalables.

Aquí cabe afirmarse que las herramientas utilizadas para la extracción de datos, *Google Keyword Planner* y *Google Trends*, (3.2.1 y 3.2.3., respectivamente) ofrecen retos y limitaciones (fundamentalmente técnicas) que exigen replantear la metodología empleada, con el fin de obtener expresiones de búsqueda más relacionadas con la temática. En particular, la lógica de *Google Keywords Planner* en 2010 ofrecía demasiadas expresiones relacionadas atendiendo a su coincidencia textual que luego no lo han sido semánticamente, donde la polisemia ha resultado ser un escollo. *Google Trends* mostró que sólo podía ofrecer datos de las expresiones con un mínimo índice de búsqueda (3.2.3), lo que ha retirado de una parte del estudio a gran parte de las expresiones de búsqueda con menor popularidad, lo que disminuye la granularidad del estudio longitudinal.

La evidencia hallada de expresiones que no podían ser desambiguadas a priori basándose sólo en los textos (3.2.2.), donde partes de la población podían tener distintas intenciones de búsqueda y utilizar distintas acepciones de la misma expresión, invita a integrar distintas técnicas de clasificación de expresiones de búsqueda. El fin último al respecto, será optimizar la escalabilidad de la metodología, y evitar sesgos en la medida de lo posible, incorporando al usuario y otros factores (como enlaces, páginas de respuesta de los buscadores, sitios web...) para mejorar la identificación de intenciones de búsqueda de expresiones que, a pesar de las ventajas de la facetación, permanecen en la ambigüedad.

De la misma forma, el análisis longitudinal ha desvelado que no hay un único patrón al que prestar atención, cuando se han observado tendencias no sólo distintas entre expresiones de tipo temporal y atemporal, sino dentro de ellas, distintas agrupaciones que atendían a conductas dispares. Estas conductas y la exploración de las posibles explicaciones causales,

estimula la necesidad de mejorar la técnica de clasificación por temporalidad utilizada (3.2.4), utilizando otros parámetros de identificación de series históricas de tipo temporal. La detección (en 3.3.5, 3.3.6 y 3.3.7) de familias de expresiones cuya tendencia no coincidía con la central (a la baja), sino que en ocasiones era incluso la contraria, ha motivado concluir que se debería ampliar el concepto de “conducta de búsqueda” hacia “conductas de búsqueda”, incidiendo en el plural.

Esta conclusión es debida a que las herramientas utilizadas, *Google Keyword Planner* y *Google Trends*, introducen en la misma segmentación (por país, lenguaje y marco temporal) a tipologías de usuarios que, aun usando el mismo lenguaje, tienen intenciones y contextos de uso distintos. Además, la riqueza y variedad de expresiones halladas llevan a pensar que, al estar examinando la evolución de las mismas, pero sólo de las detectadas en 2010, se precisa buscar en el futuro nuevas expresiones que se hayan empezado a emplear posteriormente, para que el mapa resultante no sea incompleto.

Por todo ello, las expresiones de búsqueda pueden considerarse como indicadores sólo válidos para la búsqueda de indicios de las conductas de búsqueda sobre el ámbito de estudio, pero los datos deberán utilizarse con cautela y cruzarse con otros datos y técnicas, así como ampliar la recopilación de expresiones en el futuro. La multidisciplinariedad investigadora será clave, tanto para mejorar la calidad de las exploraciones del lenguaje de recuperación de información, como para detectar y comprender los factores que aporten evidencias causales de estas conductas. El esfuerzo integrador dará como resultado el enriquecimiento del análisis y estudio de los hábitos de búsqueda de lectura y posible interés en los diversos contenidos ofrecidos por las bibliotecas en España.

4.4. Futuras líneas de investigación

Toda investigación exploratoria no constituye un final en sí misma, sino un principio. El corpus creado es pues tan sólo una semilla de nuevas posibilidades de ampliación, puesto que el sector de la lectura, el libro y las bibliotecas trabaja en muy diversos frentes temáticos, para distintas tipologías de usuarios, y en una variedad cada vez mayor de formatos y dispositivos de experiencia cultural, desde el papel a la realidad aumentada. El propio proceso de investigación ha generado nuevas preguntas, que necesitarán de ulteriores investigaciones en el futuro.

Así, el uso durante el estudio de *Google Keyword Planner* como herramienta de extracción masiva de expresiones de búsqueda, y las conclusiones obtenidas de esta experiencia, llevan a plantear un aumento en el trabajo previo de identificación y delimitación de palabras clave semilla, para extender la cobertura temática a más títulos, más autores, y más tipologías de lectura distintas, de forma que se amplíe más la exhaustividad de la muestra.

Como limitación, debe hacerse notar que, a no ser que la lógica de la herramienta cambie, siempre va a ofrecer una parte de expresiones que no resultarán relevantes en lo que a las temáticas del ámbito lector hace referencia. De hecho, lo ha hecho, aunque ciertos problemas continúen presentes (Hardwick, 2018) En este sentido, se deberá tener especial cuidado con casos de polisemia, como por ejemplo, expresiones que combinen “novela” con otras palabras, cuando en realidad están buscando información sobre una “telenovela” (Besancon, Rajman y Chappelier, 1999; Gong et. al., 2013)

Para superar este tipo de limitaciones, y para mejorar el método y las técnicas de análisis aquí expuestas, el trabajo de limpieza semántica siempre va a existir, pero se pueden buscar métodos para hacerlo de forma más escalable, menos manual y con menor sesgo. A este respecto, se plantea el uso de expresiones de búsqueda más específicas como semillas. Esta práctica podría llevar a una mayor restricción semántica de lo buscado, en lugar de usar expresiones de búsqueda abiertas, sin restricciones. El problema de no usar palabras sueltas o expresiones con menor especificación semántica, será que no por un lado no llevarán a un mayor número de expresiones de búsqueda en ocasiones no relacionadas semánticamente, pero también se podrá incurrir en ejercer un sesgo no intencionado.

También se plantea la posibilidad de trabajar a posteriori, es decir con el conjunto de expresiones recuperado más amplio posible, seleccionando sólo aquellas más relevantes. Se

identificarían los sintagmas formados por varias palabras realmente relacionados semánticamente y los no relacionados, bien con técnicas probabilísticas (Yu y Shi, 2009) bien usando los sitios webs que aparecen en las páginas de respuesta antes para las distintas expresiones de búsqueda (Tseng, Yun-Fei y Hsu, 2006; Jin, Zhou y Mobasher, 2004), bien utilizando los ejemplos ya clasificados en este estudio. De esta forma, un sistema de aprendizaje de máquina o *machine learning*, podría comparar futuras expresiones de búsqueda con los ejemplos (tanto positivos como negativos) para entrenar sus capacidades de clasificación, en el caso del uso de técnicas de clustering o agrupamiento supervisado o semisupervisado (Beitzel et. al., 2005; Yu y Ren, 2012; Ren et. al., 2015)

Esa es la razón por la que el archivo de expresiones de búsqueda ofrecido como anexo, podrá ayudar en el comienzo de esta labor, de por sí interminable, al ir incorporando nuevas expresiones en el futuro. En cualquier caso, sería deseable incorporar técnicas que no trabajen sólo con los textos de las expresiones localizadas porque, a no ser que sean trabajos cualitativos y enfocados a una variedad semántica muy controlada (Stokes y Urquhart, 2013), la mayoría de técnicas desarrolladas para segregar expresiones de búsqueda con mayor relevancia, incorporan al usuario en sus técnicas, a través del análisis de clics en los listados de respuestas, de su histórico o de su reformulación de búsquedas (Baeza-Yates, Hurtado y Mendoza, 2005; Duan, Kiciman y Zhai, 2012; Park et. al., 2012; Gong et al., 2013; Preetha y Vimal Shankar, 2015; Dou et. al., 2015) pero el problema persistirá si el investigador no tiene acceso a estudiar cantidades masivas de usuarios (Hagen et. al., 2011).

Por otro lado, la propia clasificación de intenciones de búsqueda en facetas, aunque puede superar la clasificación meramente jerárquica (Yin y Shah, 2010) debe mejorar la clasificación facetada para entender el grado de pertenencia de una expresión a una faceta o aspecto. Una posible técnica será considerar técnicas probabilísticas (Mishra et. al. 2011) como el cálculo de la frecuencia de pertenencia de varias palabras a una misma faceta, dentro de la propia frase, de forma que se pueda discernir cuánto pertenece una expresión a una faceta, en comparación con otras.

A partir de la observación de casos de subagrupaciones, como en el ejemplo del aspecto lingüístico “género”, se ha concluido que se debería aumentar la granularidad de la facetación, buscando distintas posibilidades de agrupamiento y extracción de aspectos (Paşca et. al., 2010) Esto se podría hacer interconectando dimensiones de clasificación y diferenciación, de forma que se podrían construir análisis más incisivos, reduciendo variables de análisis.

Cabría la posibilidad de emplear distintas técnicas, muy poco aplicadas al contexto del español y menos en el ámbito temático aquí estudiado. Por ejemplo, se podría hacer uso de la diferenciación gramatical o sintáctica (Herdagdelen et. al., 2010; Pinter, Reichart y Szpektor, 2016) pero aplicada al lenguaje español (Olvera-Lobo y Robinson-García, 2009); de la clasificación semántica (Guo et. al., 2009; Limam et. al., 2010); de la clasificación por volumen de palabras empleadas (Bendersky y Croft, 2009) o por la frecuencia de aparición de sintagmas (Downey, Dumais y Horvitz, 2007) Incorporando la técnica aquí utilizada de integrar la dimensión temporal, se podrían combinar las anteriores técnicas con la relación entre expresiones por su semejanza de patrones longitudinales (Alfonseca, Ciaramita y Hall, 2009)

Esa mayor granularidad y agrupación de expresiones en virtud de distintas dimensiones de relación, debería servir de base para crear una ontología de las motivaciones y conductas de búsqueda para la investigación de los hábitos lectores, de forma que se enriquecerían las relaciones entre facetas y aspectos (modificadores y lingüísticos) Esa ontología, de la misma forma que ya se usa en el análisis del lenguaje natural (Nirenburg y Raskin, 2004) en la desambiguación de nombres de persona (Delgado et. al., 2016) e incluso en el estudio de expresiones para la recuperación de información (Rodríguez et. al, 2017) permitiría conectar pares de facetas, aspectos e intenciones de búsqueda por distintos tipos de relación, de forma que el investigador podría segmentar y buscar correspondencias entre agrupaciones más concretas.

Es más, si se conectara con estudios de análisis del usuario, (clics, histórico, reformulación...) ayudaría a entender mejor las tipologías de uso de ciertas expresiones, como en el caso de aquellas con alta carga polisémica (recuérdese el ejemplo de “novela” versus “telenovela” del capítulo 1.3.3., o el de “autores” mencionado en 3.2.2.) en las que es imposible dilucidar a ciencia cierta y sólo por el puro texto de la expresión, la intención concreta de búsqueda del autor y su facetación, excepto algunos aspectos modificadores o lingüísticos.

Una de las dimensiones de análisis que suscitan mayores posibilidades de investigación futura, efecto de lo anterior, es la posibilidad de identificar co-ocurrencias entre las facetas. De esa forma, el investigador puede segmentar subconjuntos de expresiones, e incluso la propia herramienta de *Google Keyword Planner* podría ayudarse de esta funcionalidad, puesto que permitiría identificar mejor aquellas expresiones más apropiadas para cierta combinación de intención de búsqueda y aspectos (por ejemplo, aquellas que desambiguan utilizando el formato y el aspecto de tiempo, etc.). En este sentido, parece lógico avanzar el análisis de co-

ocurrencias, tanto entre pares de facetas, pares de aspectos, y pares de faceta - aspecto, expandiendo la búsqueda de coincidencias a tripletas y conjuntos de N facetas.

Otra gran dimensión de análisis que debería ser estudiada con mayor profusión sería la del estudio de series temporales en este ámbito. Para ello, una propuesta sería continuar realizando análisis sistemáticos con cierta temporalidad, para entender la evolución de este mismo conjunto de expresiones de búsqueda, utilizando *Google Trends*. Pero como se ha visto en el apartado 3.3.5., no sólo será pertinente conocer la evolución futura de los patrones de las expresiones aquí localizadas, sino que el corpus de expresiones deberá aumentar con nuevas incorporaciones de expresiones que, simplemente, todavía no se han utilizado. Por tanto, otra variante de esta propuesta será trabajar con la identificación de nuevas formas de uso del lenguaje en el entorno de la recuperación de información para este ámbito, como se ha explorado en el apartado 3.2.2. También deberá buscarse consenso investigador para homogeneizar lo que se considera “error” en la redacción de las expresiones de búsqueda en el futuro, en un contexto tecnológico que favorece el uso de errores, por entender mejor la semántica de la consulta gracias al avance en el procesamiento del lenguaje natural por parte de los motores de búsqueda (Vallez y Pedraza-Jimenez, 2007; González Villa, 2016).

Asimismo, se debería trabajar para mejorar la propuesta de identificación y clasificación de expresiones temporales aquí fraguada, utilizando los parámetros mencionados en 3.2.4 u otros por descubrir. Investigar sobre los factores que ayuden a delimitar la temporalidad de una expresión, como la diferencia de índice de búsqueda entre el valor más alto (100) y el resto de la serie, la existencia de estacionalidad, y la familia semántica a la que pertenezca la expresión, puede aportar un análisis muy rico, si se contrasta con otros datos, para buscar relaciones causales, e incluso realizar análisis predictivos más precisos (Naccarato et. al., 2018).

Por último y más relacionado si cabe con los hábitos de lectura, un fenómeno que merece la pena estudio aparte es la evolución de los clásicos, como en los casos mencionados de “El Principito” (que lo es) o de “Harry Potter” (que ¿lo es ya?, ¿puede serlo?...) El reto aquí será conectar los datos de búsqueda con los de otras fuentes, para buscar los distintos factores que pueden ejercer de variable no sólo independiente y dependiente, sino interviniente, en el aumento o descenso de la popularidad de las búsquedas relativas a ese mismo concepto u objeto de análisis.

El estudio de los hábitos de lectura, el libro y el uso de las bibliotecas es un ámbito complejo, poliédrico, en el que, como se ha apreciado, especialmente en los apartado 3.3.3 a

3.3.7.,mientras los números a nivel macro dicen una cosa (descenso generalizado de uso de los canales clásicos de consumo cultural), los datos a nivel micro dicen otra (traslado de interés hacia otros canales como el online, cambios en el lenguaje de recuperación de información, popularidad de segmentos concretos de literatura y expresiones de búsqueda, mezcla de públicos con intencionalidades distintas usando una misma frase...)

En esta exploración se ha pretendido descubrir y acercar al investigador en hábitos de lectura una nueva rama de investigación que, siguiendo a Norman (2017) debería cruzarse en el futuro con las metodologías clásicas empleadas en el sector, tanto cuantitativas como cualitativas, para comprender el sector del libro como sistema, en su creciente complejidad. El reto aquí será entender al lector, al no-lector, al usuario, al no-usuario, desde distintos puntos de vista, para servir mejor a sus intereses y necesidades presentes y futuros y, por ende, servir de ayuda al sector del libro y las bibliotecas.

5. BIBLIOGRAFÍA

BIBLIOGRAFIA

6. 20minutos. (2008). El rincón de los vagos cumple 10 años dando alegrías a través de internet. Retrieved October 15, 2018, from <https://www.20minutos.es/noticia/352126/0/rinco/vago/estudiantes/>
7. Abadal, E., & Codina, L. (2008). La diversidad cultural en Google y los motores de búsqueda: una aproximación conceptual. *El Profesional de La Información*, 17(2), 191–198. Retrieved from <http://eprints.rclis.org/11489/>
8. ABC. (2012). EE.UU. cierra Megaupload, una de las mayores web de intercambio de archivos - ABC.es. Retrieved October 15, 2018, from <https://www.abc.es/20120119/medios-redes/abci-eeuu-cierra-megaupload-201201192052.html>
9. Abedi, V., Mbaye, M., Tsvigoulis, G., Male, S., Goyal, N., Alexandrov, A. V., & Zand, R. (2015). Internet-Based Information-Seeking Behavior for Transient Ischemic Attack. *International Journal of Stroke*, 10(8), 1212–1216. <https://doi.org/10.1111/ijls.12593>
10. Acevedo Caro, E., & Mercado Lafertte, F. (2013). *Industrias creativas: ¿cómo mejoramos la industria editorial chilena?: análisis de cadena de valor del libro en Argentina, Chile y España*. Universidad de Chile. Retrieved from <http://repositorio.uchile.cl/handle/2250/112172>
11. Achananuparp, P. (2006). *Anomalous States of Knowledge in Information Retrieval: A Synthesis of Nick Belkin's Research*. Retrieved from <https://pdfs.semanticscholar.org/c854/24fd2ad59b468ab974ce7b02c695846e5f78.pdf>
12. Aguado, A., Miguel, S. E., Archuby, C., Fushimi, M. S., González, C., Pichinini, M., ... Pené, M. (2006). *Desarrollo de colecciones en bibliotecas universitarias. Metodología de evaluación. Proyecto de investigación acreditado y financiado por Universidad Nacional de La Plata, Argentina (H432) Período 2006-2007*. Retrieved from <http://eprints.rclis.org/9520/>
13. Aguilar-Lopez, D., Lopez-Arevalo, I., & Sosa-Sosa, V. (2009). Uso de ontologías para la mejora de resultados de motores de búsqueda web. *El Profesional de La Información*, 18(1). <https://doi.org/10.3145/epi.2009.ene.05>
14. Aguilera Moyano, J. de, Baños González, M., & Ramírez Perdiguero, J. (2016). Consumer Engagement: Una revisión crítica. Retrieved from https://telos.fundaciontelefonica.com/seccion=1268&idioma=es_ES&id=2016071314480003&activo=6.do
15. Alamo Felices, F. (2009). Literatura y mercado: El best-seller. Aproximaciones a su estructura narrativa, comercial e ideológica. *Espéculo: Revista de Estudios Literarios*, (43), 4. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=3099955>
16. Albornoz, S., Corda, M.-C., Gamba, V., Piovani, J. I., Pisarello, R. Z., & Ristuccia, C. (2007). Búsqueda de la información y uso de la Biblioteca por parte de los investigadores de humanidades: un estudio de caso en la Universidad Nacional de la Plata. *Revista Interamericana de Bibliotecología*, 30(1), 73–92. Retrieved from <http://eprints.rclis.org/10463/>
17. Alexandra Bota, A. M. (2014). *Previsión de la demanda turística mediante indicadores de Internet*. Valencia. Retrieved from [https://riunet.upv.es/bitstream/handle/10251/47149/Bota%2C Ana María Alexandra%2C](https://riunet.upv.es/bitstream/handle/10251/47149/Bota%2C%20Ana%20María%20Alexandra%2C)

