



Trabajo Final de Grado

Bases tecnológicas y conceptuales para el uso de asistentes virtuales como interfaces de sistemas de información geográfica

Autor

Guillermo Reloba López

Directores

Dra. Teresa Blanco Bascuas

Dr. F. Javier Zarazaga Soria

Escuela de Ingeniería y Arquitectura

2019

Bases tecnológicas y conceptuales para el uso de asistentes virtuales como interfaces de sistemas de información geográfica

Conceptual and technological basis for the use of Virtual Assistants as an interface to Geographic Information Systems

Resumen

En el contexto de un paradigma de interacción por voz en expansión con el auge de los asistentes virtuales, la información geográfica sigue, a día de hoy, siendo información mayoritariamente inaccesible mediante dicha interfaz. Este proyecto busca explorar las características y dificultades que el desarrollo de una interfaz mediante voz del usuario para un sistema de información geográfica conlleva.

Para ello se han utilizado métodos mixtos involucrando encuestas, un experimento Mago de Oz y el desarrollo y validación de un prototipo en dos iteraciones. De ello se ha obtenido un corpus de instrucciones, del que puede verse lo específico del vocabulario de estas interacciones. También se ha identificado una fuerte influencia del Sistema de Información Geográfico hegemónico, marcando dos perfiles de interacción muy distintos asociados a la experiencia en el uso frecuente del mismo. Como herramienta para la investigación y para servir como punto de partida del desarrollo posterior que este problema requiere, se ha desarrollado un prototipo modular y funcional. Algunos de estos resultados ya han sido expuestos en las Jornadas Ibéricas de la Infraestructura de Datos Espaciales.

Este proyecto surge de la continua colaboración entre Geoslab y el IAAA, que ha puesto en marcha una línea de investigación vinculada a los nuevos modelos de interacción con los sistemas de información geográfica, y que cuenta con financiación dentro del programa de ayudas Torres Quevedo del Ministerio de Ciencia e Innovación. El presente TFG se encuadra dentro de la sublínea de interacción por voz. Entre las líneas de trabajo que surgen de este proyecto puede contarse la integración de un sistema de aprendizaje automático, un sistema de recuperación de información para topónimos, realimentación más elaborada y estudios de usuario más concretos.

Declaración de autoría

TRABAJOS DE FIN DE GRADO / FIN DE MÁSTER

 Escuela de
Ingeniería y Arquitectura
Universidad Zaragoza

**DECLARACIÓN DE
AUTORÍA Y ORIGINALIDAD**

(Este documento debe entregarse en la Secretaría de la EINA, dentro del plazo de depósito del TFG/TFM para su evaluación).

D./Dª. Guillermo Reloba López ,en

aplicación de lo dispuesto en el art. 14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)

Grado (Título del Trabajo)

Bases tecnológicas y conceptuales para el uso de asistentes virtuales como interfaces de sistemas de información geográfica.

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 20 de Febrero de 2020



Fdo: Guillermo Reloba López

Índice

Bases tecnológicas y conceptuales para el uso de asistentes virtuales como interfaces de sistemas de información geográfica	2
Resumen	2
Declaración de autoría	3
1 Introducción	6
1.1 Contexto organizacional del proyecto	6
1.1.1 GeosPatiLab (Geoslab)	6
1.1.2 Grupo de Sistemas de Información Avanzados (IAAA)	7
1.2 Contexto Científico y Tecnológico	8
2 Motivación y objetivos	9
2.1 Problema que se aborda	9
2.2 Alcance y limitaciones	10
3 Trabajo desarrollado	11
3.1 Encuesta de vocabulario	11
3.1.1 Metodología	11
3.1.2 Resultados	13
3.2 Experimento Mago de Oz	15
3.1.1 Metodología	15
3.1.2 Resultados	16
3.3 Desarrollo del prototipo	17
3.1.1 Diseño	17
3.1.2 Características	18
3.4 Validación del prototipo	18
3.1.1 Metodología	18
3.1.2 Resultados	19
4 Lecciones aprendidas y conclusiones	19
4.1 Conclusiones del trabajo	19
4.2 Indicios para futuras líneas de trabajo	20
4.3 Recomendaciones para el desarrollo de una interfaz comercial	21
5 Estructura de los anexos	21
Anexo I: Manual de Usuario	21
Anexo II: Planificación del Proyecto	22

Anexo III: Documentación Mago de Oz	22
Anexo IV: Resultados de la Encuesta	22
Anexo V: Estado del Arte	22
6 Bibliografía	22
Anexos	23
Anexo I. Manual de usuario	23
<u>Anexo II. Planificación del proyecto</u>	<u>25</u>
<u>Anexo III. Documentación del Mago de Oz</u>	<u>27</u>
<u>Anexo IV. Resultados de la encuesta</u>	<u>38</u>
<u>Anexo V. Estado del Arte</u>	<u>40</u>

1 Introducción

1.1 Contexto organizacional del proyecto

1.1.1 GeosPatiuLab (Geoslab)

GeoSpatiumLab (Geoslab¹) es una empresa de base tecnológica, especializada en el tratamiento digital de la información geoespacial y georreferenciada y sus ámbitos de aplicación. GeoSLab nace como spinoff de la Universidad de Zaragoza en enero de 2.007 con el objetivo último de transferir la tecnología desarrollada por el Grupo de Sistemas de Información Avanzados² (IAAA) de la Universidad de Zaragoza a lo largo de una trayectoria de más de 10 años. Su misión es crear valor para nuestros clientes prestando servicios de gestión avanzada de la información espacial.

Geoslab cuenta con un equipo integrado por ingenieros informáticos y titulados superiores, con formación especializada y experiencia contrastada. El enfoque con el que aborda los proyectos se basa en el diseño de equipos multidisciplinares compuestos por personal propio y otros expertos externos, si el tema lo requiere, procedentes de nuestra red contactos y aliados.

Las tecnologías se aplican en los proyectos desarrollados por Geoslab para ofrecer soluciones en distintos ámbitos: Medioambiente; Turismo; Infraestructuras de Datos Espaciales; y Ordenación Territorial y Urbanística.

Geoslab ha participado en numerosos proyectos de I+D+i con financiación pública competitiva entre los que destacan:

- Agroslab-GO: geolocalización y optimización de los procesos agrarios en entornos abiertos. Programa Retos Colaboración 2016. España. Ministerio de Economía y Competitividad. RTC-2016-4790-2 (Programa Estatal de I+D+i Orientada a los Retos de la Sociedad - Retos Colaboración). 2016-2019.
- CANTOGETHER, Crops and ANimals TOGETHER. European Commission. Project ID: 289328. FP7- KBBE.2011.1.4-06. 2012 – 2015
- SMARTQUA, Advanced ICT Risk Assessment Tool to Increase Climate Resilience, Water-Use Efficiency and Environmental Sustainability of Agricultural Production. European Commission. Project ID: 744140. H2020-SMEINST-1-2016-2017.
- Pesticide Finder: Your advisor on pesticide management. European Commission. First call of FINODEX (Future Internet Open Data Expansion). <http://www.finodex-project.eu/>. 2016
- EuroGeoSource, EU Information and Policy Support System for Sustainable Supply of Europe with Energy and Mineral Resources. European Commission (CIP-ICT-PSP-2009-3). 2009 - 2012.
- España Virtual. 4th edition of program CENIT (National Strategic Consortiums of Technical Research). Ministry of Science and Innovation of Spain. 2008 - 2011.

¹ <http://www.geoslab.com>

² <http://iaaa.unizar.es>

1.1.2 Grupo de Sistemas de Información Avanzados (IAAA)

El Grupo de Sistemas de Información Avanzados (IAAA) constituye un grupo estable de investigadores y titulados universitarios que viene desarrollando una prolífica labor de investigación, creación y transferencia de tecnología en los últimos 25 años. En 2003 obtuvo el reconocimiento de Grupo de Investigación Emergente del Gobierno de Aragón, y desde entonces ha sido renovado en todas las convocatorias, consiguiendo la categoría de Grupo de Investigación Consolidado en 2005.

La actividad de investigación del grupo está enfocada en las tecnologías de software abierto, distribuido e interoperable, principalmente mediante servicios web y para sistemas de información geoespacial, abarcando áreas como Sistemas de Información Geográfica (SIG), Teledetección, Servicios Basados en la Localización (SBL) y, con una atención especial a las Infraestructuras de Datos Espaciales (IDEs).

El IAAA es un equipo técnico de referencia en el desarrollo de la Infraestructura de Datos Espaciales de España (IDEE), de la que ha sido promotor desde sus inicios. A partir de trabajos de investigación y desarrollo tecnológico promovidos por la Comisión Interministerial de Ciencia y Tecnología, viene realizando desarrollos tecnológicos en esta área y dando asesoramiento experto a las más importantes instituciones nacionales y la Comisión Europea, así como a las organizaciones nacionales e internacionales de estandarización. La tecnología desarrollada está siendo utilizada en el nodo de referencia nacional de la IDEE, responsabilidad del Instituto Geográfico Nacional, y en varias infraestructuras regionales y locales. El grupo fue también el responsable técnico del proyecto piloto de la Directiva Europea INSPIRE. Así mismo, personal del grupo ha participado como expertos en la elaboración de las reglas de implementación de la Directiva INSPIRE.

De forma resumida, entre sus logros técnicos se podrían destacar los siguientes: Más de 50 proyectos y ayudas a la investigación de convocatorias públicas (INTERREG, 6FP CE, 7FP CE, COST, CICYT, PROFIT, PETRI, TRACE, CENIT, CDTI, AVANZA, autonómicos y locales); Más de 100 contratos con empresas y administraciones públicas para la asesoría, creación y transferencia de tecnología: 8 internacionales que incluyen la ESA, EUROSTAT, EUSC o FAO, Más de 20 nacionales de especial relevancia (financiación del orden de 60.000 € o superior); Más de 300 publicaciones de investigación y divulgación técnica; Trabajos para organismos de estandarización: CEN, OGC, ISO, AENOR; 21 programas informáticos con registro de la propiedad intelectual contratados por instituciones o empresas; Promotores de una empresa spin-off de base tecnológica; Evaluadores en programas de investigación europeos, nacionales, autonómicos y locales.

