



Departamento de  
Informática e  
Ingeniería de Sistemas  
Universidad Zaragoza



Escuela de  
Ingeniería y Arquitectura  
Universidad Zaragoza



Universidad  
Zaragoza



## Proyecto Fin de Carrera

# Desarrollo de un Gestor Documental para un Medio de Comunicación

**Sandra Escudero Sánchez**

Director: María Dolores González González

Co-Director: Ángel Luis Garrido Marín

Aragón Radio

Ponente: Sergio Ilarri Artigas

Departamento de Informática e Ingeniería de Sistemas

Ingeniería Informática

Escuela de Ingeniería y Arquitectura

Universidad de Zaragoza

Junio 2014



# Desarrollo de un Gestor Documental para un Medio de Comunicación

## RESUMEN

Actualmente, las tareas de producción, gestión y manejo de grandes cantidades de información en ámbitos de documentación conllevan un gran trabajo asociado. Por ello, este proyecto desarrollará un sistema de gestión de documentación para Aragón Radio, la Radio Autonómica de Aragón, que facilite dichas tareas. Dicho proyecto deberá cubrir las necesidades de archivar, documentar y buscar los contenidos que se generan en el departamento de documentación de la empresa.

A día de hoy, Aragón Radio dispone de una herramienta que permite realizar la mayoría de las gestiones que el departamento de documentación necesita, pero esta posee una interfaz de usuario poco intuitiva, con numerosos fallos de integridad de datos, y problemas de seguridad, dificultando en todo momento el tedioso trabajo de los documentalistas. Así mismo, la herramienta no dispone de un tesoro multinivel <sup>1</sup> y bien formado, lo cual convierte en una tarea ardua la creación de nuevos documentos y la categorización de estos adecuadamente. Con la nueva aplicación a desarrollar se quiere mejorar la inteligencia y la calidad del sistema, es decir, proveer a Aragón Radio de un gestor documental que le ayude a realizar el trabajo diario del departamento de documentación de manera sencilla, y que además añada opciones muy útiles (la aplicación actual carece de ellas), como por ejemplo: un generador de informes en formato PDF, un generador de resúmenes automático, unos buscadores rápidos, posibilitar la gestión de un tesoro multinivel con relaciones padre-hijo (entre otras características imprescindibles para facilitar el trabajo diario del departamento de documentación).

Para el desarrollo de la aplicación se ha utilizado en su mayor parte el lenguaje C# .Net junto a la tecnología de Microsoft WPF (Windows Presentation Foundation) para la creación de las interfaces, exceptuando el desarrollo de un programa capaz de generar la nueva base de datos y migrar los registros antiguos, para el cual se ha utilizado el lenguaje Visual Basic .Net.

Como balance general del proyecto apuntar que se ha implementado un gestor de documentación con una interfaz intuitiva que permite realizar todas las tareas necesarias en el departamento de documentación de Aragón Radio. Han existido una serie de necesidades orientadas a la recuperación de información que han requerido del uso de herramientas I+D más específicas en la línea de generación de resúmenes automáticos. Esto ha dado fruto a varios artículos de investigación donde he participado activamente como autora. El proyecto se encuentra enmarcado en un convenio de colaboración entre Aragón Radio y el grupo de Sistemas de Información Distribuida (SID) de la Universidad de Zaragoza.

---

<sup>1</sup>Tesoro: (Del lat. *thesaurus*) Conjunto de términos empleados para representar un concepto.



*A mis padres porque sin ellos no hubiese podido cumplir mi sueño. A mi hermana por su apoyo. A Noa por darme suerte. A Adrián por estar siempre a mi lado. A mis amigos/as por los buenos ratos.*

## **Agradecimientos**

A Lola por su simpatía y su apoyo. A Ángel por confiar en mí y mostrarme el camino en mis momentos de confusión. A Sergio por sus recomendaciones y su paciencia. A Eduardo por permitirme entrar en el mundo de la investigación. A Álvaro por estar siempre disponible para ayudarme. Y en especial a María por sus consejos y por conseguir que el trabajo sea más ameno.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto . . . . .	1
1.2. Situación actual . . . . .	2
1.3. Identificación de necesidades y problemas . . . . .	4
1.4. Tecnologías utilizadas . . . . .	5
1.5. Fases del proyecto . . . . .	6
1.6. Organización de la memoria . . . . .	7
<b>2. Gestor de Documentación</b>	<b>9</b>
2.1. Ámbito del proyecto . . . . .	9
2.2. Objetivos . . . . .	9
2.3. Herramienta . . . . .	11
<b>3. Arquitectura</b>	<b>15</b>
3.1. General . . . . .	15
3.2. Unidades funcionales del sistema . . . . .	16
3.2.1. Base de datos . . . . .	17
3.2.2. Servidor . . . . .	18
3.2.3. Servicio PLN . . . . .	18
3.2.4. Servicio Windows X-FRAME . . . . .	22
<b>4. Interfaces</b>	<b>25</b>
4.1. Pantalla principal de herramienta . . . . .	25
4.2. Opciones de la herramienta . . . . .	25
4.2.1. Búsqueda de contenido . . . . .	26
4.2.2. Búsqueda de Audios . . . . .	27
4.2.3. Búsqueda de contactos . . . . .	29
4.2.4. Buscador general . . . . .	29
4.2.5. Perfil . . . . .	30
4.2.6. Crear nuevo contenido . . . . .	31
4.2.7. Gestión de usuarios . . . . .	32

4.2.8. Gestión de Tesauros . . . . .	32
4.2.9. Configuración . . . . .	35
<b>5. Planificación y conclusiones</b>	<b>37</b>
5.1. Estimación de horas de trabajo . . . . .	37
5.2. Conclusiones generales . . . . .	38
5.3. Trabajo futuro . . . . .	40
5.4. Conclusión personal . . . . .	41
<b>A. Análisis y diseño de la aplicación</b>	<b>45</b>
A.1. Catálogo de requisitos del sistema . . . . .	45
A.2. Descripción del sistema . . . . .	47
A.3. Diagramas UML de Análisis . . . . .	48
A.3.1. Casos de uso . . . . .	48
A.3.2. Diagramas de secuencia de eventos . . . . .	59
A.3.3. Clases y objetos . . . . .	65
A.3.4. Navegación . . . . .	71
A.4. Base de datos . . . . .	72
A.4.1. Migración de la base de datos . . . . .	72
A.4.2. Nueva base de datos . . . . .	73
<b>B. Unidad de Resúmenes Automáticos</b>	<b>85</b>
B.1. Factores a tener en cuenta . . . . .	86
B.2. Estado del arte . . . . .	86
B.3. Arquitectura . . . . .	88
B.3.1. Entradas y salidas del sistema . . . . .	89
B.3.2. Etapa de preproceso . . . . .	90
B.3.3. Etapa de proceso . . . . .	90
<b>C. Pruebas</b>	<b>95</b>
C.1. Pruebas unitarias a realizar . . . . .	95
C.2. Pruebas de integración . . . . .	103
C.3. Pruebas del sistema y de aceptación . . . . .	103
<b>D. Glosario</b>	<b>105</b>
<b>E. Manual de usuario</b>	<b>107</b>
E.1. Descripción . . . . .	107
E.2. Guía de instalación . . . . .	108
E.3. Tutorial completo . . . . .	111
E.3.1. Buscador de contenido . . . . .	114
E.3.2. Buscador de audios . . . . .	125

E.3.3. Buscador de contactos . . . . .	129
E.3.4. Buscador general . . . . .	131
E.3.5. Crear nuevo contenido . . . . .	132
E.3.6. Gestión de usuarios . . . . .	137
E.3.7. Gestión de tesauros . . . . .	139
E.3.8. Configuración . . . . .	141
<b>F. Artículos publicados</b>	<b>145</b>
F.0.9. TM-Gen: A Topic Map Generator from Text Documents (ICTAI 2013) (Página 164) . . . . .	145
F.0.10. Obtaining Knowledge from the Web using Fusion and Summarization Techniques (FUSION 2014) (Página 170) . . . . .	145



# Índice de figuras

3.1. Arquitectura del sistema . . . . .	16
3.2. Diagrama funcionamiento servicio Windows para X-FRAME . . . . .	23
4.1. Pantalla de acceso al sistema . . . . .	26
4.2. Pantalla de inicio del sistema . . . . .	26
4.3. Pantalla de búsqueda de contenido . . . . .	27
4.4. Pantalla listado de resultados . . . . .	28
4.5. Pantalla de búsqueda de audio . . . . .	28
4.6. Pantalla de búsqueda de contactos . . . . .	29
4.7. Pantalla del buscador general . . . . .	30
4.8. Pantalla de creación de nuevo contenido . . . . .	31
4.9. Pantalla de creación antigua . . . . .	32
4.10. Pantalla de gestión de usuarios . . . . .	33
4.11. Pantalla de gestión de tesauros . . . . .	34
4.12. Pantalla de gestión de tesauros temático . . . . .	34
4.13. Pantalla de gestión antigua de tesauros . . . . .	35
5.1. Tabla fechas de la planificación . . . . .	38
5.2. Diagrama de Gantt con el desarrollo del proyecto . . . . .	38
A.1. Diagrama de despliegue del sistema . . . . .	47
A.2. Diagrama de casos de uso de la herramienta . . . . .	49
A.3. Caso de uso <i>buscar</i> (1.1, 1.2, 1.3, 1.4) . . . . .	52
A.4. Caso de uso <i>listar</i> (1.0.3) . . . . .	54
A.5. Caso de uso <i>más información</i> (1.0.3.3) . . . . .	56
A.6. Caso de uso <i>gestión usuarios</i> (1.6) . . . . .	58
A.7. Caso de uso <i>crear contenido</i> (1.5) . . . . .	58
A.8. Caso de uso <i>gestión tesauros</i> (1.7) . . . . .	59
A.9. Ejemplo de traza de eventos para la iniciación del sistema . . . . .	59
A.10. Diagrama de secuencia <i>búsqueda</i> . . . . .	60
A.11. Diagrama de secuencia <i>listado de resultados</i> . . . . .	61
A.12. Diagrama de secuencia <i>información del resultado</i> . . . . .	62

A.13. Diagrama de secuencia <i>creación de contenido</i> . . . . .	63
A.14. Diagrama de secuencia <i>gestión de usuarios</i> . . . . .	64
A.15. Diagrama de secuencia <i>gestión de tesauros</i> . . . . .	65
A.16. Diagrama de clases ventana principal . . . . .	66
A.17. Diagrama de clases y objetos para las opciones de búsqueda . . . . .	66
A.18. Diagrama de clases y objetos para la opción <i>crear contenido</i> . . . . .	67
A.19. Diagrama de clases y objetos para la opción <i>gestión de tesauros</i> . . . . .	68
A.20. Diagrama de clases para la opción <i>configurar</i> . . . . .	69
A.21. Diagrama de clases y objetos para el resto de opciones del sistema . . . . .	70
A.22. Diagrama de navegación del sistema . . . . .	71
A.23. Esquema conceptual de la base de datos antigua . . . . .	73
A.24. Esquema relacional de la base de datos antigua . . . . .	74
A.25. Esquema conceptual de la base de datos nueva . . . . .	75
A.26. Modelo lógico de la base de datos nueva . . . . .	76
B.1. Esquema general de la generación de resúmenes . . . . .	86
B.2. Diagrama etapas del generador de resúmenes . . . . .	89
B.3. Diagrama clúster de similitud . . . . .	93
B.4. Diagrama clúster de keywords . . . . .	93
E.1. Iniciando configuración . . . . .	108
E.2. Instalación gestor documental . . . . .	109
E.3. Seleccionar directorio instalación . . . . .	109
E.4. Icono de la aplicación . . . . .	110
E.5. Pantalla de acceso a la herramienta . . . . .	111
E.6. Pantalla principal de la herramienta . . . . .	112
E.7. Flechas de navegación . . . . .	112
E.8. Opciones del botón ayuda . . . . .	113
E.9. Pantalla perfil del usuario . . . . .	113
E.10. Pantalla buscador de contenidos . . . . .	114
E.11. Pestaña búsqueda de términos . . . . .	115
E.12. Ventana añadir términos del tesauro . . . . .	115
E.13. Desplegable de categorías . . . . .	116
E.14. Pantalla con listado de resultados . . . . .	117
E.15. Pantalla guardar fichero PDF . . . . .	118
E.16. Pantalla con listado de resultados . . . . .	118
E.17. Pantalla con listado de resultados . . . . .	119
E.18. Pantalla información del documento . . . . .	120
E.19. Pantalla buscador de contenidos . . . . .	121
E.20. Pantalla pestaña tesauros en información de contenido . . . . .	122
E.21. Pantalla editar tesauros en información de contenidos . . . . .	122

