



Universidad Politécnica de Madrid

FACULTAD DE INFORMÁTICA

Trabajo fin de carrera

**Interfaz Gráfica de BaDELE3000.
Mejora de accesibilidad y visualización**

Autor

Juan José Rodríguez Sánchez

Tutora

María del Socorro Bernardos Galindo

Fecha de presentación:

25 de Julio de 2018

Índice general

Índice General	2
Índice de figuras	3
Resumen	4
Capítulo 1 – Introducción	5
Capítulo 2 – Nociones lingüísticas	7
Capítulo 3 – Trabajo desarrollado	10
3.1 – Introducción	10
3.1.1 – Casos de uso	11
3.2 – Instalación del sistema	16
3.3 – Visualización gráfica. Creación de un scroll en pantalla	19
3.4 – Visualización gráfica. Todas las etiquetas semánticas de un lema	20
3.5 – Creación de un área de descargas	21
3.6 – Activación del modo producción	24
3.7 – Modificación de la página de inicio de la aplicación	24
3.8 – Actualización del manual de usuario	25
3.9 – Pruebas realizadas	26
3.9.1 – Visualización del funcionamiento del scroll	26
3.9.2 – Verificación de la visualización de todas las etiquetas semánticas	27
3.9.3 – Verificación del grafo para el idioma inglés	28
3.9.4 – Verificación del módulo de descargas	28
Capítulo 4 – Conclusiones y trabajo futuro	30
Anexo A – ATAMIRI	33
Anexo B – Estructura del sistema	39
Referencias y bibliografía	45

Índice de figuras

Figura 1: Diagrama de casos de Uso	11
Figura 2: Panel de acceso de Hostalia	16
Figura 3: Código htaccess	16
Figura 4: Tablas de la base de datos	17
Figura 5: UTF base de datos	18
Figura 6: Código de scroll	19
Figura 7: Scroll.....	19
Figura 8: Página principal BADELE.....	20
Figura 9: Área de descargas	21
Figura 10: Página de inicio de BADELE.....	23
Figura 11: Scroll.....	25
Figura 12: Consulta etiquetas semánticas	26
Figura 13: consulta de idioma.	27
Figura 14: Página principal de BADELE (Usuario Anónimo)	28
Figura 15: Área de descargas.	28
Figura 16: Diagrama E/R.....	41

Resumen

En este Trabajo de Fin de Carrera se abordará la mejora de la interfaz web para la base de datos BaDELE3000, esta base de datos cuenta en la actualidad con más de 3300 lexemas y continúa desarrollándose. BaDELE3000 viene a modelar de una forma potente los elementos principales de la Teoría Sentido-Texto, la mejora de la interfaz sirve para incluir nuevas funcionalidades así como para intentar hacer de un modo más sencillo la instalación y la búsqueda de las versiones del software necesario para su funcionamiento, tomando el de máquinas que ya están en funcionamiento.

En la memoria se desarrollará el trabajo realizado (Véase cap 3), así como los conceptos básicos de la Teoría Sentido-Texto (Véase cap 2), además se incluirá un anexo indicando futuras mejoras para la inclusión de un idioma pivote en las traducciones como fruto del trabajo de varios grupos de trabajo alrededor del lenguaje Aymara.

Capítulo 1

Introducción

La base de datos léxica de la que partimos, hasta el momento contiene aproximadamente los 3300 sustantivos más usados del español de España, un léxico que se consideró suficiente para el aprendizaje del español en niveles intermedio o avanzado. A partir de estos sustantivos, la base de datos permitió la obtención automática de unas 9.000 relaciones léxicas. En total se obtuvieron unas 20.700 relaciones léxicas con la ayuda de diccionarios combinatorios del español. Para la herencia de glosas y valores de funciones léxicas de cada dominio, elaboraron una clasificación de sustantivos que no existía hasta la fecha.

Para conseguir que el acceso a la información de la base de datos fuese lo más sencillo y universal posible se creó una aplicación web, que además de permitir realizar consultas, posibilitaba manipular el contenido de la base de datos por aquellos usuarios con permiso. Uno de los problemas que se detectó fue que la instalación y puesta en marcha de la aplicación resultaba muy difícil, por lo que existía tan solo en modo local en un equipo con sistema operativo Windows.

Por otro lado, BaDELE3000 no disponía de la capacidad de incluir unidades léxicas en otro idioma que no fuese el español, así como existían diferentes problemas en las visualizaciones gráficas de consultas grandes y de unidades léxicas clasificadas bajo varias etiquetas semánticas. En otro orden de cosas existían problemas en la visualización de las páginas resultado de las consultas, dado que había caracteres que no se correspondían con los que deberían aparecer.

Dadas estas circunstancias, el trabajo descrito en este TFC tiene como objetivo:

1. Puesta en marcha de la aplicación en un servidor web
2. Mejorar la parte gráfica con la inclusión de las consultas de unidades léxicas clasificadas bajo varias etiquetas semánticas
3. Solución de los errores detectados en la visualización de las consultas con caracteres extraños al resultado real.
4. Creación de un área de descargas.
5. Inclusión de otros idiomas en las consultas gráficas.
6. Modificación de la página de inicio de la aplicación.
7. Actualización del manual de usuario.

Capítulo 2

Nociones lingüísticas

Cuando una persona domina una lengua con fluidez, utiliza de una forma cotidiana diferentes colocaciones en las expresiones del idioma, dichas colocaciones dan sentido a las ideas que se quieren expresar de una forma potente. Las citadas colocaciones son unidades fraseológicas formadas por dos unidades léxicas en relación sintáctica. Esas relaciones son combinaciones que no pueden ser utilizadas de cualquier manera a la hora de dar sentido a las expresiones. En estas relaciones existen distintos tipos de palabras según su significado, así a la palabra que por sí misma tiene un significado, se la denomina base y a la que se elige para dar un sentido a la expresión se la denomina colocativo.

En el caso de querer indicar un sentido de, por ejemplo, intensidad se utilizarán diferentes colocativos dependiendo de la base. En el caso de que se utilice la base hambre, se utilizará atroz, pero en el caso de utilizar miedo, se utiliza cerval.

Así pues nos encontramos con que estas colocaciones se convierten en una parte esencial de las lenguas a la hora de expresarse de una manera cotidiana por los hablantes de dicho idioma, para conseguir un dominio real sobre cualquier

idioma, será necesario tener en cuenta de una manera exhaustiva las citadas colocaciones.

En este sentido, BaDELE surgió para poder formalizar de una manera computacional dicho concepto y poder tener un nuevo recurso computacional en castellano y con las diferentes iteraciones ampliarlo a diferentes idiomas.

Para ello, se recurrió a la Teoría Sentido Texto (en adelante TST), una teoría lingüística que aborda el estudio de las colocaciones mediante las funciones léxicas. Las funciones léxicas sirven para describir la relación entre los miembros de una colocación: la base y el colocativo. A su vez, cada función representa la relación semántica y sintáctica que se establece entre estos miembros. Así, una función que podemos denominar Magn, por ejemplo, hace referencia a un adjetivo o adverbio que intensifica la base. Los valores asociados al nombre amor mediante la función Magn serán los adjetivos incondicional, superficial, apasionado, profundo, etc.; dado que todos ellos expresan de alguna manera el significado ‘intenso’ unido a la base, amor. Otra función, Oper1, vincula la base, un nombre, con el verbo que se utiliza para expresar que alguien lleva a cabo la acción implicada por la base. La función Oper1 con la base amor cobra los valores sentir (amor por alguien), tener (amor a alguien), profesar (amor por alguien).

Las funciones léxicas describen las relaciones léxicas en términos universales, aplicables a cualquier lengua. Es decir, volviendo a nuestro primer ejemplo, vemos que existe la misma relación entre el nombre inglés attention y el verbo pay que entre el español atención y prestar. En los dos casos la relación entre el nombre y el verbo se describe por la función Oper1.

Así, la función léxica relaciona unidades léxicas entre sí. Una unidad léxica será una palabra o palabras con significado propio.