1.2 Contexto Científico y Tecnológico

Los asistentes inteligentes, tipo Amazon Alexa y Google Home, se están convirtiendo en la nueva interfaz para acceder a todos nuestros dispositivos tecnológicos. Tanto es así, que en muchos foros se considera que el CES de 2019³ ha sido el congreso de los asistentes virtuales⁴. Básicamente, un asistente virtual aporta toda una serie de recursos tecnológicos destinados a entender lo que le estamos solicitando por voz y convertirlo en instrucciones específicas para nuestros dispositivos tecnológicos. Para ello ofrecen unas potentes capacidades de reconocimiento de la voz a las cuales hay que ligar unos sistemas de inteligencia artificial que permiten interpretar lo que le hemos solicitado y convertirlo en comandos de nuestros dispositivos tecnológicos.

Por otro lado, los Sistemas de Información Geográfica o SIG son ya un elemento bien establecido en la vida de cualquier nativo digital y un buen número de inmigrantes digitales, siendo el más conocido Google Maps. Estos sistemas proporcionan cada vez más información conforme se normaliza la geolocalización de todos los dispositivos móviles y se añaden más funciones a estos sistemas que originalmente respondían solamente a la necesidad de navegación.

Por último, la accesibilidad es un aspecto del diseño que por fin comenzó a extenderse desde la arquitectura al resto de campos a principios de este siglo, por ejemplo con la Declaración de Estocolmo del European Institute for Design and Disability(EIDD) en 2004, que sienta los principios del Diseño para Todos. Se han logrado grandes avances en todos los campos, incluidos los sistemas informáticos, pero está lejos de ser un problema resuelto.

En el Anexo V se presenta un pormenorizado análisis del estado del arte.

Comentado [1]: De mal gusto mencionarlo?

Comentado [2]: De mal gusto por qué? De esto refs que te pasé y en el Voz a Tierra tendrás más cosas. Puedes cortar y pegar

³ Congreso mundial de referencia en nuevos productos tecnológicos <https://www.ces.tech/>

⁴ <https://www.nytimes.com/2019/01/09/technology/ces.html>

2 Motivación y objetivos

2.1 Problema que se aborda

El problema abordado con este trabajo es la exploración del uso de las instrucciones verbales como única interfaz de entrada a Sistemas de Información Geográfica. Esta decisión está motivada por dos hechos principales:

El primero es el **paradigma emergente de interacción asociado al auge de los asistentes virtuales**. Aunque muchos de estos asistentes siguen dependiendo de interfaces tradicionales para gran parte de su entrada de datos, hay una marcada preferencia y tendencia a independizarse cada vez más de los mismos en el esfuerzo por una interacción lo más próxima a la humana posible. De este modo, hay todo un paradigma de interacción que tiende a ser únicamente verbal, y del que la información geográfica está, en este momento, excluida.

El segundo motivo que impulsa la investigación de este modo de interacción es, por supuesto, el **aumento de accesibilidad** que una nueva interfaz siempre conlleva. En el Real Decreto 1112/2018 sobre la accesibilidad de los sitios web y aplicaciones para móviles del sector público manifiesta que: *"Están EXCLUIDOS... los siguientes tipos de contenido: ... Servicios de mapas y cartografía en línea, siempre y cuando la información esencial se proporcione de manera accesible digitalmente en el caso de mapas destinados a fines de navegación..."*. Esta limitación a la accesibilidad ofrecida está siendo estudiada por distintos trabajos, ya que responde a una intrínseca dificultad de adaptar la interfaz de este tipo de sistemas(1-4).

En este marco se encuadra este trabajo, tratando de dar los primeros pasos hacia la creación de una interfaz de lenguaje natural para los sistemas de información geográfica. Por un lado, para no dejar atrás estos sistemas en el avance hacia un nuevo paradigma en los asistentes virtuales. Por otro lado, para no seguir dejando atrás a las personas con discapacidad en el acceso a esta información. Y por último, porque siguiendo los principios del **Diseño Para Todos es necesario desarrollar sistemas que respondan a todos en todas nuestras circunstancias**, y este trabajo aspira a ser una pieza más en lograr que los sistemas de información geográfica lo hagan(5).

2.2 Alcance y limitaciones

El enfoque de este trabajo es el abordaje del problema e identificación de sus características, así como de los retos que el mismo presenta. Como tal, **es de manera principal un trabajo de investigación**, siendo el prototipo desarrollado una herramienta de estudio más y no el objetivo del proyecto.

Por un lado, ha de tenerse en cuenta la limitación en la capacidad de movilización de usuarios de prueba dadas las características del proyecto.
Aunque 337 respuestas es una muestra razonable para el estudio de un Trabajo de Fin de Grado, los resultados pretenden ser extrapolables a 483 millones de hispanohablantes nativos. A nivel estadístico, todo resultado del estudio cuantitativo de este trabajo ha de considerarse un indicio más que una certeza. Por esta razón, se ha optado por usar métodos mixtos, con un estudio cuantitativo, pero utilizando principalmente **estudios cualitativos** para el diseño y la validación del prototipo.
Esta planificación se ha visto posteriormente alterada por la pandemia COVID19, reduciendo la disponibilidad de usuarios y las posibilidades de hacer pruebas presenciales, fundamentales en cualquier estudio cualitativo.

Algunas limitaciones a tener en cuenta respecto a la validez de los resultados en **grupos de especial interés** en este tipo de interfaz, son la ausencia de hablantes de castellano no nativos y personas con alteraciones del habla (disfonía, dislalia, etcétera) en los usuarios de prueba. Estando considerados distintos niveles de alfabetismo digital y distintas variantes dialectales del castellano.

Comentado [3]: Para la parte cualitativa la muestra no es modesta.

Por otro lado, has evaluado con mixed-methods (puedes citarlo en la parte de evaluación: Xassess: <https://www.tandfonline.com/doi/abs/10.1080/09544828.2016.1200717>) y el método es perfectamente válido. Un método cuantitativo y varios cualitativos.

Leido así, parece que ya dices que la evaluación ha sido una caca, y puedes predisponer al tribunal.

Como estás hablando de limitaciones sí pon la limitación, pero redacta de otra forma.

Comentado [4]: Mejor?

3 Trabajo desarrollado

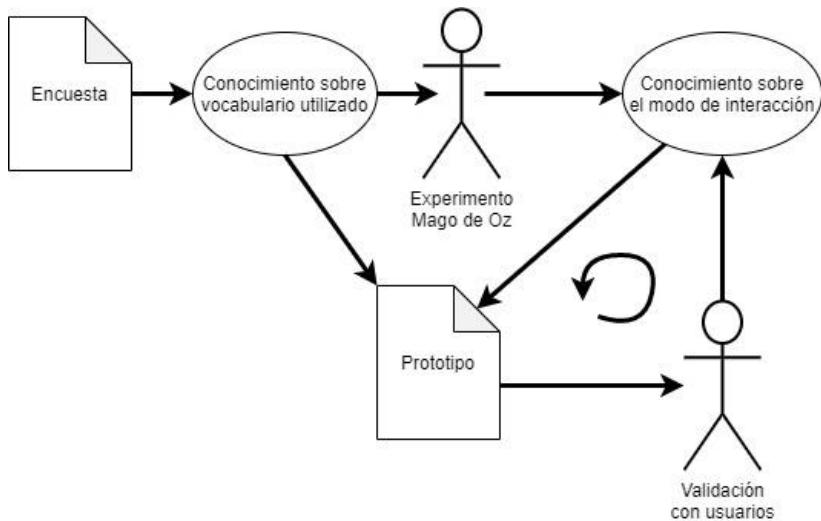


Ilustración 1 Esquema general del trabajo

3.1 Encuesta de vocabulario

3.1.1 Metodología

Como paso inicial se propuso la identificación del vocabulario que los usuarios emplearían para relacionarse con un SIG mediante la voz, y si el mismo dependía de algún factor de su perfil. El principal reto metodológico de este paso consiste en no predisponer a los encuestados, pero hacer entender de forma clara el comportamiento que ellos habrían de replicar. Otra cuestión a considerar en el diseño de la encuesta es seleccionar correctamente qué elementos preguntar para desarrollar luego perfiles de usuario.

La opción escogida para evitar condicionar respuestas fue utilizar vídeos que ejemplificasen las funcionalidades básicas de un SIG (zoom, paneo, cambio de capa, etcétera) y pedir, en pregunta abierta, que dijeran cómo indicarían a Google Maps que hiciera la acción mostrada. Esto evita mencionar por un término o por otro la acción estudiada. En cuanto al perfil se decidió preguntar por experiencia con asistentes virtuales, SIGs y competencia digital en general además de las preguntas básicas (edad, sexo, nivel de estudios).

Para la realización de la encuesta se probaron varias herramientas de encuesta online, pero acabó escogiéndose el servicio de pago Survey Monkey⁵ dado que ninguno de los gratuitos explorados cubría las necesidades de integración de vídeo de forma satisfactoria. Esta encuesta mediante vídeo fue distribuida a través de listas de difusión al mayor número y diversidad de

⁵ <https://es.surveymonkey.com/>

usuarios posibles. Más concretamente, se hizo especial esfuerzo en hacer llegar la encuesta a usuarios de América Latina, ya que forman en torno al 90% de hispanohablantes del mundo.

* 4. Frecuencia de uso de Google Maps y similares

- Diario
- Más de una vez por semana
- Más de una vez al mes
- Menos de una vez al mes

* 5. ¿Usa asistentes virtuales (Alexa, Siri, Cortana, Google Assistant, ...)?

- Sí
- No

Comentado [5]: Tendré que pasar a Office para añadir pies de foto

Ilustración 2 Fragmento de la encuesta

* 6. ¿Cómo le indicarías a Google Maps que hiciese lo que muestra el siguiente vídeo?