E.22.Pantalla fichas de audio vinculadas . . . . .	123
E.23.Pantalla datos ficha audio . . . . .	124
E.24.Pantalla reproducir ficha de audio . . . . .	124
E.25.Pantalla buscador fichas de audio . . . . .	125
E.26.Pantalla resultados búsqueda fichas audio . . . . .	126
E.27.Pantalla información ficha de audio . . . . .	127
E.28.Pantalla información documento vinculado . . . . .	128
E.29.Pantalla buscador de contactos . . . . .	129
E.30.Pantalla información del contactos . . . . .	130
E.31.Pantalla buscador general . . . . .	131
E.32.Pantalla listado resultados buscador general . . . . .	132
E.33.Pantalla desplegable seleccionar categoría . . . . .	133
E.34.Pantalla crear nuevo contenido . . . . .	133
E.35.Pantalla pestaña ficha de audio . . . . .	134
E.36.Pantalla seleccionar ficha de audio . . . . .	135
E.37.Pantalla añadir ruta . . . . .	136
E.38.Pantalla gestión de usuarios . . . . .	137
E.39.Pantalla editar usuario . . . . .	138
E.40.Pantalla añadir nuevo usuario . . . . .	138
E.41.Pantalla gestión de tesauros . . . . .	139
E.42.Pantalla inserción nuevo término . . . . .	140
E.43.Pantalla edición de un término . . . . .	140
E.44.Menú opciones de configuración . . . . .	141
E.45.Pantalla configurar tipos de documentos . . . . .	142
E.46.Pantalla configurar categorías . . . . .	143
E.47.Pantalla editar categorías . . . . .	143



# Índice de tablas

5.1. Resumen temporal . . . . .	37
A.1. Requisitos de Cliente. . . . .	47
A.2. Tabla entidad Usuario . . . . .	77
A.4. Tabla entidad Documento . . . . .	79
A.5. Tabla entidad Tipo_Doc . . . . .	79
A.6. Tabla entidad Fichero . . . . .	80
A.3. Tabla entidad Privilegio . . . . .	81
A.7. Tabla entidad Categorías . . . . .	81
A.8. Tabla entidad CamposCategoría . . . . .	83
A.9. Tabla entidad Tesauro . . . . .	83



# Capítulo 1

## Introducción

El presente documento representa la memoria de trabajo del Proyecto Final de Carrera “*Desarrollo de un Gestor Documental para un Medio de Comunicación*”, realizado por la alumna Sandra Escudero Sánchez y desarrollado dentro del marco de Aragón Radio, la Radio Autonómica de Aragón. El proyecto ha sido dirigido por María Dolores González, la actual documentalista de Aragón Radio, quién se ha encargado de aportar sus conocimientos en documentación, ayudando a tomar decisiones y supervisando las fases de implantación y pruebas. El codirector del proyecto ha sido Ángel Luis Garrido, quien ha supervisado los desarrollos a nivel técnico al ser el Ingeniero encargado del proyecto, supervisando por tanto las fases de análisis, diseño y desarrollo. El ponente del proyecto ha sido el profesor titular del Departamento de Informática e Ingeniería de Sistemas de la Universidad de Zaragoza, Sergio Ilarri. Por último, comentar que tanto mi codirector como mi ponente me han dirigido en la realización de artículos de investigación a nivel internacional. En este proyecto se ha desarrollado un gestor documental para Aragón Radio cuya finalidad principal es dotar al departamento de documentación de una herramienta sencilla, completa, y segura que le asista en su trabajo diario y les permita aumentar su productividad.

### 1.1. Contexto

Los constantes cambios tecnológicos y sociales, la creciente cantidad de documentación en todo tipo de formatos (papel o electrónico), así como la demanda informativa que los usuarios solicitan de la misma, hacen necesarias las tareas de uno o varios documentalistas. Y es que la información y la documentación que dispone una empresa será la base para la generación de conocimiento. Por ese motivo, una empresa dedicada a la información y al conocimiento, como es el caso de un medio de comunicación, necesita una herramienta que ayude al personal a

localizar la información demandada y a gestionarla en el menor tiempo posible.

Aragón Radio, la Radio Autonómica de Aragón, empezó sus emisiones en 2005 con una clara vocación pública y de proximidad. Aragón Radio tiene consolidada una amplia red de cobertura en todo el territorio de la comunidad que permite difundir la señal las 24 horas del día a través de las 46 frecuencias activas en Aragón.

## 1.2. Situación actual

Actualmente, Aragón Radio dispone de una herramienta que permite realizar la mayoría de las tareas necesarias de producción, gestión y manejo de grandes cantidades de información, haciendo uso de dos gestores de contenido, lo cual dificulta el uso de la herramienta y conlleva un gran trabajo asociado al departamento de documentación. La existencia de dos gestores de contenido se debe a que en el año 2010 la herramienta de documentación se vio modificada por una nueva, y se procedió a guardar todos los registros de la herramienta antigua en un gestor de contenido, y los registros de la nueva herramienta en otro, provocando la existencia actual de dos gestores de contenido dentro de la herramienta con distinto nombre, y distinto contenido, que hacen uso de la misma base de datos. Esta característica del sistema entorpece las búsquedas y la gestión del mismo provocando inconsistencias, dado que la base de datos del sistema contiene redundancias, y ambigüedades provocadas al fusionar dos herramientas (la antigua y la nueva) en una sola. Como ejemplo de los problemas que podemos encontrar en la herramienta provocados por esta característica del sistema son: documentos sin título, varios campos idénticos (con el mismo contenido) en la base de datos aunque con nombre distinto, documentos que tienen relacionados términos del tesauro cuyos términos en realidad no se encuentran en el tesauro, lo cual provoca fallos de integridad en los datos y por consiguiente errores al realizar búsquedas de contenido. Además de lo anterior, la base de datos permitía insertar varios usuarios con un mismo nombre, así como añadir usuarios sin contraseña, lo cual podía provocar serios problemas de seguridad dentro de la herramienta dado que cada usuario tiene unos privilegios activos los cuales permiten a un usuario acceder a ciertas opciones o tener el acceso restringido a otras.

El departamento de documentación se encarga de almacenar en el sistema todo el contenido que sale de redacción. Este contenido está formado por documentos de texto con transcripciones que pertenecen a una categoría en concreto (programas, informativos, biografías, ...), además cada documento tiene vinculada una o varias fichas de audio que contienen un fichero MP3 asociado a las mismas. Ese fichero MP3 es el que representa el contenido almacenado en la ficha de audio (por ejemplo una canción o una entrevista). Estos ficheros se gestionan con una herramienta

conocida como X-FRAME, que es utilizada por el equipo técnico y por el departamento de documentación. En concreto, el departamento de documentación la utiliza para poder documentar y archivar los ficheros de audio. Esta tarea conlleva mucho tiempo añadido dado que se debe acceder a un servidor remoto para poder realizar las gestiones, y posteriormente esperar los resultados para poder llevarlos a los ordenadores clientes del departamento de documentación.

Dado que Aragón Radio dispone de una gran cantidad de tipos distintos de contenido (documentos pertenecientes a distintas categorías, fichas de audio y fichas de contactos personales), es necesario disponer de una herramienta que permita encontrar el contenido deseado de manera rápida y sencilla. Sin embargo, la herramienta actual no permite realizar búsquedas de manera sencilla, ya que al existir dos gestores de contenido diferentes se debe conocer con exactitud en cual de los dos se encuentra el contenido a buscar antes de llevar a cabo la búsqueda, dado que la herramienta te obliga a indicar en que gestor de los dos disponibles se desea buscar. Esta característica es un problema cuando no se conoce con exactitud cual de los dos gestores contiene el contenido deseado (como pasa en la mayoría de los casos). Además en diversas ocasiones el buscador falla dependiendo del contenido introducido en la búsqueda. Estos problemas provocan una gran pérdida de tiempo cuando se quiere buscar algo en concreto, por ejemplo cuando se necesita buscar el número de teléfono de una persona para poder realizar a esta una entrevista. También destacar que si el departamento de documentación desea realizar una síntesis de noticias, para crear una biografía, un cronología, u otro tipo de contenido que necesita de la recopilación de información, se debe buscar el contenido deseado con los buscadores y copiar manualmente todo el contenido encontrado en un documento TXT, para posteriormente agrupar todas las ideas encontradas y generar un resumen con lo más relevante a mano.

Junto a lo anterior, destacar que la interfaz de usuario actual es poco intuitiva y contiene numerosos fallos, como por ejemplo en la introducción de tesauros (lista de términos que emplea Aragón Radio para catalogar el contenido) esta opción solo permite seleccionar los términos deseados del tesauro de uno en uno, obligando a entrar y salir de la gestión de tesauros tantas veces como términos de este se quieren añadir, esta opción también carece de la posibilidad de mover términos de un lugar a otro del tesauro cuando se da el caso de que un término no se encuentre en la posición correcta, además no se pueden añadir términos de forma jerárquica ordenados por niveles lo cual dificulta la búsqueda de términos dentro del tesauro y la gestión del mismo. Por ejemplo, el tesauro actual contiene los términos “Europa”, “España”, “Aragón” y “Temático” de forma independiente, cuando lo correcto sería una jerarquía de 4 niveles de la siguiente forma: Temático/Europa/España/Aragón en la cual cada término está contenido en otro. También comentar que la opción de creación de nuevo contenido obliga a llenar todos los campos disponibles en

el formulario aunque estos no pertenezcan a la categoría seleccionada, además en los casos en los que se quiere vincular varias fichas de audio a un documento es necesario llenar para cada ficha de audio toda la información del documento al que pertenece.

Por todos los motivos expuestos, se ha decidido desarrollar un sistema de gestión de documentación que ayude a realizar de manera sencilla todas las tareas necesarias para gestionar de forma correcta las grandes cantidades de información que actualmente se almacenan en Aragón Radio.