En relación a las colocaciones, Mel’cuk y Wanner en 1996, propusieron la idea del Principio de Herencia Léxica. Para ello demostraron que es posible, crear una entrada genérica y otra no genérica para «emoción». La entrada no genérica agrupa las colocaciones de la unidad léxica «emoción», mientras que la genérica

agrupa las funciones léxicas que comparten los valores de una serie de lexemas de «emoción». Por ejemplo, la entrada genérica tendría, al menos, la función léxica Oper₁, cuyo valor «sentir» es común a todos los nombres o lexemas de «emoción»: sentir alegría, odio, felicidad, etc. En el caso de que una función léxica se aplique a un grupo de unidades léxicas, se denomina glosa de la función léxica. Así pues, llevar es una glosa ya que combina con todas las prendas y complementos. A menudo, se utiliza “llevar unas gafas”, “llevar unas botas”. De otro modo, calzar será un valor específico para ciertas unidades léxicas como “calzar unos zapatos”, “calzar unas botas”.

Capítulo 3

Trabajo desarrollado

3.1. Introducción

A través de un servicio de *hosting* se ha procedido a poner en marcha la aplicación para conseguir un acceso vía web, a continuación presentaremos brevemente el diagrama de casos de uso, ya que ha cambiado respecto a la anterior iteración al añadir un nuevo caso como será la descarga del software para el montaje de la aplicación en Windows o Linux.

Los requisitos han sido conseguir la accesibilidad vía web a la base de datos, que un lema aparezca representado gráficamente con más de una etiqueta semántica, que en las consultas gráficas aparezca un scroll en pantalla que permita una visualización completa del mismo, la creación de un área de descargas, la modificación de la página de inicio de la aplicación, así como la actualización del manual de usuario.

Al añadir el caso de uso de acceso a descargas, se ha modificado el diagrama de casos de uso de la iteración anterior, procediendo a presentar el diagrama de casos de uso que existe en la actualidad.

3.1.1. Casos de Uso

Los casos de uso permiten entender y especificar los requisitos del sistema, es decir, permiten identificar todas las funciones u operaciones que el sistema deberá llevar a cabo para satisfacer las necesidades de los usuarios.

El diagrama de casos de uso, muestra los actores y el conjunto de operaciones que se han identificado para la aplicación. La Figura 1 representa el diagrama de casos de uso del proyecto utilizando la notación UML, Unified Modeling Language , donde las operaciones se representan mediante óvalos.

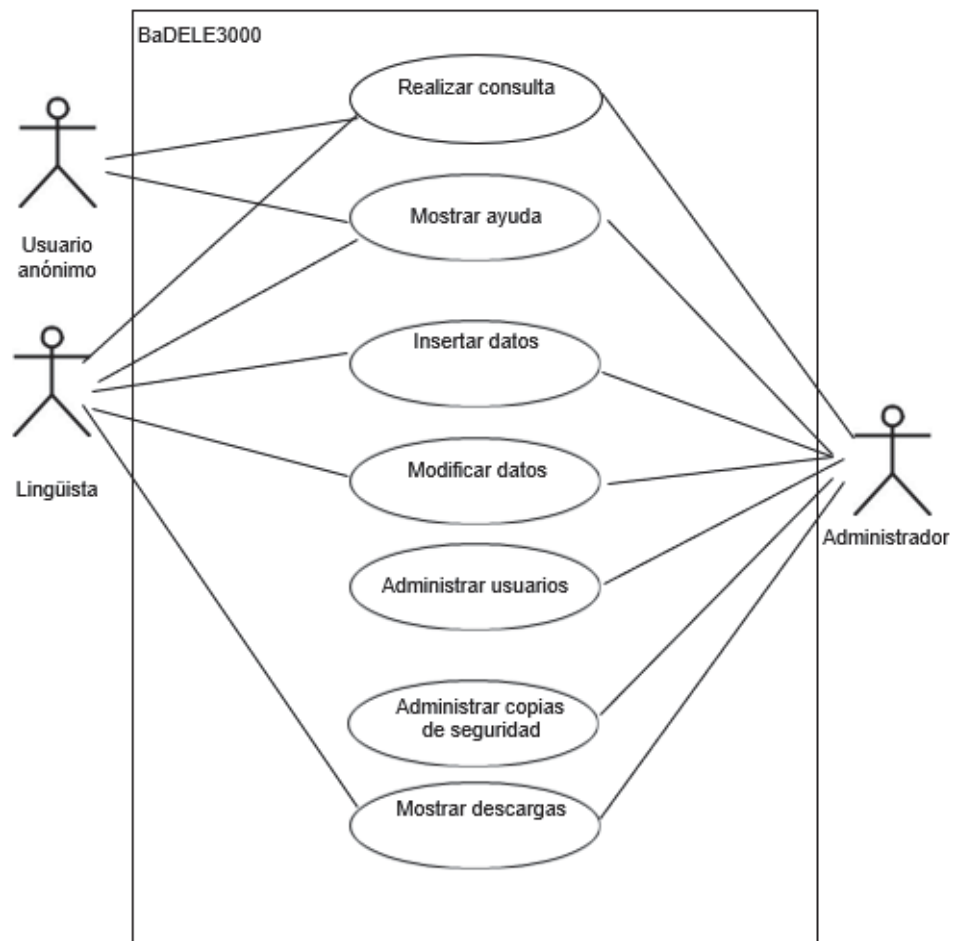


Figura 1: Diagrama de casos de Uso

Los actores van a representar aquellos agentes externos que interactúan con el sistema que se ha desarrollado. Un actor puede ser una persona, un sistema informático o una organización.

En el caso de esta aplicación existen tres tipos de actores. A continuación, se especifica cada uno de ellos ordenados de menor a mayor nivel de acceso:

- Usuario Anónimo. Es aquella persona que introduce en su navegador web la URL de la aplicación (www.badele3000fi.es), pero no se registra en ella. Este usuario podrá realizar una consulta, una consulta gráfica y ver la ayuda de la aplicación.
- Lingüista. Es aquella persona que accede al sistema, porque tiene un usuario con credenciales de lingüista. El nivel de acceso de este actor aún no le permite realizar todas las operaciones del sistema, pero podrá realizar las siguientes operaciones: realizar consulta, realizar consulta gráfica, acceder a descargas, insertar datos, modificar aquellos datos que previamente se hayan insertado y ver la ayuda de la aplicación.
- Administrador. Es aquella persona que accede al sistema con credenciales de Administrador. Este tipo de usuario tiene un acceso completo a todas las operaciones ofrecidas por el sistema: realizar consulta, realizar consulta gráfica, acceder a descargas, insertar datos, modificar datos, añadir usuario, eliminar usuario, realizar copia de seguridad, cargar copia de seguridad y ver la ayuda de la aplicación.

A continuación, se describen en formato breve cada uno de los casos de uso identificados:

- Realizar consulta. El usuario decide comenzar una consulta, para ello pulsa en el enlace correspondiente en el menú de la aplicación. Seguidamente, el sistema presenta el formulario de consulta. Una vez que el usuario haya rellenado el formulario, conforme a lo que desee consultar, este deberá pulsar el botón de «Buscar». Dependiendo de aquello rellenado por el usuario, el sistema presentara una tabla con la información asociada, o bien, un mensaje informando de que no existe información para la consulta pedida.
- Insertar datos. El usuario pulsa en el enlace «Insertar datos» del menú de la aplicación. Seguidamente, si el usuario tiene permisos, el sistema muestra un listado con todos los conceptos lingüísticos que la aplicación permite insertar. El usuario debe elegir uno de ellos. A continuación, el sistema muestra un formulario de inserción que el usuario debe de rellenar con aquello que crea conveniente. Una vez rellenado el formulario, el usuario debe pulsar en el botón Insertar. A continuación, el sistema muestra un mensaje informando si la inserción se ha realizado correctamente o no.
- Modificar datos. El usuario pulsa en el enlace «Modificar datos» del menú de la aplicación. Seguidamente, si el usuario tiene permisos, el sistema muestra un listado con todos los conceptos lingüísticos que la aplicación permite modificar. El usuario debe elegir uno de ellos. A continuación, el sistema muestra un formulario donde el usuario debe elegir aquel concepto específico que quiere modificar. Si el usuario elige un concepto que él no ha insertado, el sistema muestra un mensaje informando que no tiene los derechos suficientes para modificar ese dato. En caso contrario, el sistema muestra un último formulario para realizar la modificación. El usuario debe pulsar en el botón «Modificar» para que la modificación se realice correctamente.

