

MANUAL DE USUARIO

adquisición semi-automática del conocimiento: una arquitectura preliminar

TABLA DE CONTENIDOS

manual de instalación	7
1. manual de instalación	9
1.1. requisitos de hardware	9
1.2. requisitos de software	9
1.2.1. requisitos generales	9
1.2.2. requisitos específicos	10
1.2.3. instalación de los programas necesarios	10
1.3. instalación de la aplicación	13
manual del usuario	15
2. manual del usuario	17
2.1. arrancar la herramienta	17
2.2. vista corpus	20
2.2.1. generación	20
2.3. clasificación	22
2.3.1. clasificación por términos	22
2.3.2. clasificación por dependencias	24
2.4. vista búsqueda	29
2.4.1. búsqueda por términos	29
2.4.2. búsqueda por dependencias	32
2.5. vista configuración	32
2.5.1. gestión de temas	33
2.5.2. gestión de idiomas	35
2.5.3. gestión de extractores	37
2.5.4. gestión de analizadores sintácticos	39
3. manual del usuario experto	41
3.1. añadir un extractor	41
3.2. añadir un analizador sintáctico	43

índice de figuras 47

 4. índice figuras.....49

MANUAL DE INSTALACIÓN

1. MANUAL DE INSTALACIÓN

En esta sección del manual de usuario se describe la instalación de los recursos necesarios para poder funcionar con la herramienta desarrollada.

1.1.REQUISITOS DE HARDWARE

- CPU: 400 Mhz o superior (recomendable 1.5GHz). El tipo de procesador es indiferente, el único requisito es que exista una máquina virtual de java para dicho procesador.
- Memoria: 512MB (recomendable 1GHz).
- Disco duro: 20 GB.
- Conexión a Internet para permitir a la herramienta de traducción de Google realizar las traducciones.
- Unidad de disco.

1.2.REQUISITOS DE SOFTWARE

1.2.1.REQUISITOS GENERALES

Se describen en este apartado cuales son los requisitos para la arquitectura general implementada.

- Sistema operativo Linux para poder ejecutar aquellas herramientas desarrolladas para el caso concreto. Se recomienda una distribución basada en Debian.

- Máquina Virtual Java *jdk 1.5*.
- Herramienta ANT para la compilación e instalación del programa.
- Librería *dot*. Es una librería para crear grafos jerárquicos. Permite generar los gráficos utilizados al mostrar los nodos de la ontología.

1.2.2.REQUISITOS ESPECÍFICOS

Se describen en este apartado cuales son los requisitos para las herramientas proporcionadas al sistema en el desarrollo del caso concreto.

- Compilador de PERL, por normal general este compilador ya viene incluido en las distribuciones de Linux.
- Twig que es un módulo PERL para analizar fichero XML.
- Etiquetador TreeTagger.
- Extractor Acabit.
- Extractor Fastr.
- Analizador sintáctico Erial.

1.2.3.INSTALACIÓN DE LOS PROGRAMAS NECESARIOS

En el disco proporcionado con la documentación se encuentra todas las herramientas necesarias para instalar el sistema, salvo el propio sistema operativo y las librerías *twig* y *dot* que se instalarán con el comando *apt-get* (*apt* es un sistema de gestión de paquetes software, concretamente *apt-get* es el comando utilizado para usar dicho sistema a través de la línea de comandos).

(A) INSTALACIÓN DE LA MÁQUINA VIRTUAL DE JAVA

- En primer lugar es necesario descargar el *jdk* del sitio web de SUN (<http://java.sun.com/j2se>). Si se trabaja bajo Debian hay que bajarse el fichero con extensión *.bin*
- El proceso de instalación accederá a ciertos directorios para los cuales es necesario tener permisos de administrador, por lo tanto se abre una consola y se entra tecla *sudo -s* a continuación se solicita la contraseña, habrá que introducir la del administrador.
- Se dan permisos de ejecución al archivo que se acaba de descargar: (nota: en la notación se pone <version> el usuario hará coincidir ese valor con la versión que se descargue)

```
jdk-<version>-linux-i586.bin
```

- Se copia dicho archivo a */usr/local/jdk* (si no existe hay que crear este directorio).
- Se ejecuta el archivo desde la ubicación anterior escribiendo en la consola:

```
./jdk-<version>-linux-i586.bin
```

En la consola se solicitará aceptar la licencia.