### 1.3. Identificación de necesidades y problemas

En este apartado se van a exponer todos los problemas y necesidades detectadas en el sistema actual, y que se mejorarán con el nuevo sistema. Como necesidad primaria, destacar la necesidad de unificar en el nuevo sistema los dos gestores de contenido actuales para facilitar las búsquedas del usuario. Como necesidades secundarias señalar las siguientes:

- Mejorar el uso del tesauro, dado que la interfaz actual no es intuitiva, ni fácil de usar por cualquier usuario. Permitiendo recuperar todos los términos que se necesitan rescatar del tesauro de una sola vez, y habilitando la posibilidad de mover términos de una posición a otra dentro del tesauro.
- Modificar la estructura del tesauro permitiendo un tesauro multinivel con relaciones (padre-hijo) entre términos.
- Optimizar el buscador, permitiendo que busque utilizando todos los campos disponibles a la vez, y habilitando 3 buscadores adicionales que faciliten el filtrado de las búsquedas según el contenido que se desea buscar (contenido general (programas, informativos, ...), fichas de audios o contactos personales para entrevistas).
- Permitir la generación automática de informes en formato PDF.
- Desarrollar una unidad de generación de resúmenes automáticos que sintetice las ideas más relevantes dados varios documentos de texto. Con esta unidad se podrán crear por ejemplo biografías o cronologías de manera fácil, así como conseguir sintetizar gran cantidad de documentos de un tema determinado.
- Separar los campos por las categorías a las que estos pertenecen: informativos, música, programas...

- Posibilidad de crear y vincular documentos como cronologías, biografías, obituarios, etc, lo cual permitirá agilizar la respuesta a las demandas de información y la difusión de las mismas.
- Mejorar la presentación de todos los registros que conforman la herramienta.
- Habilitar la copia de registros, para evitar cumplimentar los mismos varias veces cuando sólo se necesita modificar uno.
- Mostrar los resultados de las búsquedas en orden cronológico por fecha de creación del contenido
- Diseñar una nueva base de datos que sustituya a la actual, la cual contiene redundancias y ambigüedades.
- Agilizar la comunicación de ficheros MP3 con X-FRAME.
- Diseñar una interfaz del usuario sencilla, e intuitiva para el usuario.
- Añadir la posibilidad de vincular todas las fichas de documentación con los documentos sonoros (uno o varios) que les relacionan.

En general, se necesita mejorar la inteligencia, la calidad y la potencia del sistema, mejorando en todo momento la productividad del departamento de documentación, ofreciendo una interfaz sencilla, eficaz, con un diseño actual y libre de errores.

## 1.4. Tecnologías utilizadas

A continuación se van a exponer las diferentes tecnologías utilizadas en el desarrollo del proyecto.

Como lenguajes de programación, comentar que se ha utilizado *C# .Net* [13] en todo el proyecto, exceptuando en la implementación del ejecutable encargado de migrar todos los datos de la antigua base de datos a la nueva, para el cual se ha utilizado *Visual Basic .Net* (VB.NET) [18]. También se ha utilizado la tecnología WPF (Windows Presentation Foundation) [22] para diseñar las interfaces de la herramienta.

En cuanto a los entornos de desarrollo utilizados, nombrar a *Visual Studio 2010* [26] para los lenguajes de programación citados anteriormente, y *SQL Server* [23] para manejar la base de datos.

Para acoplar complementos, bibliotecas y productos adicionales orientados al I+D desarrollados por el grupo de Sistemas de Información Distribuida (SID)<sup>1</sup> de

---

<sup>1</sup><http://sid.cps.unizar.es/> (Último acceso: miércoles, 20 de mayo de 2014)

la Universidad de Zaragoza para el proyecto GENIE<sup>2</sup>, se ha utilizado un servicio al cual he llamado “Servicio de Procesamiento de Lenguaje Natural (PLN)” situado en el mismo servidor en el cual se encuentra alojada la base de datos. Este servicio permite extraer información de N documentos dados. En concreto, se ha utilizado este servicio para lematizar<sup>3</sup> todos los campos introducidos en la base de datos, mejorando así las búsquedas realizadas en la herramienta dado que el rango de resultados dada una palabra concreta aumentan considerablemente (por ejemplo si buscamos la palabra “tiene” también buscará a su vez la palabra “tener”), y para acoplar en él la unidad que he desarrollado para generar de forma automática resúmenes de contenido, dado que esta unidad necesita de módulos contenidos en el servicio PLN (como el módulo de lematización) para su correcto funcionamiento. Este servicio será explicado con mayor detalle posteriormente en la subsección 3.2.3.

Para acceder a la base de datos y ejecutar consultas en ella, se va a utilizar una biblioteca de acceso a base de datos que he desarrollado. Por otro lado se ha utilizado el estándar XML [4] para generar los ficheros de configuración que contiene la aplicación.

Como sistema operativo para el desarrollo y la ejecución de la aplicación, se ha utilizado Windows XP [28] en los entornos de Aragón Radio, y fuera del mismo se ha utilizado Windows 7 [1], lo que ha permitido comprobar el correcto funcionamiento de la herramienta en ambos sistemas operativos.

Por último, la herramienta se implementará como una aplicación de escritorio, la cual se encontrará instalada en todos los ordenadores clientes de Aragón Radio que necesiten hacer uso de ella (documentalistas y redactores).

## 1.5. Fases del proyecto

En esta sección se van a indicar las diferentes fases por las que se ha ido pasando a lo largo del desarrollo del proyecto:

- Análisis de la aplicación actual de Aragón Radio.
- Estudio de las necesidades actuales y definición de los objetivos a lograr.
- Trabajo de investigación enfocado en la generación de resúmenes de forma automática.
- Análisis de requisitos del sistema.

---

<sup>2</sup>[http://sid.cps.unizar.es/SEMANTICWEB/GENIE/index\\_es.html](http://sid.cps.unizar.es/SEMANTICWEB/GENIE/index_es.html) (Último acceso: miércoles, 20 de mayo de 2014)

<sup>3</sup>Lematizar: Concentrar en una única palabra (lema) las formas de una palabra variable.

- Diseño de la nueva aplicación a desarrollar.
- Creación de un programa capaz de migrar los datos de la antigua base de datos en la nueva diseñada.
- Implementación del sistema.
- Pruebas y puesta en marcha del sistema.
- Resolución de incidencias en la puesta en marcha.

## 1.6. Organización de la memoria

La presente memoria está estructurada en una serie de capítulos y anexos, de manera que abarquen todo el proceso seguido durante la realización del proyecto. La división de la memoria contempla los siguientes apartados:

- Capítulo 1, expone el contexto en el que se realizó el proyecto, junto a los problemas detectados, y la descripción de la tecnología empleada en el desarrollo del proyecto.
- Capítulo 2, descripción de la herramienta de gestión desarrollada.
- Capítulo 3, descripción de la arquitectura utilizada en el sistema.
- Capítulo 4, explicación de las interfaces utilizadas en la herramienta.
- Capítulo 5, planificación con los tiempos estimados y reales de todo el proceso de realización del proyecto, junto a las conclusiones del autor y valoración personal, además de las posibles aportaciones de este PFC a líneas futuras de trabajo.

Además se incluyen los siguientes anexos:

- Apéndice A, análisis y diseño del sistema.
- Apéndice B, unidad de resúmenes automáticos.
- Apéndice C, pruebas realizadas.
- Apéndice D, glosario.
- Apéndice E, manual de usuario.
- Apéndice F, artículos publicados.



# Capítulo 2

## Gestor de Documentación

En este capítulo se van a explicar los aspectos más importantes de la herramienta de gestión. En el apéndice A encontramos un análisis mucho más profundo de la herramienta.

### 2.1. Ámbito del proyecto

Se parte de la necesidad de un producto software que permita archivar y documentar los contenidos que vienen de la redacción (grabaciones de la radio en formato MP3, documentos de texto con transcripciones de los programas y metadatos de estos) de manera rápida y sencilla. La herramienta dispondrá de una interfaz de usuario intuitiva, manejable y actual. Tiene como objetivo fundamental aumentar la productividad del departamento. Además de lo anterior, se deberá poder acoplar complementos, bibliotecas y productos adicionales orientados al I+D y elaborados por el grupo de investigación SID de la Universidad de Zaragoza para el proyecto GENIE. Estos añadidos mejorarán las funcionalidades básicas de la herramienta, y podrán ser objeto de trabajos futuros. En concreto, se ha decidido habilitar en la herramienta la posibilidad de generar resúmenes automáticos para poder mejorar los contenidos almacenados, así como ayudar a la creación de informes de un tema en concreto, lo cual es útil por ejemplo cuando el departamento de redacción necesita condensar información para llevar a cabo el desarrollo de una noticia que posteriormente un locutor de la radio utilizará para informar a todos los oyentes.

### 2.2. Objetivos

En esta sección se van a plasmar los distintos objetivos que se deben cumplir en el desarrollo de la nueva herramienta de gestión para que esta satisfaga todas

las necesidades impuestas por el departamento de documentación facilitando en todo momento su trabajo diario.

- Mejorar la estructura del tesauro permitiendo multinivel y relaciones entre términos.
- Mejorar el uso del tesauro, dado que la interfaz actual no es intuitiva, ni fácil de usar por cualquier usuario. Actualmente no permite recuperar todos los términos que se necesitan rescatar del tesauro de una sola vez.
- Habilitar la posibilidad de cambiar la posición actual de un término en el tesauro.
- Desarrollar una herramienta que permita archivar y documentar los contenidos que vienen de la redacción de forma sencilla.
- Diseñar una nueva base de datos.
- Migrar los datos de la base de datos antigua a la nueva.
- Depurar los datos migrados de la antigua base de datos.
- La interfaz deberá ser sencilla, manejable, e intuitiva.
- La herramienta permitirá acoplar complementos, bibliotecas y productos adicionales orientados al I+D que puedan mejorar las funcionalidades básicas de este producto.
- Añadir la posibilidad de generar resúmenes automáticos a partir de un conjunto de contenidos. Dado que Aragón Radio dispone de una gran cantidad de datos, se ha visto necesario incorporar una unidad de resúmenes automáticos. Esta posibilidad mejorará por ejemplo la creación de biografías (buscando una persona en concreto con los buscadores y generando un resumen con todos los ficheros encontrados sobre ella) que actualmente se hacen a mano, aumentando con ello la productividad.
- Se habilitará la posibilidad de generar de forma automática informes en formato PDF.
- Posibilidad de separar los campos por las áreas a las que pertenecen: informativos, música, programas.
- Posibilidad de alojar y vincular documentos como cronologías, biografías, obituarios, etc, permitiendo agilizar la respuesta a las demandas de información y su difusión.

- Mejorar la presentación de todos los registros que conforman la herramienta.
- Mostrar en los formularios de creación de contenido únicamente los campos que pertenecen al contenido que se desea crear.
- Posibilidad de copiar registros, para evitar cumplimentar los mismos campos cuando sólo se necesita modificar uno (por ejemplo al vincular varias fichas de audio a un documento en concreto).
- Mostrar los resultados de las búsqueda en orden cronológico por la fecha de emisión.
- Agilizar la comunicación de ficheros MP3 con X-FRAME.
- Añadir la posibilidad de vincular todas las fichas de documentación con sus respectivos documentos sonoros (uno o varios).

### 2.3. Herramienta

La herramienta de gestión dotará al departamento de documentación de una aplicación software de escritorio que les asistirá en su trabajo diario, y se compondrá de los siguientes elementos funcionales:

- **Acceso a la herramienta:** el acceso se realizará mediante usuario y contraseña por razones de seguridad y control.
- **Perfil:** informará al usuario de su nombre, usuario, contraseña, fecha de alta y de sus permisos activos. En caso de tener privilegios de administrador o de gestión de usuarios, el usuario podrá modificar sus credenciales y sus permisos en la base de datos.
- **Pantalla principal:** dispondrá de una pantalla sencilla donde habrá unos accesos desde los cuales se podrán realizar las diferentes acciones disponibles. Las acciones de las que dispone la herramienta son las siguientes:
  1. **Búsqueda de contenido:** esta opción permitirá realizar consultas únicamente de documentos que contenga el sistema. Los resultados se mostrarán en un listado ordenado cronológicamente por fecha de acontecimiento, en el cual se podrá acceder a la información de los documentos y editar o borrar dicha información. En caso de tener vinculada alguna ficha de audio, se podrá seleccionar la ficha deseada y escuchar y/o descargar el audio en formato MP3. Además se podrán generar informes en formato PDF y resúmenes automáticos en formato TXT de un documento o de varios. Los audios se reproducirán con el software VLC media [12].