A continuación, el sistema mostrará un mensaje informando si la modificación se ha realizado correctamente o no.

- Administrar usuarios. El usuario con permisos de administrador pulsa en el enlace «Administración de usuarios» del menú de la aplicación. Después el sistema muestra tres enlaces: añadir un usuario, eliminar un usuario y mostrar usuarios. El usuario pulsa en el enlace de añadir usuario, y el sistema muestra un formulario para realizar la inserción del usuario. Una vez rellenado el formulario, el usuario pulsa el botón «Insertar». Después el sistema mostrará un mensaje informando si la operación se ha realizado correctamente o no, en el caso de que el usuario pulse sobre el enlace de eliminar usuario, y el sistema muestra un desplegable donde aparecen todos los usuarios del sistema. El usuario selecciona aquel usuario que desea eliminar y pulsa el botón «Eliminar». Después el sistema muestra un mensaje informando si la operación se ha realizado correctamente o no, en el caso de que el usuario pulse sobre el enlace de eliminar usuario el sistema muestra una tabla donde aparecen el id del usuario, nombre del usuario y email del mismo.
- Administrar copia de seguridad. El usuario con permisos de administrador pulsa en el enlace «Administración de base de datos» del menú de la aplicación. Posteriormente, el sistema muestra dos enlaces: uno de ellos para realizar la copia de seguridad, y otro para cargar una copia de seguridad ya existente. El usuario pulsa en el enlace de realizar copia de seguridad. El sistema muestra un formulario. Una vez que el usuario rellena el formulario, pulsa en el botón «Realizar backup». Después el sistema muestra un mensaje informando si la copia de seguridad se ha realizado correctamente o no, en el caso de que pulse sobre el enlace cargar copia de seguridad, el sistema muestra un formulario. Una vez que el usuario rellena el formulario, pulsa en el botón «Realizar backup». Después el sistema

muestra un mensaje informando si la operación se ha realizado con éxito o no.

- Mostrar ayuda de la aplicación. El usuario decide ver la ayuda de la aplicación, para ello pulsa en el enlace correspondiente en el menú de la aplicación. Posteriormente, el sistema muestra el manual de ayuda de la aplicación.
- Mostrar descargas. El usuario que ha iniciado sesión decide acceder al menú de descargas, una vez dentro le aparecerán los diferentes enlaces a las descargas del software que se desee descargar, así como al manual de usuario

3.2. Instalación del sistema

Dado que BaDELE3000 es una herramienta que debería nutrirse y estar en continuo desarrollo para consulta, la idea de montarlo en un hosting y conseguir su acceso vía internet, es para que cualquier persona pueda acceder a ella para consultar y los expertos, para modificarla.

La instalación del sistema ha sido especialmente compleja, a pesar de que partíamos de un equipo que tenía instalado en local la aplicación y la base de datos. Para la puesta en marcha se eligió un proveedor de servicios de *hosting*, hostalia, configurable con versiones de los diferentes productos.

El *hosting* se montó sobre Linux como se puede ver en la siguiente figura.

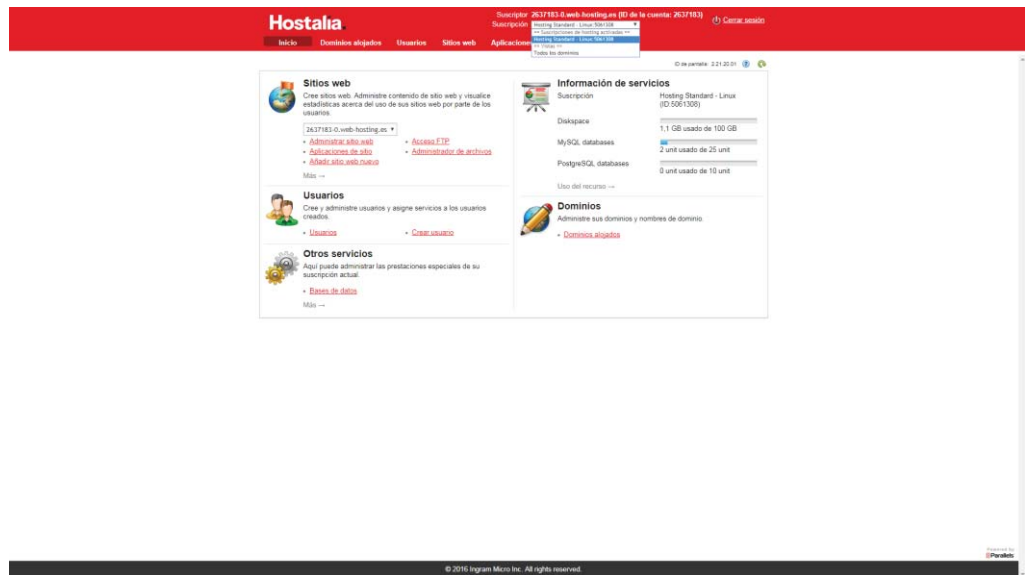


Figura 2: Panel de acceso de Hostalia

A pesar de introducir la aplicación en la carpeta adecuada (htdocs) del *hosting*, no funcionaba debido a las redirecciones internas que hace el propio hosting. Este extremo tuvo que ser solventado gracias a la configuración del archivo “htaccess” en la carpeta raíz, en el cual se le indica que apunte a la carpeta web de la aplicación (badele), donde se encuentra el archivo inicial (index.php). La Figura 3 recoge el código incluido en el archivo htaccess.

```
Options +FollowSymLinks

RewriteEngine on

RewriteCond %{REQUEST_URI} !(.*)web
RewriteRule ^(.*)$ web/$1 [L]
```

Figura 3: Código htaccess

Una vez escrito htaccess se consigue un funcionamiento correcto del sistema, esto que a primera vista puede parecer una cuestión baladí costó muchas horas de investigación, por lo que lo dejamos aquí documentado para futuras instalaciones.

En otro orden de problemas, tuvimos que adaptar los prefijos de la base de datos para que se correspondiesen con las del servidor, llamada db5061308_bd6 así como las modificaciones en el archivo *databases.yml* ubicado en la carpeta /config

El cotejamiento indica qué tipo de codificación queremos usar para los valores de ciertos campos. El cotejamiento elegido para las tablas de las bases de datos ha sido utf8_spanish_ci, debido a que los caracteres que se utilizarán en ellas se corresponden con el español moderno. Habría que tener esto en cuenta si en el futuro se añadieran idiomas con distinta codificación, por ejemplo, en el caso de introducir el idioma chino necesitaríamos codificar los datos con un tipo de cotejamiento que admitiera los símbolos chinos.

En la consola php del *hosting* existe la capacidad de elegir el cotejamiento adecuado, habilitándonos a utilizar la ñ (Véase figuras 4 y 5)

The screenshot shows the phpMyAdmin interface for a database named 'db5061308_bd6'. The main table lists various database tables with their respective actions and collation settings. The collation for all tables is 'utf8_spanish_ci'.