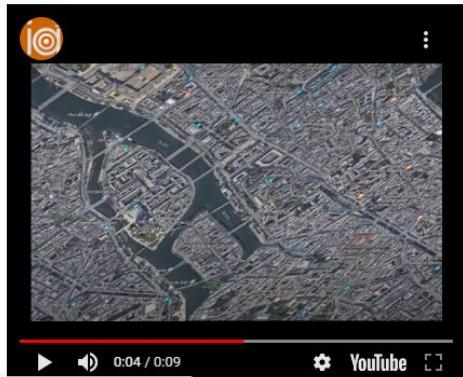


Ilustración 3 Fragmento de la encuesta

Los resultados recogidos fueron clasificados por los verbos, modos de conjugación y sustantivos utilizados. Para su clasificación se trató de emplear diversas herramientas (Tableau⁶, la biblioteca Pandas Database de Python, etcétera), pero debido a la variedad de las respuestas en ortografía, abreviaturas y conjugaciones, resultó imposible el uso de ninguna. La clasificación se hizo de manera natural, con el uso extensivo de la función de búsqueda y reemplazo del

⁶ <https://www.tableau.com/>

procesador de texto Notepad++⁷. Una vez clasificados, los datos fueron contabilizados y procesados para tratar de observar perfiles y correlaciones entre el vocabulario utilizado y las características de los usuarios. El test estadístico escogido para estudiar dichas correlaciones es la prueba chi cuadrado.

3.1.2 Resultados

El resultado más directo de la encuesta es un corpus del vocabulario utilizado para interaccionar por voz con un SIG. Analizando dicho corpus puede verse que hay una tendencia a una distribución exponencial, con unas pocas palabras reuniendo a un alto porcentaje de usos.

En términos generales, los verbos más frecuentes para instruir por voz a un SIG que mueva la pantalla son: Mover, desplazar, ir y los verbos direccionales subir y bajar, con su infinitivo o en imperativo. Con frecuencia, en caso de ser conjugado en imperativo, el verbo es acompañado de una partícula pronominal (p.ej. “Desplázate, Muéveme”).

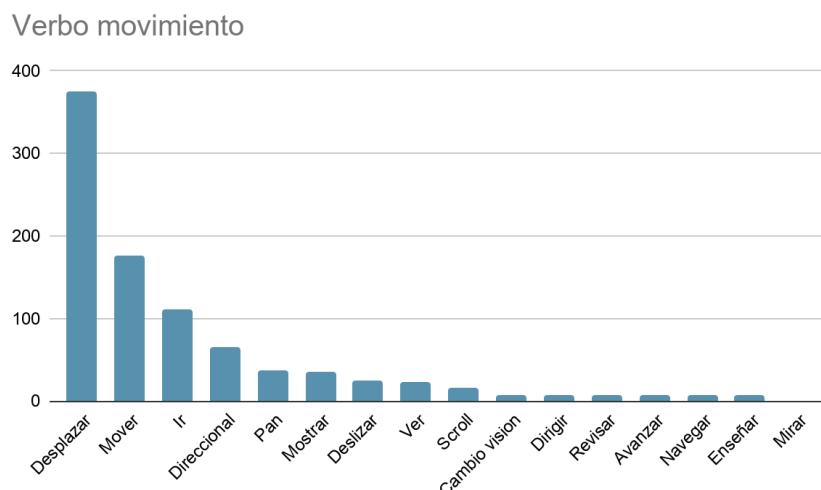


Ilustración 4 Resultado del verbo escogido en las respuestas a la acción de paneo

Los verbos más frecuentes para instruir por voz a un SIG que aumente el nivel de zoom son: Ampliar, acercar y zoom. Ampliar y acercar pueden ser conjugados en segunda persona de imperativo e ir acompañados de un complemento directo “mapa” o “imagen”, pero es más frecuente el infinitivo. Los usos de la acción *zoom in* están particularmente unificados, con solo siete verbos distintos y tres de ellos agrupando el 95% de usuarios.

⁷ <https://notepad-plus-plus.org/>

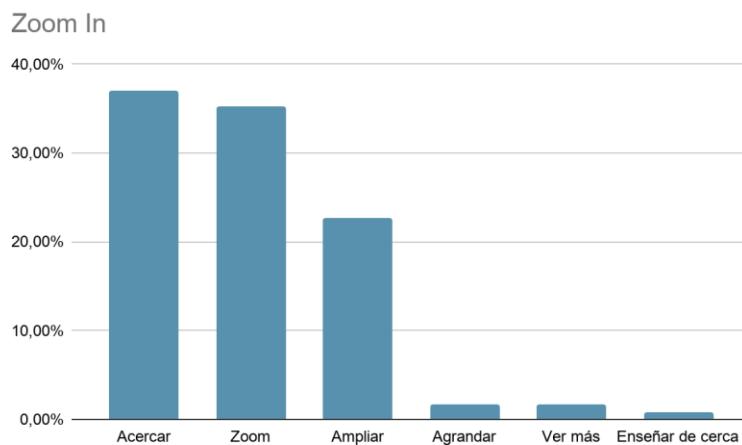


Ilustración 5 Resultado del verbo escogido en las respuestas a la acción de aumentar zoom

Los verbos más frecuentes para instruir por voz a un SIG que muestre una ortofoto son: Cambiar, mostrar y ver. Cabe destacar el uso del término *satélite* tanto como sustantivo en otras instrucciones como por instrucción completa.

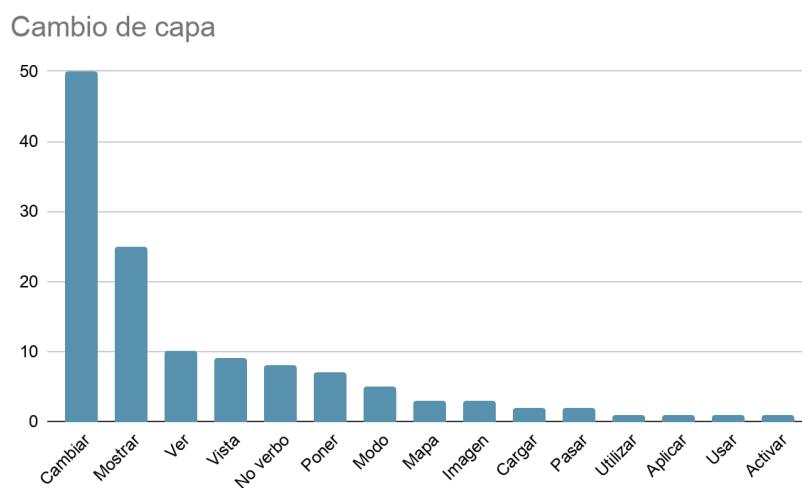


Ilustración 6 Resultado del verbo escogido en las respuestas a la acción de cambio de capa

Se han analizado todas las correlaciones entre la edad, el género, el nivel de estudios y el nivel de familiaridad con asistentes virtuales o SIG y diferencias en los diferentes vocabularios.

Respecto a los resultados de la prueba, se ha tomado una única respuesta por usuario en las direcciones a pesar de que se recogieron distintas respuestas por usuario. Esto es debido

a que la prueba chi-cuadrado asume independencia entre las entradas, lo cual no sería así de ser varias respuestas por usuario. Por otra parte, las muestras tomadas en algunas categorías (personas sin estudios, personas de edad avanzada, menores) son demasiado pequeñas como para poder afirmar con certeza ninguna correlación en esos perfiles. Son necesarios estudios más dedicados capaces de alcanzar a más usuarios de grupos específicos para observar sus características.

Un fenómeno observado en la encuesta es el de distintas formas en las que el usuario interpreta el sistema. Algunos interpretan el sistema como un mapa interactivo, otros se localizan a sí mismos en el mundo representado por el sistema, y por último otros ven el sistema como sistema y no tratan de interactuar de manera natural, sino averiguar cómo provocar la respuesta deseada.

3.2 Experimento Mago de Oz

3.1.1 Metodología

La encuesta proporciona mucha información acerca del contenido de la interacción, pero no sobre la forma de la misma. Para poder observar el enfoque y estrategia, así como el comportamiento ante errores, que emplearían los usuarios en un caso de uso real, se utilizó una evaluación con usuarios de un sistema simulado. Este tipo de experimentos reciben el nombre de Mago de Oz⁸ por referencia a la novela en la que, como en este experimento, un humano oculto tras una pantalla crea el efecto de algo que no existe. En este caso, se simuló un sistema capaz de interpretar a la perfección el lenguaje natural de los usuarios y realizar las acciones requeridas sobre el SIG, pues la pantalla del usuario estaba conectada al equipo del *mago*, que era capaz de oír lo que el usuario decía ya que se encontraba en la misma habitación tras una pared.

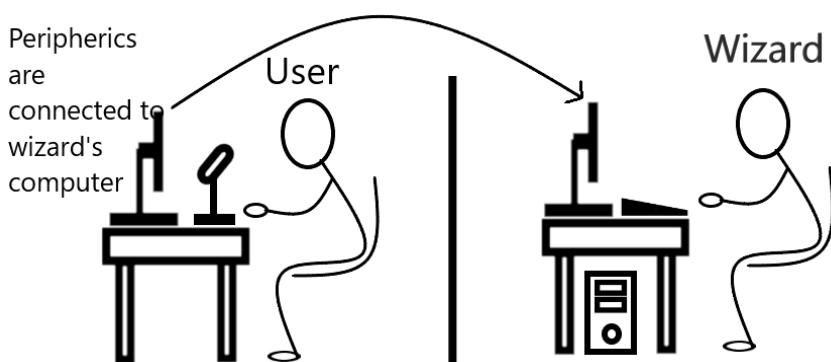


Ilustración 7 Representación del experimento Mago de Oz

Para los usuarios se diseñaron cuatro escenarios distintos con el objetivo de recorrer todas las funcionalidades típicas de un SIG que se implementarían después en el prototipo. Dichos escenarios integran de manera natural distintas funcionalidades de los SIG, pero también

⁸ Green, P. (1985). *The Wizard of Oz: a tool for rapid development of user interfaces. Final report.*

están preparados para inducir ciertos errores para poder observar toda la gama del comportamiento del usuario en las distintas situaciones de un sistema real. Para cada escenario se elaboró una lista de errores previstos y baterías de términos de búsqueda que permitiesen al mago reaccionar a los usuarios con una velocidad convincente. También para ello se especificaron estrategias para enmascarar el uso de periféricos y el sonido, todo con el objetivo de mantener la ilusión para obtener reacciones naturales de los usuarios. Las pruebas fueron grabadas en vídeo para su posterior visionado y análisis. Tras el experimento se realizó una breve entrevista de satisfacción con el sistema, una encuesta de perfil de usuario y se revelaba el engaño, ofreciendo a los usuarios eliminar la grabación del experimento en caso de que así lo prefiriesen.