2. **Búsqueda de contactos:** esta opción permitirá realizar consultas de los contactos añadidos en la base de datos. Los resultados se mostrarán en un listado ordenado cronológicamente por fecha de acontecimiento, en el cual se podrá acceder a la información de los contactos, y editar o borrar esta información. También se podrá eliminar completamente el contacto seleccionado, y se podrán generar informes en formato PDF o resúmenes en formato TXT de un contacto, o de varios.
3. **Búsqueda de audios:** esta opción permitirá realizar consultas de las fichas de audio que contenga el sistema. Los resultados se mostrarán en un listado, en el cual se podrá reproducir y/o descargar la ficha de audio deseada, así como acceder a la información de la misma permitiendo editarla o borrarla. En caso de estar vinculada a un documento, se podrá acceder al contenido de este para ver que otras fichas tiene vinculadas. Además se podrán generar informes en formato PDF y resúmenes en formato TXT de una ficha o de varias.
4. **Buscador General:** esta opción permitirá realizar consultas utilizando todos los campos disponibles en la base de datos, siendo el uso del operador booleano “AND” opcional. Las búsquedas pueden ser exactas realizando la consulta entre dobles comillas o no (en este caso buscará todo lo relacionado con la consulta introducida). Los resultados se mostrarán en un listado ordenado cronológicamente por fecha de acontecimiento, en el cual se podrá acceder a la información de los documentos, audios o contactos encontrados, pudiendo editar o borrar dicha información. Además, las fichas de audio que se encuentren, podrán ser escuchadas y/o descargadas. Como en los casos anteriores, también se pueden generar informes en formato PDF y resúmenes en formato TXT con el contenido deseado.
5. **Creación de nuevo contenido:** esta opción se encargará de mostrar unos campos u otros de forma dinámica, dependiendo de la categoría seleccionada (programas, informativos, biografías,...). Los campos que se muestren se encontrarán habilitados para su edición y posterior guardado en la base de datos. Para facilitar la creación de nuevo contenido, se podrán copiar registros y fichas de audio, evitando así la cumplimentación de los mismos de nuevo cuando sólo se necesitan modificar algunos campos.
6. **Gestión de Usuarios:** en caso de tener privilegios de administrador o de gestión de usuarios, se mostrará la información de todos los usuarios dados de alta en el sistema, junto a su nombre, usuario, contraseña, fecha de alta, estado en el sistema (activo o no), y sus permisos, permitiendo la edición de

los datos, así como la eliminación completa del usuario. Esta opción también permitirá añadir un nuevo usuario al sistema con los privilegios deseados.

7. **Gestión de Tesauros:** la herramienta dispondrá de una pantalla donde poder gestionar de forma cómoda el contenido del tesauro, teniendo en cuenta que existen tres grupos de términos generales que engloban al resto por ser estos más genéricos. Estos grupos son: onomásticos (engloba todos los términos relacionados con nombres y en especial con nombres propios), topográficos (relativo a lugares como continentes, países, ...) y temáticos (todos aquellos términos que no son calificados como onomásticos o topográficos). Dado que uno de los objetivos fundamentales del proyecto, y el más demandado por el departamento de documentación, es la necesidad de habilitar una herramienta intuitiva que permita una gestión completa del tesauro, se ha procedido a diseñar una pantalla en la cual se muestre un árbol jerárquico con todas las relaciones padre-hijo existentes en el tesauro. El tesauro se modificará con respecto al que se tenía anteriormente, habilitando un tesauro de tipo facetado por grupos que permite una estructura multinivel (sin límite de niveles), lo cual facilita la ingesta y la gestión de los términos disponibles en el tesauro. Por ejemplo dentro del grupo *onomástico*, se pueden tener atributos (en este caso personas), y dentro de cada atributo tener varias características (nombre, apellidos, ...). La pantalla de gestión permitirá añadir, editar, borrar (sólo se podrá realizar si ningún documento está etiquetado con el término seleccionado, para evitar así problemas de integridad de datos), mover y buscar términos.
8. **Configuración:** esta opción permitirá modificar las tablas de la base de datos usadas para clasificar los documentos. Se podrán añadir nuevos “tipos” de documentos o eliminar los existentes (si no existe ningún documento con ese tipo). También se podrán añadir nuevas categorías a las actuales (programas, informativos, ...) o eliminar una en concreto (si no existe ningún documento perteneciente a esa categoría). Además se podrá elegir que campos se desea tener a la hora de crear contenido para cada categoría en concreto. De esta manera la herramienta es totalmente configurable para el usuario.



# Capítulo 3

## Arquitectura

En esta sección se va a explicar toda la arquitectura del sistema, junto a la explicación de cada una de las partes que la conforman.

### 3.1. General

Tal y como se puede observar en la Figura 3.1, la arquitectura del sistema a desarrollar se compone de dos servidores. El primero de ellos, cuyo nombre es “*Sw2k3dox*”, contiene la base de datos del sistema y un servicio al que he llamado *Servicio PLN (Procesamiento de Lenguaje Natural)* que contiene varios módulos desarrollados por el grupo SID para el proyecto GENIE que permiten, entre otras opciones, lematizar (extraer los lemas de las palabras) algunos de los campos de la base de datos, como el título o el contenido, permitiendo así una mejora considerable a la hora de realizar las búsquedas deseadas desde la aplicación. Además, este servidor contiene la unidad que he desarrollado para generar resúmenes de forma automática llamada *GENIE\_SUM*, la cual hace uso de varios módulos contenidos en el servicio PLN. La herramienta de gestión es la encargada de llevar a cabo todas las peticiones necesarias para conectarse a dicho servicio, y quedarse a la espera hasta que este genere el resumen, recogerlo y enviarlo al equipo cliente para que el usuario haga uso de él.

El segundo servidor, cuyo nombre es “*Swxptrans01*”, contiene el software X-FRAME, el cual permite transcodificar<sup>1</sup> los audios almacenados en la radio. Para facilitar el manejo de los ficheros resultantes generados por X-FRAME, he desarrollado un servicio Windows que automatiza esta tarea, dado que anteriormente esta conllevaba mucho trabajo asociado al departamento de documentación. Este servicio se encontrará en ejecución en todo momento en este servidor sin que el

---

<sup>1</sup>Transcodificar: conversión directa (de digital a digital) de un códec a otro. Puede ser con o sin pérdida de calidad, dependiendo del códec usado.

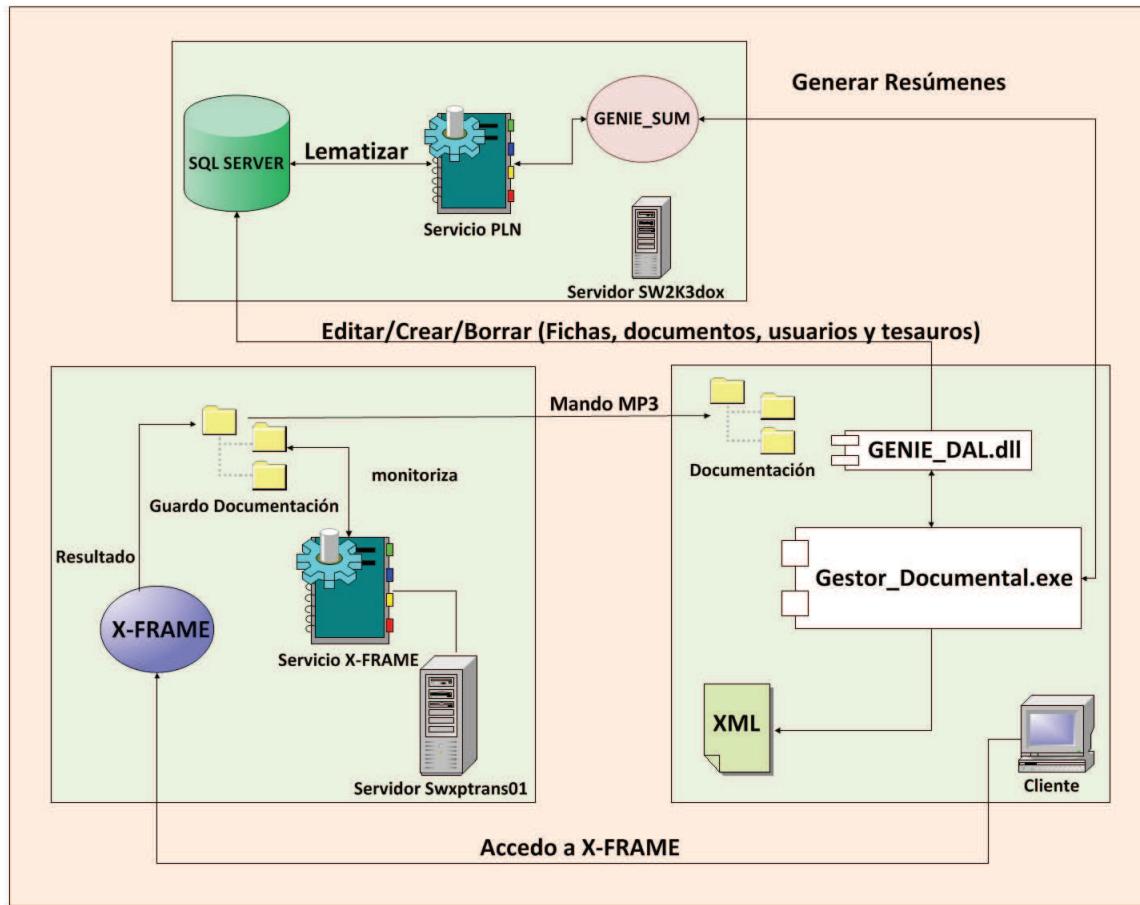


Figura 3.1: Arquitectura del sistema

cliente tenga que realizar ninguna tarea adicional para su correcto funcionamiento. Todos los ordenadores (clientes) se encontrarán conectados a ambos servidores, y podrán acceder a la herramienta de gestión mediante un ejecutable que se encontrará instalado en el mismo cliente.

### 3.2. Unidades funcionales del sistema

La herramienta de gestión desarrollada es la encargada de administrar la información que almacena Aragón Radio. Es una herramienta software de escritorio a la cual se accede mediante usuario y contraseña por motivos de seguridad y control, dado que cada usuario posee unos privilegios diferentes para interactuar con la herramienta.

Las diferentes funcionalidades de las que dispone la herramienta se encuentran en una pantalla principal a la cual se accederá cuando los credenciales de usuario sean correctos. Estas funcionalidades son: búsqueda de documentos, búsqueda de contactos (siendo contactos personas en concreto cuyos datos personales se encuentran almacenados para contactar con ellos, por ejemplo para una entrevista), búsqueda de audios, buscador general, creación de nuevo contenido, gestión de usuarios, gestión de tesauros, perfil del usuario, y configuración del sistema.