Tabla	Acción	Filas	Tipo	Cotejamiento	Tamaño
actante_c_de_I	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
actante_ul_de_I	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
clase_FL	Examinar Estructura Buscar Insertar Vaciar Eliminar	~35	InnoDB	utf8_spanish_ci	16 KE
clasificacion_de_C	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
clasificacion_de_ES	Examinar Estructura Buscar Insertar Vaciar Eliminar	~438	InnoDB	utf8_spanish_ci	32 KE
clasificacion_de_I_en_es	Examinar Estructura Buscar Insertar Vaciar Eliminar	~3,950	InnoDB	utf8_spanish_ci	432 KE
clasificacion_de_P	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
clasificacion_de_UL_en_C	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
combinaciones_de_FL	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	48 KE
concepto	Examinar Estructura Buscar Insertar Vaciar Eliminar	~3	InnoDB	utf8_spanish_ci	16 KE
correspondencia_FL_ES_UL	Examinar Estructura Buscar Insertar Vaciar Eliminar	~1,026	InnoDB	utf8_spanish_ci	224 KE
correspondencia_FL_I_UL	Examinar Estructura Buscar Insertar Vaciar Eliminar	~19,689	InnoDB	utf8_spanish_ci	4.5 ME
equivalencia_entre_C_y_ES	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
esquema	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	16 KE
esquema_lemma	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
es_G_de_ES	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	32 KE
ES_sinonima	Examinar Estructura Buscar Insertar Vaciar Eliminar	~0	InnoDB	utf8_spanish_ci	16 KE
ES_supone_FL_sin_valores	Examinar Estructura Buscar Insertar Vaciar Eliminar	~1	InnoDB	utf8_spanish_ci	32 KE
etiqueta_semantica	Examinar Estructura Buscar Insertar Vaciar Eliminar	~417	InnoDB	utf8_spanish_ci	80 KE
funcion_lexica	Examinar Estructura Buscar Insertar Vaciar Eliminar	~271	InnoDB	utf8_spanish_ci	64 KE
G_expresa_FI	Examinar Estructura Buscar Insertar Vaciar Eliminar	~9	InnoDB	utf8_spanish_ci	32 KE

Figura 4: Tablas de la base de datos

Hay que tener en cuenta que, del mismo modo que se ha actuado con las tablas, se ha de actuar con el servidor, eligiendo el cotejamiento indicado anteriormente, como se indica en la siguiente figura.

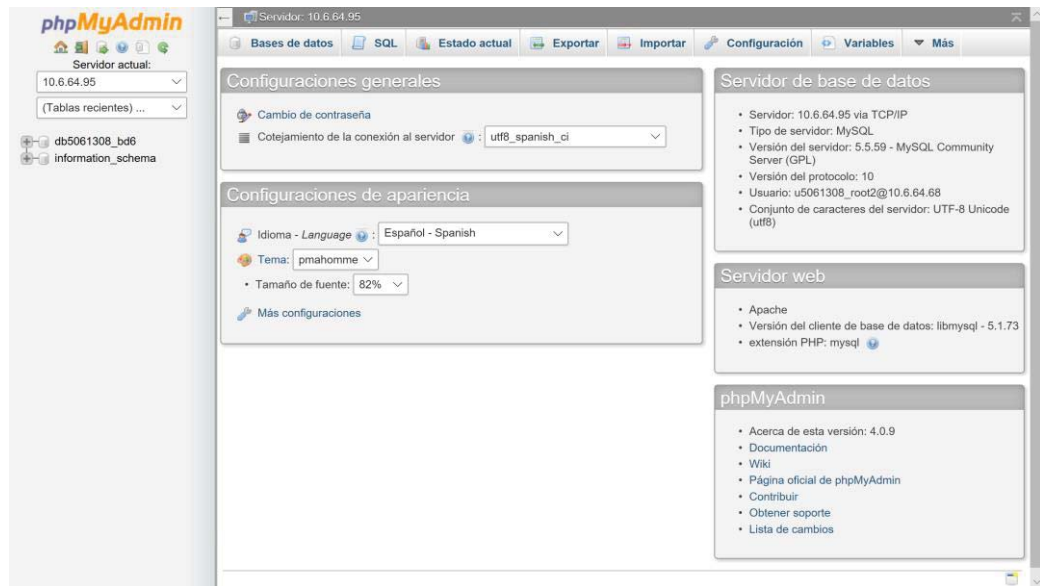


Figura 5: UTF base de datos

Para la carga de la base de datos se utilizaron los *scripts* contenidos en la carpeta /doc/scripts, mediante estos scripts se consigue la creación en primer lugar de la base de datos mediante `crea_bd.sql` y a continuación la carga de la misma a través de `insert_bd.sql`. Este último script es especialmente importante ya que en él están contenidos los valores que han sido introducidos por la lingüista para tener una base de datos coherente.

3.3 Visualización gráfica: Creación de un *scroll* en pantalla

Uno de los problemas de la consulta gráfica consistía en que en el caso de consultas grandes no aparecía el gráfico correspondiente completamente, con lo que quedaba fuera del área de visión del canvas programado en anteriores iteraciones. Para solventar esto, se modifica el canvas, que es un componente de HTML5 para dibujar de forma dinámica en una página web, dado que la consulta gráfica dibuja cada vez que sobre ella se hace una consulta.

La solución adoptada consiste en la creación de un *scroll* de tamaño fijo que se añadirá dentro del propio canvas. Esto se ha conseguido metiendo el canvas en un div con scroll de un tamaño amplio, y el propio canvas se ha redimensionado.

La Figura 6 contiene el código de modificación del canvas

```
body onload = "draw();" >
```

```
<div style="overflow: scroll;
width:1000px; height:1000px;">
    <canvas id="idCanvas" width="1000"
height="2000" style="border:1px solid #c3c3c3;">
        Version del navegador no
soportada. Por favor, actualice su navegador.
    </canvas>
</div>
```

Figura 6: Código de scroll

La ruta del fichero modificado es

```
\apps\frontend\modules\grafo\templates\salsuccess.php
```

De este modo se consigue que en el momento en el que se genere un gráfico mayor que el tamaño de la pantalla, el usuario se pueda mover a través de él para su completa visualización como se puede ver en la siguiente figura.

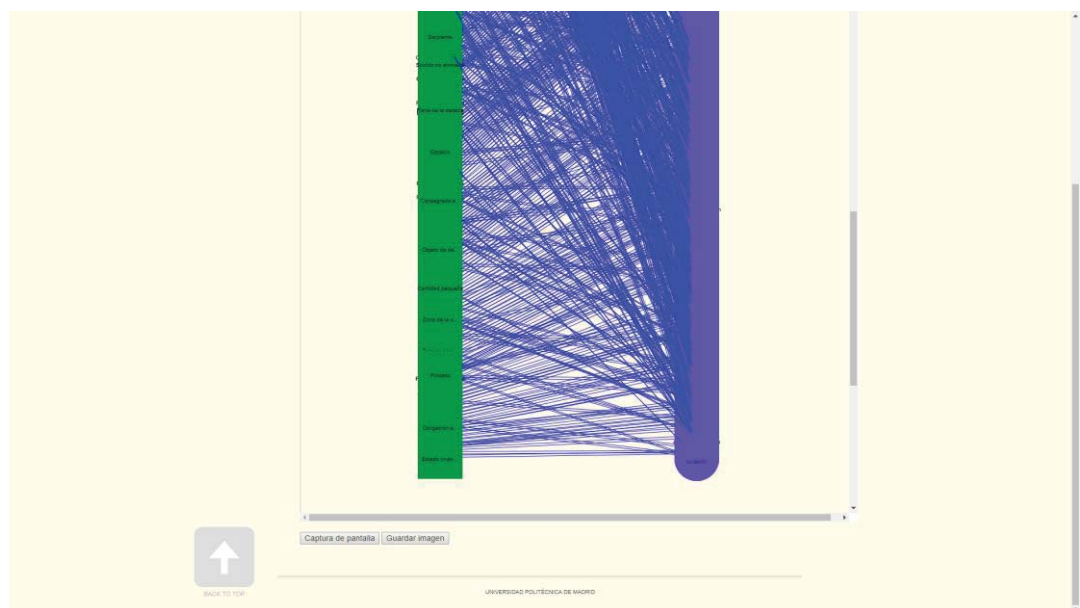


Figura 7: Scroll

3.4 Visualización gráfica: Todas las etiquetas semánticas de un lema.

La implementación de esta funcionalidad ha consistido tener en cuenta la modificación en la relación entre el lema y la etiqueta semántica, trasponiendo la consulta correspondiente desde el módulo de consultas hasta el módulo de grafo (el encargado de las consultas gráficas). El módulo grafo hace el resto, realizando el dibujo cómo ya estaba establecido.