- Una vez instalado es necesario actualizar la variable PATH del sistema añadiéndole la ruta del directorio donde se acaba de instalar el *jdk*. Para ello, el usuario puede acceder al *.bashrc* de su *home* o bien editar el fichero */etc/profile* (para lo cual necesitará permisos de administrador).

(B) INSTALACIÓN DE LA HERRAMIENTA ANT

Esta herramienta se descarga del sitio web:

```
http://ant.apache.org/bindownload.cgi
```

- Se crea el directorio */usr/local/ant* (para cuyo acceso se necesitan permisos de administrador).
- El fichero descargado debe situarse en el directorio */usr/local/ant*.

- A continuación es necesario descomprimir el archivo:

```
tar -xzvf archivo.tar.gz
```

- Por último es necesario actualizar la variable ANT_HOME con el valor */usr/local/ant*, así como la variable JAVA_HOME con el valor */usr/local/jdk*. Estas variables, al igual que ocurría con el *jdk*, se pueden actualizar bien en el fichero *.bashrc* del usuario bien en */etc/profile* para todos los usuarios

(C) INSTALACIÓN DE LA LIBRERÍA DOT

Para esta operación es necesario utilizar el comando *apt-get*, para ello se escribe en la consola:

```
apt-get install springgraph
```

(D) INSTALACIÓN DE LA LIBRERÍA TWIG

Para esta operación es necesario utilizar el comando *apt-get*, para ello se escribe en la consola:

```
apt-get install libxml-twig-perl
```

No se describe cual es el proceso de instalación de TreeTagger, Fastr, Acabit y Erial. Esto es así, porque debido a las modificaciones que se han hecho de estas herramientas, ya están incluidas en la distribución del disco que acompaña a la documentación. Se verá a continuación que las mismas ya están incorporadas en la estructura de ficheros y solamente será necesario realizar ciertas modificaciones en el fichero *etc/profile*. De cualquier forma, consultando las referencias de la web se pueden encontrar los sitios donde poder descargar estas herramientas. En estas páginas se describe el proceso de instalación que puede ser de interés para el usuario si desea hacer una prueba de las mismas con su funcionamiento estándar. En el caso de Erial no se proporciona ninguna fuente de descarga pues no es una herramienta de código libre.

1.3.INSTALACIÓN DE LA APLICACIÓN

En el disco que acompaña la documentación existe un directorio llamado programa que el usuario deberá copiar al directorio en él que desee instalar la aplicación.

- Si se supone que el directorio de instalación va a ser */home/usuario/conocimiento/* copiamos en su interior el contenido de la carpeta.
- A continuación procesamos los ficheros con *ant*, por lo que se escribe en línea de comandos:

```
ant AdConocimiento
```

- El programa en estos momentos ya está instalado. Ahora es necesario actualizar la variable CLASSPATH. En primer lugar se añade a la variable el directorio donde se sitúan los archivos del programa instalado, es decir:

```
CLASSPATH=/home/usuario/conocimiento/distribucion/bin
```

- A continuación se añaden a la variable todos los ficheros situados en */home/usuario/conocimiento/lib* . Por ejemplo:

```
CLASSPATH=/home/usuario/conocimiento/lib/lucene-core-2.1.0.jar:$CLASSPATH.
```

- Una vez hecho esto, accediendo a */home/usuario/distribucion* se puede ejecutar el script que lanza la aplicación:

```
./AdConocimiento.sh
```