Su arquitectura se compone de una base de datos *SQL Server 2008* alojada en un servidor (“*Sw2k3dox*”) y un fichero de configuración en *XML* para conectarse a la misma. Este fichero XML contiene: el nombre del servidor, usuario, contraseña y el nombre de la base de datos. Para realizar la conexión de la herramienta con la base de datos se ha utilizado una biblioteca desarrollada en Visual Basic .Net que permite acceder a los datos de dicho fichero XML y conectarse a la base de datos dada.

Para gestionar la comunicación de los ficheros MP3 con X-FRAME, se ha creado un servicio que se mantendrá activo en todo momento en el servidor en el que se encuentra el software X-FRAME (“*Swxptrans01*”), para automatizar las tareas de acceso a carpetas y obtención de ficheros transcodificados.

Las bibliotecas externas que se han utilizado a lo largo de todo el proyecto han sido: “*itextsharp.dll*” y “*GENIE\_DAL.dll*”. La primera es una biblioteca que permite crear, adaptar, inspeccionar y mantener documentos en formato PDF para C# desarrollada por uno de los principales desarrolladores de *iText*, Paulo Soares.

La segunda biblioteca, en cuyo desarrollo participé, permite realizar todas las conexiones a la base de datos SQL Server. Esta biblioteca se ha reutilizado en proyectos del grupo SID de la Universidad de Zaragoza.

Además para poder cambiar el formato fuente de la interfaz, se han utilizado los ficheros “*Segoeui Segoeui.ttf*” y “*Segoe Ui Semilight.ttf*”, estos dos últimos han sido desarrollados por *Microsoft Corporation*.

### 3.2.1. Base de datos

La base de datos es el recurso donde se almacena toda la información del sistema. Para modelarla se va a utilizar un modelo Entidad-Relación (E-R), dado que es una herramienta que permite representar las entidades relevantes de un sistema de información así como sus interrelaciones y propiedades. La base de datos del sistema se almacenará en un servidor *SQL Server 2008*. Su descripción en detalle se realizará en el anexo A.

### **3.2.2. Servidor**

Recurso que constituye la máquina donde se almacena toda la información y da soporte a las herramientas que componen el sistema. En este caso, la arquitectura se compone de dos servidores explicados anteriormente, uno que aloja la base de datos junto al servicio PLN desarrollado por el grupo SID y la unidad de generación de resúmenes, y otro que aloja todo lo referente a la conexión con X-FRAME. Todos los usuarios que utilicen la aplicación harán uso de los dos servidores, pero en especial se hará uso del primer servidor para almacenar y/o modificar el contenido de la base de datos que compone el sistema.

### **3.2.3. Servicio PLN**

Dado que uno de los objetivos deseados es que la herramienta permita acoplar complementos, bibliotecas y productos adicionales orientados al I+D, con la intención de poder mejorar las funcionalidades básicas de la herramienta, se ha decidido añadir un servicio llamado PLN que se compone de un conjunto de módulos que implementan tareas imprescindibles para poder llevar a cabo procesos de extracción de información independientes del contexto. Este servicio ha sido desarrollado por el grupo SID de la Universidad de Zaragoza en el marco del proyecto GENIE. En concreto, los módulos por los que está constituido el servicio son:

- Procesador de lenguaje natural multilingüe.
- Analizador sintáctico y semántico.
- Identificador de palabras clave.
- Extractor de información geográfica.
- Analizador estadístico.
- Etiquetador automático.
- Unidad de generación de resúmenes.
- Motor de búsqueda semántica.
- Interfaces que permiten interconectar los distintos módulos.

De todos estos módulos, los que se han utilizado en este proyecto han sido:

1. **Unidad de procesamiento de lenguaje natural (PLN).** Esta unidad lleva a cabo procesos de análisis del lenguaje natural, pudiéndose hacer en diferentes idiomas. Estos procesos pueden ser: lematización (hallar los lemas de una palabra dada), obtención de nombres de entidad, eliminación de ruido, etc. Para realizar los procesos de análisis, se ejecuta un analizador que hace uso de la aplicación llamada Freeling [5], la cual analiza textos en lenguaje natural. El proyecto Freeling fue creado como un medio para poner a disposición los resultados de la investigación llevada a cabo del lenguaje natural. *Freeling* está diseñado para ser utilizado como una biblioteca externa de cualquier aplicación que requiera servicios de procesamiento del lenguaje natural. Aprovechando las características que ofrece este módulo, este ha sido utilizado para poder lematizar algunos de los campos que forman la base de datos como: título, contenido o descripción. La lematización ofrece una mejora considerable a la hora de realizar búsquedas en la base de datos, dado que se recuperan más resultados que antes podrían ser obviados. Por ejemplo, dado el título: “Javier Bardem, actor español, dedica el Óscar obtenido”, este se lematizaría y se almacenaría en un campo de la base de datos llamado título\_lematizado con lo siguiente: “Javier Bardem actor español dedicar el Óscar obtener”. Así a la hora de realizar las búsquedas se buscaría en ambos campos (título y título\_lematizado) aumentando el número de resultados posibles. Para que los usuarios no tengan que preocuparse de lanzar la orden de lematización desde la herramienta, este módulo ha sido modificado para que se active por si solo a una hora programable por Aragón Radio y se encargue de lematizar los nuevos registros añadidos a la base de datos.
2. **Unidad de identificación de palabras clave o keywords.** Esta unidad lleva a cabo procesos de análisis de textos y de obtención de las palabras clave o *keywords* presentes en los mismos, aplicando un conjunto de métodos estadísticos. Este proceso se encarga de obtener todas las palabras que componen un texto previamente lematizado (con sus lemas correspondientes) por la unidad de procesamiento del lenguaje natural. A continuación, se realizan dos cálculos que expresan la relevancia de una palabra para un texto dentro de una colección: el TF-IDF (Term Frequency - Inverse Document Frequency) [20] y el TF-WP (Term Frequency - Word Position) de todas las palabras del texto. Para esto, se necesitará la información sobre las frecuencias de las palabras que es obtenida en la unidad de preprocesado. De todas aquellas palabras que poseen mayor frecuencia final, se selecciona un número máximo de ellas y se les cataloga como *keywords*.

3. **Unidad de generación de resúmenes.** Hoy en día, toda organización posee gran cantidad de documentos de texto en formato digital, y además estos documentos están en constante aumento, por ello tenemos que tener en cuenta que no está mejor informado quien posee más datos, sino quien dispone de mejores medios para obtener exclusivamente los datos que necesita y asimilarlos. Por este motivo, es necesario poseer de una unidad capaz de generar resúmenes, dado que los resúmenes automáticos de documentos supone un nuevo paso hacia la optimización del tratamiento de la información en formato digital y su adaptación a las necesidades de los usuarios, facilitando a estos la lectura de gran cantidad de documentos y el acceso de manera rápida a la información relevante de los mismos. Dado el volumen de datos de los que dispone Aragón Radio, así como la necesidad de poder ofrecer contenidos de calidad a la gente, se ha visto necesario incorporar una unidad de resúmenes automáticos que permita seleccionar varios documentos con transcripciones de categorías o de fichas de audio, y generar un resumen de todos ellos. De esta manera, se pueden crear por ejemplo biografías de manera sencilla (buscando una persona en concreto y generando un resumen con todos los ficheros encontrados sobre ella), así como conseguir sintetizar gran cantidad de documentos de un tema determinado, aumentando con ello la productividad.

El módulo de elaboración de resúmenes que he desarrollado, cuyo nombre es GENIE\_SUM, hace uso de un doble clúster o agrupación para seleccionar las oraciones más relevantes dentro del texto. En concreto estos clústers son: clúster de similitud y clúster de keywords.

- **Clúster de similitud:** calcula la similitud existente entre todas las oraciones que forman el texto o textos introducidos. Cuando se ha calculado la similitud para todas las oraciones se procede a escoger aquella que posee un valor de similitud mayor. Si este valor de similitud está por encima de un umbral dado entonces se procede a añadir esta frase al clúster de similitud, en caso contrario se añade esta oración a un nuevo clúster. Las oraciones con las puntuaciones más altas de cada clúster se denominan centroides.
- **Clúster de keywords:** agrupa los centroides anteriormente extraídos basándose en las apariciones de sus keywords.

Para poder calcular la puntuación de cada oración se ha utilizado el algoritmo estadístico tf-idf [20], el cual es muy conocido en ámbitos de recuperación de información y minería de textos. Este algoritmo permite obtener una medida numérica que expresa cuán relevante es una palabra para un documento

dentro de una colección. El valor tf-idf aumenta proporcionalmente al número de veces que una palabra aparece en el documento, pero es compensada por la frecuencia de la palabra en la colección de documentos dados, lo que permite manejar el hecho de que algunas palabras son generalmente más comunes que otras. Una vez generado el resumen se procede a simplificar el mismo eliminando todo aquel contenido que se considera irrelevante y que los clústers no han conseguido filtrar, por ejemplo, todas aquellas palabras que se encuentran entre comas. Para ello se usa un módulo de simplificación llamado GENIE\_SIMPLY que también ha sido desarrollado dentro del grupo SID para el proyecto GENIE. Además de lo anterior, se realiza un análisis sintáctico que permite comprobar que todas las oraciones que componen el resumen están bien formadas. Para llevar a cabo este análisis, se usa el analizador de lenguaje natural Freeling.

Esta unidad de resúmenes fue utilizada para desarrollar la arquitectura de un sistema dedicado a la creación de topic maps [19] a partir de textos en lenguaje natural. Dicho trabajo dio lugar a un artículo de investigación desarrollado con el grupo SID, en cuya elaboración participé activamente colaborando como un autor más. Fue seleccionado en la conferencia internacional ICTAI 2014 (CORE B) celebrada en Washington en noviembre de 2013 [7]. En el anexo F se adjunta dicho artículo.

Además, también ha sido utilizada en otra aplicación que consiste en dada una consulta introducida por un usuario, la aplicación busca en distintos motores de búsqueda web configurables todo aquel contenido que está relacionado con la consulta realizada. Con todo el contenido encontrado en la web, se genera un resumen que condensa las ideas más relevantes. Dicho trabajo también dio lugar a un artículo de investigación desarrollado junto al grupo SID, en cuya elaboración participe como autora principal. Este artículo ha sido seleccionado en la conferencia internacional FUSION 2014 (CORE C) que se celebrará en Salamanca en julio de 2014 [6] y en la cual presentaré este trabajo. El artículo se adjunta en el anexo F.

Para comunicar la herramienta de gestión con este servicio se han utilizado varios ficheros XML de configuración que permiten asignar las rutas a las cuales se debe acceder para poder recorrer todos los módulos que contiene el servicio. Una vez que la herramienta de gestión ha mandado al servidor los datos que se desean resumir, la herramienta permanece “monitorizando” el directorio de salida situado en el servidor para recoger el resumen resultante y almacenarlo en el equipo cliente. En el anexo B se explicará con más detalle el funcionamiento de esta unidad, la cual hace uso de los dos módulos anteriormente explicados en esta subsección (unidad PLN y unidad de identificación de keywords).

En general, el servicio PLN ofrece multitud de funcionalidades como análisis semántico, control de contenidos, geolocalización, auto etiquetado, elaboración de resúmenes, minería de datos<sup>2</sup>, clusterización, generación automática de sistemas de clasificación, y desambiguación en consultas. Aunque en este caso, únicamente he hecho uso de la elaboración de resúmenes por la necesidad de los mismos dentro del departamento de documentación, otras opciones como el etiquetado automático de documentos sería muy útil como trabajo futuro.