3.5 Creación de un área de descargas

En la instalación de BaDELE3000 se ha detectado el inconveniente de utilizar versiones antiguas de software tanto para la implementación de la base de datos, como para el desarrollo de la aplicación. Se ha añadido un módulo para la descarga del software necesario tanto para el montaje de la aplicación en un sistema Windows como en Linux. Se añade una entrada al menú a los usuarios que han iniciado sesión, ya sean administradores o lingüistas, en la barra de navegación superior.



Figura 8: Página principal BADELE

Se modifican los ficheros en los que van definidos los enlaces y se define la seguridad y el propio módulo de descargas.

El módulo es un listado con los enlaces de descarga en la carpeta web de la app, los ficheros están autocontenidos en la propia aplicación haciendo referencia a la carpeta de \web\download, ya que estos ficheros son ajenos a la programación.

Se añaden las entradas al menú con niveles de seguridad para que solo lo vean los usuarios logados dentro de \apps\frontend\templates\menu.php

El elemento de asignación de URL es el elemento de configuración principal para la mayoría de las aplicaciones de PrettyFaces [http://www.ocpsoft.org/prettyfaces/]. En nuestro caso al construir un servidor de descargas y querer que nuestros usuarios accedan a la URL en la que está ubicado, solo necesitamos las características más básicas de la asignación de URL para lograr esto. Pretty path hace un mapeo de los enlaces que se incluyen en la descarga.

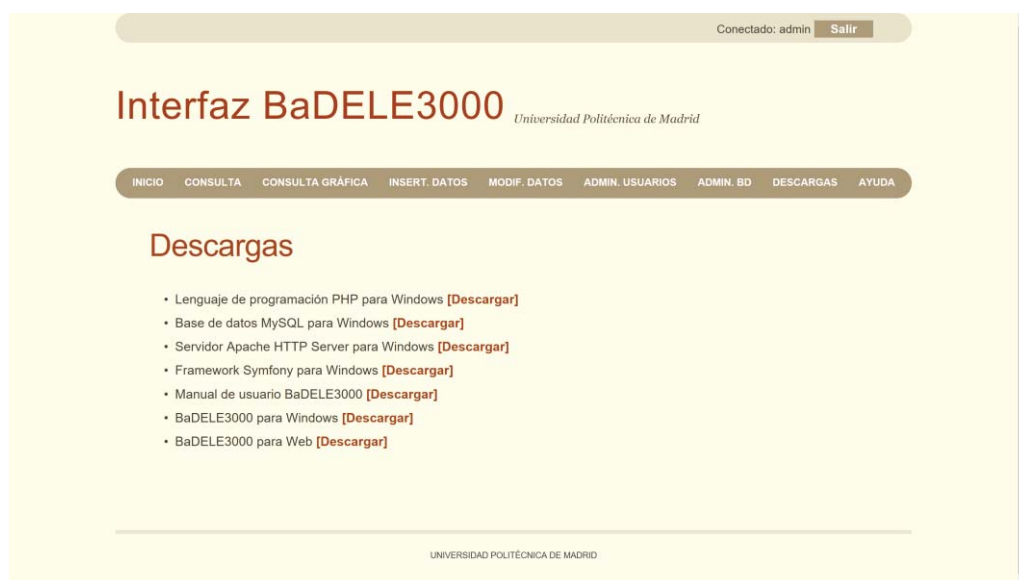


Figura 9: Área de descargas

Así pues, en el área de descargas aparecerán los siguientes apartados:

- Lenguaje de programación PHP para Windows

Es la versión 5.3.27 de PHP, utilizada en la versión de BaDELE3000 utilizada por el equipo en el que la aplicación está en funcionamiento. Si bien esta versión es bastante antigua se ha comprobado el funcionamiento de la misma en el equipo de referencia.

- Base de datos MySQL para Windows

Es la versión de `my_sql_installercommunity` 5.6.15.0, que por la misma razón anterior se ha estimado la necesidad de mantener debido al funcionamiento de la misma con el equipo de referencia.

- Servidor Apache HTTP Server para Windows

Es la versión de Apache 2.2.25, también en funcionamiento en el equipo de referencia al que estamos haciendo mención.

- Framework Symfony para Windows

Se utiliza la versión 1.4.2 de Symfony, esto condiciona toda la aplicación, ya que la estructura de la misma está basada en esta versión.

- Manual de usuario BaDELE3000

Es el mismo manual que se encontrará en el área de Ayuda para descargar en pdf, en este caso se ha direccionado al mismo archivo ya contenido en la aplicación

- BaDELE3000 para Windows

Se encuentra toda la estructura de directorios del equipo en el cual BaDELE3000 está montado en local, hay que tener en cuenta que se trataría de un equipo con un sistema operativo Windows 10 de 64 bits

- BaDELE3000 para Web

Se encuentra en un archivo zip toda la estructura de directorios existente en la web www.badele3000fi.es, el motivo de incluir íntegramente la estructura es con el fin de simplificar en el futuro la instalación del mismo por futuros administradores de la misma

3.6 Activación del modo producción

Se elimina la barra de depuración que dispone symfony para eliminar errores, para ello hay que modificar el archivo `\apps\frontend\modules\download\acciones` modificando el fichero cambiando el valor de la variable `$_SERVER['APP_ENV']` a `prod`.

3.7 Modificación de la página de inicio de la aplicación

Se ha incluido un texto explicando la finalidad de la aplicación. Para ello se modifica el archivo `\apps\frontend\modules\index\actions\actions.class.php` para contener además del texto indicado, sus referencias bibliográficas. Estas referencias se han tomado de internet, redireccionándonos a las mismas pulsando sobre ellas.



Figura 10: Página de inicio de BADELE

Es en este punto en el que se pudo comprobar que existía un problema que había sido detectado en los resultados de algunas consultas, al realizarlas los caracteres que arrojaban sus resultados no eran los que en realidad tendrían que aparecer, siendo estos una forma aleatoria de resultado.

Al utilizar php se emplea un tipo de codificación específica, en algunas páginas existía el error consistente en que al comienzo de la misma había un espacio a principio del fichero, por lo que el código no se visualizaba adecuadamente, es debido a que la codificación utilizada es UTF8 sin BOM, es la forma en la que se guarda el fichero, si no comienza por “<?php” no interpreta correctamente el código que aparecen en dicho fichero, las páginas de php dependiendo de la versión comienzan y terminan con una determinada etiqueta y en el caso que nos ocupa la etiqueta de comienzo y de fin del mismo ha de ser esta y de ese modo se consigue la correcta visualización de la misma.

Dependiendo de la configuración del editor utilizado puede que añada de forma predeterminada el BOM a los ficheros UTF-8. Esto pasa principalmente en aplicaciones Windows que al ejecutarlas en sistemas Linux suelen interferir con el proceso normal de éstas, como es nuestro caso, por ejemplo, en PHP comenzando a enviar el BOM al navegador, y evitando que las cabeceras se especifiquen correctamente.

La solución simplemente ha sido editar los ficheros con un editor de texto que nos permita el cambio de configuración y guardaremos este con formato UTF-8 sin BOM. En este caso ha sido utilizado el editor gratuito Notepad++.

3.8 Actualización del manual de usuario

Se han agregado nuevas entradas al manual de usuario con las mejoras realizadas, acompañados de capturas de pantalla, para la modificación del manual de usuario ha habido que modificar el archivo ubicado en la carpeta /web/download de BaDELE3000

Al estar el hosting web basado en una plataforma Linux, nos encontramos que tanto Linux como el lenguaje php distinguen entre mayúsculas y minúsculas, con lo que nos encontramos con el problema de que no aparecían determinadas imágenes al estar la extensión en mayúsculas, las fotos a las que se hace referencia están en la carpeta /web/images

3.9 Pruebas realizadas

A continuación se incluyen una serie de pruebas realizadas con la aplicación con el fin de comprobar si las funcionalidades han sido correctamente implementadas. Si bien las pruebas han sido muchas más en muchos aspectos de la instalación del sistema, se incluyen las cuatro más significativas.