- Previsión de la demanda turística mediante indicadores de internet.pdf?sequence=1&isAllowed=y
18. Alfonseca, E., Ciaramita, M., & Hall, K. (2009). Gazpacho and summer rash: lexical relationships from temporal patterns of web search queries. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, (August), 1046–1055. Retrieved from <http://www.aclweb.org/anthology/D/D09/D09-1109>
 19. Allen, B. (1991). Topic Knowledge and Online Catalog Search Formulation. *The Library Quarterly*, 61(2), 188–213. <https://doi.org/10.1086/602333>
 20. Allen, B. L. (1991). Cognitive Research in Information Science: Implications for Design. *Annual Review of Information Science and Technology*, 26, 3–37. Retrieved from <https://eric.ed.gov/?id=EJ441682>
 21. Alon, U. (2009). How To Choose a Good Scientific Problem. *Molecular Cell*, 35(6), 726–728. <https://doi.org/10.1016/j.molcel.2009.09.013>
 22. Alonso-Arévalo, J., & Cordon-García, J.-A. (2011). Lectura digital: evolución o revolución. *Anuario ThinkEPI*, 6, 200–206. Retrieved from <http://hdl.handle.net/10760/16919>
 23. Alonso-Arévalo, J., Cordon-García, J.-A., & Gómez-Díaz, R. (2014). Comparación de los hábitos y perfil del lector digital entre Estados Unidos y España. *Anales de Documentación*, 17(1). <https://doi.org/10.6018/analesdoc.17.1.193111>
 24. Alonso-Arévalo, J., Cordon-García, J.-A., & Maltrás-Barba, B. (2016). Altmetrics: medición de la influencia de los medios en el impacto social de la investigación. *Cuadernos de Documentación Multimedia*, 27(1). Retrieved from <http://eprints.rclis.org/29502/>
 25. Alspaugh, S., Chen, B., Lin, J., Ganapathi, A., Hearst, M. A., & Katz, R. (2014). Analyzing log analysis: an empirical study of user log mining. In *Proceedings of the 28th USENIX conference on Large Installation System Administration* (pp. 53–68). USENIX Association. Retrieved from <http://dl.acm.org/citation.cfm?id=2717495>
 26. Alvira Martín, F. (2011). *La encuesta: una perspectiva general metodológica*. Retrieved from <https://libreria.cis.es/libros/la-encuesta-una-perspectiva-general-metodologica/9788474765564/>
 27. Amitay, E. (2006). Search Performance and User ... - Google Patent Search. USA. Retrieved from <http://www.google.com/patents?id=JbyiAAAAEBAJ>
 28. Amitay, E., & Broder, A. (2008). Introduction to special issue on query log analysis. *ACM Transactions on the Web*, 2(4), 1–2. <https://doi.org/10.1145/1409220.1409221>
 29. Amitay, E., Carmel, D., Herscovici, M., Lempel, R., & Soffer, A. (2004). Trend detection through temporal link analysis. *Journal of the American Society for Information Science and Technology*, 55(14), 1270–1281. <https://doi.org/10.1002/asi.20082>
 30. Angiolillo, J. (2003). Search Log Analysis as a Usability Engineering Tool. In *CHI'03 Workshop on Best practices and future visions for search user interfaces*. Ft. Lauderdale, Florida, USA.
 31. Arnold, S. E. (2006). Search: the new application platform. *The Electronic Library*, 24(2), 121–125. <https://doi.org/10.1108/02640470610660314>
 32. Arriola Navarrete, O., & Butrón Yáñez, K. (2008). Un acercamiento a la evaluación de bibliotecas. *Biblioteca Universitaria: Revista de La Dirección General de Bibliotecas de La UNAM*, 11(2), 99–114. Retrieved from <http://eprints.rclis.org/18415/>
 33. Ashkan, A., & Clarke, C. L. A. (2009). Characterizing commercial intent. In *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09* (p. 67). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1645953.1645965>
 34. Asociación de las Cámaras del Libro de España. (2015). *Plan integral para el fomento del libro y de la lectura 2015*. Retrieved from <http://www.lecturalab.org/uploads/website/docs/5866-1-PLAN INTEGRAL PARA EL FOMENTO DEL LIBRO Y DE LA LECTURA 2015.pdf>

35. Audemar, N. (2016). Buscadores de internet ¿Existe una alternativa a Google? Retrieved September 12, 2018, from <http://agenciaseomarketingonline.es/blog/buscadores-de-internet-mas-usados/#google>
36. Audunson, R. (1999). Can institutional theory contribute to our understanding of information seeking behaviour? Towards a general model of information seeking behaviour with institutionalised, information-relevant norms as an intermediate category of variables. In D. K. (David K. Allen & T. D. Wilson (Eds.), *Exploring the contexts of information behaviour* (p. 625). Sheffield, UK.: Taylor Graham. Retrieved from <https://dl.acm.org/citation.cfm?id=315852>
37. Avanzabus. (2018). Bienvenido a Avanzabus.com - Avanza Autobuses. Retrieved October 16, 2018, from <https://www.avanzabus.com/>
38. Badia, G. (2010). Google Scholar Out-Performs Many Subscription Databases when Keyword Searching. *Evidence Based Library and Information Practice*, 5(3), 39–41. Retrieved from <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/8543/7347>
39. Baeza-Yates, R. (2017). Ten years of wisdom and beyond. In *WSDM 2017: Web Search & Data Mining*. Retrieved from <http://www.wsdm-conference.org/2017/wp-content/uploads/2016/07/Ten-Years-of-Wisdom-2017.pdf>
40. Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval. *New York*, 9, 513. <https://doi.org/10.1080/14735789709366603>
41. Baeza-Yates, R., Calderón-Benavides, L., & González-Caro, C. (2006). The Intention Behind Web Queries. In *Proceedings of the 13th international conference on String Processing and Information Retrieval* (pp. 98–109). Springer-Verlag. https://doi.org/10.1007/11880561_9
42. Baeza-Yates, R., Hurtado, C., & Mendoza, M. (2004). Query Clustering for Boosting Web Page Ranking. In *Advances in Web Intelligence. AWIC 2004. Lecture Notes in Computer Science, vol 3034*. (pp. 164–175). Retrieved from <http://www.springerlink.com/content/p4v0vatky85aj2f7>
43. Baeza-Yates, R., Hurtado, C., & Mendoza, M. (2005). Query Recommendation Using Query Logs in Search Engines. In *International Conference on Extending Database Technology* (pp. 395–397). Retrieved from <http://www.springerlink.com/content/xrxlkhx07h376dvd>
44. Baeza-Yates, R., Hurtado, C., & Mendoza, M. (2007). Improving search engines by query clustering. *Journal of the American Society for Information Science and Technology*, 58(12), 1793–1804. <https://doi.org/10.1002/asi.20627>
45. Baidu. (2018). Baidu Index. Retrieved September 12, 2018, from <http://index.baidu.com/#/>
46. Bakker, K. M., Martinez-Bakker, M. E., Helm, B., & Stevenson, T. J. (2016). Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), 6689–6694. <https://doi.org/10.1073/pnas.1523941113>
47. Ball, P. (2013, April 26). Counting Google searches predicts market movements. *Nature*. <https://doi.org/10.1038/nature.2013.12879>
48. Barba Pérez, A.-S. (2009). Indicadores de calidad en la biblioteca pública. In *I Jornada de las Bibliotecas Públicas Municipales de Arona*. Arona (Spain). Retrieved from <http://eprints.rclis.org/13095/>
49. Barba Pérez, A.-S., Palacios Serrano, A., Arroyo Vázquez, N., & Villoldo Gómez, F. J. (2006). INE, 30 años de estadísticas bibliotecarias. *El Profesional de La Información*, 15(5), 397–403. Retrieved from <http://eprints.rclis.org/9630/>
50. Barber, W., & Badre, A. (1998). Culturability: The Merging of Culture and Usability. In *Proceedings of the 4th Conference on Human Factors and the Web*. Retrieved from https://www.usj.edu.lb/moodle/stephane.bazan/obs_interculturelle/Barber and Badre.pdf

51. Bar-Ilan, J., Zhu, Z., & Levene, M. (2009). Topic-specific analysis of search queries. In *WSCD '09 Proceedings of the 2009 workshop on Web Search Click Data* (pp. 35–42). Barcelona, Spain: ACM. <https://doi.org/10.1145/1507509.1507515>
52. Barr, C., Jones, R., & Regelson, M. (2008). The Linguistic Structure of English Web-Search Queries. In Association for Computational Linguistics (Ed.), *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 1021–1030). Honolulu. Retrieved from <http://www.aclweb.org/anthology/D08-1107>
53. Barry, C., & Lardner, M. (2011). A Study of First Click Behaviour and User Interaction on the Google SERP. In *Information Systems Development* (pp. 89–99). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9790-6_7
54. Basanta Reyes, A. (2010). *La Lectura*. (A. Basanta Reyes, Ed.). Madrid: Consejo Superior de Investigaciones Científicas; Los libros de la Catarata. Retrieved from <https://editorial.csic.es/publicaciones/libros/11882/978-84-00-09097-5/la-lectura.html>
55. Bates, M. (1985). An Exploratory Paradigm for Online Information Retrieval. In *Proceedings of the Sixth International Research forum in Information Science (IRFIS 6)* (pp. 91–99). Frascati, Italy.
56. Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science*, 30(4), 205–214. <https://doi.org/10.1002/asi.4630300406>
57. Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5), 407–424. <https://doi.org/10.1108/eb024320>
58. Bauckhage, C., & Manshaei, K. (2014). Kernel Archetypal Analysis for Clustering Web Search Frequency Time Series. In *2014 22nd International Conference on Pattern Recognition* (pp. 1544–1549). IEEE. <https://doi.org/10.1109/ICPR.2014.274>
59. Baye, M. R., Morgan, J., Scholten, P., Ancarani, F., Arbatskaya, M., Chevalier, J., Taylor, B. (2006). Information, Search, and Price Dispersion. *Handbook on Economics and Information Systems*, 1, 323–377. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.482.7461&rep=rep1&type=pdf>
60. Baye, M., Santos, B. de los, & Wildenbeest, M. (2013). *Searching for Physical and Digital Media: The Evolution of Platforms for Finding Books*. Cambridge, MA. <https://doi.org/10.3386/w19519>
61. Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation*, 53(1), 8–19. <https://doi.org/10.1108/EUM0000000007187>
62. Beaulieu, M., Robertson, S., & Rasmussen, E. (1996). Evaluating interactive systems in TREC. *Journal of the American Society for Information Science*, 47(1), 85–94. [https://doi.org/10.1002/\(SICI\)1097-4571\(199601\)47:1<85::AID-ASIS8>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(199601)47:1<85::AID-ASIS8>3.0.CO;2-Z)
63. Beeferman, D., & Berger, A. (2000). Agglomerative clustering of a search engine query log. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00* (pp. 407–416). New York, New York, USA: ACM Press. <https://doi.org/10.1145/347090.347176>
64. Beeferman, D., & Berger, A. (2000). Agglomerative clustering of a search engine query log. In *KDD '00 Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 407–416). Boston, Massachusetts, United States: ACM. <https://doi.org/10.1145/347090.347176>
65. Beel, J., Gipp, B., & Wilde, E. (2010). Academic Search Engine Optimization (ASEO). *Journal of Scholarly Publishing*, 41(2), 176–190. <https://doi.org/10.3138/jsp.41.2.176>
66. Beitzel, S. M., Jensen, E. C., Chowdhury, A., Frieder, O., & Grossman, D. (2007). Temporal Analysis of a Very Large Topically Categorized Web Query Log. *Journal of the American Society for Information Science and Technology*, 58(2), 1532–2890. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.146.1848>

67. Beitzel, S. M., Jensen, E. C., Frieder, O., Lewis, D. D., Chowdhury, A., & Kolcz, A. (2005). Improving Automatic Query Classification via Semi-Supervised Learning. In *Fifth IEEE International Conference on Data Mining (ICDM'05)* (pp. 42–49). IEEE.
<https://doi.org/10.1109/ICDM.2005.80>
68. Belam, M. (2003). How search can help you understand your audience - part 1. Retrieved March 12, 2018, from http://www.currybet.net/cbet_blog/2003/05/how-search-can-help-you-under-2.php
69. Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *The Canadian Journal of Information Science*, 5, 133–143. Retrieved from http://faculty.washington.edu/harryb/courses/INFO310/Belkin1980_ASK.pdf
70. Belkin, N. J. (1991). A methodology for taking account of user tasks, goals and behavior for design of computerized library catalogs. *SIGCHI Bull.*, 23(1), 61–65.
<https://doi.org/10.1145/122672.122687>
71. Belkin, N. J., Bierig, R., Buscher, G., Elst, L. van, & Gwizdka, J. (2009). Understanding the User-Logging and Interpreting User Interactions in Information Search and Retrieval (UIIR-2009). In *Understanding the User - Logging and Interpreting User Interactions in Information Search and Retrieval (UIIR-2009)*. Retrieved from https://www.researchgate.net/profile/Jacek_Gwizdka/publication/228544999_Understanding_the_User-Logging_and_Interpreting_User_Interactions_in_Information_Search_and_Retrieval_UIIR-2009/links/09e415058849969411000000.pdf#page=42
72. Bendersky, M., & Croft, W. B. (2009). Analysis of long queries in a large scale search log. In *Proceedings of the 2009 workshop on Web Search Click Data - WSCD '09* (pp. 8–14). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1507509.1507511>
73. Besancon, R., Rajman, M., & Chappelier, J.-C. (1999). Textual similarities based on a distributional approach. *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, 180–184.
<https://doi.org/10.1109/DEXA.1999.795163>
74. Biblioteca Digital Mundial. (2009). Información general - Biblioteca Digital Mundial. Retrieved October 15, 2018, from <https://www.wdl.org/es/background/>
75. Biblioteca Nacional de España. (2013). *Manual de indización de Encabezamientos de Materia Biblioteca Nacional de España*. Retrieved from <http://www.bne.es/es/Micrositios/Guias/ManualMaterias/Docs/EMBNE.pdf>
76. Biblioteca Nacional de España. (2014). *Manual de Autoridades*. Biblioteca Nacional de España. Retrieved from http://www.bne.es/es/Micrositios/Publicaciones/AUTORIDADES/003_Normativa/
77. Biblioteca Nacional de España. (2018). Dostoievski - Datos BNE. Retrieved October 16, 2018, from <http://datos.bne.es/find?s=dostoievski&type=C1005>
78. Biggs, M. (1990). Information Overload and Information Seekers: *The Reference Librarian*, 11(25–26), 411–429. https://doi.org/10.1300/J120v11n25_18
79. Bilder, G. (2006). In Google We Trust? *Journal of Electronic Publishing*, 9(1).
<https://doi.org/http://dx.doi.org/10.3998/3336451.0009.101>
80. Bing. (2018). Bing - Keyword Research Tool. Retrieved September 12, 2018, from <https://www.bing.com/toolbox/keywords>
81. Black, E. L. (2009). Web Analytics: A Picture of the Academic Library Web Site User. *Journal of Web Librarianship*, 3(1), 3–14. <https://doi.org/10.1080/19322900802660292>
82. Blanco, E. F. (2014). *Herramientas de Big Data: ¿Podemos aprovechar Google Trends para pronosticar algunas variables macro relevantes? * Estacionalidad View project Nowcast Actividad View project*. <https://doi.org/10.13140/RG.2.1.1514.2644>

83. Blázquez, J. (2016, April 16). "El Ministerio del Tiempo", una máquina de rescatar literatura | Cultura | EL MUNDO. *Elmundo.Es*. Retrieved from <http://www.elmundo.es/cultura/2016/04/16/57117d60e5fdea33238b4580.html>
84. Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, 130, 99–113. <https://doi.org/10.1016/j.techfore.2017.07.027>
85. Blecic, D. D., Bangalore, N. S., Dorsch, J. L., Henderson, C. L., Koenig, M. H., & Weller, A. C. (1998). Using Transaction Log Analysis to Improve OPAC Retrieval Results. *College & Research Libraries*, 59(1), 39–50. <https://doi.org/VL - 59>
86. Boock, M. (2007). The Future of the Catalog: an Overview of Current Developments and Trends. In *Oregon Library Association Annual Conference*. Corvallis, Oregon. Retrieved from <http://ir.library.oregonstate.edu/jspui/handle/1957/4589>
87. Borges, J. L. (1964). *Obra poética, 1923-1964*. (Emecé, Ed.). Retrieved from https://books.google.es/books/about/Obra_poética_1923_1964.html?id=5OISAAAAYAAJ&redir_esc=y
88. Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information-retrieval studies. *Journal of the American Society for Information Science*, 37(6), 387–400. Retrieved from <https://pantherfile.uwm.edu/mux/www/sois110/details/materils/borgman.pdf>
89. Borgman, C. L. (1989). All users of information retrieval systems are not created equal: An exploration into individual differences. *Information Processing & Management*, 25(3), 237–251. [https://doi.org/10.1016/0306-4573\(89\)90042-3](https://doi.org/10.1016/0306-4573(89)90042-3)
90. Bosca, A., & Dini, L. (2009). The role of logs in improving cross language access in digital libraries. In *Proceedings of the International Conference on Semantic Web and Digital Libraries*. Retrieved from <http://research.celi.it/attach/Publications/icsd2.pdf>
91. Bowden, J. L.-H. (2009). The Process of Customer Engagement: A Conceptual Framework. *The Journal of Marketing Theory and Practice*, 17(1), 63–74. <https://doi.org/10.2753/MTP1069-6679170105>
92. Bradford, C., & Marshall, I. W. (1999). Analysing Users WWW Search Behaviour. In *Proc IEE colloquium 99/149 - navigating the web* - (p. 4). Retrieved from [http://www.research.lancs.ac.uk/portal/en/publications/analysing-users-www-search-behaviour\(038832a5-cbb3-490d-9c11-cea809235726\).html](http://www.research.lancs.ac.uk/portal/en/publications/analysing-users-www-search-behaviour(038832a5-cbb3-490d-9c11-cea809235726).html)
93. Brenes, D. J., & Gayo-Avello, D. (2009). Stratified analysis of AOL query log. *Information Sciences*, 179(12), 1844–1858. <https://doi.org/10.1016/j.ins.2009.01.027>
94. Breslau, L., Cao, P., Fan, L., Phillips, G., & Shenker, S. (1999). Web Caching and Zipf-like Distributions: Evidence and Implications. *IEEE INFOCOM*. Retrieved from <https://pdfs.semanticscholar.org/337e/4b7f57ccbb7485950b93da9c5bb4ec4dc9ad.pdf>
95. Bridwell, T. (2011). Hourly analysis of navigational, transactional, and informational user-intents in search engine queries. *Library Student Journal*, 9. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=70317604&site=ehost-live>
96. Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, 36(2), 3. <https://doi.org/10.1145/792550.792552>
97. Broder, A., Maarek, Y. S., Bharat, K., Dumais, S. T., Papa, S., Pedersen, J., & Raghavan, P. (2005). Current trends in the integration of searching and browsing. In *Special interest tracks and posters of the 14th international conference on World Wide Web - WWW '05* (p. 793). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1062745.1062751>
98. Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer Engagement. *Journal of Service Research*, 14(3), 252–271. <https://doi.org/10.1177/1094670511411703>
99. Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2–3), 87–129. <https://doi.org/10.1007/BF00143964>

100. Brynjolfsson, E., Hu, Y. (Jeffrey), & Simester, D. (2011). Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science*, 57(8), 1373–1386. <https://doi.org/10.1287/mnsc.1110.1371>
101. Brynjolfsson, E., Hu, Y. (Jeffrey), & Smith, M. D. (2011). La cola más larga: La curva de distribución de Amazon on JSTOR. *Trama & Texturas*, 15, 57–70. Retrieved from http://www.jstor.org/stable/41349407?seq=1#page_scan_tab_contents
102. Butler, D. (2013). When Google got flu wrong. *Nature*, 494(7436), 155–156. <https://doi.org/10.1038/494155a>
103. Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled Web we wove. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99* (pp. 544–551). New York, New York, USA: ACM Press. <https://doi.org/10.1145/302979.303154>
104. Cahill, K., & Chalut, R. (2009). Optimal Results: What Libraries Need to Know About Google and Search Engine Optimization. *The Reference Librarian*, 50(3), 234–247. <https://doi.org/10.1080/02763870902961969>
105. Calderón-Benavides, L. (2011, September 27). *Unsupervised Identification of the User's Query Intent in Web Search. TDX (Tesis Doctorals en Xarxa)*. Universitat Pompeu Fabra. Retrieved from <https://repositori.upf.edu/handle/10230/13032>
106. Camarero, L. (2002). Acerca de las medidas de asociación en investigación social: un viejo problema que conviene no olvidar. In J. M. Arribas & M. Barbut (Eds.), *Estadística y Sociedad* (pp. 377–397). Madrid: UNED y L'ecole des Hautes Etudes en Sciences Sociales. Retrieved from https://www2.uned.es/dpto-sociologia-l/departamento_sociologia/luis_camarero/asociacion.pdf
107. Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 49(10), 1557–1564. <https://doi.org/10.1086/630200>
108. Carrière-Swallow, Y., & Labbé, F. (2010). Nowcasting With Google Trends in an Emerging Market. *Working Papers Central Bank of Chile*. Retrieved from <https://ideas.repec.org/p/chb/bcchwp/588.html>
109. Carsen, T. M. (1991). Los métodos cuantitativos en bibliotecología y documentación como herramienta de apoyo a la investigación científica. In *XXVI Reunión Nacional de Bibliotecarios*. Buenos Aires (Argentina). Retrieved from <http://eprints.rclis.org/6914/>
110. Casado Sánchez, A., Lloret-Romero, N., & Serrano-Cobos, J. (2011). *Conducta de búsqueda de usuarios de internet en google.com*. Universidad Politécnica de Valencia. Retrieved from <https://riunet.upv.es/bitstream/handle/10251/12130/Memoria.pdf?sequence=1>
111. Case, D., & Given, L. M. (2002). *Looking for information: a survey of research on information seeking, needs, and behavior*.
112. Castellanos, M. (2003). Hot Miner: Discovering Hot Topics from Dirty Text. In M. W. Berry (Ed.), *Survey of Text Mining* (1st ed., pp. 123–157). Springer New York LLC. Retrieved from <http://books.google.com/books?hl=en&lr=&id=WNxbDFbmO-8C&oi=fnd&pg=PA123&dq=%22search+log%22&ots=YSXFDbbJ5p&sig=MnzJFHy2fP0RaXCYDjk8qWOCuc8#v=onepage&q=%22search+log%22&f=false>
113. Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Comput. Netw. ISDN Syst.*, 27(6), 1065–1073. Retrieved from <http://portal.acm.org/citation.cfm?id=206540&dl=ACM&coll=ACM&CFID=22427616&CFTOKEN=67770541>
114. Cavazos-Rehg, P. A., Krauss, M. J., Spitznagel, E. L., Lowery, A., Grucza, R. A., Chaloupka, F. J., & Bierut, L. J. (2015). Monitoring of non-cigarette tobacco use using