### 3.2.4. Servicio Windows X-FRAME

X-FRAME es la aplicación que se ocupa de garantizar la emisión, y realizar toda la gestión de escaletas<sup>3</sup> o parrillas<sup>4</sup> en Aragón Radio. Esta aplicación incluye módulos de grabación y edición de audio y video. Los ficheros de audio que están relacionados con un documento en concreto del sistema se obtienen a partir de esta aplicación. El equipo técnico encargado de crear y editar estos ficheros genera cierta cantidad de ellos al día; por ello, el departamento de documentación debe documentar y archivar estos ficheros en la base de datos.

Actualmente, si un documentalista quiere recuperar un fichero MP3 en concreto para añadirlo a una ficha de audio nueva, o simplemente necesita buscar un audio porque el personal de redacción lo necesita para su retransmisión en directo, los documentalistas necesitan conectarse al servidor remoto que contiene el software X-FRAME (el cual permite la transcodificación de los ficheros almacenados en MP3), y allí seleccionar los ficheros que se desean transcodificar (Ver Figura 3.2).

Posteriormente se debe esperar a que se terminen de procesar todos los ficheros seleccionados. Una vez procesados, estos son mandados a un directorio conocido del escritorio del servidor cuyo nombre es “Guardo Documentación”. Cuando todos los ficheros son procesados, los documentalistas tienen que acceder a dicho directorio, copiar todos los ficheros generados y pegarlos en otro directorio conocido del servidor cuyo nombre es “Documentación”, el cual es un directorio compartido con los ordenadores clientes de los documentalistas. Para mejorar esta situación se ha desarrollado un servicio Windows que se mantiene en ejecución constantemente en el servidor que aloja X-FRAME. Este servicio se encarga de estar constantemente “monitorizando” en el directorio del escritorio “Guardo Documentación”, de esta manera, cuando le llegan los ficheros transcodificados, los detecta y estos son enviados automáticamente a la carpeta “Documentación”. Con este servicio, los

---

<sup>2</sup>Minería de datos: proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

<sup>3</sup>Escaletas: boceto, lista o minuta donde los productores escriben en orden, los pasos, segmentos o secciones que componen un programa.

<sup>4</sup>Parrilla: disposición en el tiempo de los programas de un canal de televisión o una emisora de radio. Indica la hora de comienzo y fin de cada programa a lo largo del día.

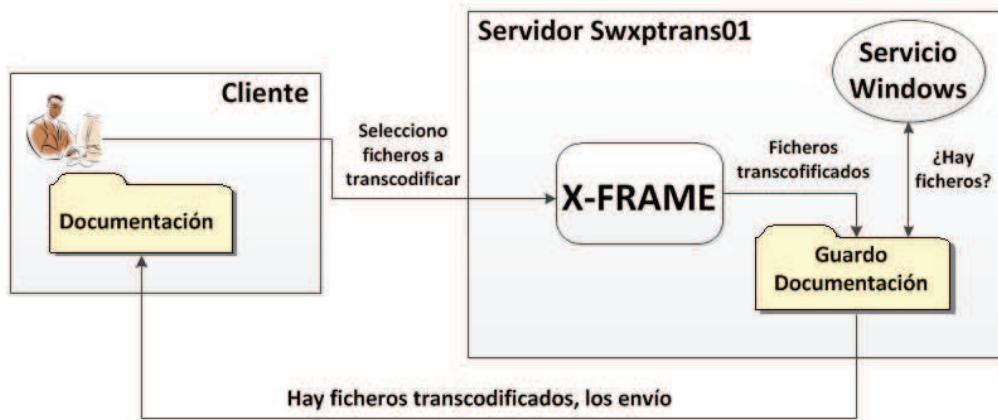


Figura 3.2: Diagrama funcionamiento servicio Windows para X-FRAME

documentalistas solo tienen que conectarse al servidor remoto para seleccionar los ficheros que desean transcodificar, y cuando terminen pueden volver a su ordenador para continuar con las tareas que se estaban realizando, y únicamente mirar el directorio “Documentación” de sus ordenadores, en el cual aparecerán los ficheros resultantes tras la transcodificación.



# Capítulo 4

## Interfaces

En este capítulo se van a especificar las interfaces entre el gestor de documentación y los usuarios finales del mismo: formatos de pantallas, diálogos e informes. El diseño de todas las ventanas de la interfaz es intuitivo, de tal forma que el usuario entienda inmediatamente dicha ventana y sepa qué pasos debe realizar para la ejecución de las tareas.

### 4.1. Pantalla principal de herramienta

Para acceder a la herramienta de gestión, se deberá introducir un usuario y una contraseña válidos a través de una pantalla típica de acceso, cuyo formato se presenta en la Figura 4.1

Una vez que los credenciales introducidos sean válidos se mostrará la pantalla principal de la herramienta.

### 4.2. Opciones de la herramienta

La pantalla principal de la herramienta (Ver Figura 4.2) posee una interfaz donde se muestran un conjunto de enlaces dentro de un menú, que llevan a otras pantallas donde estas permiten realizar todas las acciones disponibles en el sistema. Las opciones disponibles son: búsqueda de documentos, búsqueda de audios, búsqueda de contactos, buscador general, creación de nuevo contenido, gestión de usuarios, gestión de tesauros, información de usuario y configuración.



Figura 4.1: Pantalla de acceso al sistema

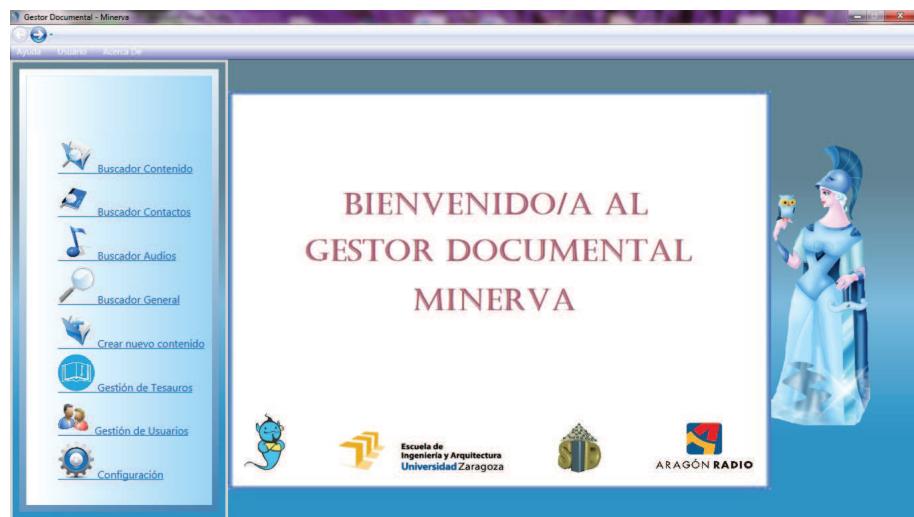


Figura 4.2: Pantalla de inicio del sistema

#### 4.2.1. Búsqueda de contenido

Esta pantalla permite realizar búsquedas sobre todo el contenido almacenado en la base de datos. Para ello se muestra un formulario con los campos disponibles (título, programa, descripción...), y una pestaña específica para añadir a la bús-

queda términos del tesauro. El formato de esta pantalla es como el que se presenta en la Figura 4.3.



**Búsqueda de Contenidos**

**Campos** **Tesaurios**

Categoría:  Tipo:  Autor:

Título:  Tipología:

Acontecimiento:  Hasta  Desde  Seleccione una fecha  Seleccione una fecha  Hora:  :

Fecha creación:  Seleccione una fecha  Seleccione una fecha  Lugar:

Programa:  Producción:

Cuadro Técnico:

Notas:  Voces:

Descripción:

Contenido:

**Buscar**

Figura 4.3: Pantalla de búsqueda de contenido

El resultado de la búsqueda se muestra en un listado (Ver Figura 4.4) ordenado de forma descendiente por fecha de emisión. Desde esta ventana se puede acceder a la información detallada del contenido encontrado, escuchar y/o descargar los ficheros de audio vinculados al contenido (en caso de disponer de ellos), generar un fichero PDF que recopile toda la información del contenido seleccionado, y generar un resumen que unifique todas las ideas relevantes del contenido encontrado facilitando la tarea de generación de informes dado un tema en concreto.

#### 4.2.2. Búsqueda de Audios

El formato de esta pantalla es análogo al formato de la pantalla anterior (búsqueda de contenido), pero en este caso los campos para realizar la búsqueda son distintos, dado que esta búsqueda se debe ajustar a los campos disponibles para un fichero de audio (intérprete, grupo, ISRC, disco,...). El formato de esta pantalla es como el que se presenta en la Figura 4.6. Los resultados son mostrados en un listado con las mismas características que se han explicado anteriormente para la búsqueda de contenido.

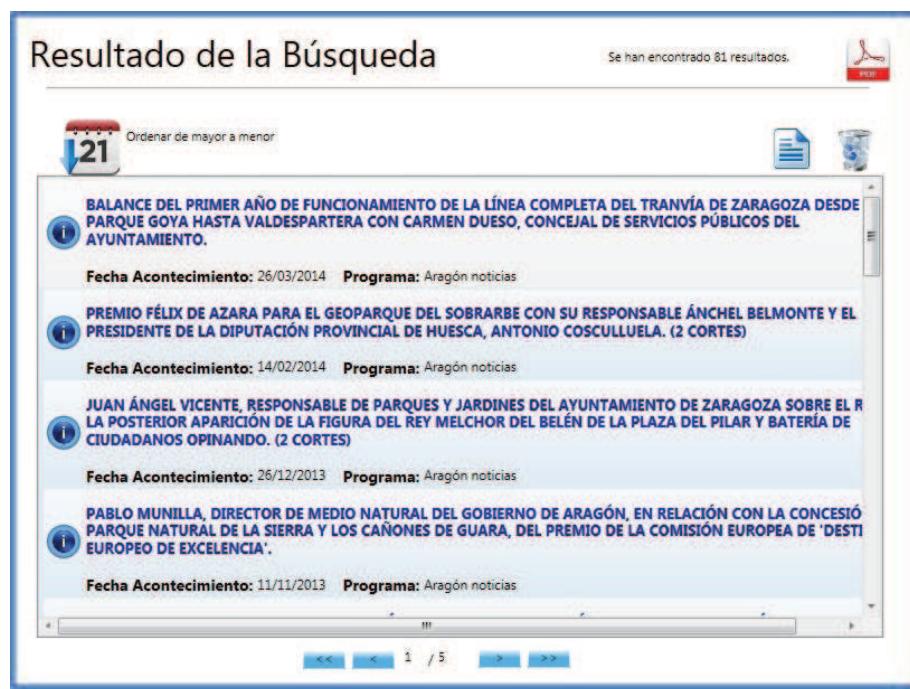


Figura 4.4: Pantalla listado de resultados

Búsqueda de Audios

Nombre del fichero  Tipo:

Intérprete:  Autor:  Grupo:

Disco:  Discográfica:  Duración:  :  :

Idioma:  Género:  Enlace:

Integrantes:  Signatura Topográfica:

Fecha publicación:  Lugar publicación:

Fecha grabación:  Lugar grabación:

Sello Discográfico:  ISRC:  Depósito Legal:

Figura 4.5: Pantalla de búsqueda de audio

#### 4.2.3. Búsqueda de contactos

El formato es análogo a las dos pantallas anteriores, pero al igual que ocurre con la búsqueda de documentos en este caso los campos para realizar la búsqueda también son distintos, ajustándose a los campos disponibles para un contacto o también conocido como pauta (nombre del contacto, datos del contacto, observaciones,...). El formato de esta pantalla se puede ver en la Figura 4.6. Los resultados son mostrados en un listado con las mismas características que las explicadas anteriormente para la búsqueda de contenido y la búsqueda de audio.