3.9.1 Verificación del funcionamiento del *scroll*

Se realiza una consulta que arrojará muchos resultados para comprobar el funcionamiento del *scroll*: todos los vocablos que correspondan con la etiqueta semántica complemento

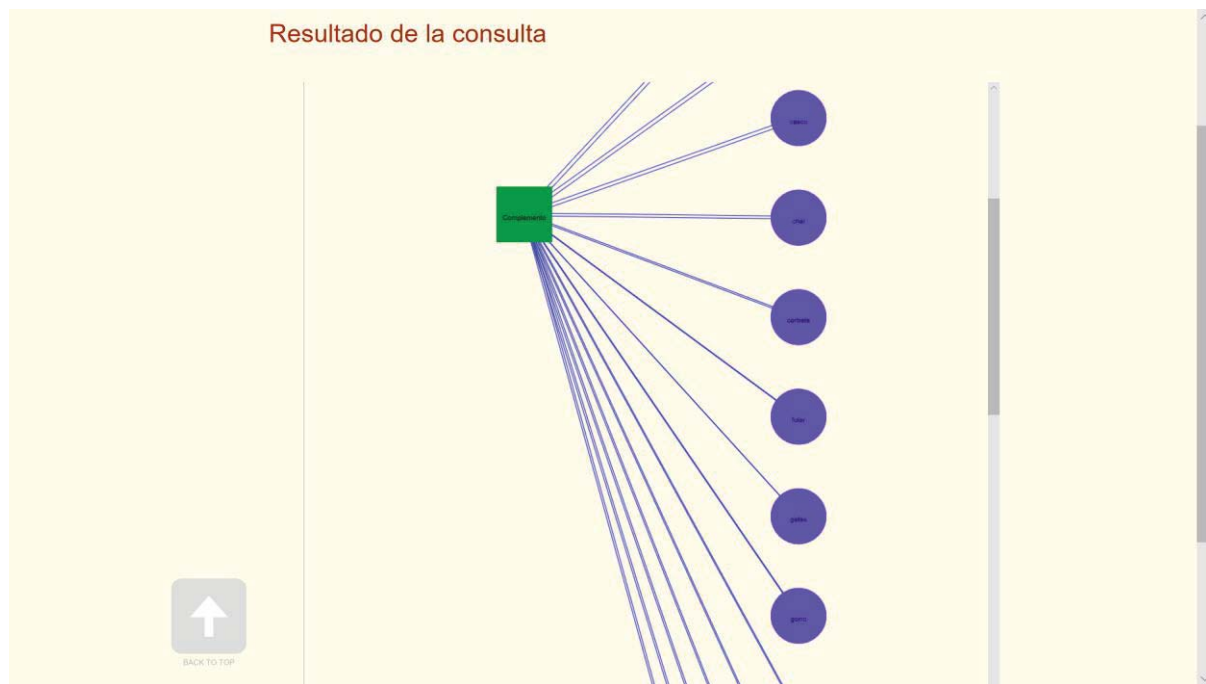


Figura 11: Scroll

Se comprueba el funcionamiento correcto del mismo

3.9.2 Verificación de la visualización de todas las etiquetas semánticas de un lema

Se comprueba con un lema conocido al cual se ha añadido más de una etiqueta semántica: el lema abandono en concreto tiene dos, acción inmaterial y estado físico.

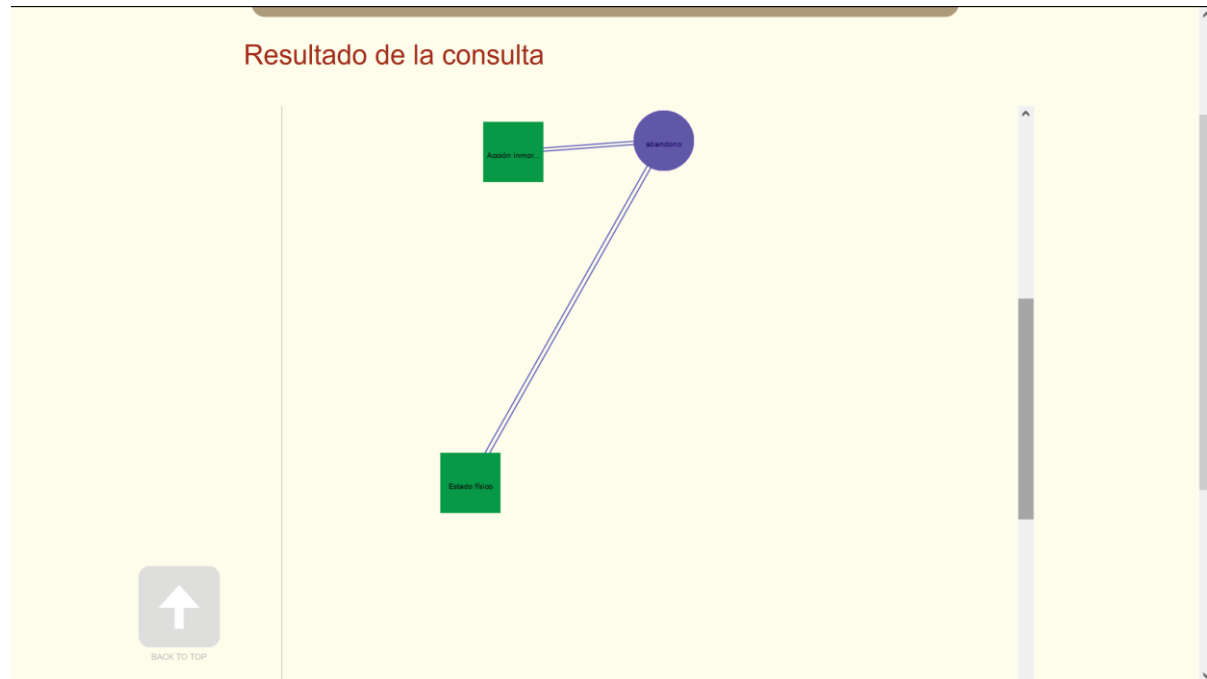


Figura 12: Consulta etiquetas semánticas

Se comprueba el funcionamiento correcto de la misma

3.9.3 Verificación del grafo para el idioma inglés

Se realiza la consulta indicando todos los vocablos con una etiqueta semántica en inglés



Figura 13: consulta de idioma.

En este caso aparece la única que por el momento ha sido insertada, vocablo *rain* con su etiqueta semántica *weather*

3.9.4 Verificación del módulo de descargas

Para ello en un primer lugar habrá que iniciar sesión en la aplicación, ya que de otro modo no aparecerá en la barra de menú al ser una funcionalidad restringida a los usuarios autenticados



Figura 14: Página principal de BADELE (Usuario Anónimo)

Una vez hecho esto comprobamos que efectivamente aparece en el menú la opción de descargas



Figura 15: Área de descargas.

Accedemos al módulo y al pinchar sobre cada uno de los enlaces nos dará la opción de realizar la descarga, con lo que se comprueba el funcionamiento correcto del módulo.

Capítulo 4

Conclusiones y trabajo futuro

En este Trabajo de Fin de Carrera se ha plasmado el trabajo que se ha realizado en la aplicación web BaDELE3000, la tarea y el reto principal consistió en poner en funcionamiento la aplicación en un servicio de alojamiento para que fuese accesible por cualquier usuario.

Se ha procedido a implementar todos los requisitos planteados al final del capítulo 1 que en su momento han sido planteados por la tutora, además de otros que sobre la marcha del proyecto han ido surgiendo, si bien se han visto limitados por la extrema complejidad en la instalación y puesta en funcionamiento de la web

No es una aplicación que haya sido desarrollada desde cero y en todas las iteraciones que se han empleado sobre BaDELE3000, las personas que han trabajado sobre ella hacen hincapié sobre este particular, ha sido lo más tedioso y complicado para la realización del posterior trabajo, por lo que convendría mantener el sitio web abierto así como tratar de realizar una instalación automática que fuese multiplataforma para esta aplicación.