- En el directorio */home/usuario/conocimiento/distribucion/* existen dos carpetas *Extract* y *Analyzer*. Ambas carpetas contienen respectivamente, los extractores y los analizadores sintácticos concretos utilizados. Estarán listos para funcionar en el momento en él que se los referencie en el fichero */etc/profile*. Es necesario publicarlos en dicho fichero para que la máquina virtual java sea capaz de localizarlos. En dicho fichero se añadirá:

export

PERL5LIB=/home/usuario/conocimiento/distribucion/Extract/frances/Acabit/Flemm

*PATH=/home/usuario/conocimiento/distribucion/Extract/frances/FastrFr/fastr/bin
/:\$PATH*

PATH=/home/usuario/conocimiento/distribucion/Extract/Tagger/cmd/:\$PATH

export PATH=/home/usuario/conocimiento/distribucion/Extract/Tagger/bin/:\$PATH

Se describe a continuación el significado de lo añadido al fichero:

- La primera variable, PERL5LIB, es necesaria para la ejecución de Acabit.
- A la variable PATH en primer lugar se le añade la referencia a Fastr, esto lo utilizarán tanto la versión en español como la versión en francés.
- A continuación se añade la referencia al etiquetador TreeTagger.

En cuanto a las clases que implementan cada una de estas herramientas ya están incluidas, por defecto en la distribución.

Si se decide utilizar estas herramientas será necesario aportar la información correspondiente en el fichero de configuración que originalmente está vacío.

MANUAL DEL USUARIO

2. MANUAL DEL USUARIO

2.1 .ARRANCAR LA HERRAMIENTA

En un primer instante, una vez se ha lanzado la aplicación la pantalla de acogida es la que se puede ver en la siguiente figura