En la selección de usuarios, se buscó y logró diversidad en los factores pertinentes a la interacción con sistemas mediante la voz (edad, experiencia con sistemas informáticos, variedad dialectal, nivel de estudios, etcétera). Gracias al reducido número de usuarios en comparación con la encuesta, adecuado ya que se trata de una prueba cualitativa, pudo estudiarse cada grabación en detalle. Cada funcionalidad que el usuario ha tratado de usar, estuviera prevista y el mago actuase acorde a la petición o no, ha sido registrada para observar cual es la forma natural de interacción en la que basar el prototipo.

3.1.2 Resultados

El principal resultado esperado de este experimento era la observación de la estrategia seguida por los usuarios para la resolución de los problemas presentados. Pudo comprobarse que en esto hay dos perfiles completamente diferenciados y opuestos. La estrategia mayoritaria es la del uso de la funcionalidad de búsqueda, tratando de resolver el problema en un solo paso. Estos usuarios han mantenido esa estrategia incluso cuando diferentes búsquedas consecutivas han fallado en la resolución del escenario. Una minoría de usuarios, menos versada en la interacción con sistemas informáticos, incluyendo los SIG y los asistentes virtuales, trató de resolver los problemas mediante el movimiento y zoom de manera iterativa.

La interacción observada corroboró en gran medida el vocabulario observado en la encuesta. También se observaron varios aspectos críticos que debieran ser implementados en el prototipo, como la necesidad de un feedback explícito o la persistencia de contexto. También, aunque minoritario, trató de utilizarse elementos externos al SIG mediante la voz (ver las fotos, imprimir,...).

Comentado [6]: se me queda corta la parte de resultados- Había cuestiones que corregían los resultados de la encuesta, creo recordar. Puedes añadir Verbatines, citas literales de comentarios que hicieron los usuarios en la entrevista. Al finalizar alguna de ellas comentamos ejemplos.

3.3 Desarrollo del prototipo

3.1.1 Diseño

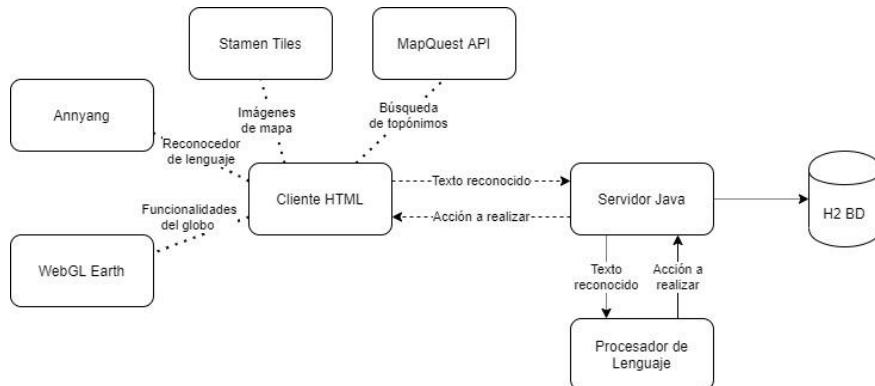


Ilustración 8 Esquema de la arquitectura del prototipo

El prototipo consta de tres partes diferenciadas. Por un lado el servidor que recibe y responde las peticiones de procesos de instrucciones y almacena las interacciones en una base de datos. Este se ha desarrollado mediante Spring Boot y utiliza una base de datos h2⁹.

Por otro lado, el paquete que procesa las instrucciones en lenguaje natural y responde cuál es la instrucción adecuada en cada caso. En este prototipo este paquete ha sido desarrollado mediante JavaCC, un generador de parsers, mediante las especificaciones obtenidas de los dos experimentos anteriores.

Por último, el cliente es el que se encarga de mostrar el mapa proporcionado por WebGL Earth¹⁰, una aplicación de planeta virtual de código abierto, y mandar peticiones al servidor con el texto reconocido mediante la librería Annyang¹¹ que expone las funcionalidades de la interfaz Web Speech API¹². Web Speech API es una interfaz de JavaScript ya soportada por la mayoría de navegadores para el reconocimiento y síntesis de voz.

Este desacoplamiento, que resultaría innecesario en un proyecto de este tamaño, es intencional. El propósito del mismo es crear un prototipo fácilmente ampliable y del que puedan sustituirse partes de manera modular a medida que avanza el estudio de este tipo de interacción.

⁹ <https://www.h2database.com/html/main.html>

¹⁰ <https://www.webglearth.com/>

¹¹ <https://www.talater.com/annyang/>

¹² <https://wicg.github.io/speech-api/>

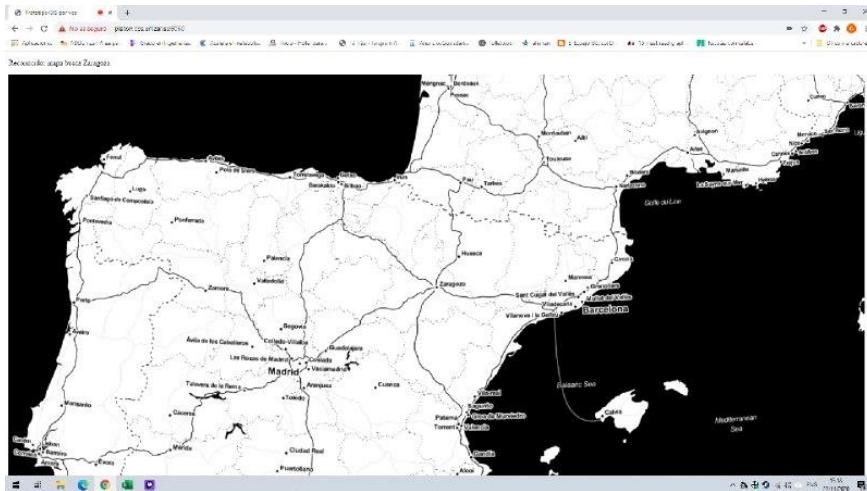


Ilustración 9 Captura de pantalla del prototipo en funcionamiento

3.1.2 Características

El prototipo es capaz de responder a instrucciones y ser manejado como cualquier SIG usando únicamente la voz. Proporciona realimentación en forma del texto reconocido, lo cual resulta fundamental en caso de que el sistema esté confundiendo instrucciones por palabras similares (por ejemplo, nueve y mueve). También es capaz de buscar topónimos a través de MapQuest API, pero el reconocimiento correcto de topónimos es un problema pendiente en el caso de topónimos en otros idiomas o de pronunciación difícil o variable.

...

3.4 Validación del prototipo

3.4.1 Metodología

Teniendo el prototipo desarrollado, se procedió a su validación mediante dos métodos distintos. Por un lado, la validación de que el prototipo desarrollado responde a lo aprendido de los experimentos anteriores mediante un experimento de escenarios basado en las tareas del Mago de Oz. Siempre y cuando los escenarios cubran de manera exhaustiva las funcionalidades, comparar los resultados de la prueba con el sistema ideal simulado por un humano pone de manifiesto el éxito en el desarrollo del prototipo acorde a las especificaciones propuestas en un primer momento. Como en el experimento anterior, se mantiene la intervención del observador al mínimo necesario, tratando de replicar una interacción natural con el sistema. Una ventaja de tener el prototipo real en funcionamiento es que las interacciones quedan registradas en el sistema, haciendo mucho más fácil su contabilización y la obtención de un ratio de éxito de instrucciones.

La muestra para este experimento se vio limitada por el efecto de la pandemia COVID-19. Debido a la limitación de reunirse en persona, se hicieron pruebas de manera telemática hasta llegar a un tamaño de muestra aceptable. Esto limitó la capacidad de observación durante el experimento, pero se tomaron medidas para tratar de mitigarlo utilizando dos herramientas

Comentado [7]: lo mismo que en las anteriores, sigue el esquema de descripción de métodos en cada fase, muestra, etc.

Faltan también alusiones al cognitive walkthrough, escenarios, etc.

Comentado [8]: _Marcado como resuelto_

Comentado [9]: _Reabierto_

de comunicación a la vez, una para la observación de la cámara web y otra para la pantalla del usuario.

Como parte de la validación del prototipo, también se ha enviado a expertos en Interacción Persona-Ordenador para la evaluación heurística de varios principios de diseño y usabilidad. Se ha enviado un enlace al prototipo y a un formulario de evaluación en el que han valorado el nivel de error en cada uno de los principios. Las heurísticas escogidas han sido: visibilidad del estado del sistema, control y libertad del usuario, prevención de errores, diseño estético y minimalista, ayudar a reconocer, diagnosticar y recuperarse de los errores y consistencia y estándares. Esta lista ha sido extraída de entre aquellas heurísticas aplicables al prototipo de la lista de Principios Heurísticos de Nielsen y las Ocho Reglas de Oro de Ben Shneiderman(6,7).

Comentado [10]: referencias bibliográficas
@g.relobo.lopez@gmail.com
Asignado a Guillermo Reloba López

3.1.2 Resultados

La validación resultó exitosa, confirmando la captura de los requisitos identificados en los experimentos anteriores. Se optó por implementar únicamente los verbos más frecuentes para las acciones, y se observó que, en el caso de un fallo debido a el uso de un término no reconocido, los usuarios pasaban rápidamente a uno de los términos frecuentes. Solo un usuario de los entrevistados expresó que no lo usaría en su vida diaria y la valoración media fue de 3.8 sobre 5.