Con la puesta en marcha en la web se consigue acceder desde los navegadores más populares (Microsoft Edge, Internet Explorer, Google Chrome, Mozilla Firefox), así como desde dispositivos móviles basados tanto en IOS y en Android, la accesibilidad universal está garantizada.

La estructura de la aplicación permite una relativamente sencilla ampliación de las funcionalidades, como ha sido el caso de tener en cuenta en la visualización gráfica la inclusión del idioma y la posibilidad de tener lemas con más de una etiqueta semántica. Lo realmente complicado es comprender su funcionamiento interno, al estar organizado de una forma modular, la comprensión acerca de la interconexión entre los diferentes módulos resulta compleja..

Si bien se han depurado muchos errores que aún persistían en la aplicación (por ejemplo, utf-8), en el futuro habría que continuar por esta senda, especialmente en el control de errores o en el tipo de consulta que se quiere hacer, para acotar cuales se pueden realizar y cuales no en un principio y para definir aquellas que no ofrecen resultados de una forma exhaustiva.

Alguna de las mejoras que se proponen para la aplicación y nuevas funcionalidades serían las siguientes:

- La inclusión del área de descargas debería ser ampliada con las diferentes versiones que de la aplicación se han venido haciendo a lo largo del tiempo, para poder tener así un histórico y poder contar con un control de versiones efectivo centralizado a modo de repositorio en la misma web.
- Contemplar para el administrador un control de accesos, si bien en la base de datos se puede ver las inserciones y los accesos, se debería poder ver el usuario que hizo cualquier cambio y la fecha en la que se realizó mediante una interfaz controlada.

Sería especialmente interesante abordar en el futuro la adaptación de la aplicación a symfony 4, es necesaria esta actualización ya que simplifica el árbol de directorios, creando una estructura más plana. Si bien el cambio sería bastante tedioso debido a la complejidad de la aplicación, sí que el esfuerzo realizado en un primer momento resultaría de gran interés para el futuro a la hora de mejorar el rendimiento de la misma.

Uso de varias lenguas. A pesar de que esta funcionalidad ha quedado programada, se debería tener en cuenta un modo diferente de realizar la traducción de estos conceptos a través de un idioma pivote. Si bien esto no está directamente relacionado con la aplicación tal y como está concebida y tendría relación más bien con los conceptos de la TST, los estudios de Ivan Guzmán de Rojas en el campo de las traducciones tomando como base el lenguaje Aymara y la potencia que desarrolla el sistema ATAMIRI (Véase Anexo A), sería una potente herramienta en el campo de la cuestión que nos atañe, consiguiendo ya no un traductor universal como pretende ATAMIRI, sino una herramienta mediante la cual las traducciones no dependerían de la corrección de la base de conocimiento de las que se vaya nutriendo sino de las reglas y de las colocaciones que se incluyan en la base de conocimiento de las funciones léxicas y las relaciones incluidas en BaDELE3000.

Finalmente, interesaría destacar que la experiencia adquirida en este Trabajo Fin de Carrera y el esfuerzo realizado en el mismo han resultado especialmente productivo en el campo de la comprensión de la TST, cumpliendo así con la transversalidad que se le quiere dar a los mismos.

Anexo A

ATAMIRI

El sistema ATAMIRI MT ha sido desarrollado por un grupo de investigación en ingeniería de idiomas (IGRAL), dirigidos por su creador Iván Guzmán de Rojas.

El sistema está compuesto de un conjunto de programas de procesamiento de lenguaje natural con su base de datos léxica y gramatical, ambos diseñados para servir como herramienta de soporte para un centro de traducción,. ATAMIRI también se ha sido adaptado para operar en una red de comunicaciones como un traductor interactivo de respuesta inmediata.

En Washington, en marzo de 1985, bajo los auspicios del Secretario General de la OEA, Iván Guzmán de Rojas presentó el primer prototipo de sistema capaz de traducir simultáneamente de inglés a español, francés y alemán y viceversa del español a los otros tres idiomas. Es el primer sistema de traducción automática con una capacidad genuinamente multilingüe cuya base de idioma se puede ampliar fácilmente, es decir, puede "aprender" enriqueciendo su base de datos léxica y gramatical, con un mínimo de esfuerzo de programación adicional. Una

vez que se ha "aprendido" un idioma, ATAMIRI puede usarlo como fuente o como idioma de destino en el entorno multilingüe implementado.

Debido a que su modelo lingüístico se basa en la representación del lenguaje formal del antiguo idioma aymara andino, este logro en la ingeniería del lenguaje es muy peculiar. Debido a su estructura matricial algorítmica, una representación formal aymara es útil como un puente sintáctico o interlingua, lo que posibilita la traducción simultánea del idioma de origen a otros idiomas de destino.

Tanto para la comunidad hispanohablante como para la antigua cultura aymara ATAMIRI representa un resultado singular en la ingeniería del lenguaje, ya que surge precisamente del tratamiento integrado de estructuras lingüísticas muy diferentes, como el español y el aymara.

Experiencia operacional

Desde 1985 hasta 1988, el sistema operó en un entorno de producción en la Oficina de Traducción del Canal de Panamá (traducción del inglés al español), así como en el Centro de Traducción Wang de Panamá, traduciendo manuales técnicos. En 1987 el sistema trabajó en una operación de prueba traduciendo textos técnicos del inglés al alemán, holandés, francés e italiano en los Centros de Traducciones Wang en Europa. En 1988 se probó el sistema para traducir del inglés al sueco en el Centro de Traducción de Wang en Estocolmo. Debido al cierre de los Centros de Traducción de Wang en Europa a fines de 1989, la versión de la máquina Wang VS ha sido descontinuada.

En 1990, la nueva versión de PC para redes Novell se instaló como una prueba en la oficina de internacionalización de productos BULL en París. En base a los

resultados, esta firma hizo una propuesta para comprar la tecnología; a pesar de esto, no hubo acuerdo, y la operación fue cancelada.

En este momento, también se instaló la misma versión para ser probada en la Oficina de Consultoría de David Stanton Associates en Bournemouth, Inglaterra. La cobertura reducida de la terminología técnica en el diccionario ATAMIRI para campos especializados no permitió a esta oficina de consultoría usar el sistema comercialmente.

En 1993, invitado por CompuServe, el autor instaló ATAMIRI en Data Technologies Office en Cambridge, con el objetivo de ejecutar una evaluación para comparar ATAMIRI con otros diez sistemas de traducción. La evaluación se basó en la traducción del inglés al francés de textos típicos que suelen aparecer en el intercambio de mensajes en la red CompuServe. Los resultados fueron positivos, a pesar de que el francés todavía no era un lenguaje fuerte en ATAMIRI y las pruebas fueron especialmente difíciles. En cualquier caso, ATAMIRI demostró ser el único sistema genuinamente multilingüe que puede operar en un entorno de trabajo real. Es un sistema muy rápido que permite una respuesta en tiempo real y tiene altas posibilidades de mejora.

ATAMIRI dejó de tener un uso comercial durante casi una década, aún así se realizó un considerable trabajo de programación para salvar el sistema de la obsolescencia debido a su plataforma operativa. Ahora funciona bajo Windows o UNIX. Gracias a una iniciativa de Unión Latina, una organización de la UE, durante enero y febrero de 2001, en La Paz, se llevó a cabo una operación piloto para probar las capacidades multilingües de ATAMIRI con la implementación del rumano en el sistema. El autor presentó el prototipo resultante en París, en la UNESCO. Este evento lanzó nuevamente ATAMIRI en la arena de la tecnología multilingüe MT.

Con el patrocinio de Unión Latina, París, y NEOTEC, La Paz, la versión actual ofrece un servicio de mensajería con traducción simultánea en 30 idiomas en www.atamiri.cc/qopuchawi

Mientras tanto, ATAMIRI opera a una escala muy reducida por gestionada por IGRAL, que ofrece servicios de traducción por correo electrónico desde La Paz. Esto contribuye a enriquecer la base de datos léxica del sistema y mejora su capacidad gramatical.