- Google Trends. *Tobacco Control*, 24(3), 249–255. <https://doi.org/10.1136/tobaccocontrol-2013-051276>
115. Celaya, J. (2017). ¿Fatiga visual o ceguera sectorial? Retrieved March 21, 2018, from <http://www.dosdoce.com/2017/05/10/fatiga-visual-ceguera-sectorial/>
 116. Centro de Investigaciones Sociológicas de España. (1998). Distribuciones de frecuencia marginales del estudio 2280. Retrieved May 19, 2017, from http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2280_2299/2280/e228000.html
 117. Centro de Investigaciones Sociológicas de España. (2003). DISTRIBUCIONES DE FRECUENCIA MARGINALES DEL ESTUDIO 2478 CUESTIONARIO 0 MUESTRA 0. Retrieved May 19, 2017, from http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2460_2479/2478/e247800.html
 118. Centro de Investigaciones Sociológicas de España. (2009). Frecuencias marginales unidimensionales del estudio 2806.0.0. Retrieved May 19, 2017, from http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2800_2819/2806/e280600.html
 119. Centro de Investigaciones Sociológicas de España. (2014). 3047/0-0 BARÓMETRO DE DICIEMBRE 2014. Retrieved May 19, 2017, from http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/3040_3059/3047/es3047mar.html
 120. Centro de Investigaciones Sociológicas de España. (2016). Banco de Datos del CIS. Retrieved May 19, 2017, from <http://www.analisis.cis.es/cisdb.jsp>
 121. Cervellin, G., Comelli, I., & Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *Journal of Epidemiology and Global Health*, 7(3), 185–189. <https://doi.org/10.1016/j.jegh.2017.06.001>
 122. Chan, W. S., Leung, K. W.-T., & Lee, D. L. (2004). Clustering Search Engine Query Log Containing Noisy Clickthroughs. In *International Symposium on Applications and the Internet* (p. 305). Los Alamitos, CA, USA: IEEE Computer Society. <https://doi.org/http://doi.ieeecomputersociety.org/10.1109/SAINT.2004.1266134>
 123. Cheuk, B. W.-Y. (1998). Modelling the Information Seeking and Use Process in the Workplace. *Information Research*, 4(2). Retrieved from <http://informationr.net/ir/4-2/jisic/cheuk.html>
 124. Cheuk, B. W.-Y., & Dervin, B. (1999). A qualitative sense-making study of the information seeking situations faced by professionals in three workplace contexts. *Electronic Journal of Communication*, 9(2,3,4). Retrieved from <http://www.cios.org/EJCPUBLIC/009/2/009215.html>
 125. Chevalier, J., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics*, 1(2), 203–222. Retrieved from <http://faculty.chicagobooth.edu/austan.goolsbee/research/amzn.pdf>
 126. Chiang-Martín, D., & et. al. (2014). Diseño de un plan de marketing interno: factor de impulso para la adquisición del libro electrónico. In *13º Workshop de Rebiun sobre proyectos digitales, Los libros electrónicos en las bibliotecas*. Zamora, Spain. Retrieved from <http://eprints.rclis.org/28917/>
 127. Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(1), 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
 128. Choo, C. W., Detlor, B., & Turnbull, D. (1998). A Behavioral Model of Information Seeking on the Web--Preliminary Results of a Study of How Managers and IT Specialists Use the Web. *First Monday*. Retrieved from <http://choo.ischool.utoronto.ca/fis/respub/asis98/>

129. Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information seeking on the Web: An integrated model of browsing and searching. *First Monday*, 5(2).
<https://doi.org/10.5210/fm.v5i2.729>
130. Choo, C. W., Detlor, B., & Turnbull, D. (2000). Models of Information Seeking on the World Wide Web. In *Web Work: Information Seeking and Knowledge Work on the World Wide Web* (pp. 133–158). Springer Netherlands. https://doi.org/10.1007/978-94-015-9405-9_5
131. Churick, N. (2018). How to Use Google Trends for Keyword Research: 7 Effective Ways. Retrieved September 18, 2018, from <https://ahrefs.com/blog/how-to-use-google-trends-for-keyword-research/>
132. Clough, P., Sanderson, M., Abouammoh, M., Navarro, S., & Paramita, M. (2009). Multiple approaches to analysing query diversity. In *SIGIR '09 Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 734–735). Boston, MA, USA: ACM. <https://doi.org/10.1145/1571941.1572102>
133. Cole, C. (2014). Information Need and the Beginning of Information Search. In *Encyclopedia of Information Science and Technology* (Third Edit, pp. 4117–4128). IGI Global. <https://doi.org/10.4018/978-1-4666-5888-2.ch405>
134. Conchi Jiménez Fernández. (2005). la entrevista MB: Manuel Carrión. *Mi Biblioteca*, (Abril), 28–33. Retrieved from https://gredos.usal.es/jspui/bitstream/10366/119821/1/MB1_N1_P28-33.pdf
135. Cook, T. D., & Reichardt, C. S. (1986). *Métodos cualitativos y cuantitativos en investigación evaluativa* (1ª). Madrid: Sage Publications. Retrieved from http://www.fceia.unr.edu.ar/geii/maestria/2014/DraSanjurjo/12de20/Cook_Reichardt.pdf
136. Cooper, A. (2008). A survey of query log privacy-enhancing techniques from a policy perspective. *ACM Transactions on the Web (TWEB)*, 2(4), 1–27.
<https://doi.org/10.1145/1409220.1409222>
137. Corda, M.-C., & Viñas, M. (2015). Modelo de evaluación para sitios Web de bibliotecas universitarias. *Informatio: Revista Del Instituto de Información de La Facultad de Información y Comunicación*, 20(2). Retrieved from <http://eprints.rclis.org/29400/>
138. Cordón-García, J.-A. (2006). Lecturas sobre la lectura: los circuitos de compra y lectura de libros y otros aspectos de recepción actual en Europa. In Arco Libros (Ed.), *Lecturas sobre la lectura*. (pp. 735–763). ICC Press. Retrieved from <http://eprints.rclis.org/8905/>
139. Cordón-García, J.-A. (2011). Los libros electrónicos en la biblioteca: nuevas lecturas. In *IV Jornada Profesional de la Red de Bibliotecas del Instituto Cervantes: «Bibliotecas para el lector digital: relación, espacio y tecnología»* (pp. 1–12). Madrid: Red de Bibliotecas del Instituto Cervantes. Retrieved from <http://gredos.usal.es/jspui/handle/10366/113066>
140. Cordón-García, J.-A. (2011). Los libros en la encrucijada digital. Nuevas formas de edición, nuevas formas de lectura. In *Nuevas tecnologías en bibliotecas y archivos (Narón, 25-26 de noviembre, 2010)* (pp. 15–33). Universidade da Coruña, Servicio de Publicacións. Retrieved from http://ruc.udc.es/dspace/bitstream/handle/2183/13112/CC-116_art_2.pdf?sequence=1&isAllowed=y
141. Cordón-García, J.-A. (2013). Las lecturas fingidas: entre la economía de la atención y la tentación del abandono. Retrieved March 2, 2017, from <http://diarium.usal.es/jcordon/2013/07/25/las-lecturas-fingidas-entre-la-economia-de-la-atencion-y-la-tentacion-del-abandono/>
142. Cordón-García, J.-A., & Fernández, A. O. J. (2015). ¿Se está transformando la lectura y la escritura en la era digital? *Revista Interamericana de Bibliotecología*, 38(2), 137–145. Retrieved from <https://aprendeenlinea.udea.edu.co/revistas/index.php/RIB/article/view/22585/18658>

143. Cordón-García, J.-A., Alonso-Arévalo, J., Gómez-Díaz, R., & Linder, D. (2014). El sector del libro en España: las paradojas de un modelo dual. *Electronic Library*, 32(4). Retrieved from <http://eprints.rclis.org/29659/>
144. Cordón-García, J.-A., Gómez-Díaz, R., & Alonso-Arévalo, J. (2011). Libros electrónicos: oferta comercial y redes p2p. *El Profesional de La Información*, 20(2), 149–158. Retrieved from <https://gredos.usal.es/jspui/handle/10366/82972>
145. Cordón-García, J.-A., Gómez-Díaz, R., & Alonso-Arévalo, J. (2013). *Documentos electrónicos y textualidades digitales: nuevos lectores, nuevas lecturas, nuevos géneros*. Retrieved from <http://eprints.rclis.org/22615/>
146. Costa, R., & Seco, N. (2008). Hyponymy Extraction and Web Search Behavior Analysis Based on Query Reformulation. In D. N. Geffner H., Prada R., Machado Alexandre I. (Ed.), *Advances in Artificial Intelligence – IBERAMIA 2008* (pp. 332–341). Springer Berlin / Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-540-88309-8_34
147. Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>
148. Cothey, V. (2002). A longitudinal study of World Wide Web users’ information-searching behavior. *Journal of the American Society for Information Science and Technology*, 53(2), 67–78. <https://doi.org/10.1002/asi.10011>
149. Croft, W. B., Bendersky, M., Li, H., & Xu, G. (2010). Query Representation and Understanding Workshop. In *The Query Representation and Understanding Workshop*. Geneva, Switzerland. Retrieved from <https://pdfs.semanticscholar.org/83cc/be1e5c7066c9170334b237c1e3feb939e43e.pdf>
150. Cruz Gil, M. del C. (2014, September 16). *Sistemas de recuperación de la información en el Ministerio de Fomento: características y propuesta de un modelo de análisis*. Universidad Complutense de Madrid. Retrieved from <http://eprints.ucm.es/30489/>
151. Cuesta Rodríguez, F., Montiel Valdés, A., Cano Inclán, A., Cabrera Morales, I., & Patiño Díaz, J. (2014). Instrumento para medir los Servicios Bibliotecarios en la Educación Superior. *Bibliotecas. Anales de Investigación*, 10(10), 82–93. Retrieved from <http://eprints.rclis.org/25189/>
152. Cueto-Álvarez-De-Sotomayor, L., Soler-Herreros, C., Entremonzaga-Ayuso, R., Martínez, E., Carrasco, A., Aguillo, I. F., ... Sanz-de-las-Heras, J. (2004). Técnicas de análisis de posicionamiento. *El Profesional de La Información*, 14(1), 14–20. Retrieved from <http://eprints.rclis.org/19415/>
153. Delgado, E., Martínez, A. D. ;, Montalvo, R. ;, & Fresno, S. ; (2016). Procesamiento del Lenguaje Natural Sociedad Española para el Procesamiento del Lenguaje Natural. *Procesamiento Del Lenguaje Natural*, Núm, 57, 117–124. Retrieved from <http://www.redalyc.org/articulo.oa?id=515754424013>
154. Deloitte. (2017). *Estudio de consumo navideño*. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/es/Documents/bienes-comsumo-distribucion-hosteleria/Estudio-Consumo-Navideno-2017.pdf>
155. Dervin, B. (1983). *An Overview of Sense-making Research: Concepts, Methods, and Results to Date*. Retrieved from https://books.google.es/books/about/An_Overview_of_Sense_making_Research.html?id=wlnhAAAAMAAJ&redir_esc=y
156. Dholakia, U. M., & Bagozzi, R. (2001). Consumer Behavior in Digital Environments. In Wiley (Ed.), *Digital Marketing* (pp. 163–200). New York.
157. Dholakia, U. M., Kahn, B. E., Reeves, R., Rindfleisch, A., Stewart, D., Taylor, E. (2010). Consumer Behavior in a Multichannel, Multimedia Retailing Environment. *Journal of Interactive Marketing*, 24(2), 86–95. <https://doi.org/10.1016/J.INTMAR.2010.02.005>

158. Díaz-Jatuf, J., & Panzer, S. (2014). Uso didáctico del cuestionario para detectar necesidades de información en alumnos/usuarios de pregrado de la Biblioteca “Prof. Dr. José Arce” de la Facultad de Odontología de la Universidad de Buenos Aires. In *X Jornadas de Material Didáctico y Experiencias Innovadoras en Educación Superior*. Retrieved from <http://eprints.rclis.org/23596/>
159. Digital Analytics Association. (2011). Digital Analytics Association Mission / Vision. Retrieved September 20, 2018, from <https://www.digitalanalyticsassociation.org/mission-vision>
160. Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human-computer interaction*. Prentice Hall. Retrieved from <http://dl.acm.org/citation.cfm?id=249491>
161. Dotras, O. (2015, April). ¿Cuáles son los hábitos de lectura de los españoles? *La Vanguardia*. Retrieved from <http://www.lavanguardia.com/vangdata/20150423/54430790670/habitos-lectura-espanoles.html>
162. Dou, Z., Hu, S., Luo, Y., Song, R., & Wen, J.-R. (2011). Finding dimensions for queries. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management - CIKM '11*, 1311–1320. Retrieved from <http://dl.acm.org/citation.cfm?doi=2063576.2063767>
163. Dou, Z., Jiang, Z., Hu, S., Wen, J.-R., & Song, R. (2015). Automatically Mining Dimensions for Queries from Their Search Results. *IEEE Transactions on Knowledge and Data Engineering, PP(99)*, 1–1. <https://doi.org/10.1109/TKDE.2015.2475735>
164. Downey, D., Dumais, S. T., & Horvitz, E. (2007). Heads and tails: studies of web search with common and rare queries (pp. 847–848). Amsterdam, The Netherlands: ACM. <https://doi.org/10.1145/1277741.1277939>
165. Duan, H., Kiciman, E., & Zhai, C. (2012). Click Patterns: An Empirical Representation of Complex Query Intents. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 1035–1044. Retrieved from <http://doi.acm.org/10.1145/2396761.2398400>
166. Dumais, S. T., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding User Behavior Through Log Data and Analysis. In *Ways of Knowing in HCI* (pp. 349–372). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-0378-8_14
167. Edelman, B. (2012). Using Internet Data for Economic Research. *Journal of Economic Perspectives*, 26(2), 189–206. <https://doi.org/10.1257/jep.26.2.189>
168. Efthimiadis, E. N. (2008). How do Greeks search the web?: a query log analysis study. In *iNEWS '08 Proceedings of the 2nd ACM workshop on Improving non english web searching* (pp. 81–84). Napa Valley, California, USA: ACM. <https://doi.org/10.1145/1460027.1460041>
169. Ellis, D. (1984). The effectiveness of information retrieval systems: the need for improved explanatory frameworks. *Social Science Information Studies*, 4(4), 261–272. [https://doi.org/10.1016/0143-6236\(84\)90002-4](https://doi.org/10.1016/0143-6236(84)90002-4)
170. Ellis, D. (1984). Theory and explanation in information retrieval research. *Information Scientist*, 8(1), 25–38. <https://doi.org/10.1177/016555158400800105>
171. ELLIS, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, 45(3), 171–212. <https://doi.org/10.1108/eb026843>
172. Ellis, D. (1989). A BEHAVIOURAL APPROACH TO INFORMATION RETRIEVAL SYSTEM DESIGN. *Journal of Documentation*, 45(3), 171–212. <https://doi.org/10.1108/eb026843>
173. Ellis, D., & Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53(4), 384–403. <https://doi.org/10.1108/EUM0000000007204>
174. Ellis, D., Wilson, T. D., Ford, N., Foster, A., Lam, H. M., Burton, R., & Spink, A. (2002). Information seeking and mediated searching. Part 5. User-intermediary interaction.

- Journal of the American Society for Information Science and Technology*, 53(11), 883–893.
<https://doi.org/10.1002/asi.10133>
175. Enge, E., Spencer, S. M., & Stricchiola, J. (2015). *The Art of SEO : Mastering Search Engine Optimization* (3rd ed.). O`Reilly Media.
 176. España, I. N. de E. de. (2017). CULTURABase. Retrieved from <http://www.mcu.es/culturabase/cgi/um?L=0>
 177. Expansión. (2017). España - Inmigración 2017 | datosmacro.com. Retrieved November 21, 2018, from <https://datosmacro.expansion.com/demografia/migracion/inmigracion/espana>
 178. Fagan, J. C. (2014). The suitability of web analytics key performance indicators in the academic library environment. *Journal of Academic Librarianship*, 40(1), 25–34.
<https://doi.org/10.1016/j.acalib.2013.06.005>
 179. Federación de Gremios de Editores de España. (2004). *Hábitos de lectura y compra de libros Año 2004*. Retrieved from federacioneditores.org/img/documentos/Habitos_LecturaYcompraLibros_2004.ppt
 180. Federación de Gremios de Editores de España. (2012). *Hábitos de lectura y compra de libros en España 2012*. Retrieved from <https://www.mecd.gob.es/dam/jcr:a5511a51-6ff8-4d36-a087-4859bf122495/habitos-lectura-2012.pdf>
 181. Federación de Gremios y Editores de España. (2017). *Barómetro de hábitos de Lectura y Compra de Libros en España 2017*. Retrieved from <http://federacioneditores.org/img/documentos/HabitosLecturaCompraLibros2017.pdf>
 182. Fernández Blanco, V. M., García Díez, M., & Prieto Rodríguez, J. (1999). Los hábitos de lectura en España: características sociales, educativas y ambientales. In *VI Encuentro de Economía Pública: el gasto social y su financiación: 4-6- de febrero de 1999, 1999, pág. 43* (p. 43). Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=3143456>
 183. Fernández Marcial, V. (2014). Segmentación y posicionamiento de los servicios de información. *Revista Academia y Virtualidad*, 7(2), 88–101. Retrieved from <http://eprints.rclis.org/28161/>
 184. Fernández Valdés, M. de las M., & Núñez Paula, I. A. (2007). Metodología para el estudio de las necesidades de información, conocimiento y aprendizaje en las bibliotecas y centros de documentación de salud. *ACIMED*, 15(3). Retrieved from <http://eprints.rclis.org/9506/>
 185. Fernández, E. (2015). Medición “user centric” vs. “site centric”: comScore MyMetrix | Analítica web. Retrieved May 30, 2017, from <http://www.analiticaweb.es/medicion-user-centric-vs-site-centric-comscore-mymetrix/>
 186. Fernandez-Blanco, V., Prieto-Rodriguez, J., & Suarez-Pandiello, J. (2015). *A quantitative analysis of reading habits* (ACEI working paper series). Retrieved from <http://www.culturaleconomics.org/awp/AWP-05-2015.pdf>
 187. Ferrada-Cubillos, M., & Jaña-Monsalve, C. (2007). Animación a la lectura weblog: Experiencias y reflexiones centradas en el libro y la lectura. *Bibliotecología y Gestión de Información*, 27, 1–69. Retrieved from <http://eprints.rclis.org/10145/>
 188. Ferrer I Cancho, R., & Sole, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B*, 268(1482). <https://doi.org/10.1098/rspb.2001.1800>
 189. Feuer, A., Savev, S., & Aslam, J. A. (2007). Evaluation of phrasal query suggestions. In *CIKM '07 Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 841–848). Lisbon, Portugal: ACM.
<https://doi.org/10.1145/1321440.1321556>
 190. Fidel, R. (1985). Moves in online searching. *Online Review*, 9(1), 61–74. Retrieved from <http://faculty.washington.edu/fidelr/RayaPubs/MovesinOnlineSearching.pdf>
 191. Fidel, R., & Efthimiadis, E. N. (1999). Web searching behavior of aerospace engineers (poster abstract). In *Proceedings of the 22nd annual international ACM SIGIR conference on*