4 Lecciones aprendidas y conclusiones

4.1 Conclusiones del trabajo

Se han identificado varias características del problema abordado y se ha desarrollado un prototipo fácilmente ampliable sobre el que iterar soluciones al mismo. Resultados parciales de este trabajo han sido presentados en las XI Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE2020), celebradas del 26 al 30 octubre de 2020 en Évora (Portugal), bajo el título “*Primeros pasos para manejar por voz la visualización de datos geográficos*”.¹³

Una conclusión alentadora para la resolución del problema es que el vocabulario utilizado en la interacción es fácil de acotar. En todas las funcionalidades ha sido necesario menos de cinco posibles opciones verbales, eso sí, en distintas conjugaciones, para cubrir más del 95% de los usos. Incluso en los casos restantes, se ha observado que los usuarios cambian rápidamente a una de las opciones mayoritarias después del primer fallo. Por otra parte, esto no nos libra de la necesidad de un sistema de aprendizaje automático en una versión definitiva, ya que la interpretación de instrucciones ambiguas y algunos aspectos de la persistencia de contexto esperadas por los usuarios son difíciles de implementar de otro modo.

¹³ <https://www.dgterritorio.gov.pt/jiide2020/>

También se ha identificado la realimentación como elemento clave en este tipo de interacción. La ausencia de realimentación sensorial en la acción del usuario (el click del ratón, el ruido del teclado), aumenta la ansiedad del usuario ante la duda de si su acción ha sido registrada. A esto se suma que la situación de la tecnología de reconocimiento de voz y de los micrófonos obliga a que la interacción siga siendo torpe en este momento. Una realimentación adecuada no solamente avisa al usuario de que ha de repetir su acción porque se ha malentendido su instrucción, sino que además indica al usuario cómo se ha malentendido para que pueda cambiar su enunciación para tratar de evitar errores sucesivos.

La estrategia seguida por la mayoría de usuarios es la única ofrecida por Google Maps mediante voz en este momento, la de búsqueda en un solo paso. Además, a ello se une que la característica que permite separar los dos perfiles de usuario, la experiencia en el uso de sistemas informáticos (incluyendo SIGs), sugiere que se trata de otro caso en el que la tecnología ya presente ha influido fuertemente en la interacción con futuros sistemas. Se induce de la observación, aunque no se ha realizado un experimento específico para confirmarlo, que los usuarios han internalizado las capacidades y limitaciones de SIGs ya existentes.

Por último, de la evaluación del prototipo puede extraerse que un sistema de estas características es una necesidad real. En comparación con la opinión expresada sobre la calidad del prototipo, resulta notable el porcentaje de gente que ha expresado que lo usaría en su vida diaria.

4.2 Indicios para futuras líneas de trabajo

De este estudio inicial se observan varias líneas de trabajo en la futura investigación de este tipo de interacción. Por un lado, y con la accesibilidad como motivador principal de esta investigación, el estudio de los grupos a los que no se ha podido acceder y que más riesgo tienen de aislamiento en el paradigma de interacción hablada. En particular, es necesario investigar las dificultades específicas que personas con trastorno de comunicación social, dislalia, disfemia, desartria y disglosia, y hasta qué punto pueden hacer uso de los estándares de reconocimiento de voz. Por otra parte, en personas con afasia o discapacidades cognitivas, es necesario el desarrollo de soluciones específicas que permitan su accesibilidad a estos sistemas. También es relevante el estudio de vocabulario e interacción, específicamente con SIGs, de menores y personas de edad avanzada.

Otra posible línea es la de la profundización en indicios encontrados durante el estudio. Por ejemplo, del grado de interiorización de la interfaz de Google Maps en el uso de SIGs mediante otras interfaces o las distintas formas en las que los distintos usuarios se relacionan a nivel cognitivo con el sistema: situándose en la localización mostrada, observándolo como una representación virtual del planeta o como una mera herramienta.

Por último, y desde un punto de vista más técnico, este trabajo proporciona las bases y herramientas para el estudio y desarrollo de un procesamiento de lenguaje natural más elaborado en el contexto de la interacción con SIGs. Permitiendo quizás salvar alguna de las dificultades inherentes a la accesibilidad de este tipo de información que le han llevado a estar excluida en primer lugar.

4.3 Recomendaciones para el desarrollo de una interfaz comercial

El prototipo desarrollado cumple de manera satisfactoria las necesidades mínimas de manejo de un SIG mediante la voz. Permitir una interacción íntegra por voz, sin necesidad de intervención de ratón y teclado. Como se ha mencionado, ofrece realimentación con el contenido del texto reconocido, lo cual resulta fundamental en la usabilidad tanto por mantener al usuario informado como para permitirle aprender.

Una interfaz comercial habría de ampliar sobre esto, y se han observado durante las pruebas de evaluación como hay mejoras que resultarían directamente aplicables. Una de ellas es la persistencia de contexto, es decir, que el sistema tenga en cuenta las últimas instrucciones y la situación en la que está a la hora de interpretar las instrucciones. Algunos ejemplos de situaciones afectadas por esta característica es el poder deshacer acciones, añadir detalles a una instrucción anterior o ajustar la cantidad de movimiento o zoom acorde a las últimas instrucciones.

Otra mejora que podría facilitar y acelerar la interacción es un sistema de adaptación al usuario. A su vocabulario específico, a su enunciación y pronunciación, a sus patrones de uso, etcétera. Probablemente para algunos de estos casos esto requiere de un sistema de aprendizaje automático, pero incluso permitir ajustes predefinidos al sistema puede suponer una mejora sustancial que mejoraría la usabilidad para usuarios avanzados.

Un problema analizado con especial interés fue el de los topónimos. Aunque pueden encontrarse listas muy comprehensivas de los mismos, muchos son homófonos o difíciles de pronunciar en nuestro idioma. Este problema ha sido menos prevalente de lo previsto, habiendo dado errores sólo en casos de un error claro de pronunciación que, sin sistemas de recuperación de información mucho más avanzados, son simplemente irresolubles.

Por último, varios usuarios trataron de controlar mediante voz elementos de la interfaz que no pertenecen al SIG. Es importante considerar que la mayoría de usuarios espera una forma de interacción consistente en el sistema en su conjunto, por lo que esta interfaz debe formar parte de un dispositivo que ya responda mediante voz o que la interfaz debe ser diseñada de manera que delimite claramente con qué partes puede interactuarse con la voz y con qué partes no.

En resumen, una interfaz comercial debe dedicar especial atención a la realimentación e implementar la persistencia de contexto entre instrucciones. También debe marcar claramente el límite, en caso de haberlo, de la interacción por voz. Sistemas de aprendizaje automático que mejoren el reconocimiento y procesado de instrucciones y sistemas de recuperación de información que permitan errores en topónimos serían adiciones interesantes.

5 Estructura de los anexos

Anexo I: Manual de Usuario

Incluye las instrucciones requeridas para conectarse y utilizar el prototipo mientras siga alojado en el servidor del grupo IAAA.

Anexo II: Planificación del Proyecto

Incluye la planificación organizativa del proyecto y la distribución de las horas de trabajo dedicadas al mismo.

Anexo III: Documentación Mago de Oz

Incluye la documentación generada y utilizada para la correcta realización del experimento Mago de Oz, con instrucciones para el entrevistador, el mago, fichas de recogida de notas, los escenarios utilizados, etcétera.

Anexo IV: Resultados de la Encuesta

Incluye las tablas de contingencias generadas a raíz de los resultados de la encuesta.

Anexo V: Estado del Arte

Análisis del estado del arte de los campos relevantes para el proyecto realizado antes de comenzarlo.

6 Bibliografía

1. Calle Jiménez, T., & Luján-Mora, S. (2016). Web accessibility barriers in geographic maps.
2. Calle-Jimenez, T., Orellana-Alvear, B., & Prado-Imbacuan, R. (2019, November). GIS and User Experience in Decision Support for Retail Type Organizations. In 2019 International Conference on Information Systems and Software Technologies (ICI2ST) (pp. 156-161). IEEE Computer Society.
3. Chan, A., & Dascalu, S. (2017, May). Using Brain Computer Interface Technology in Connection with Google Street View. In 2017 21st International Conference on Control Systems and Computer Science (CSCS) (pp. 571-576). IEEE.
4. Götzelmann, T., & Eichler, L. (2016, July). Blindweb maps—an interactive web service for the selection and generation of personalized audio-tactile maps. In International Conference on Computers Helping People with Special Needs (pp. 139-145). Springer, Cham.
5. Aragall, F. (2003). European concept of accessibility.
6. Nielsen, J. (1995). How to conduct a heuristic evaluation. *Nielsen Norman Group*, 1, 1-8.
7. Shneiderman B. (2016) The Eight Golden Rules of Interface Design

Anexos

Anexo I. Manual de usuario

Las interacciones con el sistema son registradas con la frase reconocida en texto y la respuesta dada por el sistema de manera anónima. En ningún momento se graba la voz. Los pasos necesarios para abrir el prototipo son los siguientes:

-La conexión al prototipo es a través de >><https://platon.cps.unizar.es:6060><< El certificado SSL de la página es autofirmado, así que es posible que el navegador advierta de que la conexión no es segura. Para acceder a la página en Chrome, pulsar Configuración avanzada y la opción Acceder a platon.cps.unizar.es (sitio no seguro).

-Permitir uso de micrófono.

-Dar instrucciones al sistema diciendo Mapa ..., por ejemplo, Mapa mueve a la derecha.

-En la margen superior de la página aparecerán las instrucciones que el sistema reconozca (aquellas en las que haya reconocido "Mapa").

Comentado [11]: Sugerencia de anexos, no hay que ceñirse a ellos.

Comentado [12]: Aquí podrías poner el "Manual de uso del prototipo"
Básicamente las instrucciones que mandaste para usarlo y algún pantallazo donde se vea el mensaje y explicarlo. Por ejemplo, una pantalla completa y luego un detalle resaltando el texto informativo
@g.reloba.lopez@gmail.com
Asignado a Guillermo Reloba López

Comentado [13]: Incluyendo el link?

Anexo II. Planificación del proyecto



Este diagrama de Gantt recoge la planificación inicial del proyecto, pero debido a imprevistos acontecidos hubo de ser modificado por completo. El diagrama modificado es el siguiente.