El origen de ATAMIRI se remonta al estudio de Los problemas lógico-lingüísticos de la comunicación social con el pueblo aymara (Problemática Lógico-Lingüística de la Comunicación Social con el Pueblo Aymara) que el autor del sistema escribió en 1979 y se publicó en 1982 y 1985 bajo los auspicios del IDRC de Canadá y distribuido en tres idiomas: español, inglés y francés. Aquí se demuestra el carácter algorítmico de la sintaxis aymara, que posee una lógica trivalente debajo. Aunque el sistema utiliza una representación formal del aymara como interlingua solo a niveles más profundos que la superficie morfémica, los textos también pueden traducirse en este lenguaje, siempre que haya suficiente lexema aymara en la base de datos lexicográficos del sistema; además, si se implementara un analizador morfo-sintáctico para los aymaras, este lenguaje también podría usarse como idioma de origen. En 1984, con el apoyo inicial de la UNESCO (oficina de Montevideo), IGRAL comenzó a introducir el lexema aymara en el sistema, probando y evaluar la capacidad de traducción en el idioma aymara para producir textos educativos utilizando ATAMIRI. El propósito era desarrollar textos escolares para la educación bilingüe y así revitalizar esta antigua lengua andina. A pesar de los resultados iniciales exitosos, este proyecto fue discontinuado debido a la falta de fondos y también porque IGRAL dio prioridad

a la traducción del inglés al español para cumplir los contratos firmados con la Comisión del Canal de Panamá y con Wang.

Diez años después (1994) Gracias a la donación de hardware del Sr. Sean Morriss de Vancouver, Canadá, IGRAL comenzó a enriquecer la base de datos lexicográficos del sistema con el lexema aymara nuevamente. Esta vez, se utilizó un escáner para colocar textos seleccionados a nivel escolar y universitario, en inglés, tratando de hacer un análisis léxico que ayude a encontrar la terminología aymara que aún por completar.

Como resultado de este análisis, cada uno de los lexemas fueron introducidos de manera ordenada, almacenando términos de diccionarios antiguos, especialmente del diccionario de Bertonio (elaborado entre 1585 y 1608 en la región del Lago Titicaca), y en muchos casos generando neologismos adecuados. basado en raíces y sufijos del idioma Aymara. Para ello, Iván Guzmán de Rojas diseñó técnicas de ingeniería lingüística, que permiten el control efectivo de la consistencia léxica y la adaptación sintagmática necesaria para obtener una estandarización funcional del idioma. Para realizar la tarea de traducción masiva de textos educativos en Aymara, en diferentes niveles y en diversos campos temáticos.

Las pruebas comparativas han demostrado que ATAMIRI es uno de los sistemas de traducción más rápidos del mundo. Con un procesador Pentium de 800 MHz, se conseguían traducir 700 palabras por segundo. Es más apropiado para la traducción de correo electrónico y páginas web en tiempo real. Debido a su capacidad multilingüe, ATAMIRI puede distinguir entre la sintaxis de hipertexto y la sintaxis normal del idioma en el que se escribe el texto incrustado. De esta forma, las páginas de hipertexto entregadas por los sitios web se pueden traducir inmediatamente en línea .

ATAMIRI es un sistema multilingüe de traducción automática y su diseño es pionero en el uso de una interlingua que utiliza una representación matricial del lenguaje natural, que permite la traducción simultánea de un idioma de origen en varios idiomas de destino. La versión de Atamiri 2001 utiliza el español, el francés, el portugués, el italiano, el rumano o el inglés como idioma de origen y destino.

Una implementación de 30,000 lexemas de un nuevo idioma objetivo requeriría de aproximadamente ocho meses. Para hacer que el idioma sea el idioma de origen, se requerirán cuatro meses adicionales. De esta forma, el nuevo idioma ya puede usarse para combinar pares de traducción con todos los demás idiomas que se han introducido previamente. Estas tareas se pueden realizar simultáneamente para más de un idioma, ya que la expansión a otros idiomas no se realiza en pares. En lugar de administrar varios diccionarios, solo se administra una única base de datos léxica multilingüe, lo que permite una entrada fácil de la propia terminología del usuario. Las capacidades multilingües de ATAMIRI permiten ahorros significativos en cuanto al dinero y el tiempo de implementación para nuevas implementaciones de idiomas.

Anexo B

Estructura del sistema

B.1 Introducción

En este apartado se va a tratar de explicar una visión global del sistema implementado, ya que no es un desarrollo que parte desde cero conviene hacerlo con el fin de poder tener una comprensión del mismo.

B.2 Diseño de la Base de datos

En esta aplicación se ha utilizado un sistema de gestión de datos que permite almacenar, recuperar, actualizar y eliminar la información de los elementos que pueden ser persistentes en el sistema.

Para el diseño del modelo de datos se creó en primer lugar un diagrama entidad-relación. Dado que se optó por seguir un modelo relacional para la base de datos, este diagrama después se transformó en un conjunto de tablas que correspondiera al esquema de la base de datos. El paso a tablas no ha sido incluido en esta memoria.

B.2.1. Diagrama entidad-relación

En este diseño partimos de un diagrama entidad-relación el cual ha sufrido modificaciones para la inclusión de la funcionalidad de más de un lema por etiqueta semántica, así como para la inclusión del idioma. A continuación se muestra el diagrama entidad-relación final. Los rectángulos representan las entidades, los círculos atributos de entidades o relaciones, los rombos las relaciones entre entidades y los números, la cardinalidad de las relaciones, es

decir, el número de ocurrencias de las entidades con las que puede estar relacionada la ocurrencia de una entidad. En los siguientes párrafos se describen los elementos más importantes.

- **Unidad léxica.** Se define unidad léxica como forma lingüística o conjunto de formas lingüísticas que posee significado propio. Una unidad léxica puede ser base de alguna función léxica con uno o varios lemas como valores. De la misma manera, puede ser base de alguna función léxica con uno o varias etiquetas semánticas. Además, una unidad léxica puede pertenecer a uno o varios conceptos y puede ser actante de uno o varios lemas.
- **Lema.** Representa cada una de las entradas de un diccionario. Puede tener uno o varios rasgos, semantemas, esquemas o locuciones. Puede ser actante de uno o varios conceptos y unidades léxicas. Por último, puede pertenecer a una o varias etiquetas semánticas.
- **Concepto.** Esta entidad almacena un sistema de categorías o conceptos de un dominio, en concreto el mundo en el que vivimos. Es una entidad que tiene una relación reflexiva en la que un concepto puede pertenecer a uno o varios conceptos. De esta forma, se obtiene una tabla que sirve para clasificar los conceptos del dominio. Un concepto puede formar parte de ninguna o varias etiquetas semánticas, unidades léxicas o lemas. Y de igual manera le ocurre a una etiqueta semántica, unidad léxica o lema.
- **Etiqueta semántica.** Una etiqueta semántica identifica, generalmente, el inmediato genérico que se emplea en su definición. Tiene una relación reflexiva en la que una etiqueta semántica puede pertenecer a una o varias etiquetas semánticas. De esta manera, se obtiene una clasificación de etiquetas semánticas. Además una etiqueta semántica

se relaciona con uno o varios conceptos. Una etiqueta semántica puede suponer una o varias funciones léxicas.

- **Función léxica.** Una función léxica pone en relación dos unidades léxicas, una de ellas recibe el nombre de base y la otra se llama valor.
- **Paráfrasis.** Algunas se corresponden con funciones léxicas que admiten alguna explicación en lenguaje natural. Cada una de las cuales constituiría una paráfrasis de la función léxica.
- **Idioma.** Se corresponde con los diferentes idiomas que se le puedan añadir a la base de datos, se relacionará en una relación 1:N con las paráfrasis, etiquetas semánticas y lema, con lo que al menos un idioma ha de tener cada una de ellas. Esta ha sido la nueva funcionalidad añadida en esta iteración de BaDELE.