Figura 4.6: Pantalla de búsqueda de contactos

#### 4.2.4. Buscador general

El formato de esta pantalla (Ver Figura 4.7) es diferente al resto; su estilo es similar al de un buscador web conocido, como por ejemplo Google. Este buscador permite realizar búsquedas por todos los campos disponibles en la base de datos sin distinguir si estos son audios, documentos o contactos. Esta búsqueda es útil si se desconoce el tipo al que pertenece el posible contenido a buscar, o el campo en el que se encontraría la información (título, descripción, contenido...). Un ejemplo de uso sería, si queremos buscar todo aquello referente a “Adolfo Suárez”, introduciríamos la consulta en el buscador y este nos devolvería todos los contenidos

(documentos, audios o contactos) que en alguno de sus campos se pueden encontrar las palabras “Adolfo” y “Suárez”. Si la búsqueda se realiza mediante dobles comillas, la búsqueda será exacta. Dado que el departamento de documentación está acostumbrado a realizar este tipo de consultas haciendo uso del operador booleano *AND*, se ha habilitado la posibilidad de introducir la consulta de la siguiente forma “Adolfo AND Suárez” obteniendo los mismos resultados que si se prescinde del operador booleano. Los resultados son mostrados en un listado con las mismas características que los explicados anteriormente para los otros buscadores.

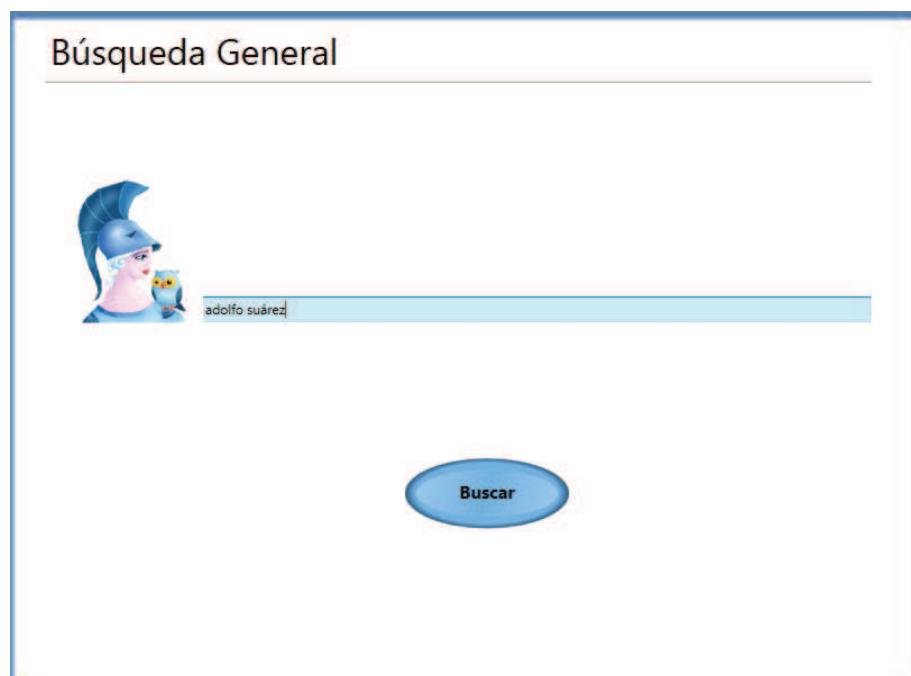


Figura 4.7: Pantalla del buscador general

#### 4.2.5. Perfil

Esta pantalla muestra los datos del usuario conectado. Estos datos son: nombre, usuario, contraseña, fecha de alta en el sistema, permisos activados, y situación dentro de la empresa (activo o no). También se ofrece un enlace al contenido que este usuario ha editado o creado. En caso de poseer privilegios de gestión de usuarios y cambio de clave, desde esta ventana se podrá editar cualquier dato almacenado del usuario.

#### 4.2.6. Crear nuevo contenido

Esta pantalla permite añadir nuevo contenido a la base de datos. Este contenido varía según la categoría seleccionada en la cabecera de la pantalla (biografía, programa, informativos,...). El motivo de que el contenido varíe según la categoría seleccionada, se debe a que cada categoría tiene activos unos campos u otros para agilizar la inserción de nuevo contenido, los cuales se pueden configurar desde la opción configuración de la pantalla principal. Si se quiere vincular un fichero de audio al contenido creado, se dispone de una pestaña de nombre “Fichas de audio” que permite introducir los campos de la nueva ficha. En caso de que se quieran vincular varias fichas de audio al nuevo contenido, se pueden introducir una a una, sin tener que volver a llenar todos los campos del nuevo contenido tal y como ocurría en la herramienta anterior. Además, gracias a la navegación que ofrece esta ventana, se pueden copiar y pegar campos llenados en otras fichas de audio sin tener que volver a introducir los mismos datos. El formato de esta pantalla es como el que se presenta en la Figura 4.8.



The screenshot shows a web-based form titled "Crear Contenido" (Create Content). At the top, a dropdown menu shows "Categoría: Informativos". Below the title, there are four tabs: "Campos" (selected), "Tesauros", "A", and "Datos Contactos". The main content area includes fields for "Título", "Programa", "Tipo", "Acontecimiento" (with a dropdown showing "Selección 15"), "Tipología", "Producción", and a "Cuadro técnico" section with "Voces" and "Notas" fields. There are also "Contenido" and "Descripción" text areas. In the bottom right corner, there is a blue "Crear" button.

Figura 4.8: Pantalla de creación de nuevo contenido

Para poder comparar el diseño entre la pantalla actual y la que tenían, se presenta en la Figura 4.9 la pantalla antigua.

Usuario:	Documentacion
Fecha de ingestá:	20140520
Fecha de creación:	19/05/2014 (AAAA/MM/DD)
Tipo:	Audio
Código:	
Descripción:	
Notas:	
Clase:	1 BIOGRAFÍAS
Buzón:	A-DOX (Nuevos)
Fichero:	
<b>DOX - PROGRAMAS</b>	
PROGRAMA:	
CUADRO TÉCNICO:	
PRODUCCIÓN:	
FECHA GRABACIÓN:	
LUGAR GRABACIÓN:	
FECHA PRIMERA EMISIÓN:	
GÉNERO/TOPOLOGÍA:	
VOZES:	
DESCRIPCIÓN TOPOGRÁFICO:	
DESCRIPCIÓN ONOMÁSTICO:	
DESCRIPCIÓN TEMÁTICO:	
NOTAS:	
CONTENIDO:	
RESUMEN:	
SITUACIÓN:	
SIGNATURA TOPOGRÁFICA:	

Figura 4.9: Pantalla de creación antigua

#### 4.2.7. Gestión de usuarios

En caso de poseer privilegios de gestión de usuarios, esta pantalla permite observar los datos de todos los usuarios añadidos al sistema (nombre, usuario, contraseña, fecha de alta en el sistema, situación y permisos), editarlos, borrarlos, e incluso añadir un nuevo usuario al sistema. El formato de esta pantalla se puede ver en la Figura 4.10

#### 4.2.8. Gestión de Tesauros

Esta pantalla permite realizar toda la gestión necesaria para una buena clasificación de los documentos. El formato de esta pantalla es como el que se presenta en la Figura 4.11. El tesauro es multinivel con relaciones padre-hijo (*EUROPA/ESPAÑA/ARAGÓN*), lo cual permite una representación sencilla de usar y fácil de entender. Para que la representación del tesauro sea intuitiva, se ha ge-

Gestión Usuarios								
Nombre	Login	Clave	Fecha de Alta	Activo	Permisos	Editar	Borrar	
Administrador	sa		18-09-2009					
Documentación	reda		04-05-2010					
Consulta Usuarios	usuario		19-05-2010					
Mario San Miguel	mсанmiguel		24-10-2012					
Alicia Zarazaga	azarazaga		24-10-2012					

Añadir Usuario

Figura 4.10: Pantalla de gestión de usuarios

nerado un árbol en el cual se puede observar fácilmente las relaciones jerárquicas que existen entre los diferentes términos (etiquetas) que forman el tesauro. Desde esta pantalla se pueden eliminar, añadir, editar y mover términos. La mejora de esta pantalla era uno de los requisitos fundamentales por parte del departamento de documentación, dado que el tesauro del que disponían no permitía multínivel, y además contenía redundancias, y fallos como se puede ver en la Figura 4.13, en la cual se observa cómo el término *ENTIDADES* es independiente cuando debería formar parte de *ONOMÁSTICOS* puesto que onomástico engloba todos aquellos términos relativos a nombres y especialmente nombres propios, y el término *ENTIDADES* contiene nombre de entidades como por ejemplo, AENA (Aeropuertos Españoles y Navegación Aérea), Banesto, DGA (Diputación General de Aragón)... También se puede observar cómo los términos *ACTOS OFICIALES*, *ACTOS SOCIALES*,...,*BOTÁNICA*,...deberían formar parte del término *TEMAS* (se corresponde con el término *TEMÁTICO*), que es el padre de estos (dado que no pertenecen ni a topográfico ni a onomástico), tal y como puede verse en el nuevo tesauro (Ver figura 4.12).

Por ese motivo, además de diseñar una pantalla de gestión de tesauros más completa se ha tenido que modificar el contenido almacenado en la base de datos

referente a los términos que forman el tesauro, mejorando todas las relaciones existentes y añadiendo nuevas.

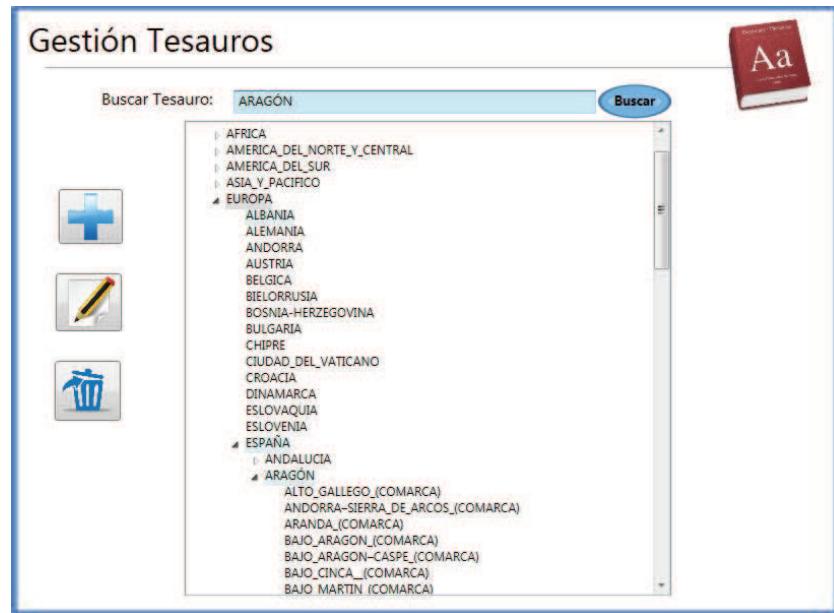


Figura 4.11: Pantalla de gestión de tesauros



Figura 4.12: Pantalla de gestión de tesauros temático

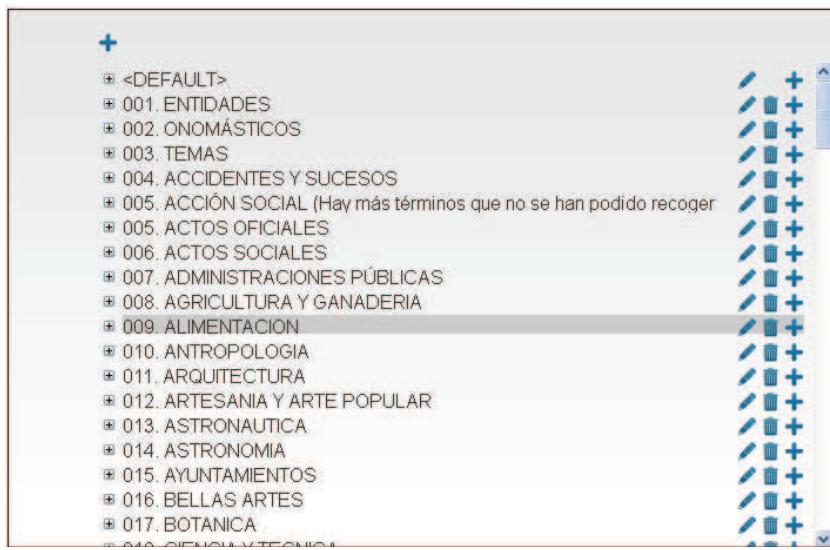


Figura 4.13: Pantalla de gestión antigua de tesauros

#### 4.2.9. Configuración

Pantalla orientada al personal con privilegios para poder modificar el sistema (añadir nuevas categorías, tipos, etc.).