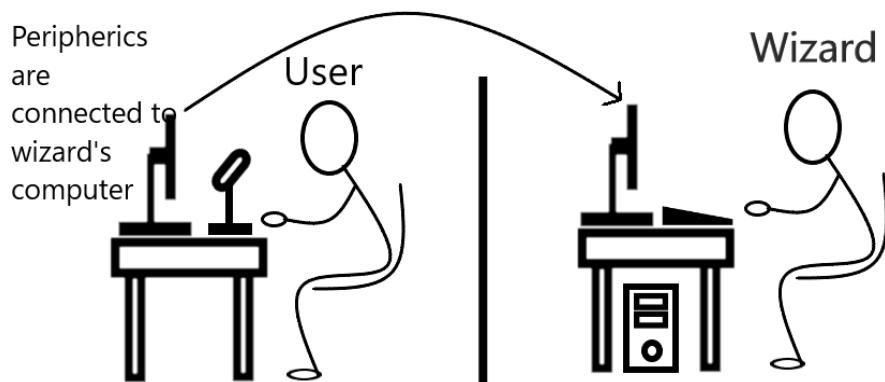
Trimestres 2019	Junio 3	Julio 1	Julio 2	Julio 3	Agosto 1	Agosto 2	Agosto 3	Septiembre 1	Septiembre 2	Septiembre 3-Diciembre 3	Enero 2	Enero 3		
Diseño de encuesta														
Lanzamiento de encuesta														
Valoración de resultados encuesta														
Diseño de Mago de Oz-Entrevista														
Realización de Mago de Oz-Entrevista														
Valoración de entrevista														
Diseño del prototipo														
Programación del prototipo														
Diseño de Entrevista validación														
Realización de Validación-Entrevista														
Valoración de entrevista														
Redacción de la memoria														
Trimestres 2020	Febrero 1	Febrero 2	Febrero 3-Mayo 3	Junio 1	Junio 2	Junio 3-Agosto 2	Agosto 3	Septiembre 1	Septiembre 2	Septiembre 3	Octubre 1	Octubre 2	Octubre 3	Noviembre 1
Diseño de encuesta														
Lanzamiento de encuesta														
Valoración de resultados encuesta														
Diseño de Mago de Oz-Entrevista														
Realización de Mago de Oz-Entrevista														
Valoración de entrevista														
Diseño del prototipo														
Programación del prototipo														
Diseño de Entrevista validación														
Realización de Validación-Entrevista														
Valoración de entrevista														
Redacción de la memoria														

Anexo III. Documentación del Mago de Oz

Los siguientes documentos fueron producidos para la realización de la prueba Mago de Oz.

Esquema del experimento

El usuario estará sentado frente a una pantalla, sin periféricos excepto un micrófono. Con la aplicación de datos geográficos ya abierta. Toda interacción del usuario será por voz. En otro punto de la sala, al otro lado del elemento de separación en caso de ser necesario, un miembro del equipo, el “mago”, estará sentado frente al equipo al que dicha pantalla está conectada, con todos los periféricos a su disposición. Una vez el entrevistador introduzca las tareas al usuario, el mago responderá a las órdenes verbales simulándolas mediante los periféricos. El entrevistador tomará nota de cada orden y reacción percibida en los usuarios.



Materiales

- Un equipo informático y sus periféricos (ratón, teclado, pantalla).
- Una pantalla adicional, conectada a dicho equipo. Quizá sea necesario equipo adicional para realizar la conexión (extensión de cable o un segundo equipo).
- Micrófono.
- Cámara.
- Bolígrafo.
- Copia para los entrevistados (presentación, escenarios)
- Copia para el mago (escenarios y errores previstos)
- Plantillas para entrevistador (recogida de observaciones)
- Plantillas de recogida de datos del entrevistado
- Mesa y silla para el entrevistado, el micrófono, la cámara y su pantalla.
- Mesa y silla para el mago y su equipo.
- Elemento de separación (como un biombo) en caso de ser necesario.
- Copias del guión para el equipo.

Protocolo

Al inicio de la actividad el entrevistador recibirá al usuario y le comunicará las instrucciones de bienvenida. Le mostrará su estación y el usuario comenzará con los escenarios.

La única interacción directa con el usuario una vez iniciados los escenarios será en respuesta a preguntas del usuario. Si las dudas planteadas son acerca de cómo resolver los escenarios, se responderá que simplemente actúe como si estuviese en esa situación. En respuesta a dudas sobre ambigüedades de los escenarios, se pedirá que lo interprete como mejor considere. Si las preguntas son sobre la actividad en sí misma (p. ej. “¿Puedo parar cuando yo quiera?”, “¿Puedo mover el micrófono?”, “¿No puedo usar un teclado?”,...) se aclararán las dudas.

La otra forma de interacción con el usuario, a través de la pantalla, será llevada a cabo por el mago. El objetivo es simular un sistema de reconocimiento de voz que reconozca órdenes (básicas o complejas). Por tanto, el mago tratará de cumplir todas las órdenes del usuario sin tener en cuenta gramática o léxico. El mago empleará atajos de teclado siempre que sea posible, y el ratón lo usará solamente de manera orthogonal. Enmascarará el ruido de los periféricos de entrada mediante música con la excusa de poner dificultad al micrófono con ruido ambiental.

No obstante, si el usuario expresa una orden no considerada en el inventario, el mago hará sonar un sonido de error. La primera vez que dicha situación ocurra se explicará al usuario que el sistema no ha reconocido la orden, que lo intente de otra manera.

Tras la actividad, se entregará al usuario un formulario donde rellenará sus datos y responderá a las preguntas de evaluación en la misma estación.

Por último, se informará al usuario de las características del experimento, haciendo hincapié en el engaño de la simulación y la grabación de su voz a lo largo del experimento, que será destruida si el usuario así lo decide.

Guión de equipo

1. Entregar una copia de este guión a todo el equipo.
2. Organizar mesas y sillas de las estaciones del usuario y el mago de manera que no estén en contacto visual pero sí se oiga.
3. Conectar la pantalla del usuario al equipo del mago y comprobar que funciona correctamente.
4. Conectar el micrófono y la salida de audio si la pantalla carece de una al equipo del mago y colocarlos en la mesa del usuario. Asegurarse de que la conexión está suficientemente disimulada.
5. Comprobar el correcto funcionamiento de la grabación y el sonido de error. Ajustar volumen.
6. Colocar el guión de escenarios en la estación del usuario. Entregar al mago su copia de los escenarios. Preparar en un punto accesible las plantillas de recogida de datos.
7. Rellenar la plantilla del entrevistador con los datos de la siguiente entrevista (Nombre/Número, Fecha, Hora).
8. Hacer pasar al siguiente entrevistado, darle la bienvenida.

9. El mago simula la actividad del prototipo mientras el usuario desarrolla el escenario. El entrevistador anota sus observaciones sobre la reacción del usuario en cada momento en su plantilla.
10. Una vez el usuario termine o abandone las tareas, el entrevistador le proporciona un formulario de recogida de datos que incluye las preguntas de evaluación.
11. El entrevistador agradece de nuevo la participación, informa al usuario y lo acompaña fuera.
12. Repetir pasos 7-11 hasta haber concluido con los entrevistados.

Bienvenida

(Usuario de pie)

Bueno, lo primero de todo muchas gracias por venir y ayudarnos en este experimento.

¿Sabes qué es una aplicación de control de voz para mapas? Es una aplicación para controlar aplicaciones de mapas, como Google Maps, sin usar teclado, ni ratón, ni pantalla táctil.

Hemos desarrollado el prototipo de una aplicación de control por voz para mapas y queremos ver qué tal funciona. Ten en cuenta que es sólo un prototipo, así que puede que funcione lenta o tenga algún fallo.

Lo que queremos saber es qué tal lo hemos hecho nosotros, así que **recuerda siempre** que cualquier fallo es del prototipo y nuestro, no tuyo. Es muy importante que sepas que no hay respuestas correctas ni incorrectas, y que no te estamos evaluando a ti sino al prototipo. Si tienes dudas, si el sistema no responde... ya sabes, culpa nuestra.

Te sentarás en este ordenador que tiene el prototipo y te pediremos que lo pruebes. Detrás tienes una cámara que solo graba tu espalda y la pantalla. Aunque no se te va a ver la cara, te pedimos que firmes esta hoja de autorización para el uso de la grabación y protección de datos personales. (*Darle la hoja al usuario. Que la firme*).

La prueba será la siguiente. Te pasaremos 4 fichas con 4 tareas diferentes en 4 escenarios diferentes en los que te puedes encontrar en la vida real. Lee la tarea con calma y si tienes alguna duda no dudes en preguntar. Si puedo contestarte, te contestaré encantado y si no te lo haré saber. Te pedimos que intentes ejecutar la tarea solo con tu voz. Imagina que estás en el escenario descrito, tú solo y con el tiempo que necesites: no hay prisa. Simplemente haz lo que harías si contaras con esta aplicación. Eso sí, recuerda que la aplicación funciona sólo a través de la voz.

Cuando te parezca empezamos. Lee el escenario y cuando quieras puedes empezar a hablar a este micrófono, ya está funcionando.

Escenarios

Escenario 1: Un familiar tuyo se ha ido de viaje a Sri Lanka, y tienes curiosidad por ver la isla. Localiza la isla de Sri Lanka (al sureste de India).

Errores previstos: Mala pronunciación de una ciudad.

Batería de búsqueda: Colombo/Kandy/Jaffna/Batticaloa/Galle//Srilanka/Sri Lanka//India//Asia

Escenario 2: Estás buscando un sitio donde hacer un picnic en París, y quieres comparar qué zona es más frondosa. Localiza la zona de París que se vea más verde desde satélite.

Errores previstos: Buscar parque parís lleva a madrid.

Batería de búsqueda: París/Paris Francia/Parque París/París parque/ Parque//Bois de Boulogne/Bois de Vincennes

Escenario 3: Has quedado con un amigo en recogerle con el coche en la primera rotonda de Vitoria entrando por la A-132. Localízala en el mapa para saber dónde dirigirte.