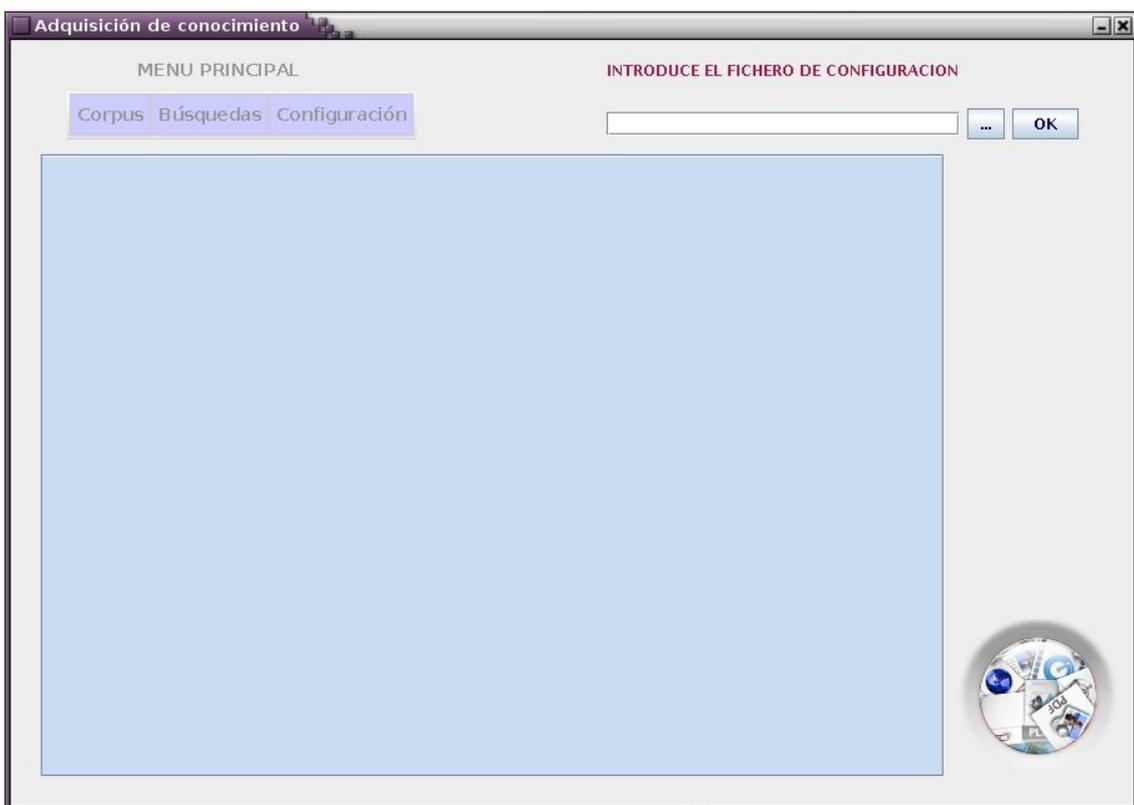


Figura 1 entrada a la aplicación

En la esquina superior derecha se encuentra una caja de texto donde se deberá introducir la ruta en la que se ubica el fichero de configuración. El botón cuyo símbolo son tres puntos abrirá una caja de diálogo que permitirá elegir