- Research and development in information retrieval - SIGIR '99* (pp. 319–320). New York, New York, USA: ACM Press. <https://doi.org/10.1145/312624.312751>
192. Figueroa, A. (2015). Exploring effective features for recognizing the user intent behind web queries. *Computers in Industry*, 68, 162–169.
 193. Fisher, K. E., Erdelez, S., & McKechnie, L. (2005). *Theories of information behavior*. Published for the American Society for Information Science and Technology by Information Today. Retrieved from https://www.amazon.com/Theories-Information-Behavior-Asist-Monograph/dp/157387230X/ref=pd_rhf_dp_s_cp_2?_encoding=UTF8&pd_rd_i=157387230X&pd_rd_r=MVJ3RSXE3QEK3HDEY03C&pd_rd_w=N5Y8z&pd_rd_wg=Qc9s8&psc=1&refRID=MVJ3RSXE3QEK3HDEY03C
 194. Fishkin, R. (2017). The State of Searcher Behavior Revealed Through 23 Remarkable Statistics. Retrieved March 20, 2017, from <https://moz.com/blog/state-of-searcher-behavior-revealed>
 195. Fishkin, R. (2018). Why Google AdWords' Keyword Volume Numbers Are Wildly Unreliable - Whiteboard Friday - Moz. Retrieved October 2, 2018, from <https://moz.com/blog/unreliable-google-adwords-keyword-volume>
 196. Floria, A. (2000). Recopilación de métodos de usabilidad. Retrieved June 25, 2006, from <http://www.entrelinea.com/usabilidad/>
 197. Forman, C., Ghose, A., & Goldfarb, A. (2008). Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live. *MANAGEMENT SCIENCE*, 55(1), 47–57. <https://doi.org/10.1287/mnsc.1080.0932>
 198. Foss, E., & Druin, A. (2014). Children's Internet Search: Using Roles to Understand Children's Search Behavior. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 6(2), 1–106. <https://doi.org/10.2200/S00591ED1V01Y201408ICR034>
 199. Foss, E., Druin, A., Yip, J., Ford, W., Golub, E., & Hutchinson, H. (2013). Adolescent search roles. *Journal of the American Society for Information Science and Technology*, 64(1), 173–189.
 200. Fouces González, C. G. (2004). La comunicación persuasiva y el mercado del best seller de ficción narrativa en la sociedad digital - Monografias.com. In *II Congreso Online del Observatorio para la CiberSociedad*. Retrieved from <http://www.monografias.com/trabajos901/comunicacion-persuasiva-best-seller-cibersociedad/comunicacion-persuasiva-best-seller-cibersociedad.shtml>
 201. Franganillo, J. (2008). La industria editorial frente al libro electrónico. *El Profesional de La Información*, 17(4), 416–417. <https://doi.org/10.3145/epi.2008.jul.08>
 202. Fu, H. (2014). *Efficient, Effective, and Scalable Personalized Keyword Query Interpretation for RDF Databases*. ProQuest Dissertations and Theses. North Carolina State University. Retrieved from http://search.proquest.com/docview/1660170701?accountid=26642%5Cnhttp://link.periodicos.capes.gov.br/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Globa
 203. Fujii, A. (2008). Modeling anchor text and classifying queries to enhance web document retrieval. In *WWW '08 Proceedings of the 17th international conference on World Wide Web* (pp. 337–346). Beijing, China: ACM. <https://doi.org/10.1145/1367497.1367544>
 204. Future Today Institute. (2016). *2016 Trend Report. Emerging technology trends that will influence consumer behavior, our society and your organization's strategy in the coming year*. Retrieved from <http://futuretodayinstitute.com/wp-content/uploads/2015/11/WebbmediaGroup-2016-TechTrends.pdf>

205. Gabrilovich, E., Broder, A., Fontoura, M., Joshi, A., Josifovski, V., Riedel, L., & Zhang, T. (2009). Classifying search queries using the Web as a source of knowledge. *ACM Transactions on the Web (TWEB)*, 3(2), 1–28. <https://doi.org/10.1145/1513876.1513877>
206. Gajani, A. (2017). The key differences between SQL and NoSQL DBs. - Monitis Blog. Retrieved September 18, 2018, from <http://www.monitis.com/blog/cc-in-review-the-key-differences-between-sql-and-nosql-dbs/>
207. Gamboa, C., & Reina, M. (2006). *Hábitos de lectura y consumo de libros en Colombia*. Retrieved from <http://www.repository.fedesarrollo.org.co/handle/11445/803>
208. García Campos, J. M. (2015, July). ¿Cuáles son los escritores con más ingresos del mundo? *La Vanguardia*. Retrieved from <http://www.lavanguardia.com/vangdata/20150728/54434173747/escritores-ingresos-mundo.html>
209. García Canclini, N. (2015). *Hacia una antropología de los lectores*. Madrid. Retrieved from http://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/469/
210. García Ferrando, M. (1999). *Socioestadística: introducción a la Estadística en Sociología*. Alianza. Retrieved from <https://dialnet.unirioja.es/servlet/libro?codigo=293433>
211. Garfield, E. (1984). Current Comments number 7. *Essays of an Information Scientist*, 7, 45–49. Retrieved from <http://www.garfield.library.upenn.edu/essays/v7p045y1984.pdf>
212. Gaslikova, I. (1999). Information Seeking in Context and the development of information systems. *Information Research*, 5(1). Retrieved from <http://www.informationr.net/ir/5-1/paper67.html>
213. Gide, A. 1869-1951. (2016). *Dostoievski: artículos y charlas*. Ediciones del Subsuelo.
214. Gil Espin, M., & Jiménez, F. J. (2008). *El nuevo paradigma del sector del libro* (1ª). Trama. Retrieved from <http://www.tramaeditorial.es/libro/9788489239913/>
215. Gil Espin, M., & Rodríguez López, J. (2011). *El paradigma digital y sostenible del libro*. Trama Editorial. Retrieved from <http://www.tramaeditorial.es/libro/9788492755493/>
216. Gil, M. (2015). Nueva metodología de encuestas y hábitos de lectura – Antinomia libro. Retrieved May 10, 2017, from <https://antinomiaslibro.wordpress.com/2015/04/20/nueva-metodologia-de-encuestas-y-habitos-de-lectura/>
217. Giménez Toledo, E. (2017). *La Edición Académica Española. Indicadores y características*. Madrid: Federación de Gremios de Editores de España. Retrieved from <http://ilia.cchs.csic.es/SPI/spi-fgee/docs/EAEV2.pdf>
218. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014. <https://doi.org/10.1038/nature07634>
219. Giraldo-Giraldo, Y.-N., Álvarez-Zapata, D., Ocampo-Molina, N.-Y., Guerra-Sierra, L.-M., Melgar-Estrada, L., & Gómez-Vargas, M. (2008). La promoción de la lectura en las bibliotecas públicas de Medellín: informe de investigación. *Revista Interamericana de Bibliotecología*, 31(1), 161–205. Retrieved from <http://eprints.rclis.org/24916/>
220. Given, L. M., Julien, H., & Case, D. (2012). The evolution of information behavior research: Looking back to see the future. In *Proceedings of the American Society for Information Science and Technology* (Vol. 49, pp. 1–2). Wiley Subscription Services, Inc., A Wiley Company. <https://doi.org/10.1002/meet.14504901020>
221. Glänzel, W. (2004). *Bibliometrics as a research field. A course on theory and application of bibliometric indicators - Semantic Scholar*. Retrieved from <https://www.semanticscholar.org/paper/BIBLIOMETRICS-AS-A-RESEARCH-FIELD-A-course-on-theo-Glänzel/54add1634fe0518686cdd85bc87794d756f51867>
222. Godoy, D., & Amandi, A. (2006). Modeling user interests by conceptual clustering. *Information Systems*, 31(4–5), 247–265.

223. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), 17486–17490. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2955127&tool=pmcentrez&endertype=abstract>
224. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2009). What Can Search Predict? Retrieved from <http://jay.cam.cornell.edu/~sharad/papers/searchpreds.pdf>
225. Gómez, N.-D., & Colombo, F. (1997). Como llegar a nuestros usuarios en las puertas del siglo XXI? In *II Encuentro de Bibliotecas Universitarias* (pp. 14–18). Buenos Aires, Argentina. Retrieved from <http://eprints.rclis.org/9510/1/usuariosweb.pdf>
226. Gómez-Escalonilla, G. (2005). *El libro y la industria editorial*. Fundación Alternativas. Retrieved from <https://dialnet.unirioja.es/servlet/libro?codigo=616971>
227. Gómez-Hernández, J.-A. (1996). Los usuarios. In *Manual de biblioteconomía* (pp. 229–245). Síntesis. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=1005695>
228. Gómez-Tarragona, D. (2010). *Marketing editorial: cómo satisfacer las necesidades de los lectores de libros*. (Pirámide, Ed.) (1ª ed.). Madrid: Pirámide. Retrieved from http://www.edamel.com/wp-content/uploads/2016/04/1as_paginas_marketing_editorial.pdf
229. Gong, S., Xiong, J., Zhang, C., & Liu, Z. (2013). Identifying semantic-related search tasks in query log. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7808 LNCS, pp. 518–525).
230. Gong, X., Guo, X., Zhang, R., He, X., & Zhou, A. (2013). Search behavior based latent semantic user segmentation for advertising targeting. In *Proceedings - IEEE International Conference on Data Mining, ICDM* (pp. 211–220).
231. González Fernández-Villavicencio, N. (2009). Bibliotecas y marketing en red. *BiD : Textos Universitaris de Biblioteconomia i Documentació*, 23. Retrieved from <http://eprints.rclis.org/14143/>
232. González Fernández-Villavicencio, N. (2014). *La rentabilidad de la biblioteca en la web social*. Universidad de Salamanca. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=44590>
233. González Fernández-Villavicencio, N., Menéndez Novoa, J. L., Seoane García, C., & San Millán Fernández, M. E. (2013). Revisión y propuesta de indicadores (KPI) de la Biblioteca en los medios sociales. *Revista Española de Documentación Científica*, 36(1). <https://doi.org/10.3989/redc.2013.1.919>
234. González Teruel, A. (2005). *Los Estudios de necesidades y usos de la información: fundamentos y perspectivas actuales*. Trea. Retrieved from <http://vufind.uniovi.es/Record/1103695>
235. González Teruel, A., & Barrios Cerrejón, M. (2012). *Métodos y técnicas para la investigación del comportamiento informacional: fundamentos y nuevos desarrollos*. Ediciones Trea. Retrieved from <http://www.trea.es/books/metodos-y-tecnicas-para-la-investigacion-del-comportamiento-informacional-fundamentos-y-nuevos-desarrollos>
236. González Villa, J. (2016). El Futuro del SEO: ¿Cómo afectarán RankBrain y las búsquedas por voz? Retrieved March 1, 2017, from <http://useo.es/busquedas-voz-rank-brain/>
237. González, L. (2017). Hábitos lectores y políticas habituales de lectura. In Federación de Gremios de Editores de España (Ed.), *La lectura en España Informe 2017* (pp. 81–91). Madrid. Retrieved from <http://www.fge.es/lalectura/2017/default.html>
238. González-Caro, C. (2011, September 27). *Supervised Identification of the User Intent of Web Search Queries. TDX (Tesis Doctorals en Xarxa)*. Universitat Pompeu Fabra. Retrieved from <https://repositori.upf.edu/handle/10230/13029>

239. González-Caro, C., & Baeza-Yates, R. (2011). A Multi-faceted Approach to Query Intent Classification. In S. F. Grossi R., Sebastiani F. (Ed.), *SPIRE 2011: String Processing and Information Retrieval* (pp. 368–379). Berlin, Heidelberg: Springer. Retrieved from <https://pdfs.semanticscholar.org/347c/eff02e071826be3be6ab6162e693cd76cbf1.pdf>
240. González-Solar, L. (2016). *La biblioteca universitaria orientada a la investigación: propuesta de un modelo de servicio centrado en el usuario desde la perspectiva del marketing*. Retrieved from <http://eprints.rclis.org/30313/>
241. González-Teruel, A. (2011). Investigación sobre usuarios y realidad de la gestión de unidades de información. Interacciones posibles y necesarias. In *2º Seminário de Estudos da Informação*. Niteroi, RJ (Brasil). Retrieved from <http://eprints.rclis.org/18526/>
242. González-Teruel, A. (2015). Estrategias metodológicas para la investigación del usuario en los medios sociales: análisis de contenido, teoría fundamentada y análisis del discurso. *El Profesional de La Información*, 24(3), 321–328. Retrieved from <http://eprints.rclis.org/25273/>
243. González-Teruel, A. (2016). Beyond indicators and measures. Understanding the user's reality through a qualitative approach. In *The identity of the contemporary public library. Principles and methods of analysis, evaluation, interpretation*. (pp. 75–95). Ledizioni LediPublishing. Retrieved from <http://eprints.rclis.org/29573/>
244. González-Teruel, A. (2016). Referentes teóricos y dimensiones aplicadas en el estudio del usuario de la información. In: *Informação: agentes e intermediação*. Brasília: IBICT. *Informação: Agentes e Intermediação*. IBCT. Retrieved from <http://eprints.rclis.org/29365/>
245. González-Teruel, A., & Abad-García, M.-F. (2007). Information needs and uses: an analysis of the literature published in Spain, 1990–2004. *Library & Information Science Research*, 29(1), 30–46. Retrieved from <http://eprints.rclis.org/10766/>
246. González-Teruel, A., & et.al. (2015). Mapping recent information behavior research: an analysis of co-authorship and co-citation networks. *Scientometrics*, 103(2), 687–705. <https://doi.org/10.1007/s11192-015-1548-z>
247. Goodreads. (2013). Goodreads Blog Post: What Makes You Put Down a Book? Retrieved May 6, 2017, from <http://www.goodreads.com/blog/show/424-what-makes-you-put-down-a-book>
248. Google Ads Help. (2010). Keyword Planner: Search for new keywords - Google Ads Help. Retrieved November 21, 2018, from <https://support.google.com/google-ads/answer/6325025?hl=en>
249. Google Labs. (2011). Google Correlate. Retrieved May 13, 2017, from <https://www.google.com/trends/correlate/tutorial>
250. Google News Lab. (2016). The Rhythm of Food. Retrieved May 13, 2017, from <http://rhythm-of-food.net/>
251. Google News Lab. (2017). Visualizing Google data - Google Trends. Retrieved May 13, 2017, from https://trends.google.com/trends/story/US_cu_6fXtAFIBAABWdM_en
252. Google News Lab. (n.d.). The Year in Language. Retrieved May 13, 2017, from <https://googletrends.github.io/year-in-language/#>
253. Google Research. (2018). Research - Google AI. Retrieved October 1, 2018, from <https://ai.google/research>
254. Google Support. (2018). Cómo se ajustan los datos de Trends - Ayuda de Tendencias de búsqueda. Retrieved September 18, 2018, from https://support.google.com/trends/answer/4365533?hl=es&ref_topic=6248052
255. Google Trends. (2017). Google Trends. Retrieved September 18, 2018, from <https://trends.google.es/trends/?geo=ES>
256. Google. (2018). Google Keyword Planner. Retrieved from <https://adwords.google.com/KeywordPlanner#search>

- Universidad de Alicante. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=60723>
274. Hancock-Beaulieu, M. (1993). A comparative transaction log analysis of browsing and search formulation in online catalogues. *Program: Electronic Library and Information Systems*, 27(3), 269–280. <https://doi.org/10.1108/eb047145>
275. Hansen, P. (1998). Evaluation of IR User Interface. *Human IT*, 2. Retrieved from <http://etjanst.hb.se/bhs/ith//2-98/ph.htm>
276. Hardwick, J. (2018). How to Use Google Keyword Planner in 2018 (and Beyond). Retrieved October 19, 2018, from <https://ahrefs.com/blog/google-keyword-planner/>
277. Hartel, J., Fuller, S., Szostak, R., & Bonnici, L. (2012). The interdisciplinary study of information. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–3. <https://doi.org/10.1002/meet.14504901025>
278. He, Q.-C., Jiang, D., Liao, Z., Hoi, S. C. H., Chang, K., Lim, E.-P., & Li, H. (2009). Web Query Recommendation via Sequential Query Prediction. In *IEEE 25th International Conference on Data Engineering, 2009. ICDE '09*. (pp. 1443–1454). IEEE Computer Society. Retrieved from <http://portal.acm.org/citation.cfm?id=1546683.1547330>
279. Hearst, M. A. (2009). *Search user interfaces*. Cambridge University Press. Retrieved from <http://searchuserinterfaces.com/>
280. Herdagdelen, A., Ciaramita, M., Mahler, D., Holmqvist, M., Hall, K., Riezler, S., & Alfonseca, E. (2010). Generalized syntactic and semantic models of query reformulation. *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 283. Retrieved from <http://portal.acm.org/citation.cfm?doid=1835449.1835498>
281. Hernández Meléndrez, E. (2006). *Cómo escribir una tesis*. Retrieved from http://biblioteca.ucv.cl/site/servicios/documentos/como_escribir_tesis.pdf
282. Hernández Salazar, P., Nicholas, D., & Rowland, I. (2009). Acceso y uso de libros electrónicos por comunidades universitarias del Reino Unido. *Revista Interamericana de Bibliotecología*, 32(2), 13–58. Retrieved from <http://eprints.rclis.org/16982/>
283. Higuero, G., & Paredes, P. P. (2018). Ventas DC: Enero 2018 | Zona Negativa. Retrieved October 15, 2018, from <https://www.zonanegativa.com/ventas-dc-enero-2018/>
284. Hildreth, C. R. (1997). The use and understanding of keyword searching in a university online catalog. *Information Technology and Libraries*, 16(2), 52–63. Retrieved from <http://proquest.umi.com/pqdlink?did=16474829&Fmt=7&clientId=39490&RQT=309&VName=PQD>
285. Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. In *Computer Networks* (Vol. 33, pp. 337–346). Amsterdam, The Netherlands: North-Holland Publishing Co. [https://doi.org/10.1016/S1389-1286\(00\)00031-1](https://doi.org/10.1016/S1389-1286(00)00031-1)
286. Houghton-Jan, S. (2007). Twenty Steps to Marketing Your Library Online. *Journal of Web Librarianship*, 1(4), 81–90. <https://doi.org/10.1080/19322900802111445>
287. Hu, J., Wang, G., Lochovsky, F., Sun, J., & Chen, Z. (2009). Understanding user's query intent with wikipedia. In *WWW '09 Proceedings of the 18th international conference on World wide web* (pp. 471–480). Madrid, Spain: ACM. <https://doi.org/10.1145/1526709.1526773>
288. Hu, Y., & Et.al. (2012). Mining query subtopics from search log data. In *SIGIR '12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 305–314). Portland, Oregon, USA: ACM New York, NY, USA. <https://doi.org/10.1145/2348283.2348327>
289. Huang, C.-K., Chien, L.-F., & Oyang, Y.-J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638–649. <https://doi.org/10.1002/asi.10256>