Errores previstos:

Batería de búsqueda: Vitoria/a132/a132 Vitoria/Vitoria a132

Escenario 4: Tienes que llevar un paquete a la plaza del Pilar, y estás en el centro comercial Grancasa. Encuentra la ruta más directa y apunta las calles que seguirías.

Errores previstos: No encontrar nombres de calles, confundir Gomez de Avellaneda y Maria Zambrano.

Batería de búsqueda: Plaza del Pilar/Basílica del Pilar/el Pilar//Grancasa/Centro Comercial Grancasa/CC Grancasa//Puente de Santiago//Calle Gertrudis Gomez de Avellaneda/Gertrudis Gomez de Avellaneda/Calle Gomez de Avellaneda/Gomez de Avellaneda/Avenida de los Pirineos/Paseo Echegaray/Paseo Echegaray y Caballero/ Echegaray/Echegaray y Caballero/ Avenida Ranillas/ Ranillas

Entrevista posterior

-¿Cómo valorarías el sistema (de una a cinco estrellas)?

-¿Lo utilizarías en tu vida diaria?

-¿Te ha resultado natural la interacción con el sistema?

-¿Te has sentido nervioso durante la prueba?

-¿Has sentido frustración en algún momento? ¿En cual? ¿Por qué?

Anexo 1: Fichas para usuario

Vamos a plantearte unos escenarios, haz como si te encontraras en la situación descrita tranquilamente y fueras a hacer uso de la aplicación. Simplemente haz lo que harías si contaras con esta aplicación. Recuerda que la aplicación funciona sólo a través de la voz. Gracias por participar.

Escenario 1: Un familiar tuyo se ha ido de viaje a Sri Lanka, y tienes curiosidad por ver la isla. Localiza la isla de Sri Lanka (al sureste de India).

Escenario 2: Estás buscando un sitio donde hacer un picnic en París, y quieres comparar qué zona es más frondosa. Localiza la zona de París que se vea más verde desde satélite.

Escenario 3: Has quedado con un amigo en recogerle con el coche en la primera rotonda de Vitoria entrando por la A-132. Localízala en el mapa para saber dónde dirigirte.

Escenario 4: Tienes que llevar un paquete a la plaza del Pilar, y estas en el centro comercial Grancasa. Encuentra la ruta más directa y decide las calles que seguirías.

Anexo 2: Ficha Observador

Número de entrevista: ____

Nombre:

Notas:

Lenguaje no verbal, expresiones de frustración.

Una vez hayas acabado con los escenarios, por favor, contesta estas preguntas:

-¿Cómo valorarías el sistema (de una a cinco estrellas)?

★☆☆☆★☆

-¿Lo utilizarías en tu vida diaria?

Sí.

No.

-¿Te ha resultado natural la interacción con el sistema?

Sí.

No.

-¿Te has sentido nervioso durante la prueba?

Sí.

No.

-¿Has sentido frustración en algún momento? ¿En cuál? ¿Por qué?

Anexo IV. Resultados de la encuesta

A continuación se muestran las tablas de contingencia de los datos obtenidos en la encuesta.

	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
edad70	0	1	7	0	0	0	0	0	0	8	0	0	0	0	0	0
edad60	0	34	8	41	7	1	8	0	0	0	0	0	0	0	14	7
edad50	1	32	100	24	0	17	15	8	0	0	0	0	0	8	17	0
edad40	0	61	112	38	8	1	0	0	8	8	0	8	0	0	21	0
edad30	0	32	72	8	15	9	0	0	0	16	0	0	0	0	14	0
edad20	0	17	77	0	8	0	0	0	8	1	0	0	8	0	0	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Hombre	1	114	203	42	16	10	15	8	8	8	8	0	8	8	52	7
Mujer	0	63	173	69	22	26	8	0	8	17	0	0	8	0	14	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Universidad	1	160	347	103	30	27	16	0	16	24	8	8	8	8	66	7
Bachillerato	0	8	24	8	8	9	7	8	0	0	0	0	0	0	0	0
Obligatorio	0	9	5	0	0	0	0	0	0	1	0	0	0	0	0	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Diarario	0	45	117	24	23	1	0	0	8	9	0	8	8	8	20	0
Más de semana	1	72	101	38	15	17	23	8	8	16	0	8	0	0	18	7
Más de mes	0	59	90	33	0	17	0	0	0	0	0	0	0	0	21	0
Menos de mes	0	1	8	16	0	1	0	0	0	0	0	0	0	0	7	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Usa asistente	0	38	100	33	7	17	7	0	8	25	0	8	8	8	20	0
o usa asistente	1	139	276	78	31	19	16	8	8	0	8	0	0	0	46	7
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Estudios	1	94	147	30	31	24	8	0	8	16	0	0	8	8	18	7
Trabajo	0	75	204	57	12	15	0	8	9	0	8	0	0	48	0	0
Recreativo	0	8	25	16	0	0	0	8	0	0	0	0	0	0	0	0
Nuoso	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Instruction	1	101	262	64	7	18	0	0	0	25	8	8	8	52	0	0
Neutral	0	76	107	47	31	9	8	8	8	0	0	0	0	8	14	0
Películon	0	0	7	0	0	9	15	0	0	0	0	0	0	0	7	0
	Mirar	Mover	Desplazar	Ir	Pán	Mostar	Ver	imbio visi	Scroll	Desistar	Dirigir	Revisar	Avanzar	Navegar	Direccional	Enseñar
Relativo	1	95	109	50	0	9	0	0	16	25	0	8	8	8	59	0
Caidinal	0	80	266	61	38	27	23	8	0	0	8	0	0	0	7	0

Anexo V. Estado del Arte

Anexo V. I. Accesibilidad Web

La primera definición que se puede atribuir a la accesibilidad proviene del sector de la geografía y se refiere a la facilidad de acceder a un lugar, una persona o una cosa. Con el advenimiento de la sociedad de la información, el concepto de accesibilidad ha evolucionado a fin de tener en consideración nuevas realidades. En efecto, se observa que la movilidad, la proximidad y la distancia ya no son elementos esenciales de la definición de accesibilidad, o más bien, que la accesibilidad en el espacio físico se halla ahora complementada por la accesibilidad en el espacio virtual, desafiando los principios de la distancia, de la proximidad o de la interacción espacial.

La accesibilidad al medio físico se refiere a la calidad que tienen los espacios para que cualquier persona, incluso las afectadas de discapacidades de movilidad o comunicación, puedan:

- Llegar a todos los lugares y edificios sin sobreesfuerzos y con autonomía.
- Acceder a los establecimientos de uso público y los servicios que presten en condiciones de seguridad y autonomía.

De forma paralela a la accesibilidad al medio físico, la accesibilidad a la web y a Internet en general (medio electrónico), se refiere al conjunto de elementos que facilitan el acceso a la información web de todas las personas en igualdad de condiciones, y ello independientemente de la tecnología que utilicen (ordenador, PDA, teléfono y otros) y de la discapacidad del usuario (física, psíquica, sensorial y otras).

En ISO/TC 16027, se define accesibilidad como la facilidad de uso de forma eficiente, eficaz y satisfactoria de un producto, servicio, entorno o instrumento por personas que poseen diferentes capacidades. Por tanto, accesibilidad electrónica hace referencia a que los productos y servicios electrónicos puedan ser utilizados por los usuarios con efectividad, eficiencia y satisfacción en un contexto de uso determinado. Por ejemplo: accesibilidad de los equipos informáticos (hardware y software), accesibilidad web, accesibilidad de la televisión digital, accesibilidad de la telefonía móvil, accesibilidad de los productos y servicios de domótica, así como otros servicios característicos de la sociedad de la información.

La accesibilidad web trata de aspectos relacionados con la codificación y la presentación de información en el diseño de un sitio web, que va a permitir que las personas con algún tipo de limitación puedan percibir, entender, navegar e interactuar de forma efectiva con la Web, así como crear y aportar contenido. Actualmente, la mayoría de los sitios web presentan barreras de accesibilidad, lo que hace difícil e incluso imposible su utilización. Sin embargo, si los sitios web y el software web, fueran accesibles, las personas con discapacidad podrían utilizar estos servicios de forma eficaz.

La accesibilidad es una problemática tratada y considerada ampliamente tanto en ámbitos nacionales como internacionales. Por ello se han creado una serie de recursos que permiten especificar las características que han de cumplir los contenidos disponibles mediante tecnologías Web en Internet, Intranets y otro tipo de redes informáticas, para que puedan ser

utilizados por la mayor parte de las personas, incluyendo personas con discapacidad y personas de edad avanzada, de forma autónoma o mediante las ayudas técnicas pertinentes.

Entre estos recursos destacan:

- Norma EN 301 549:2018, versión 2.1.2¹⁴, "Requisitos de accesibilidad para productos y servicios TIC", especifica los requisitos funcionales de accesibilidad aplicables a los productos y servicios que incluyan TIC (sitios web, software, apps nativas, documentos, hardware, etc). Además de describir los procedimientos de prueba y la metodología de evaluación a seguir para cada requisito de accesibilidad.
- Pautas de Accesibilidad para el Contenido Web (WCAG) 2.1. Las Web Content Accessibility Guidelines (WCAG) 2.1¹⁵ son la última versión de las pautas de accesibilidad del contenido en la Web del W3C (World Wide Web Consortium). La nueva versión 2.1.2 de la EN 301-549:2018 está alineada con las WCAG 2.1 y por lo tanto introduce todos los cambios incorporados a esta última: nuevos criterios de conformidad, nueva pauta "2.5 Modalidades de entrada", etc.
- Norma UNE 139803:2012¹⁶. Requisitos de Accesibilidad para contenidos en la web. Es una norma española que establece los requisitos de accesibilidad para los contenidos web. En cuanto a sus requisitos referencia completamente a las Pautas de Accesibilidad para el contenido web WCAG2.0 de la Iniciativa para la Accesibilidad Web (WAI) del W3C. Por lo tanto, hay una equivalencia directa entre ellas.
- UNE-EN 301 549¹⁷, "Requisitos de accesibilidad de productos y servicios TIC aplicables a la contratación pública en Europa" – "Accessibility requirements suitable for public procurement of ICT products and services in Europe". Esta norma establece los requisitos funcionales que garantizarán que los productos y servicios TIC sean accesibles para todas las personas; por ejemplo, desde un teléfono móvil, hasta ordenadores, pasando por páginas web; así, los requisitos para la web se basan en las directrices de accesibilidad de contenidos web WCAG 2.0, desarrolladas por W3C, estando en línea con los indicados en la norma española UNE 139803, 'Requisitos de accesibilidad para contenidos en la Web'.
- WAI-ARIA¹⁸ Accessible Rich Internet Applications, es una recomendación internacional que define cómo hacer el contenido y las aplicaciones webs más accesibles para personas con discapacidad. Es de especial ayuda para la generación de contenido dinámico y para interfaces de usuario avanzadas desarrolladas con AJAX, HTML y Javascript. Se engloba en la Iniciativa para la Accesibilidad Web (WAI) del W3C.