el fichero de configuración navegando por el disco. Se muestra a continuación dicha caja.

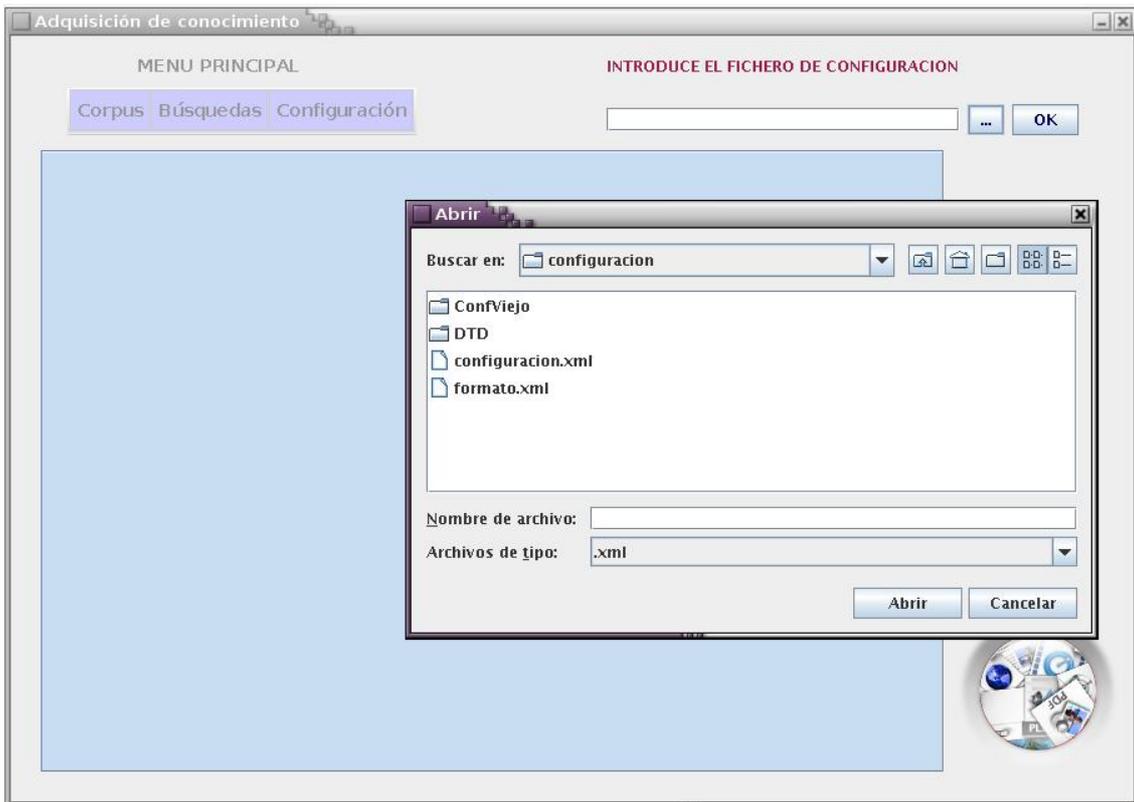


Figura 2 elegir un fichero de configuración

Esta caja de dialogo sólo permite elegir ficheros con extensión *.xml*, que es la que debe poseer el fichero de configuración.