290. Huang, J., Gao, J., Miao, J., Li, X., Wang, K., Behr, F., & Giles, C. L. (2010). Exploring web scale language models for search query processing. In *Proceedings of the 19th international conference on World wide web - WWW '10* (p. 451). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1772690.1772737>
291. Iberlibro. (n.d.). La historia de la primera edición de El Principito en español | Entre libros y letras. Retrieved October 15, 2018, from <https://www.iberlibro.com/blog/index.php/2016/06/16/la-historia-de-la-primera-edicion-de-el-principito-en-espanol/>
292. Igarza, R., & Monak Salinas, L. (2014). *Metodología común para explorar y medir el comportamiento lector: El encuentro con lo digital*. (José Diego González M., Ed.) (Primera ed). Bogotá, Colombia: CERLALC. Retrieved from http://www.lacult.unesco.org/docc/Metodologia_Comportamiento_Lector.pdf
293. IMDb. (2010). La saga Crepúsculo: Eclipse (2010) - IMDb. Retrieved October 15, 2018, from <https://www.imdb.com/title/tt1325004/>
294. INE. (2017). Población extranjera por país de nacionalidad, edad (grupos quinquenales) y sexo. Retrieved November 21, 2018, from <http://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p04/provi/l0/&file=00000008.px&L=0>
295. Ingwersen, P. (1992). *Information retrieval interaction*. Taylor Graham. Retrieved from <http://dl.acm.org/citation.cfm?id=130637>
296. Ingwersen, P. (1996). Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation*, 52(1), 3–50. <https://doi.org/10.1108/eb026960>
297. Instone, K. (1997). Site Usability Heuristics for the Web. Retrieved from [Webreview.com](http://www.webreview.com)
298. Irazú Hernández, D., Gupta, P., Rosso, P., & Rocha, M. (2012). A Simple Model for Classifying Web Queries by User Intent. In Rafael Berlanga & Paolo Rosso (Eds.), *Congreso español de recuperación de la información* (pp. 235–240). Castellón de la Plana: Universidad Jaime I. Retrieved from http://users.dsic.upv.es/~proso/resources/HernandezEtAl_CERI12.pdf
299. Islamaj Dogan, R., Murray, G. C., Névéal, A., & Lu, Z. (2009). Understanding PubMed user search behavior through log analysis. *Database*, bap018. <https://doi.org/10.1093/database/bap018>
300. Jamali, H. R., & Asadi, S. (2010). Google and the scholar: the role of Google in scientists' information-seeking behaviour. *Online Information Review*, 34(2), 282–294. <https://doi.org/10.1108/14684521011036990>
301. Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, 28(3), 407–432. <https://doi.org/10.1016/j.lisr.2006.06.005>
302. Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, 28(3), 407–432. <https://doi.org/10.1016/j.lisr.2006.06.005>
303. Jansen, B. J. (2009). *The Methodology of Search Log Analysis. Handbook of Research on Web Log Analysis*. <https://doi.org/10.4018/978-1-59904-974-8.ch006>
304. Jansen, B. J. (2009). *Understanding User-Web Interactions via Web Analytics*. (G. Marchionini, Ed.), *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 1). Morgan & Claypool Publishers. <https://doi.org/10.2200/S00191ED1V01Y200904ICR006>
305. Jansen, B. J., & Booth, D. L. (2010). Classifying web queries by topic and user intent. In *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/1753846.1754140>

306. Jansen, B. J., & Rieh, S. Y. (2010). The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and Technology*, 61(8), 1517–1534. <https://doi.org/10.1002/asi.21358>
307. Jansen, B. J., & Spink, A. (2000). Methodological approach in discovering user search patterns through Web log analysis. *Bulletin of the American Society for Information Science*, 27(1), 15–17. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/bult.185/full>
308. Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42(1 SPEC. ISS), 248–263. <https://doi.org/10.1016/j.ipm.2004.10.007>
309. Jansen, B. J., Booth, D. L., & Spink, A. (2007). Determining the user intent of web search engine queries. In *16th International World Wide Web Conference, WWW2007*. <https://doi.org/10.1145/1242572.1242739>
310. Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management*, 44(3), 1251–1266. <https://doi.org/10.1016/j.ipm.2007.07.015>
311. Jansen, B. J., Brown, A., & Resnick, M. L. (2007). Factors relating to the decision to click on a sponsored link. *Decision Support Systems*, 44(1), 46–59. <https://doi.org/10.1016/j.dss.2007.02.009>
312. Jansen, B. J., Campbell, G., & Gregg, M. (2010). Real time search user behavior. In E. Mynatt & D. Schoner (Eds.), *CHI EA '10 CHI '10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3961–3966). Atlanta, Georgia, USA: ACM New York, NY, USA. <https://doi.org/10.1145/1753846.1754086>
313. Jansen, B. J., Spink, A., & Taksa, I. (2008). *Handbook of research on web log analysis*. (B. J. Jansen, A. Spink, & I. Taksa, Eds.), *Handbook of Research on Web Log Analysis*. IGI Global. <https://doi.org/10.4018/978-1-59904-974-8>
314. Jansen, B. J., Zhang, M., Booth, D., Park, D., Zhang, Y., Kathuria, A., & Bonner, P. (2009). To What Degree Can Log Data Profile a Web Searcher? *Proceedings of the American Society for Information Science and Technology*, 46(1). Retrieved from https://faculty.ist.psu.edu/jjansen/academic/jansen_profiling_searcher_log_data.pdf
315. Játiva-Miralles, V. (2004). Indicadores de calidad aplicables al análisis, evaluación y comparación de opacs. *El Profesional de La Información*, 13(1), 28–46. Retrieved from <http://eprints.rclis.org/19499/>
316. Jethava, V., Calderón-Benavides, L., Baeza-Yates, R., Bhattacharyya, C., & Dubhashi, D. (2011). Scalable multi-dimensional user intent identification using tree structured distributions. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 395–404. Retrieved from <http://doi.acm.org/10.1145/2009916.2009971>
317. Jiang, J., He, D., & Allan, J. (2014). Searching, Browsing, and Clicking in a Search Session: Changes in User Behavior by Task and over Time. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 607–616). Gold Coast, Queensland, Australia: ACM New York, NY, USA. <https://doi.org/10.1145/2600428.2609633>
318. Jin, X., Zhou, Y., & Mobasher, B. (2004). Web usage mining based on probabilistic latent semantic analysis. In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04* (p. 197). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1014052.1014076>
319. Joachims, T. (2002). Optimizing search engines using clickthrough data. In *KDD '02 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 133–142). Edmonton, Alberta, Canada: ACM. <https://doi.org/10.1145/775047.775067>

320. Johnson, A. K., & Mehta, S. D. (2014). A Comparison of Internet Search Trends and Sexually Transmitted Infection Rates Using Google Trends. *Sexually Transmitted Diseases*, 41(1), 61–63. <https://doi.org/10.1097/OLQ.0000000000000065>
321. Joho, H., Jatowt, A., & Blanco, R. (2015). Temporal information searching behaviour and strategies. *Information Processing and Management*, 51(6), 834–850. <https://doi.org/10.1016/j.ipm.2015.03.006>
322. Jones, R., & Diaz, F. (2007). Temporal profiles of queries. *ACM Transactions on Information Systems*, 25(3), 14. <https://doi.org/10.1145/1247715.1247720>
323. Jones, R., Kumar, R., Pang, B., & Tomkins, A. (2007). “I know what you did last summer”: query logs and user privacy. In *CIKM '07 Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 909–914). Lisbon, Portugal: ACM. <https://doi.org/10.1145/1321440.1321573>
324. Jones, R., Zhang, W. V., Rey, B., Jhala, P., & Stipp, E. (2008). Geographic intention and modification in web search. *International Journal of Geographical Information Science*, 22(3), 229–246. <https://doi.org/10.1080/13658810701626186>
325. Jones, S., Cunningham, S. J., McNab, R., & Boddie, S. (2000). A transaction log analysis of a digital library. *International Journal on Digital Libraries*, 3(2), 152–169. <https://doi.org/10.1007/s007999900022>
326. Jorquera-Vidal, I. (2011). Aplicación de herramientas estadísticas gratuitas para sitios web a la evaluación de recursos y servicios virtuales de bibliotecas universitarias. In *II Jornada de Intercambio y Reflexión acerca de la Investigación en Bibliotecología*. La Plata, Argentina. Retrieved from <http://eprints.rclis.org/17190/>
327. Joshi, H., Ito, S., Kanala, S., Hebbar, S., & Bayrak, C. (2007). Concept set extraction with user session context. In *ACM-SE 45 Proceedings of the 45th annual southeast regional conference* (pp. 455–460). Winston-Salem, North Carolina: ACM New York, NY, USA. <https://doi.org/10.1145/1233341.1233423>
328. Juárez-Urquijo, F. (2008). La web es divertida y la utilizo como me apetece: una introducción pedestre a la web social. In *Congreso Nacional de Bibliotecas Públicas*. A Coruña: España. Ministerio de Cultura. Retrieved from <http://eprints.rclis.org/13128/>
329. Jun, S.-P., Park, D.-H., & Yeom, J. (2014). The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. *Technological Forecasting and Social Change*, 86, 237–253. <https://doi.org/10.1016/j.techfore.2013.10.021>
330. Jun, S.-P., Yeom, J., & Son, J.-K. (2014). A study of the method using search traffic to analyze new technology adoption. *Technological Forecasting and Social Change*, 81, 82–95. <https://doi.org/10.1016/j.techfore.2013.02.007>
331. Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130, 69–87. <https://doi.org/10.1016/J.TECHFORE.2017.11.009>
332. Kathuria, A., Jansen, B. J., Hafernik, C., & Spink, A. (2010). Classifying the user intent of User intent of web queries web queries using k-means clustering. *Internet Research*, 20(5), 563–581. <https://doi.org/10.1108/10662241011084112>
333. Kelly, D. (2006). Measuring online information seeking context, Part 1: Background and method. *Journal of the American Society for Information Science and Technology*, 57(13), 1729–1739. <https://doi.org/10.1002/asi.20483>
334. Kim, S., & Ahn, K. (2012). User-Intent Custom Adaptive Search Service based on Ontology and Similar-Word Retrieval Algorithm Composition. In *2012 INTERNATIONAL CONFERENCE FOR INTERNET TECHNOLOGY AND SECURED TRANSACTIONS* (pp. 121–126). IEEE.

335. Koch, T. (2017). Los editores alertan en un informe sobre el estancamiento de la lectura. *El País*. Retrieved from http://cultura.elpais.com/cultura/2017/01/06/actualidad/1483721992_502985.html
336. Koch, T., Ardö, A., & Golub, K. (2004). Log Analysis of User Behaviour in the Renardus Web Service. In *Proceedings of Libraries in the digital age 2004 (LIDA), Human Information Behaviour and Competences For Digital Libraries* (pp. 175–177). Lund: Lund University Library. Retrieved from <http://lup.lub.lu.se/record/606262>
337. Kong, W., & Allan, J. (2013). Extracting Query Facets from Search Results. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 93–102). Dublin, Ireland: ACM New York, NY, USA. <https://doi.org/10.1145/2484028.2484097>
338. Konkiel, S., & Madjarevic, N. (2016). Altmetrics for Librarians: 100+ tips, tricks, and examples. Retrieved from <https://ndownloader.figshare.com/files/6219537>
339. Korolova, A., Kenthapadi, K., Mishra, N., & Ntoulas, A. (2009). Releasing search queries and clicks privately. In *WWW '09 Proceedings of the 18th international conference on World wide web* (pp. 171–180). Madrid, Spain: ACM. <https://doi.org/10.1145/1526709.1526733>
340. Kotler, P., Armstrong, G., Moreno López, Y., García de Madariaga, J., & Flores Zamora, J. (2012). *Principios de marketing* (14a. ed.). México: Pearson Educación. Retrieved from https://profdariomarketing.files.wordpress.com/2014/03/marketing_kotler-armstrong.pdf
341. Kralisch, A., & Berendt, B. (2004). Cultural Determinants of Search Behaviour on Websites *. In *Proceedings of the IWIPS 2004 Conference on Culture, Trust, and Design Innovation*. Vancouver, Canada. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.8988&rep=rep1&type=pdf>
342. Kronqvist-Berg, M. (2014). *Social media and public libraries exploring information activities of library professionals and users*. Åbo Akademis University. Retrieved from <http://www.doria.fi/handle/10024/94661>
343. Krug, S. (2000). *Don't make me think! : a common sense approach to Web usability*. Que Corporation. Retrieved from https://www.amazon.es/Common-Approach-Usability-Circle-Com-Library/dp/0789723107/ref=pd_lpo_sbs_14_t_2/262-5371513-4773367?_encoding=UTF8&psc=1&refRID=VD7FQG62XJ7PNVNY53V4
344. Krug, S., & Rosenfeld, L. (2004). Krug and Rosenfeld on loc.gov: An Interactive Evaluation. Retrieved June 25, 2004, from <http://www.loc.gov/rr/program/lectures/sklr.html>
345. Kuhlthau, C. (1990). ERIC - Validating a Model of the Search Process: A Comparison of Academic, Public and School Library Users., *Library and Information Science Research*, 1990. *Library and Information Science Research*, 12(1), 5–31. Retrieved from <https://eric.ed.gov/?id=EJ413650>
346. Kuhlthau, C. C. (1988). Perceptions of the information search process in libraries: a study of changes from high school through college. *Information Processing & Management*, 24(4), 419–427. [https://doi.org/10.1016/0306-4573\(88\)90045-3](https://doi.org/10.1016/0306-4573(88)90045-3)
347. Kuhlthau, C. C. (1989). Information Search Process: A Summary of Research and Implications for School Library Media Programs. *SLMQ*, 18(1). Retrieved from http://www.ala.org/aasl/sites/ala.org.aasl/files/content/aaslpubsandjournals/slr/edchoice/SLMQ_InformationSearchProcess_InfoPower.pdf
348. Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE.* , 42(5), 361–371. Retrieved from <http://bogliolo.eci.ufmg.br/downloads/kuhlthau.pdf>
349. Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science and Technology*,

- 42(5), 361–371. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.2997&rep=rep1&type=pdf>
350. Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339–355. <https://doi.org/10.1108/eb026918>
351. Kuhlthau, C. C. (1999). Accommodating the User's Information Search Process: Challenges for Information Retrieval System Designers. *Bulletin of the American Society for Information Science*, 25(3). Retrieved from <https://www.asis.org/Bulletin/February/kuhlthau.html>
352. Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically Significant Detection of Linguistic Change. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15* (pp. 625–635). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2736277.2741627>
353. Kurland, O., Lewenstein, M., & Porat, E. (2007). String Processing and Information Retrieval. In S. Inenaga, K. Sadakane, & T. Sakai (Eds.), *Springer* (p. 428). Springer. Retrieved from <http://link.springer.com/content/pdf/10.1007/978-3-319-02432-5.pdf>
354. Larrañaga, E., & Yubero, S. (2005). El hábito lector como actitud. El origen de la categoría de “falsos lectores.” *Ocnos: Revista de Estudios Sobre Lectura*, 1, 43–60. Retrieved from <http://www.redalyc.org/pdf/2591/259120382004.pdf>
355. Lau, T., & Horvitz, E. (1999). Patterns of Search: Analyzing and Modeling Web Query Refinement. *Proceedings of the Seventh International Conference on User Modeling*, 119–128. Retrieved from <http://research.microsoft.com/en-us/um/people/horvitz/queryrefine.htm>
356. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>
357. Lee, U., Liu, Z., & Cho, J. (2005). Automatic identification of user goals in Web search. In *Proceedings of the 14th international conference on World Wide Web - WWW '05* (pp. 391–400). Chiba, Japan: ACM. <https://doi.org/10.1145/1060745.1060804>
358. Li, H., & Xu, J. (2014). Semantic Matching in Search. *Foundations and Trends® in Information Retrieval*, 7(5), 343–469. <https://doi.org/10.1561/15000000035>
359. Li, X., Wang, Y., & Acero, A. (2008). Learning query intent from regularized click graphs. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 339–346. Retrieved from <http://portal.acm.org/citation.cfm?doid=1390334.1390393>
360. Li, Y., Zheng, Z., & Dai, H. (Kathy). (2005). KDD CUP-2005 report. *ACM SIGKDD Explorations Newsletter*, 7(2), 91–99. <https://doi.org/10.1145/1117454.1117466>
361. Limam, L., Coquil, D., Kosch, H., & Brunie, L. (2010). Extracting user interests from search query logs: A clustering approach. In *Proceedings - 21st International Workshop on Database and Expert Systems Applications, DEXA 2010* (pp. 5–9).
362. Liu, Y.-Q., Cen, R.-W., Zhang, M., Ru, L.-Y., & Ma, S.-P. (2009). Automatic Search Engine Performance Evaluation Based on User Behavior Analysis. *Journal of Software*, 19(11), 3023–3032. <https://doi.org/10.3724/SP.J.1001.2008.03023>
363. Liu, Y.-Q., Zhang, M., Ma, S., & Ru, L. (2009). User browsing graph: Structure, evolution and application. In *WSDM'09: Proceedings of the Second ACM International Conference on Web Search and Data Mining*. WSDM. Retrieved from http://www.wsdm2009.org/liu_2009_user_browsing_graph.pdf
364. Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation*, 61(6), 700–712. <https://doi.org/10.1108/00220410510632040>
365. Lopes, M. I. (1995). Principles Underlying Subject Heading Languages: An International Approach. In *61st IFLA General Conference*. Istanbul, Turkey. Retrieved from <https://archive.ifla.org/IV/ifla61/61-lopm.htm>

366. Lucas, A. (2014). La historia muda de los analfabetos. *El País*. Retrieved from http://elpais.com/elpais/2014/09/08/planeta_futuro/1410171192_063867.html
367. Lurie, E., & Mustafaraj, E. (2018). Investigating the Effects of Google's Search Engine Result Page in Evaluating the Credibility of Online News Sources. In *Proceedings of the 10th ACM Conference on Web Science - WebSci '18* (pp. 107–116). New York, New York, USA: ACM Press. <https://doi.org/10.1145/3201064.3201095>
368. Macias-Alegre, A. (2016). ¿Tienen sentido las palabras clave de los artículos científicos? *COMeIN. Revista de Los Estudios de Ciencias de La Información y La Comunicación*, 55. Retrieved from <http://eprints.rclis.org/29337/>
369. Magadán-Díaz, M., & Rivas-García, J. (2018). *El profesional de la información information world en español. El Profesional de la Información* (Vol. 27). Swets & Zeitlinger. Retrieved from <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2018.nov.16/41619>
370. Mahoui, M., & Jo Cunningham, S. (2009). A Comparative Transaction Log Analysis of Two Computing Collections. In J. Borbinha & T. Baker (Eds.), *Research and Advanced Technology for Digital Libraries. ECDL 2000* (pp. 418–423). Springer Berlin / Heidelberg. Retrieved from http://dx.doi.org/10.1007/3-540-45268-0_53
371. Ma-Kellams, C., Or, F., Baek, J. H., & Kawachi, I. (2016). Rethinking Suicide Surveillance. *Clinical Psychological Science*, 4(3), 480–484. <https://doi.org/10.1177/2167702615593475>
372. Mangen, A., Walgermo, B. R., & Brønneick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61–68. <https://doi.org/10.1016/J.IJER.2012.12.002>
373. Mangano, V., Beaulieu, M., & Robertson, S. (1998). Evaluation of interfaces for IRS: modelling end-user searching behaviour. In Association for Computing Machinery (Ed.), *Proceedings of the 20th Annual BCS-IRSG conference on Information Retrieval Research* (pp. 7–7). Aufrans, France: BCS Learning & Development Ltd. Retrieved from <https://dl.acm.org/citation.cfm?id=2228005>
374. Maniega-Legarda, D. (2002). Aplicación de un estudio de usabilidad en bibliotecas digitales: la Biblioteca Virtual de la UOC. In *Contenidos y Aspectos Legales en la Sociedad de la Información (CALSI)*. Valencia (Spain): Universidad Politécnica de Valencia. Retrieved from <http://eprints.rclis.org/4297/>
375. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>
376. Manrique Sabogal, W. (2015). Dos librerías se cierran cada día en España | Cultura | EL PAÍS. Retrieved September 6, 2018, from https://elpais.com/cultura/2015/03/04/actualidad/1425453103_819705.html
377. Manrique Sabogal, W. (2015). Los editores piden una movilización social urgente para salvar el libro. *El País*. Retrieved from http://cultura.elpais.com/cultura/2015/04/22/actualidad/1429697404_328666.html
378. Manuel Carrión Gútiérrez. (1981). Hacia una definición del Bibliotecario en España. *Boletín de La ANABAD, ISSN 0210-4164, Tomo 31, N° 4, 1981, Págs. 565-578, 31(4), 565–578*. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=964897>
379. Marchal, S., Francois, J., State, R., & Engel, T. (2014). PhishStorm: Detecting Phishing With Streaming Analytics. *IEEE Transactions on Network and Service Management*, 11(4), 458–471. <https://doi.org/10.1109/TNSM.2014.2377295>
380. Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge University Press. Retrieved from <http://www.cambridge.org/catalogue/catalogue.asp?isbn=0521586747>
381. Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41. <https://doi.org/10.1145/1121949.1121979>

382. Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library and Information Science Research*, 15(1), 35–69. Retrieved from <https://uncch.pure.elsevier.com/en/publications/information-seeking-in-full-text-end-user-oriented-search-systems>
383. Marcos Blázquez, A. (2012). *Metodología para la elaboración de un plan de marketing: propuesta de aplicación de las herramientas de marketing digital en la biblioteca IE*. Universidad Complutense de Madrid. Retrieved from <http://eprints.ucm.es/20722/>
384. Marcos, J. R. (2010). Mario Vargas Llosa, Premio Nobel de Literatura | Cultura | EL PAÍS. Retrieved October 15, 2018, from https://elpais.com/cultura/2010/10/07/actualidad/1286402403_850215.html
385. Marcos, M.-C., & González-Caro, C. (2010). Comportamiento de los usuarios en la página de resultados de los buscadores. Un estudio basado en eye tracking. *El Profesional de La Información*, 19(4). <https://doi.org/10.3145/epi.2010.jul.03>
386. Marnier, J. C. (1993). Measuring the success of keyword search strategy in an online catalog. *Technical Services Quarterly*, 11, 1–11.
387. Maroto-Galán, E., Niño-Herranz, A., & Ruiz-Bautista, E. (2010). A cada lector su libro: las desideratas en la Biblioteca Pública de Cuenca. *Boletín de La ANABAD*, 6, 99–124. Retrieved from <http://eprints.rclis.org/18113/>
388. Martzoukou, K. (2005). A review of Web information seeking research: considerations of method and foci of interest. *Information Research*, 10(2). Retrieved from <http://www.informationr.net/ir/10-2/paper215.html>
389. Marvel. (2018). Marvel Movies | Marvel Cinematic Universe (MCU) | Marvel Studios Films. Retrieved October 15, 2018, from <https://www.marvel.com/movies>
390. Massísimo i Sánchez de Boado, À. (2002). Evaluación de colecciones en las bibliotecas universitarias (I) Métodos basados en el estudio de la colección. *Anales de Documentación*, 5, 245–272. Retrieved from <http://eprints.rclis.org/11993/>
391. Massísimo i Sánchez de Boado, À. (2004). Evaluación de colecciones en las bibliotecas universitarias (II). Métodos basados en el estudio de la colección. *Anales de Documentación*, 7, 171–183. Retrieved from http://biblioteca.universia.net/html_bura/ficha/params/title/evaluacion-colecciones-bibliotecas-universitarias-ii-metodos-basados-estudio-coleccion/id/34102323.html
392. Mavragani, A., & Tсарarakis, K. P. (2016). YES or NO: Predicting the 2015 GReferendum results using Google Trends. *Technological Forecasting and Social Change*, 109, 1–5. <https://doi.org/10.1016/j.techfore.2016.04.028>
393. Mccallum, M., & Gwendolynn, B. (2014). Public interest in the environment is falling: A response to Ficetola (2013). *Biodiversity and Conservation*, 23(4), 1057–1062. <https://doi.org/10.1007/s10531-014-0640-7>
394. Meyer, S. (2009). *Eclipse*. Alfaguara. Retrieved from <https://www.megustaleer.com/libros/eclipse-saga-crepsculo-3/MES-063720>
395. Miguel, S. E., Caprile, L., & Jorquera-Vidal, I. (2008). Análisis de co-términos y de redes sociales para la generación de mapas temáticos. *El Profesional de La Información*, 17(6), 637–646. Retrieved from <http://eprints.rclis.org/14265/>
396. Millán, J. A. (2002). *La lectura en España: informe 2002*. Federación de Gremios de Editores de España. Retrieved from <https://dialnet.unirioja.es/servlet/libro?codigo=4566>
397. Millán, J. A. (2008). *La lectura en España: informe 2008 leer para aprender*. Fundación Germán Sánchez Ruipérez. Retrieved from <http://www.lalectura.es/docs/informe2008.pdf>
398. Millán, J. A. (2017). *La lectura en España. Informe 2017*. Federación de Gremios de Editores de España. Retrieved from http://www.lalectura.es/docs/La_Lectura_en_Espana.pdf

399. Millsap, L., & Ferl, T. E. (1993). Search Patterns of Remote Users: An Analysis of OPAC Transaction Logs. *Information Technology and Libraries*, 12(3), 321–343. Retrieved from http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=EJ471113&ERICExtSearch_SearchType_0=no&accno=EJ471113
400. Mimi, Z., Jansen, B. J., & Spink, A. (2006). Information searching tactics of web searchers. In *Proceedings of the ASIST Annual Meeting* (Vol. 43, pp. 1–14). Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/meet.14504301105/full>
401. Ministerio de Cultura y Deporte. (2016). Estadística de Bibliotecas Públicas Españolas. Retrieved from <http://www.cultura.gob.es/alziraweb/alziraweb.cmd?command=GetEvolucionIndicador&id=8>
402. Ministerio de Cultura. (2011). Mapa de ferias del libro en España. Retrieved October 18, 2018, from <http://www.mcu.es/mapaapf/feriaMapa.do?cache=init&layout=mapaferia&language=es>
403. Ministerio de Educación Cultura y Deporte. (2015). *eBiblio 2015*. Retrieved from http://www.mecd.gob.es/dam/jcr:96341376-d855-4630-9097-c8ab84baedf4/eBiblio_2015_est.pdf
404. Ministerio de Educación Cultura y Deporte. (2016). *eBiblio 2016*. Retrieved from http://www.mecd.gob.es/dam/jcr:0c2a78a9-72c9-4a1c-864e-79c6e39b97e9/eBiblio_2016_est.pdf
405. Ministerio de Educación Cultura y Deporte. (2017). *eBiblio 2017*. Retrieved from http://www.mecd.gob.es/dam/jcr:18fd9845-7170-40c4-a27d-7f881ccffb52/eBiblio_2017_ok.pdf
406. Ministerio de Educación, & Cultura y Deporte, S. G. T. (2017). *Panorámica de la edición de libros 2016. Análisis sectorial del libro*. Madrid. Retrieved from <https://sede.educacion.gob.es/publiventa/detalle.action?cod=21061C>
407. Ministerio de Educación, C. y D. (2007). *Encuesta de hábitos y prácticas culturales 2006-2007*. Retrieved from http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/cultura/mc/ehc/2006-2007/tablas-de-resultados/0103_LECTURA_BIBLIOTECAS.pdf
408. Ministerio de Educación, C. y D. (2011). *Encuesta de hábitos y prácticas culturales 2010-2011*. Retrieved from http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/cultura/mc/ehc/2010-2011/tablas-de-resultados/c3_lectura_y_bibliotecas_2010-2011.pdf
409. Ministerio de Educación, C. y D. (2011). Lista de Encabezamientos de Materia. Retrieved April 4, 2017, from <http://id.sgcb.mcu.es/>
410. Ministerio de Educación, C. y D. (2014). Estadística de Bibliotecas Públicas Españolas (2010-2014). Retrieved May 16, 2017, from <http://www.cultura.gob.es/alziraweb/alziraweb.cmd?command=GetMapa&Cagrupar=N&codigo=&id=12&ejercicio=2014>
411. Ministerio de Educación, C. y D. (2015). *Encuesta de hábitos y prácticas culturales 2014-2015*. Retrieved from http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/cultura/mc/ehc/2014-2015/Encuesta_de_Habitos_y_Practicas_Culturales_2014-2015_Lectura_y_bibliotecas.pdf
412. Ministerio de Educación, C. y D. (2015). Encuesta de Hábitos y Prácticas Culturales en España 2014 – 2015. Retrieved February 14, 2017, from http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/cultura/mc/ehc/2014-2015/Encuesta_de_Habitos_y_Practicas_Culturales_2014-2015.pdf

413. Ministerio de Educación, C. y D. (2016). *El libro digital: aproximación a la realidad de las editoriales presentes en LIBER 2015*. Ministerio de Educación. Retrieved from <http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/libro/mc/observatoriolect/redirige/destacados/2016/Febrero/mundo-libro/LibroDigital-LIBER2015/Informe-sobre-el-Libro-Digital-en-LIBER-2015/Informe sobre el Libro Digital en LIBER 201>
414. Ministerio de Educación, C. y D., & Fundación Autor. (2003). *Encuesta de hábitos y prácticas culturales 2002-2003. 3. LECTURA Y ASISTENCIA A BIBLIOTECAS*. Retrieved from <https://www.mecd.gob.es/servicios-al-ciudadano/estadisticas/cultura/mc/ehc/2002-2003/presentacion.html>
415. Ministerio de Educación, Cultura y Deporte, O. de la L. y el L. (2016). *Panorámica de la edición española de libros 2015 Análisis sectorial del libro*. Madrid. Retrieved from <http://transparencia.gob.es/transparencia/dam/jcr:acb408ef-1060-4702-a154-496d0c1845a7/Panoramica-edicion-libros-2015.pdf>
416. Ministerio de Educación, Cultura y Deporte, S. G. de E. y E. (Secretaría G. T. (2016). *Anuario de Estadísticas Culturales 2016*. Retrieved from http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/cultura/mc/naec/2016/Anuario_de_Estadisticas_Culturales_2016.pdf
417. Mishra, N., Saha Roy, R., Ganguly, N., Laxman, S., & Choudhury, M. (2011). Unsupervised query segmentation using only query logs. In *Proceedings of the 20th international conference companion on World wide web - WWW '11* (p. 91). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1963192.1963239>
418. Moulahi, B., Tamine-lechani, L., & Yahia, S. Ben. (2016). When time meets information retrieval: Past proposals, current plans and future trends. *Journal of Information Science*, 42(6), 725–747. <https://doi.org/10.1177/0165551515607277>
419. Moz.com. (2018). Moz - What are SERP features? Retrieved October 2, 2018, from <https://moz.com/learn/seo/serp-features>
420. Mukherjee, P., & Jansen, B. J. (2014). Performance analysis of keyword advertising campaign using gender-brand effect of search queries. *Electronic Commerce Research and Applications*, 13(2), 139–149. <https://doi.org/10.1016/j.elerap.2014.01.001>
421. Mukherjee, P., & Jansen, B. J. (2016). Correlation of Brand Mentions in Social Media and Web Searching before and after Real Life Events: Phase Analysis of Social Media and Search Data for Super Bowl 2015 Commercials. In *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* (pp. 21–26). <https://doi.org/10.1109/ICDMW.2015.60>
422. Muramatsu, J., & Pratt, W. (2001). Transparent Queries: investigation users' mental models of search engines. In *SIGIR '01 Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 217–224). New Orleans, Louisiana, United States: ACM New York, NY, USA. <https://doi.org/10.1145/383952.383991>
423. Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, 130, 114–122. <https://doi.org/10.1016/j.techfore.2017.11.022>
424. Navarro, B. (2016). La sostenibilidad de la creación literaria. *TELOS - Fundación Telefónica*, (104). Retrieved from https://telos.fundaciontelefonica.com/seccion=1268&idioma=es_ES&id=2016071314520001&activo=6.do
425. Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in Web Searching. In *Conference on Human Factors & the Web*. Retrieved from <https://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/>

426. Nettleton, D., Calderón-Benavides, L., & Baeza-Yates, R. (2006). Analysis of Web Search Engine Query Session and Clicked Documents. In *WebKDD'06 Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis* (pp. 207–226). Philadelphia, USA. Retrieved from <http://www.springerlink.com/content/28314348q5852n51/>
427. Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of Documentation*, 61(2), 248–280. <https://doi.org/10.1108/00220410510585214>
428. Nielsen, J. (1999). Reputation Managers are Happening. Retrieved September 20, 2018, from <https://www.nngroup.com/articles/reputation-managers-are-happening/>
429. Nielsen, J. (2006). F-Shaped Pattern For Reading Web Content. Retrieved June 8, 2017, from <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>
430. Nirenburg, S., & Raskin, V. (2004). *Ontological semantics*. MIT Press.
431. Norman, D. (2017). Don't Drop The Patient! Examine the Entire System (Don Norman) - YouTube. Retrieved October 19, 2018, from <https://www.youtube.com/watch?v=xfOpyGqgx8>
432. Núñez Paula, I. A. (1992). Guía metodológica para el estudio de las necesidades de formación e información de los usuarios o lectores. *Ciencias de La Información*, 23(2), 118–130. Retrieved from <http://eprints.rclis.org/5311/>
433. Núñez Paula, I. A. (2004). AMIGA: una metodología integral para la determinación y la satisfacción dinámica de las necesidades de formación e información en las organizaciones y comunidades. *ACIMED*, 12(4). Retrieved from http://bvs.sld.cu/revistas/aci/vol12_4_04/aci02404.htm
434. Núñez Paula, I. A., & Zayas Caballero, I. (2016). Análisis de modelos sobre Comportamiento Informacional, desde un enfoque socio-psicológico. *Bibliotecas. Anales de Investigación*, 12(1). Retrieved from <http://revistas.bnjm.cu/index.php/anales/article/view/3637>
435. Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The Use of Google Trends in Health Care Research: A Systematic Review. *PLoS ONE*, 9(10), e109583. <https://doi.org/10.1371/journal.pone.0109583>
436. Nyman, R., Ormerod, P., & Tuckett, D. (2015). Measuring Financial Sentiment to Predict Financial Instability: A New Approach based on Text Analysis. Retrieved March 27, 2017, from <https://arxiv.org/ftp/arxiv/papers/1508/1508.05357.pdf>
437. Oddy, R. N., Belkin, N. J., & Brooks, H. M. (1982). ASK FOR INFORMATION RETRIEVAL: PART I. BACKGROUND AND THEORY. *Journal of Documentation*, 38(2), 61–71. <https://doi.org/http://dx.doi.org/10.1108/eb026722>
438. Olson, J., & et. al. (2014). *Ways of Knowing in HCI*. (J. S. Olson & W. A. Kellogg, Eds.). New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4939-0378-8>
439. Olvera-Lobo, M.-D., & Robinson-García, N. (2009). Tratamiento lingüístico de las preguntas en español en los sistemas de búsqueda de respuestas. *El Profesional de La Información*, 18(2), 180–187. <https://doi.org/10.3145/epi.2009.mar.08>
440. Orduña, E., & Aguillo, I. F. (2014). *Cibernetría. Midiendo el espacio red*. Barcelona: Editorial UOC - Editorial de la Universitat Oberta de Catalunya. Retrieved from http://www.editorialuoc.cat/cibernetria-midiendo-el-espacio-red_1
441. Orduña-Malea, E., López-Cózar, E. D., Serrano-Cobos, J., & Romero, N. L. (2015). Disclosing the network structure of private companies on the web: The case of Spanish IBEX 35 share index. *Online Information Review*, 39(3). <https://doi.org/10.1108/OIR-11-2014-0282>
442. Orduña-Malea, E., Ontalba-Ruipérez, J.-A., & Serrano-Cobos, J. (2010). Bibliometric analysis of scientific production and collaboration in the middle east from 1998 to 2007 |

- Análisis bibliométrico de la producción y colaboración científica en oriente próximo (1998-2007). *Investigacion Bibliotecologica*, 24(51).
443. Orduña-Malea, E., Serrano-Cobos, J., & Lloret-Romero, N. (2009). Las universidades públicas españolas en Google Scholar: Presencia y evolución de su publicación académica web. *El Profesional de La Información*, 18(5). <https://doi.org/10.3145/epi.2009.sep.02>
444. Orduña-Malea, E., Serrano-Cobos, J., Ontalba-Ruipérez, J.-A., & Lloret-Romero, N. (2010). Web presence and visibility of Spanish public universities | Presencia y visibilidad web de las universidades públicas españolas. *Revista Española de Documentación Científica*, 33(2). <https://doi.org/10.3989/redc.2010.2.740>
445. Ormerod, P., Nyman, R., Smith, R. E., & Tuckett, D. (2013). Socio-Psychological Theory, Algorithmic Text Analysis, and Predicting the Michigan Consumer Sentiment Index. Retrieved from <https://pdfs.semanticscholar.org/18a4/0377ba8ab09633d7ca9eb1a076bac520db9f.pdf>
446. Ortega Blake, J. A. (2013). *El gran libro de las frases célebres*. Penguin Random House. Retrieved from https://books.google.es/books?id=QJIAVIKP1dgC&pg=PT460&lpg=PT460&dq=John+Steinbeck,+a+que+“por+el+grosor+del+polvo+en+los+libros+de+una+biblioteca+pública&source=bl&ots=ri_QwfxnuT&sig=otvoD0CUqeAyXZV1_QRnbJxrjms&hl=es&sa=X&ved=0ahUKEwjxteDM5t
447. Palii, I. (2017). How come the AdWords keyword planner is different with Google trends? Retrieved September 18, 2018, from <https://www.quora.com/How-come-the-AdWords-keyword-planner-is-different-with-Google-trends>
448. Pan, B., Litvin, S. W., & O'Donnell, T. (2007). Accommodation Search Query Formulation: Implications for Search Engine Marketing. *Journal of Vacation Marketing*, 13(4), 371–381. Retrieved from <http://www.ota.cofc.edu/pan/Excite-engines2.pdf>
449. Parikh, N., Sriram, P., Al Hasan, M., Parikh, N., Sriram, P., & Al Hasan, M. (2013). On segmentation of eCommerce queries. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13* (pp. 1137–1146). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2505515.2505721>
450. Park, K., Jee, H., Lee, T., Jung, S., & Lim, H. (2012). Automatic extraction of user's search intention from web search logs. *Multimedia Tools and Applications*, 61(1), 145–162. <https://doi.org/10.1007/s11042-010-0723-8>
451. Park, S. (2009). Analysis of characteristics and trends of Web queries submitted to NAVER, a major Korean search engine. *Library & Information Science Research*, 31(2), 126–133. <https://doi.org/10.1016/j.lisr.2009.01.003>
452. Paşca, M., Alfonseca, E., Robledo-Arnuncio, E., Martin-Brualla, R., & Hall, K. (2010). The role of query sessions in extracting instance attributes from web search queries. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5993 LNCS, pp. 62–74).
453. Peacock, D. (2002). *Statistics, Structures & Satisfied Customers: Using Web Log Data to Improve Site Performance*. Boston, Massachusetts, United States: For full text: <http://www.archimuse.com/mw2002/papers/peacock/peacock.html/>. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED482104>
454. Pérez López, A. (2002). *El desarrollo de las colecciones: estudio evaluativo en bibliotecas universitarias españolas*. Universidad de Granada. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=12063>
455. Perez, C. (1994). Mercado editorial y hábitos de lectura. *Distribución y Consumo*, 136–146. Retrieved from http://mercasa.es/files/multimedios/1307430672_DYC_1994_17_136_146.pdf