El uso de asistentes virtuales se ha conceptualizado como una herramienta destacada de interacción con los sistemas de información. Se trata de un nuevo modelo de acceso para

¹⁴ <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0055846>

¹⁵ <https://www.w3.org/TR/WCAG21/>

¹⁶ <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0049614>

¹⁷ https://ec.europa.eu/eip/ageing/standards/ict-and-communication/accessibility-and-design-all/en-3015492015_en

¹⁸ <https://www.w3.org/WAI/standards-guidelines/aria/>

mejorar la accesibilidad de muchas personas a los mismos que, de otro modo, no podrían contar con esta posibilidad de interacción.

Anexo V. II Accesibilidad en los sistemas de información geográfica

La información geográfica, por su propia naturaleza, ha resultado siempre un elemento de difícil entronque con las normativas y estándares de accesibilidad. De forma explícita, en el ámbito español, el Real Decreto 1112/2018 sobre accesibilidad de los sitios web y aplicaciones para dispositivos móviles del sector público¹⁹ manifiesta que:

"Están EXCLUIDOS del ámbito de este Real Decreto los siguientes tipos de contenido:

...

Servicios de mapas y cartografía en línea, siempre y cuando la información esencial se proporcione de manera accesible digitalmente en el caso de mapas destinados a fines de navegación.

..."

No obstante, no se trata de una batalla que haya sido dada por perdida y en este sentido hay trabajos de investigación en marcha desatinados a evaluar las limitaciones y cómo sortearlas (Calle-Jimenez & Luján-Mora, Sieber et al. 2016, Calle-Jimenez et al. 2019, Chan & Dascalu 2017, Götzelmann & Eichler 2016, Brulé et al. 2016).

Una forma de soslayar estas limitaciones de acceso es mediante el uso de otro tipo de interfaces de acceso tales como el acceso por voz.

Anexo V. III. Sistemas de aprendizaje

En el marco de este proyecto será necesario convertir las peticiones por voz en comandos de los sistemas de información geográfica. Para ello se propone utilizar técnicas de aprendizaje automático para el etiquetado de las solicitudes de acuerdo a los posibles comandos.

El aprendizaje automático es una rama de la inteligencia artificial que permite el desarrollo de sistemas capaces de generalizar patrones a partir de ejemplos. A partir de colecciones de datos clasificadas, estos sistemas permiten identificar automáticamente la categoría de nuevos elementos no vistos anteriormente. Estas técnicas se han utilizado durante años con éxito en campos como la clasificación de imágenes (Rosebrock, 2017), reconocimiento del habla (Yu et. al., 2016), procesamiento de lenguaje natural (Steven et. al., 2009) y minería de datos (Leskovec et. al., 2014).

En este proyecto se va a trabajar con una rama del aprendizaje automático que se denomina aprendizaje profundo (o deep learning). Esta rama se basa en asimilar

¹⁹ <https://www.boe.es/buscar/act.php?id=BOE-A-2018-12699>

representaciones de datos. Una observación (por ejemplo, una imagen) puede ser representada en muchas formas (por ejemplo, un vector de píxeles), pero algunas representaciones hacen más fácil aprender tareas de interés (por ejemplo, "¿es esta imagen una cara humana?") sobre la base de ejemplos. La investigación en esta área intenta definir qué representaciones son mejores y cómo crear modelos para reconocer estas representaciones.

No existe una única definición de aprendizaje profundo. En general se trata de una clase de algoritmos ideados para el aprendizaje automático. A partir de este punto común, diferentes publicaciones se centran en distintas características, por ejemplo:

- Usar una cascada de capas con unidades de procesamiento no lineal para extraer y transformar variables. Cada capa usa la salida de la capa anterior como entrada. Los algoritmos pueden utilizar aprendizaje supervisado o aprendizaje no supervisado, y las aplicaciones incluyen modelización de datos y reconocimiento de patrones.
- Estar basados en el aprendizaje de múltiples niveles de características o representaciones de datos. Las características de más alto nivel se derivan de las características de nivel inferior para formar una representación jerárquica.
- Aprender múltiples niveles de representación que corresponden con diferentes niveles de abstracción. Estos niveles forman una jerarquía de conceptos.

Todas estas maneras de definir el aprendizaje profundo tienen en común: múltiples capas de procesamiento no lineal; y el aprendizaje supervisado o no supervisado de representaciones de características en cada capa. Las capas forman una jerarquía de características desde un nivel de abstracción más bajo a uno más alto.

Los algoritmos de aprendizaje profundo contrastan con los algoritmos de aprendizaje poco profundo por el número de transformaciones aplicadas a la señal mientras se propaga desde la capa de entrada a la capa de salida. Cada una de estas transformaciones incluye parámetros que se pueden entrenar como pesos y umbral. No existe un estándar de facto para el número de transformaciones (o capas) que convierte a un algoritmo en profundo, pero la mayoría de investigadores en el campo considera que aprendizaje profundo implica más de dos transformaciones intermedias.

Con los avances en el aprendizaje profundo, los modelos de redes neuronales han alcanzado resultados de vanguardia en muchas tareas de etiquetado de secuencias (Lample et al., 2016; Ma y Hovy, 2016). Diferentes arquitecturas neuronales como la Red Neural Convolutacional (CNN), las redes neuronales recurrentes (Goller y Kuchle, 1996), junto con sus variantes como la "Long short-term Memory" (LSTM) (Hochreiter y Schmidhuber, 1997) han ganado prominencia en los últimos años. Collobert et al., 2011 propuso una arquitectura neural seminal para el etiquetado de secuencias. Captura la información de la secuencia de palabras con una CNN de una capa basada en incrustaciones de palabras previamente entrenadas, seguida de una capa de salida CRF (Lafferty y Mccallum, 2001). Dos Santos et al., 2015 extendió este modelo integrando características de CNN a nivel de caracteres. Los LSTM han mostrado un gran éxito resolviendo tareas de etiquetado de secuencias y logrando un rendimiento competitivo frente a los modelos tradicionales (Graves et al., 2013; Huang et al., 2015; Hu et al., 2016).

Para representar palabras individuales, los modelos de etiquetado supervisado requieren grandes cantidades de datos etiquetados manualmente para lograr un buen

rendimiento. Estos datos son difíciles de obtener. Sin embargo, es posible aprovechar datos muy grandes sin etiquetar para aprender características de palabras para enriquecer modelos más simples obtenidos a partir de pequeños estándares de oro. Estas representaciones se denominan “word embeddings” (Mikolov et al., 2011) y funcionan bien en una variedad de tareas (Collobert et al., 2011). El uso de “word embeddings” es un tipo de mapeo que permite que las palabras con significado similar tengan una representación similar. Una aproximación muy interesantes desde el punto de vista de este proyecto es la basada en el uso de Word2vec con Skipgram (Rong, 2016) para inicializar los “word embeddings”. Las representaciones de secuencias de palabras se concatenan con las representaciones de secuencias de caracteres.

Similar a las secuencias de caracteres, podemos modelar información de secuencias de palabras a través de estructuras LSTM o CNN. Los LSTM son una familia de redes neuronales que operan con datos secuenciales que capturan eficazmente las dependencias de largo alcance (Liu et al., 2017; Lample et al., 2016; Ma y Hovy, 2016). Los LSTM unidireccionales sufren de la debilidad de no utilizar la información contextual futura. Sin embargo, el LSTM bidireccional (BLSTM) aborda este problema utilizando dos LSTM independientes (hacia adelante y hacia atrás) en los que uno procesa la secuencia de entrada en la dirección hacia adelante, mientras que el otro procesa la secuencia de entrada en la dirección inversa.

Por otro lado, las CNN combinan diferentes ideas arquitectónicas para extraer características horizontalmente de múltiples palabras, lo que permite a la red extraer un estilo lingüístico de mayor nivel. El tamaño del “kernel” en la capa convolucional define el número de palabras a considerar, proporcionando un parámetro de agrupación (Collobert et al., 2011; dos Santos et al., 2015; Strubell, 2017).

Finalmente, la capa de inferencia toma las representaciones de la secuencia de palabras extraídas como características y asigna etiquetas a la secuencia de palabras. Un modelo de etiquetado muy simple pero eficaz es utilizar la capa oculta como características para tomar decisiones de etiquetado independientes para cada salida utilizando una capa de tipo Softmax (Ling et al., 2015).

A pesar del éxito de Softmax en problemas sencillos como el “Part-Of-Speech tagging” (Collobert et al., 2011), la suposición de independencia en las etiquetas de salida limita su aplicación en otras tareas comunes de procesamiento de lenguaje natural en las que existen fuertes dependencias entre las etiquetas de salida (p. ej., reconocimientos de entidades nombradas, etiquetado semántico, etc.) (Huang et al., 2015; Yang et al., 2018). En otras palabras, una “gramática” que capture las correlaciones entre etiquetas adyacentes impone restricciones imposibles de modelar con Softmax incluso cuando se utiliza BLSTM. Por esta razón, es conveniente utilizar modelos estadísticos lineales como el CRF. En última instancia es recomendable comparar el uso de Softmax y CRF como implementación de capas de inferencia.