Una vez se ha introducido el fichero se debe pulsar la tecla OK. Se podrán producir a continuación los siguientes tres eventos:

- Si el fichero de configuración no es válido, el sistema mostrará un diálogo como el siguiente:



Figura 3 error en la estructura del fichero de configuración

- Si el fichero de configuración tiene nodos repetidos, el sistema mostrará un diálogo como el siguiente:

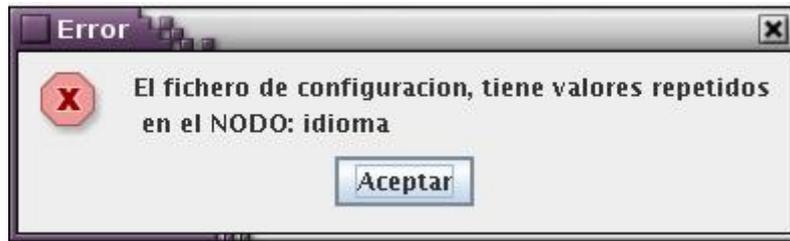


Figura 4 elementos repetidos en el fichero de configuración

- Si el fichero de configuración es correcto, el sistema activará el menú principal, mostrando una vista como la siguiente:

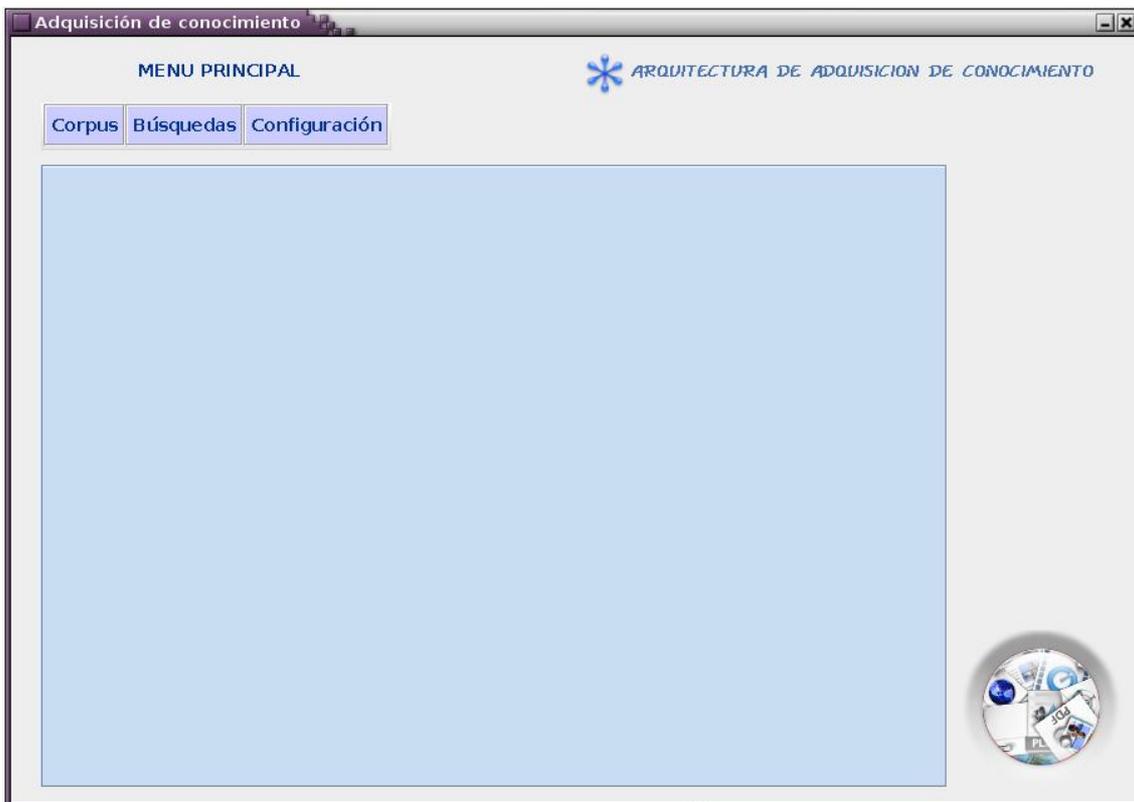


Figura 5: menú principal

2.2.VISTA CORPUS

Se van a describir las operaciones en la secuencia lógica en la que deberían de ser hechas si se estuviese trabajando con elementos nuevos en el sistema, es decir, si se está procediendo con un tema e idioma recién introducidos.

2.2.1.GENERACIÓN

Una vez se haya introducido el fichero de configuración correctamente, la vista que se muestra, como ya se ha comentado, es la de la Figura 5 descrita en el apartado anterior.

En primer lugar se elige en el menú principal (situado arriba a la derecha) la *vista corpus*. Una vez en esa vista, se elige la *generación de corpus*. Con esta opción se van a poder introducir nuevos documentos al sistema y por consiguiente se va a actualizar el corpus, se realizará el análisis sintáctico de cada documento y se actualizará el índice asociado.

El usuario elegirá el tema, una vez elegido un tema se muestran los idiomas disponibles para el mismo. Una vez el idioma elegido pueden suceder dos cosas:

- Si el idioma no tiene ningún recurso de tipo analizador sintáctico asignado. En este caso el sistema muestra un mensaje de error como el siguiente:



Figura 6 error idioma sin recurso asociado

- Si los datos son correctos se muestran los analizadores sintácticos disponibles para ese idioma.

Si se ha producido el segundo caso, el usuario ya puede indicar la ruta del nuevo documento que se va añadir asociado al idioma y tema elegido. También cabe la posibilidad de que se introduzca un directorio. Con todos los datos disponibles se selecciona el botón *Actualizar corpus* y empieza el proceso. Una vez éste ha finalizado, es posible realizar el resto de las tareas. Se muestra a continuación la vista correspondiente a la operación descrita.