# Capítulo 5

## Planificación y conclusiones

En esta sección se pretende dar una visión de los tiempos planificados así como de los invertidos, con objeto de que el lector pueda hacerse una idea del trabajo empleado y de la distribución del mismo en las distintas etapas empleadas a lo largo del desarrollo del sistema. Finalmente se reflexiona sobre las conclusiones generales del desarrollo del proyecto, se apuntan algunas posibilidades de trabajo futuro, y finalmente se aporta una conclusión a nivel personal.

### 5.1. Estimación de horas de trabajo

En los diagramas de la Figuras 5.1 y 5.2 se resume el desarrollo temporal del proyecto. También se incluye en la Tabla 5.1 un resumen de las horas que han sido necesarias para completar cada una de las partes del proyecto.

Tiempo dedicado por tarea	
Tarea	Tiempo
Estudio Previo	20 horas
Análisis y Diseño	50 horas
Implementación	310 horas
Pruebas	180 horas
Documentación	110 horas
<b>Total</b>	<b>670 horas</b>

Tabla 5.1: Resumen temporal

Nombre	Fecha de inicio	Fecha de fin
• Estudio Previo	8/01/14	12/02/14
• Análisis y Diseño	15/01/14	12/02/14
• Especificación de Requisitos	25/01/14	4/02/14
• Implementación	13/02/14	17/05/14
• Pruebas	18/05/14	28/05/14
• Documentación	15/05/14	3/06/14
• Elaboración manual	27/05/14	2/06/14

Figura 5.1: Tabla fechas de la planificación

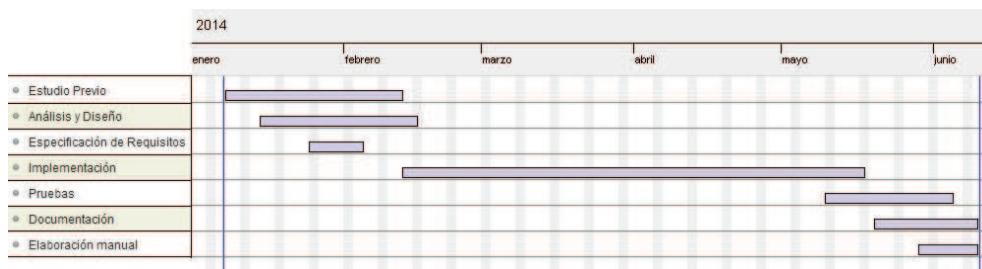


Figura 5.2: Diagrama de Gantt con el desarrollo del proyecto

## 5.2. Conclusiones generales

Una de las conclusiones más importantes del desarrollo de este proyecto es que se ha conseguido desarrollar una herramienta de gestión documental para una empresa reconocida a nivel local como es Aragón Radio. Dicha herramienta permite realizar todas las necesidades requeridas por el departamento de documentación favoreciendo y agilizando las mismas. En resumen, con la herramienta de gestión desarrollada se ha conseguido:

- Modificar el tesauro que tenía Aragón Radio (el cual solo permitía un nivel de jerarquía y no existían prácticamente relaciones entre términos), por un tesauro facetado multinivel (sin límite de niveles) con relaciones entre conceptos (padre-hijo), el cual ha sido modificado siguiendo las directrices y necesidades mostradas por el departamento de documentación después de varias reuniones en las cuales se decidió como querían estructurar el nuevo tesauro.

- Eliminar los dos gestores de contenido que tenía Aragón Radio dejando solo uno (lo cual favorece las búsquedas y la creación de contenido).
- Desarrollar una nueva base de datos sin redundancias.
- Habilitar una opción de gestión completa del tesauro, que permite: añadir nuevos términos, editar/borrar los términos existentes y mover términos de una posición del tesauro a otra.
- Realizar búsquedas de contenido mediante tres buscadores específicos que actúan como filtros de búsqueda: buscador de contenido, buscador de fichas de audio y buscador de contactos. Junto a un cuarto buscador el cual permite buscar por todos los campos de la base de datos, como si de un buscador web se tratase.
- Crear nuevo contenido vinculado a varias fichas de audio.
- Separar los campos por las categorías a las que cada contenido pertenece, mejorando con ello la presentación de todos los formularios que conforman la herramienta.
- Gestionar completamente la información almacenada de los usuarios desde la herramienta.
- Configurar el sistema de manera sencilla desde la opción de configuración.
- Agilizar la comunicación de ficheros MP3 con el software X-FRAME mediante un servicio Windows.
- Desarrollar una interfaz con un diseño actual, eficaz y ágil.
- Habilitar la generación de forma automática de informes en formato PDF del todo el contenido deseado.
- Habilitar la generación de resúmenes automáticos en formato TXT del contenido deseado.
- Acoplar complementos, bibliotecas y productos adicionales orientados al I+D desarrollados en el grupo SID para el proyecto GENIE, los cuales mejoran las funcionalidades básicas de la herramienta, en particular mejora las búsquedas gracias a la lematización de los campos que forman la base de datos y la generación de resúmenes de forma automática.

Por tanto, se ha conseguido implementar todos los requisitos que se habían propuesto en el tiempo esperado.

### 5.3. Trabajo futuro

A continuación se exponen algunos de los posibles trabajos a realizar continuando con la línea del proyecto:

- Dado que el tesauro actual utiliza una estructura multinivel con relaciones entre conceptos, este se podría transformar en una potente ontología [9], es decir en un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados que contribuya a: mejorar los procesos de archivo, la recuperación de información, y la posterior incorporación de nuevas herramientas (como generación automática de informes) que agilicen en gran medida la labor del departamento de documentación, repercutiendo así en una mejora de la calidad del servicio ofrecido, dado que la conversión de un tesauro a una ontología aumenta la precisión de la semántica en las relaciones existentes en el tesauro. Respecto a esto existen varias propuestas para migrar un tesauro a una ontología, un ejemplo es [27].
- También podría modificarse el tesauro actual para convertirlo en un tesauro reticular, es decir, un tesauro constituido por descriptores entrecruzados en forma de red. Cada descriptor puede tener varios descriptores genéricos y específicos y la información generada es más completa. Este tipo de tesauro se podría asemejar a una ontología, donde podríamos establecer por ejemplo reglas de inferencia entre términos, lo que permitiría modelar conocimiento y obtener inferencias (conclusiones a partir de unas características dadas). Un ejemplo de funcionamiento para este tipo de tesauro sería el siguiente: dado el término “Ayuntamiento\_de\_Zaragoza”, este tendría una relación “is-a” con el término “Ayuntamiento” el cual a su vez tiene los términos “Alcaldes” y “Alcalde\_actual”, y este último una relación con el término “Juan\_Alberto\_Belloch”. Por tanto se podría conocer que dentro del Ayuntamiento de Zaragoza ha habido N alcaldes que tendrían relación “is-a” con el término “Alcaldes” (donde podríamos tener sus nombres, los años que estuvieron como alcaldes,...), así como el nombre del alcalde actual mediante la relación “is-a” que cuelga del término “Alcalde\_actual”.
- El servicio PLN ofrece un módulo de etiquetado automático que en el proyecto actual no ha sido utilizado. Este módulo podría incorporarse para categorizar los documentos almacenados en Aragón Radio, dado que este componente se encarga de etiquetar cualquier tipo de archivo que lleve asociado un texto. Esta mejora sería interesante dado que la categorización automática de documentos permite encontrar la información buscada en un tiempo mínimo.

Los trabajos expuestos como posible trabajo futuro mejorarían la recuperación de información del sistema, e incrementarían la productividad en los trabajos diarios del departamento de documentación e incluso repercutirían al departamento de redacción, dado que en muchas ocasiones este departamento pide cierta información a los documentalistas.

## 5.4. Conclusión personal

El desarrollo de este proyecto me ha aportado una serie de conocimientos que me serán muy útiles en mi vida profesional. En cuanto a los conocimientos técnicos adquiridos destacar que he aprendido bastante acerca de C# .Net y la tecnología WPF, los cuales durante la carrera nunca había utilizado, y por ello he necesitado hacer un uso exhaustivo de la documentación ofrecida acerca de ambos. También he adquirido un serie de conocimientos al realizar un software a medida para satisfacer las necesidades de una empresa, como por ejemplo: tener trato con los clientes (ver que necesidades manifestaba el departamento de documentación en conjunto y buscar la solución óptima para solucionarlas), mantener reuniones semanales con el director (en las que se ha presentado el trabajo desarrollado hasta el momento, de forma que se pudieran encontrar fallos y buscar las soluciones adecuadas para cumplir las necesidades exigidas), implantar mi desarrollo en los equipos del departamento de documentación y ser ellos quienes han terminado de probar mi herramienta en su trabajo diario sin encontrar problemas.

En resumen, el proyecto desarrollado me ha permitido adquirir conocimientos interesantes de cara a mi futuro profesional, así como aprender a enfrentarme a problemas que aparecen en el transcurso de un proyecto real. Por ello recomiendo ampliamente realizar el PFC en una empresa para poder tomar contacto con el mundo real, y acercarnos al futuro laboral que nos espera. Además saber que una empresa demanda una necesidad y tú has conseguido satisfacerla con éxito, así como ver que el departamento de documentación te agradece el desarrollo de la nueva herramienta, llena de orgullo y satisfacción. Por todo lo anterior, he de mostrar mi agradecimiento al grupo SID de la Universidad de Zaragoza que me ha acercado y permitido realizar mi PFC para una empresa tan importante como es Aragón Radio a la cual he podido ayudar a mejorar sus condiciones de trabajo, así como previamente haberme permitido realizar mis prácticas de empresa con ellos, y ayudarme a publicar varios artículos de investigación ([7] y [6]) para conferencias internacionales con los cuales he adquirido experiencia a la hora de leer, escribir y buscar artículos, mejorar mi inglés y me han acercado a un terreno que para mí era totalmente desconocido, el cual me parece muy interesante, a la par de enriquecedor ya que tu trabajo se ve recompensado con una publicación y una posterior conferencia en la cual puedes mostrar tus ideas y conocer a gente.