456. Pernas Alonso, J. (2017). La búsqueda semántica y como hacer SEO en la era de Hummingbird. Retrieved March 1, 2017, from <https://es.semrush.com/blog/busqueda-semantica/>
457. Perrin, A. (2016). ¿Quiénes no leen libros? | Universo Abierto. Retrieved February 14, 2017, from <https://universoabierto.org/2016/11/23/quienes-no-leen-libros/>
458. Pervaiz, F., Pervaiz, M., Abdur Rehman, N., & Saif, U. (2012). FluBreaks: early epidemic detection from Google flu trends. *Journal of Medical Internet Research*, *14*(5), e125. <https://doi.org/10.2196/jmir.2102>
459. Petersen, A. M., Tenenbaum, J., Havlin, S., & Stanley, H. E. (2011). Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death. *Nature*. <https://doi.org/10.1038/srep00313>
460. Pharo, N., & Järvelin, K. (2004). The SST method: a tool for analysing Web information search processes. *Information Processing & Management*, *40*(4), 633–654. <https://doi.org/10.1016/j.ipm.2003.08.005>
461. Pinter, Y., Reichart, R., & Szpektor, I. (2016). Syntactic Analysis of Web Queries with Question Intent. In *The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, CA, USA. Retrieved from http://www.openu.ac.il/iscol2015/downloads/ISCOL2015_submission13_c_6.pdf
462. Pinto, F. J. (2008). *Problemática sobre el impacto de la expansión automática de consultas y desambiguación semántica en la evaluación de los sistemas de recuperación de información*. Universidade da Coruña. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=19455>
463. Pirolli, P., & Card, S. (1995). Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95* (pp. 51–58). New York, New York, USA: ACM Press. <https://doi.org/10.1145/223904.223911>
464. Pirolli, P., Fu, W.-T., Chi, E., & Farahat, A. (2005). Information Scent and Web Navigation: Theory, Models, and Automated Usability Evaluation. In Mahwah NJ (Ed.), *Proceedings of Human Computer Interaction International 2005*. Las Vegas; NV; USA.: Erlbaum. Retrieved from <https://pdfs.semanticscholar.org/5a23/2dc72dae072b852c1a93df971368d3b7c2ac.pdf>
465. Piwowarski, B., Dupret, G., & Jones, R. (2009). Mining user web search activity with layered bayesian networks or how to capture a click in its context. In *WSDM '09 Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 162–171). Barcelona, Spain: ACM. <https://doi.org/10.1145/1498759.1498823>
466. Pouliot Madero, M. C. del C. (2009). *Análisis metodológico de sondeos de opinión sobre prácticas de lectura de libros impresos y otros soportes del libro: los indicadores de lectura en España (2000-2007)*. Retrieved from <http://gredos.usal.es/jspui/handle/10366/76352>
467. Preetha, S., & Vimal Shankar, K. N. (2015). Personalized search engines on mining user preferences using clickthrough data. In *2014 International Conference on Information Communication and Embedded Systems, ICICES 2014*. Institute of Electrical and Electronics Engineers Inc.
468. Preis, T., Moat, H. S., Stanley, H. E., Bishop, S. R., & Havlin, S. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, *3*, 81–89. <https://doi.org/10.1038/srep01684>
469. Puente Lanzarote, L., Campo Hilario, C. del, & Ruiz de Luzuriaga Peña, M. (2001). Indicadores de rendimiento para la evaluación de un servicio de bases de datos en línea. *Scire*, *7*(1), 89–114. Retrieved from <http://eprints.rclis.org/14991/>

470. Quezada, J. P. A. (2016). *Hábitos de lectura en los mexicanos: un comparativo de tres encuestas nacionales*. Retrieved from <http://www.bibliodigitalibd.senado.gob.mx/handle/123456789/2112>
471. Radlinski, F., Szummer, M., & Craswell, N. (2010). Inferring query intent from reformulations and clicks. *Proceedings of the 19th International Conference on World Wide Web*, 1171–1172. Retrieved from <http://dl.acm.org/citation.cfm?id=1772859>
472. Rainie, L. (2012). The rise of e-reading. Retrieved May 5, 2017, from <http://www.pewinternet.org/2012/04/05/the-rise-of-e-reading-3/>
473. Ramírez Leyva, E. M. (2008). La poca lectura de libros: Una trayectoria con dos sentidos. *Investigacion Bibliotecologica*. Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2008000100003
474. Real Academia de la Lengua. (2016). En 2016 se conmemora el IV Centenario de la muerte de Cervantes | Real Academia Española. Retrieved October 15, 2018, from <http://www.rae.es/noticias/en-2016-se-conmemora-el-iv-centenario-de-la-muerte-de-cervantes>
475. Real Academia de la Lengua. (2017). Definición de libro -. Retrieved May 10, 2017, from <http://dle.rae.es/?id=NG3kctc6>
476. Real Academia Española de la Lengua. (n.d.). Qué es el Diccionario panhispánico de dudas de la Real Academia Española. Retrieved April 4, 2017, from <http://www.rae.es/diccionario-panhispanico-de-dudas/que-es>
477. Real Academia Española. (2005). Diccionario panhispánico de dudas. 1.^a edición, 2.^a tirada. Retrieved October 17, 2018, from <http://buscon.rae.es/dpd/srv/search?id=nypwukqZ6D64M23Yos>
478. Real Academia Española. (2018). Cómo se hace el Diccionario | Real Academia Española. Retrieved October 16, 2018, from <http://www.rae.es/diccionario-de-la-lengua-espanola/como-se-hace-el-diccionario>
479. Red de Universidades Lectoras. (2014). Diccionario Digital de Nuevas Formas de Lectura y Escritura. Retrieved September 19, 2018, from <http://dinle.usal.es/searchword.php?valor=Hábito lector>
480. Redacción de La Vanguardia. (2015). El negocio del libro en España, en cifras. *La Vanguardia*. Retrieved from <http://www.lavanguardia.com/vangdata/20150422/54430767844/negocio-libro-espana-cifras.html>
481. Reisinger, J., & Pasca, M. (2007). Fine-Grained Class Label Markup of Search Queries. In *Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011)* (pp. 1200–1209). Association for Computational Linguistics. Retrieved from <http://www.cs.utexas.edu/users/ai-lab/?reisinger:aclhlt11>
482. Ren, P., Chen, Z., Ma, J., Wang, S., Zhang, Z., & Ren, Z. (2015). Mining and ranking users' intents behind queries. *Information Retrieval*, 18(6), 504–529. <https://doi.org/10.1007/s10791-015-9271-1>
483. Research, M. (2018). Web and AI Sciences. Retrieved from <https://www.microsoft.com/en-us/research/group/web-ai-sciences/>
484. Reyes, T., Majluf, N., & Ibáñez, R. (2018). Using Internet Search Data to Measure Changes in Social Perceptions: A Methodology and an Application*. *Social Science Quarterly*, 99(2), 829–845. <https://doi.org/10.1111/ssqu.12449>
485. Rice, R. E., Mccreadie, M., Chang, S.-J. L., & Shaw, D. (2001). *Accessing and Browsing Information and Communication*. Cambridge, MA: MIT Press. <https://doi.org/10.1080/01972240290108221>
486. Richardson, M. (2008). Learning about the world through long-term query logs. *ACM Transactions on the Web (TWEB)*, 2(4), 1–27. <https://doi.org/10.1145/1409220.1409224>

487. Rodríguez, A. G., Gómez-Díaz, R., & Cordon-García, J.-A. (2007). De Alicia en el país de las maravillas a las maravillas de Alicia: o de cómo cambia la selección de libros en papel a libros electrónicos infantiles y juveniles. *Ibersid: Revista de Sistemas de Información y Documentación*, 7(0), 21–32. Retrieved from <http://www.iversid.eu/ojs/index.php/iversid/article/view/4092>
488. Rodríguez, M. L. (2010). El desarrollo de una encuesta: aspectos metodológicos y procedimentales. Retrieved April 27, 2017, from <https://aulasvirtuales.wordpress.com/2010/04/18/el-desarrollo-de-una-encuesta-aspectos-metodologicos-y-procedimentales/>
489. Rodríguez, T., Aguilar, J., Rodríguez, T., & Aguilar, J. (2017). Implementación del Marco Ontológico Dinámico Semántico. *Ingeniare. Revista Chilena de Ingeniería*, 25(3), 430–448. <https://doi.org/10.4067/S0718-33052017000300430>
490. Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th conference on World Wide Web - WWW '04* (p. 13). New York, New York, USA: ACM Press. <https://doi.org/10.1145/988672.988675>
491. Rosenfeld, L. (2011). *Search analytics for your site: conversations with your customers*. Rosenfeld Media. Retrieved from <http://rosenfeldmedia.com/books/search-analytics-for-your-site/>
492. Rosenfeld, L., & Morville, P. (1998). *Information architecture for the World Wide Web*. Cambridge: O'Reilly. Retrieved from <http://www.worldcat.org/title/information-architecture-for-the-world-wide-web/oclc/470572261?referer=di&ht=edition>
493. Rushton, E. E., Kelehan, M. D., & Strong, M. A. (2008). Searching for a New Way to Reach Patrons: A Search Engine Optimization Pilot Project at Binghamton University Libraries. *Journal of Web Librarianship*, 2(4), 525–547. <https://doi.org/10.1080/19322900802484248>
494. Ruthven, I., Lalmas, M., & van Rijsbergen, K. (2003). Incorporating user search behavior into relevance feedback. *Journal of the American Society for Information Science and Technology*, 54(6), 529–549. <https://doi.org/10.1002/asi.10240>
495. Saha Roy, R., & Rishiraj. (2013). Analyzing linguistic structure of web search queries. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 395–400). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2487788.2487949>
496. Saha Roy, R., Ganguly, M., Choudhury, N., & Singh, N. K. (2011). Complex network analysis reveals kernel-periphery structure in web search queries. In *Query Representation and Understanding Workshop 2011 (QRU '11) ACM SIGIR 2011*. Beijing, China. Retrieved from http://people.mpi-inf.mpg.de/~rsaharo/qru11slides_rsrngmcnks.pdf
497. Saint-Exupéry, A. de, & Carril, B. del. (2009). *El principito: con las acuarelas del autor*. Retrieved from <http://salamandra.info/libro/principito-tapa-dura>
498. Sáinz Peña, R. M. (2016). TELOS - Fundación Telefónica - Detalle Anteriores - Radiografía de la actividad digital de la sociedad española. Retrieved February 14, 2017, from https://telos.fundaciontelefonica.com/seccion=1268&idioma=es_ES&id=201607131458003&activo=6.do
499. Salazar Ayllón, S., & Ponce, D. (1999). Hábitos de lectura. *Biblios*, 2. Retrieved from <http://eprints.rclis.org/5727/>
500. Salton, G. (1968). *Automatic information organization and retrieval*. McGraw-Hill. Retrieved from <http://dl.acm.org/citation.cfm?id=1096906>
501. Sánchez Ramos, B. (2016). ¿No sabes qué libro leer? Prueba a buscar aquí. *El País*. Retrieved from http://economia.elpais.com/economia/2016/12/01/el_lazarillo_digital/1480619233_221443.html

502. Sánchez, A. (2015). Antena 3 estrenará The Flash en España el lunes 20 de julio | Todas las noticias de DC Comics | Espacio DC. Retrieved October 15, 2018, from <http://www.espaciopc.com/2015/07/17/television/antena-3-estrenara-the-flash-en-espana-el-lunes-20-de-julio/>
503. Santiago Méndez, G. (2011). *Nuevas estrategias de la prensa para la captación de lectores*. Universidad de Sevilla. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=43947>
504. Sanz Casado, E., & Gómez Caridad, I. (1994). *Manual de estudios de usuarios*. Fundación Germán Sánchez Ruipérez. Retrieved from <https://www.casadellibro.com/libro-manual-de-estudios-de-usuarios/9788486168933/476911>
505. Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3), 177–196. [https://doi.org/10.1002/\(SICI\)1097-4571\(198805\)39:3<177::AID-ASI3>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-4571(198805)39:3<177::AID-ASI3>3.0.CO;2-F)
506. Saracevic, T., & Kantor, P. (1988). A Study of Information Seeking and Retrieving. III. Searchers, Searches, and Overlap. *Journal of the American Society for Information Science*, 39(3), 197. Retrieved from <http://search.proquest.com/openview/8b007260407ba8488a4ba3e6c2c4bb2a/1?pq-origsite=gscholar&cbl=1818555>
507. Schavelzon, G. (2015). Vender más libros no significa crear lectores. Retrieved May 5, 2017, from https://elblogdeguillermoschavelzon.wordpress.com/2015/05/15/vender-mas-libros-no-significa-crear-lectores/?_ga=2.18950264.1141006301.1493999144-449257544.1487093690
508. Schoefegger, K., Tammet, T., & Granitzer, M. (2013). A survey on socio-semantic information retrieval. *COMPUTER SCIENCE REVIEW*, (8), 25–46. <https://doi.org/10.1016/j.cosrev.2013.03.001>
509. Schöfegger, K., Körner, C., Singer, P., & Granitzer, M. (2012). Learning user characteristics from social tagging behavior. In *Proceedings of the 23rd ACM conference on Hypertext and social media - HT '12* (p. 207). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2309996.2310031>
510. Schootman, M., Toor, A., Cavazos-Rehg, P., Jeffe, D. B., McQueen, A., Eberth, J., & Davidson, N. O. (2015). The utility of Google Trends data to examine interest in cancer screening. *BMJ Open*, 5(6), e006678. <https://doi.org/10.1136/bmjopen-2014-006678>
511. Selgas Gutiérrez, J. (2004). Las estadísticas de bibliotecas públicas en España: pasado, presente y perspectivas de futuro. In S. G. de I. y P. Ministerio de Cultura (Ed.), *Congreso Nacional de Bibliotecas Públicas* (pp. 172–181). Salamanca: Ministerio de Cultura. Subdirección General de Información y Publicaciones. Retrieved from <http://travesia.mcu.es/portaln/jspui/handle/10421/768>
512. Sellés-Carot, A., & Serrano-Cobos, J. (2011). From extended opac to expanded library: Finally a reality | Del opac extendido a la biblioteca expandida: Al fin una realidad. *El Profesional de La Información*, 20(4). <https://doi.org/10.3145/epi.2011.jul.15>
513. Sellés-Carot, A., Orduña-Malea, E., Serrano-Cobos, J., & Romero, N. L. (2010). Proposal of a goal-oriented shared catalog model. In *Proceedings - 2010 IEEE 4th International Conference on Semantic Computing, ICSC 2010*. Pittsburgh, USA. <https://doi.org/10.1109/ICSC.2010.43>
514. Serrano-Cobos, J. (2003). El álgebra booleana de google. Retrieved September 26, 2018, from <http://trucosdegoogle.blogspot.com/2003/02/#89845355%2389845355>
515. Serrano-Cobos, J. (2009). Combinación de logs internos y externos en la predicción de estacionalidad de búsquedas para el rediseño de webs. *El Profesional de La Información*, 18(1). <https://doi.org/10.3145/epi.2009.ene.02>

516. Serrano-Cobos, J. (2010). The future of information and libraries. Interview with Eric Lease Morgan. *El Profesional de La Información*, 19(2).
<https://doi.org/10.3145/epi.2010.mar.14>
517. Serrano-Cobos, J. (2014). Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos. *El Profesional de La Información*, 23(6), 561–566.
<https://doi.org/10.3145/epi.2014.nov.01>
518. Serrano-Cobos, J. (2015). *SEO: introducción a la disciplina del posicionamiento en buscadores*. UOC. Retrieved from <https://www.casadellibro.com/libro-seo-introduccion-a-la-disciplina-del-posicionamiento-en-buscadores/9788490649565/2738701>
519. Serrano-Cobos, J. (2016). Internet technology trends: Towards a paradigm shift. *Profesional de La Informacion*, 25(6). <https://doi.org/10.3145/epi.2016.nov.01>
520. Serrano-Cobos, J., Calduch-Losa, Á., De Dios, J. G., & Aleixandre-Benavent, R. (2016). Scientific communication (XXXIV). How to make a social media strategy for paediatricians (I). Strategy design for a local internet positioning. *Acta Pediatrica Espanola*, 74(6).
521. Serrano-Cobos, J., Font-Julián, C. I., González De Dios, J., & Aleixandre-Benavent, R. (2016). Scientific communication (XXXVIII). How to make a social media strategy for paediatricians (V). Academic SEO through Google Scholar. *Acta Pediatrica Espanola*, 74(10).
522. Serrano-Cobos, J., Gonzalo-Penela, C., González De Dios, J., & Aleixandre-Benavent, R. (2016). Scientific communication (XXXVII). How to make a social media strategy for paediatricians (IV). Customer relationship management techniques in social networks. *Acta Pediatrica Espanola*, 74(9).
523. Serrano-Cobos, J., Ontalba-Ruipérez, J. A., De Dios, J. G., & Aleixandre-Benavent, R. (2016). Scientific communication (XXXVI). How to make a social media strategy for paediatricians (III). Responsible research & innovation as demand analysis of citizens and peers. *Acta Pediatrica Espanola*, 74(8).
524. Shah, C., Hendahewa, C., & González-Ibáñez, R. (2016). Rain or shine? Forecasting search process performance in exploratory search tasks. *Journal of the Association for Information Science and Technology*, 67(7), 1607–1623. <https://doi.org/10.1002/asi.23484>
525. Shi, X., & Yang, C. C. (2006). Mining related queries from search engine query logs. In L. Carr, D. De Roure, & A. Iyengar (Eds.), *WWW '06 Proceedings of the 15th international conference on World Wide Web* (pp. 943–944). Edinburgh, Scotland: ACM New York, NY, USA. <https://doi.org/10.1145/1135777.1135956>
526. Shim, S., Eastlick, M. A., Lotz, S. L., & Warrington, P. (2001). An online prepurchase intentions model: The role of intention to search: Best Overall Paper Award—The Sixth Triennial AMS/ACRA Retailing Conference, 2000☆11☆ Decision made by a panel of Journal of Retailing editorial board members. *Journal of Retailing*, 77(3), 397–416.
[https://doi.org/10.1016/S0022-4359\(01\)00051-3](https://doi.org/10.1016/S0022-4359(01)00051-3)
527. Shiri, A. (2008). Information Retrieval from Digital Libraries: Assessing the Potential Utility of Thesauri in Supporting Users' Search Behaviour in an Interdisciplinary Domain. In *Culture and Identity in Knowledge Organization Proceedings of the Tenth International ISKO Conference* (pp. 184–190). Montréal, Canada. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.134.1233>
528. Shiri, A. A., & Documentation, I. F. for I. and. (1998). *Cybermetrics: A new horizon in information research*. New Delhi: INSDOC. Retrieved from <https://www.tib.eu/en/search/id/BLCP%3ACN032645419/Cybermetrics-A-new-horizon-in-information-research/>
529. Shneiderman, B., & et.al. (2016). Grand challenges for HCI researchers. *Interactions*, 23(5), 24–25. <https://doi.org/10.1145/2977645>

530. Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search A User-Interface Framework for Text Searches. *D-Lib Magazine*. Retrieved from <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>
531. Shneiderman, B., Byrd, D., & Croft, W. B. (1997). D-Lib magazine. *D-Lib Magazine*. Retrieved from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html#case_studies
532. Shtykh, R. Y., & Jin, Q. (2011). A human-centric integrated approach to web information search and sharing. *Human-Centric Computing and Information Sciences*, 1(1), 2. <https://doi.org/10.1186/2192-1962-1-2>
533. Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33(1), 6–12. <https://doi.org/10.1145/331403.331405>
534. Simmons, G. (2013). Keywords Are Dead! Long Live User Intent. *Search Engine Watch*. Retrieved from <https://searchenginewatch.com/sew/how-to/2234448/keywords-are-dead-long-live-user-intent>
535. Sloan, M., Yang, H., & Wang, J. (2015). A term-based methodology for query reformulation understanding. *Information Retrieval*, 18(2), 145–165. <https://doi.org/10.1007/s10791-015-9251-5>
536. Song, R., Luo, Z., Nie, J.-Y., Yu, Y., & Hon, H.-W. (2009). Identification of ambiguous queries in web search. *Information Processing & Management*, 45(2), 216–229. <https://doi.org/10.1016/j.ipm.2008.09.005>
537. Spink, A., & Cole, C. (2006). Human information behavior: Integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology*, 57(1), 25–35. <https://doi.org/10.1002/asi.20249>
538. Spink, A., Ozmutlu, S., Ozmutlu, H. C., & Jansen, B. J. (2002). U.S. versus European web searching trends. *ACM SIGIR Forum*, 36(2), 32. <https://doi.org/10.1145/792550.792555>
539. Stamou, S., & Efthimiadis, E. N. (2009). Queries without Clicks: Successful or Failed Searches? In *SIGIR 2009 Workshop on The Future of IR Evaluation* (pp. 23–24). Boston, USA. Retrieved from <http://turing.science.uva.nl/~kamps/publications/2009/geva:futu09.pdf#page=23>
540. Stamou, S., & Efthimiadis, E. N. (2010). Interpreting User Inactivity on Search Results. In Gurrin C. (Ed.), *Advances in Information Retrieval. ECIR 2010. Lecture Notes in Computer Science, vol 5993*. (pp. 100–113). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12275-0_12
541. Stenmark, D., & Jadaan, T. (2006). Intranet users' information-seeking behaviour: An analysis of longitudinal search log data. *Proceedings of the American Society for Information Science and Technology*, 43(1), 1–19. <https://doi.org/10.1002/meet.1450430172>
542. Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, 69, 213–225. Retrieved from <http://home.uchicago.edu/~vlima/courses/econ200/spring01/stigler.pdf>
543. Stokes, P., & Urquhart, C. (2013). Qualitative interpretative categorisation for efficient data analysis in a mixed methods information behaviour study. *Information Research*, 18(1). Retrieved from <http://www.informationr.net/ir/18-1/paper555.html#.WNKhMIXhDcs>
544. Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). Adaptive web search based on user profile constructed without any effort from users. In *Proceedings of the 13th conference on World Wide Web - WWW '04* (pp. 675–684). New York, New York, USA: ACM Press. <https://doi.org/10.1145/988672.988764>

545. Taksa, I., Spink, A., & Jansen, B. J. (2009). Web log analysis: Diversity of research methodologies. In I. Taksa, A. Spink, & B. J. Jansen (Eds.), *Handbook of Research on Web Log Analysis* (p. 17). <https://doi.org/10.4018/978-1-59904-974-8.ch025>
546. Tan, B., & Peng, F. (2008). Unsupervised query segmentation using generative language models and wikipedia. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (p. 347). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1367497.1367545>
547. Tan, B., Shen, X., & Zhai, C. (2006). Mining long-term search history to improve search accuracy. In *KDD '06 Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 718–723). Philadelphia, PA, USA: ACM. <https://doi.org/10.1145/1150402.1150493>
548. Taycher, L. (2010). Inside Google Books: Books of the world, stand up and be counted! All 129,864,880 of you. Retrieved September 19, 2018, from <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>
549. Taylor, R. S. (2015). Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries*, 76(3), 251–267. <https://doi.org/10.5860/crl.76.3.251>
550. Teevan, J., Dumais, S. T., & Liebling, D. J. (2008). To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent. *Sigir*, 163–170. Retrieved from <http://dl.acm.org/citation.cfm?id=1390364>
551. Terrapin. (2010). Keyword Research | Google Adwords Keyword Tool | 2010 - YouTube. Retrieved September 12, 2018, from <https://www.youtube.com/watch?v=sue16hjoXFg>
552. Tillotson, J. (1994). Is Keyword Searching the Answer? *College & Research Libraries*, 56(3), 199–206. Retrieved from <http://eric.ed.gov/?id=EJ510309>
553. Tolle, J. E. (1983). Transaction log analysis online catalogs. *SIGIR Forum*, 17(4), 147–160. <https://doi.org/10.1145/1013230.511816>
554. Torre, C. (2009). *Satisfacción de los Usuarios en la Biblioteca General Pico de la Universidad Nacional de La Pampa durante el período Diciembre – Marzo – Abril (2004 – 2005)*. Universidad Nacional de Mar del Plata. Retrieved from <http://eprints.rclis.org/13659/>
555. Torres Rodríguez, A. (2016, December 23). Ropa, libros y cosméticos, los regalos que más compramos en Navidad. Retrieved October 11, 2018, from <https://www.bbva.com/es/ropa-libros-cosmeticos-los-regalos-mas-compramos-navidad/>
556. Torres-Salinas, D., Cabezas-Clavijo, Á., & Jiménez-Contreras, E. (2013). Altmetrics: New Indicators for Scientific Communication in Web 2.0. *Comunicar*, 21(41), 53–60. <https://doi.org/10.3916/C41-2013-05>
557. Tramullas, J., & Garrido, P. (2011). Comportamiento informacional y estudios de usuario. In J. (coords). . Cacheda, F., Fernández-Luna, J.M., Huete (Ed.), *Recuperación de información: un enfoque práctico y multidisciplinar* (pp. 457–474). Madrid: RA-MA. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=5905909>
558. Tran, U. S., Andel, R., Niederkrotenthaler, T., Till, B., Ajdacic-Gross, V., & Voracek, M. (2017). Low validity of Google Trends for behavioral forecasting of national suicide rates. *PloS One*, 12(8), e0183149. <https://doi.org/10.1371/journal.pone.0183149>
559. Travieso Rodríguez, C., Alonso-Arévalo, J., & Vivancos Sevilla, J. M. (2007). Usabilidad de los catálogos en Bibliotecas Universitarias: propuesta metodológica de evaluación. *ACIMED*, 16(2). Retrieved from <http://eprints.rclis.org/10374/>
560. Troumbis, A. Y. (2017). Declining Google Trends of public interest in biodiversity: semantics, statistics or traceability of changing priorities? *Biodiversity and Conservation*, 1–11. <https://doi.org/10.1007/s10531-017-1294-z>
561. Tseng, C., Yun-Fei, W., & Hsu, C.-C. (2006). On the Topic Discovery Using Query Logs and Hyperlink. In *Proceedings of the 2006 International Conference on Internet Computing*

- & *Conference on Computer Games Development, ICOMP 2006*. Las Vegas, Nevada, USA. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.90.1228>
562. Tsur, G., Pinter, Y., Szpektor, I., & Carmel, D. (2016). Identifying Web Queries with Question Intent. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16* (pp. 783–793). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2872427.2883058>
563. Turner, N. (2016). Understanding Student Use of a Disciplinary Library: A Focus Group Script. Retrieved from <http://eprints.rclis.org/29690/>
564. Turner, S. J. (2010). Website Statistics 2.0: Using Google Analytics to Measure Library Website Effectiveness. *Technical Services Quarterly*, 27(3), 261–278. <https://doi.org/10.1080/07317131003765910>
565. Uribe, R., Charria, M. E., Monak, L., Roa, P., Castellanos, A., & Barrero, J. (2011). *Metodología común para explorar y medir el comportamiento lector* (1ª). Bogotá. Retrieved from https://cerlalc.org/wp-content/uploads/publicaciones/olb/PUBLICACIONES_OLB_Metodologia-comun-para-explorar-y-medir-el-comportamiento-lector_v1_010111.pdf
566. Vakkari, P. (2008). Trends and approaches in information behaviour research. *Information Research*, 13(4). Retrieved from <http://www.informationr.net/ir/13-4/paper361.html>
567. Vallez, M., & Pedraza-Jimenez, R. (2007). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines. *Hipertext.Net.*, 5. Retrieved from <https://www.upf.edu/hipertextnet/numero-5/pln.html>
568. Van Belle, G., Fisher, L., & Lumley, T. (2004). Longitudinal Data Analysis. In *Biostatistics: a methodology for the health sciences* (2nd ed., p. 871). John Wiley & Sons. Retrieved from <https://faculty.washington.edu/heagerty/Courses/VA-longitudinal/private/LDAchapter.pdf>
569. van Gisbergen, M. S., van der Most, J., & Aelen, P. (2003). *Visual attention to Online Search Engine Results*. Retrieved from http://akademor.webd.pl/download/eyetracking_research.pdf
570. Van Gysel, C., Kanoulas, E., & de Rijke, M. (2016). Lexical Query Modeling in Session Search. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval - ICTIR '16* (pp. 69–72). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2970398.2970422>
571. Vanderkam, D., Schonberger, R., Rowley, H., & Kumar, S. (2011). *Nearest Neighbor Search in Google Correlate*. Retrieved from <https://www.google.com/trends/correlate/nnsearch.pdf>
572. Vaughan, L., & Romero-Frías, E. (2014). Web search volume as a predictor of academic fame: An exploration of Google trends. *Journal of the Association for Information Science and Technology*, 65(4), 707–720. <https://doi.org/10.1002/asi.23016>
573. Vecchione, A., Brown, D., Allen, E., & Baschnagel, A. (2016). Tracking User Behavior with Google Analytics Events on an Academic Library Web Site. *Journal of Web Librarianship*. <https://doi.org/10.1080/19322909.2016.1175330>
574. Veilumuthu, A., & Ramachandran, P. (2009). Intent based clustering of search engine query log. In *2009 IEEE International Conference on Automation Science and Engineering, CASE 2009* (pp. 647–652).
575. Veríssimo, N. M. E. (2009). *Spatio-Temporal Search Log Analysis*. Universidade Técnica de Lisboa. Retrieved from <https://dspace.ist.utl.pt/bitstream/2295/574401/1/dissertacao.pdf>
576. Viet Bang, N. (2007). *Functional Faceted Web Query Classification*. Singapore. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.547.5975&rep=rep1&type=pdf>

577. Vilares, J. (2009). *Procesamiento del Lenguaje Natural en Recuperación de Información*. Retrieved from http://www.grupolys.org/docencia/eri/ERI_4_PLN_en_Recuperacion_de_Informacion.pdf
578. Villaseñor-Rodríguez, I. (2017). *Estudios de usuarios de información*. Barcelona: Editorial UOC. Retrieved from <http://www.editorialuoc.cat/estudios-de-usuarios-de-informacion>
579. Villén-Rueda, L. (2006). Indización y recuperación por materias en los opacs de las bibliotecas españolas: ¿dos décadas de evaluación? *El Profesional de La Información*, 15(2), 87–98. Retrieved from <http://eprints.rclis.org/9339/>
580. Viñas Piquer, D. (2009). *El enigma best-seller: fenómenos extraños en el campo literario*. Ariel. Retrieved from <https://www.casadellibro.com/libro-el-enigma-best-seller-fenomenos-extranos-en-el-campo-literario/9788434425156/1565443>
581. Vundavalli, S., & Hyderabad, I. (2008). Mining the behaviour of users in a multilingual information access task. In *CLEF 2008 Workshop Notes*. Aarhus, Denmark. Retrieved from http://clef-campaign.org/2008/working_notes/vundavalli-paperCLEF2008.pdf
582. Wang, P., Berry, M. W., & Yang, Y. (2003). Mining longitudinal web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology*, 54(8), 743–758. <https://doi.org/10.1002/asi.10262>
583. Wang, P., Wolfram, D., Zhang, J., Hong, N., Wu, L., Canevit, C., & Redmon, D. (2008). Mining web search behaviors: Strategies and techniques for data modeling and analysis. *Proceedings of the American Society for Information Science and Technology*, 44(1), 1–8. <https://doi.org/10.1002/meet.1450440247>
584. Wang, X., & Zhai, C. (2008). Mining term association patterns from search logs for effective query reformulation. In *CIKM '08 Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 479–488). Napa Valley, California, USA: ACM. <https://doi.org/10.1145/1458082.1458147>
585. Wang, X., Broder, A., Gabrilovich, E., Josifovski, V., & Pang, B. (2008). Cross-lingual query classification: a preliminary study. In *iNEWS '08 Proceedings of the 2nd ACM workshop on Improving non english web searching* (pp. 101–104). Napa Valley, California, USA: ACM. <https://doi.org/10.1145/1460027.1460046>
586. Wang, X., Tan, B., Shakery, A., & Zhai, C. (2009). Beyond hyperlinks: organizing information footprints in search logs to support effective browsing. In *ACM 18th International Conference on Information and Knowledge Management, CIKM 2009* (pp. 1237–1246). Hong Kong, China: ACM. <https://doi.org/10.1145/1645953.1646110>
587. Wang, Y., Liu, J., Chen, J., & Huang, Y. L. (2014). Finding similar queries based on query representation analysis. *World Wide Web*, 17(5), 1161–1188.
588. Welch, J. M. (2005). The Electronic Welcome Mat: The Academic Library Web Site as a Marketing and Public Relations Tool. *The Journal of Academic Librarianship*, 31(3), 225–228. <https://doi.org/10.1016/j.acalib.2004.01.014>
589. Whang, M. (2007). Measuring the Success of the Academic Library Website Using Banner Advertisements and Web Conversion Rates. *Journal of Web Librarianship*, 1(1), 93–108. https://doi.org/10.1300/J502v01n01_07
590. Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1), 3–15. <https://doi.org/10.1108/eb026702>
591. Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270. <https://doi.org/10.1108/EUM0000000007145>
592. Wilson, T. D. (2000). Human information behavior. *Informing Science: The International Journal of an Emerging Transdiscipline*, 3, 49–56. Retrieved from <http://go.galegroup.com/ps/anonymou?id=GALE%7CA205363366&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=15479684&p=AONE&sw=w&authCount=1&isAnonymouEntry=true>

593. Wolfram, D., Wang, P., & Zhang, J. (2008). Modeling Web session behavior using cluster analysis: A comparison of three search settings. *Proceedings of the American Society for Information Science and Technology*, 44(1), 1–13.
<https://doi.org/10.1002/meet.1450440232>
594. Wolk, A., & Wöber, K. (2008). A Comparative Study of Information Needs of City Travelers in Europe. *Information Technology & Tourism*, 10(2), 119–131.
<https://doi.org/10.3727/109830508784913077>
595. Wu, F., Madhavan, J., & Halevy, A. (2011). Identifying aspects for web-search queries. *Journal of Artificial Intelligence Research*, 40(1), 677–700. Retrieved from <https://www.jair.org/media/3182/live-3182-5568-jair.pdf>
596. Wu, M., Turpin, A., & Zobel, J. (2008). An investigation on a community's web search variability. In *ACSC '08 Proceedings of the thirty-first Australasian conference on Computer science - Volume 74* (pp. 117–126). Wollongong, Australia: Australian Computer Society, Inc. Retrieved from <http://portal.acm.org/citation.cfm?id=1378279.1378300>
597. Wu, S. (2017). Explore Adventure. Retrieved May 13, 2017, from <http://explore-adventure.com/>
598. Xie, I., & Matusiak, K. K. (2016). User needs and search behaviors. In *Discover Digital Libraries* (pp. 231–253). Elsevier. <https://doi.org/10.1016/B978-0-12-417112-1.00008-9>
599. Xie, K., Yu, H., & Cen, R. (2011). Using log mining to analyze user behavior on search engine. *Frontiers of Electrical and Electronic Engineering in China*, 7(2), 254–260.
<https://doi.org/10.1007/s11460-011-0177-4>
600. Xue, G., Zeng, H.-J., Chen, Z., Ma, W.-Y., & Lu, C. J. (2002). Log mining to improve the performance of site search. In *WISEW '02 Proceedings of the Third International Conference on Web Information Systems Engineering* (p. 238). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.4287&rep=rep1&type=pdf>
601. Yandex. (n.d.). Yandex Wordstat Keyword statistics. Retrieved September 12, 2018, from <https://wordstat.yandex.com/>
602. Yang, L., & Perrin, J. M. (2014). Tutorials on Google Analytics: How to Craft a Web Analytics Report for a Library Web Site. *Journal of Web Librarianship*.
<https://doi.org/10.1080/19322909.2014.944296>
603. Yang, S., & Ghose, A. (2011). Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, 30(1). <https://doi.org/10.1287/mksc.1100.0552>
604. Yin, X., & Shah, S. (2010). Building taxonomy of web search intents for name entity queries. In *Proceedings of the 19th international conference on World wide web - WWW '10* (p. 1001). Raleigh, North Carolina, USA: ACM New York, NY, USA.
<https://doi.org/10.1145/1772690.1772792>
605. Yin, X., & Shah, S. (2010). Building taxonomy of web search intents for name entity queries. *Proceedings of the 19th International Conference on World Wide Web*, (c), 1001–1010. Retrieved from <http://portal.acm.org/citation.cfm?id=1772792>
606. Yu, H., & Ren, F. (2012). Role-explicit query identification and intent role annotation. In *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12* (p. 1163). New York, New York, USA: ACM Press.
<https://doi.org/10.1145/2396761.2398416>
607. Yu, X., & Shi, H. (2009). Query segmentation using conditional random fields. In *Proceedings of the First International Workshop on Keyword Search on Structured Data - KEYS '09* (p. 21). New York, New York, USA: ACM Press.
<https://doi.org/10.1145/1557670.1557680>
608. Yubero, S., & et. al. (2009). *Sociedad educadora, sociedad lectora*. Ediciones de la Universidad de Castilla-La Mancha. Retrieved from <https://play.google.com/store/books/details?id=j0qCAGAAQBAJ&rdid=book->

- j0qCAgAAQBAJ&rdot=1&source=gbs_vpt_read&pcampaignid=books_booksearch_viewport
609. Yubero, S., & Larrañaga, E. (2013). *Lenguaje y Textos. Núm. 37, mayo*. Retrieved from http://www.sedll.org/sites/default/files/journal/el_proceso_de_construccion_del_habito_lector_aportaciones_desde_las_tic_yubero_s._larranaga_e.pdf
 610. Yubero, S., & Larrañaga, E. (2015). Lectura y universidad: hábitos lectores de los estudiantes universitarios de España y Portugal. *El Profesional de La Información, 24*(6), 717. <https://doi.org/10.3145/epi.2015.nov.03>
 611. Zaid, G., & Goldin, D. (2006). *Encuesta nacional de lectura: informes y evaluaciones*. (C. N. para la C. y las A. (Mexico), Ed.). Universidad Autónoma de México. Retrieved from <https://books.google.es/books?hl=es&lr=&id=h4T5Vr-P3TkC&oi=fnd&pg=PA9&dq=estudio+lectores+españa+%22libros+mas%22+comprados+librerias+&ots=S8maogmIBI&sig=pdx3JHuNooEnG-ucdTjyplmZL28#v=onepage&q&f=false>
 612. Zamoszczyk, C., De Luca, S., Ruiz Martínez, S., & Iturbide, L. (2012). Human Query Language. *Ciencia y Tecnología, 12*, 37–50. Retrieved from http://www.palermo.edu/ingenieria/pdf2013/12/12CyT_03humanquerylanguage.pdf
 613. Zhang, Z., & Nasraoui, O. (2008). Mining search engine query logs for social filtering-based query recommendation. *Applied Soft Computing Journal, 8*(4), 1326–1334. <https://doi.org/10.1016/j.asoc.2007.11.004>
 614. Zhang, Z., Zheng, X., Zeng, D. D., & Leischow, S. J. (2016). Tracking Dabbing Using Search Query Surveillance: A Case Study in the United States. *Journal of Medical Internet Research, 18*(9), e252. <https://doi.org/10.2196/jmir.5802>
 615. Zhao, Q., Bhowmick, S. S., & Gruenwald, L. (2005). WAM-Miner: in the search of web access motifs from historical web log data. In *CIKM '05 Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 421–428). Bremen, Germany: ACM. <https://doi.org/10.1145/1099554.1099679>
 616. Zhu, J. J. H., Wu, L., Wang, X., & Qin, J. (2012). Assessing Public Opinion Trends based on User Search Queries: Validity, Reliability, and Practicality. In *World Association for Public Opinion Research* (pp. 1–7). Retrieved from <https://www.researchgate.net/publication/265946891>
 617. Zipf, G. K. (1950). Human Behaviour and the Principle of Least Effort. *The Economic Journal, 60*(240), 808. <https://doi.org/10.2307/2226729>



Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd): No se permite un uso comercial de la obra original ni la generación de obras derivadas.

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Universidad
Zaragoza

Tesis Doctoral

HÁBITOS DE RECUPERACIÓN DE INFORMACIÓN
EN MOTORES DE BÚSQUEDA SOBRE LECTURA,
LIBRO Y BIBLIOTECAS EN ESPAÑA (2004-2016)

Autor

Jorge Serrano Cobos

Director/es

Tramullas Saz, Jesús

UNIVERSIDAD DE ZARAGOZA

Ciencias de la Documentación e Historia de la Ciencia

2019