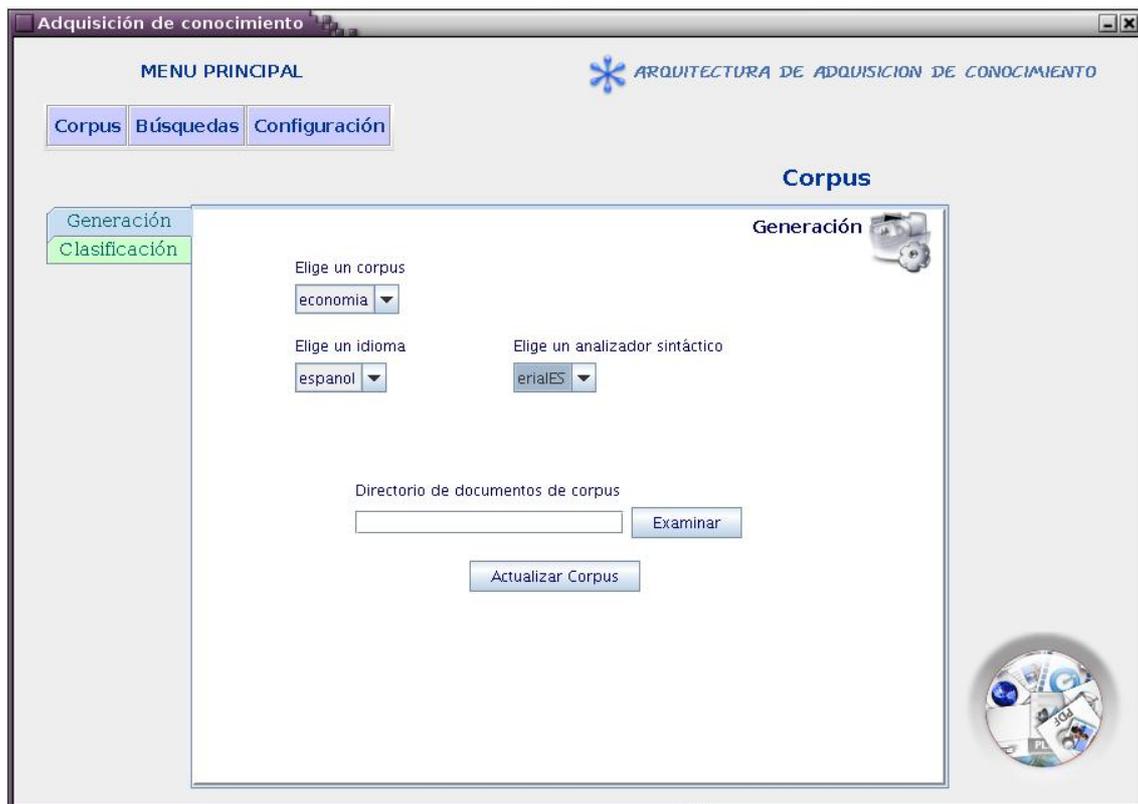


Figura 7 vista generación

NOTA: En el caso concreto desarrollado y dadas las características de las herramientas incluidas, los textos deben ser proporcionados con el formato UTF-8.

2.3. CLASIFICACIÓN

Para acceder a la vista de clasificación se selecciona la opción correspondiente en el menú principal.

2.3.1. CLASIFICACIÓN POR TÉRMINOS

En primer lugar se va a detallar la clasificación por términos. Ahora bien, no tiene porque llevarse esta tarea a cabo antes que la clasificación por dependencias. Ambas tareas son independientes la una de la otra.

Se selecciona la opción de clasificación y a continuación la pestaña términos. La vista disponible es la siguiente:

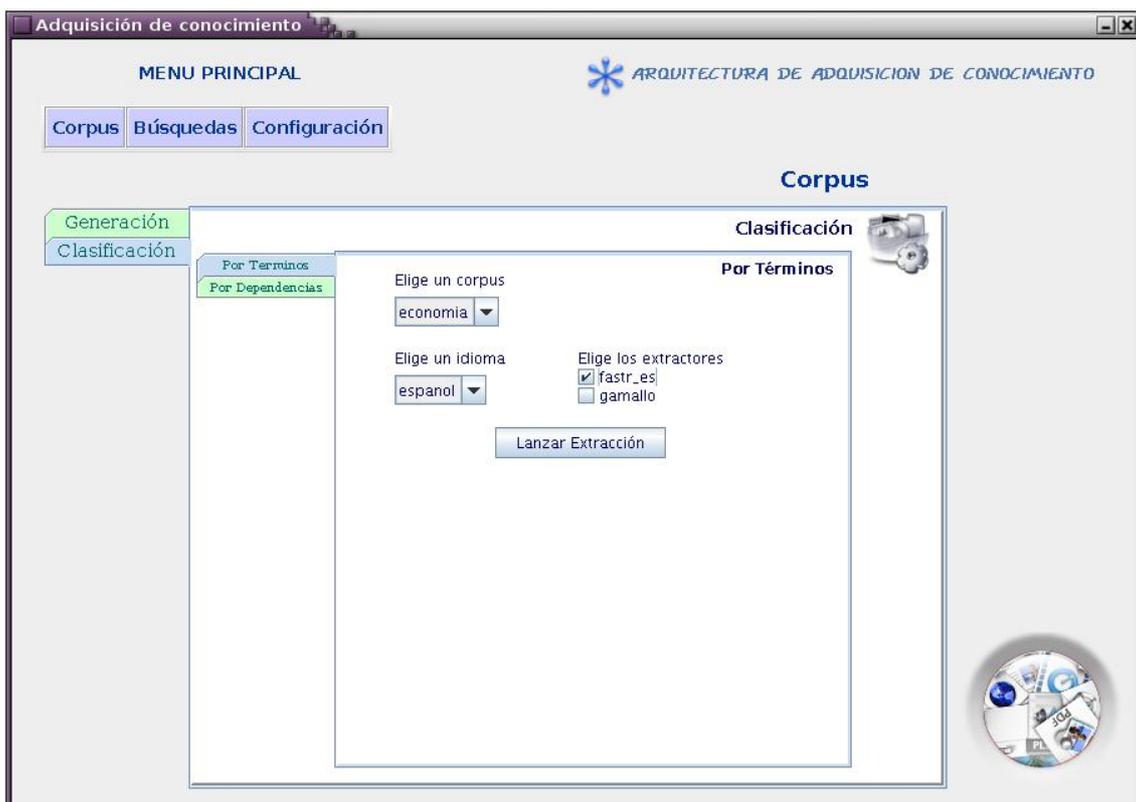


Figura 8 vista principal de clasificación por términos

Para realizar la clasificación por términos en primer lugar es necesario lanzar la extracción de términos. De este modo, el usuario debe introducir el tema e idioma, se pueden dar tres casos:

- Si el idioma no tiene ningún extractor asignado, el sistema mostrará una caja de diálogo similar a la Figura 6, donde se indicará que el idioma no tiene ningún extractor asignado.
- Si no existe ningún corpus asociado al idioma y al tema, el sistema muestra una caja de diálogo como la siguiente:



Figura 9 error no existe corpus

- El idioma tiene al menos un extractor asignado, el sistema muestra esos extractores. En ese momento el usuario puede elegir el extractor con el que desea trabajar (puede elegir más de un extractor a la vez). Una vez elegido el extractor (o extractores) puede pulsar el botón *Lanzar Extracción*. Entonces se le muestra una caja de diálogo informándole de la operación que ha solicitado. El usuario deberá confirmar dicho diálogo para proceder a la extracción.

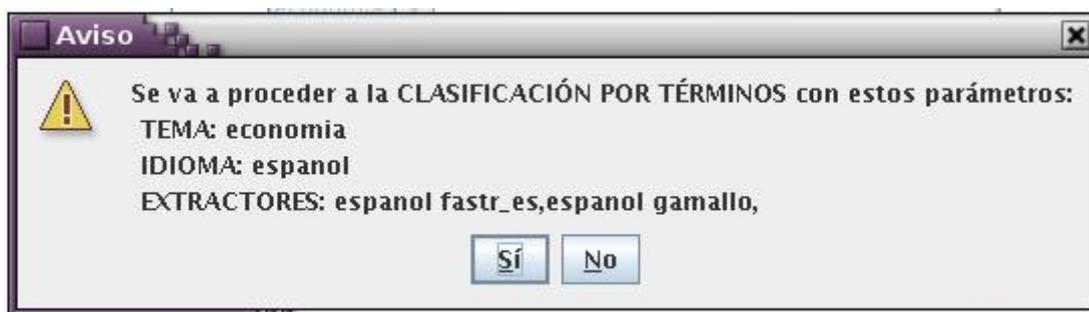


Figura 10 aviso clasificación por términos

Una vez la extracción ha finalizado se muestra el botón de *Lanzar Clasificación* que el usuario deberá pulsar. Cuando la clasificación retorne, se activará el botón de *Crear Ontología* el usuario deberá pulsarlo para generar la ontología correspondiente a los términos. En este último caso se generará la ontología correspondiente.

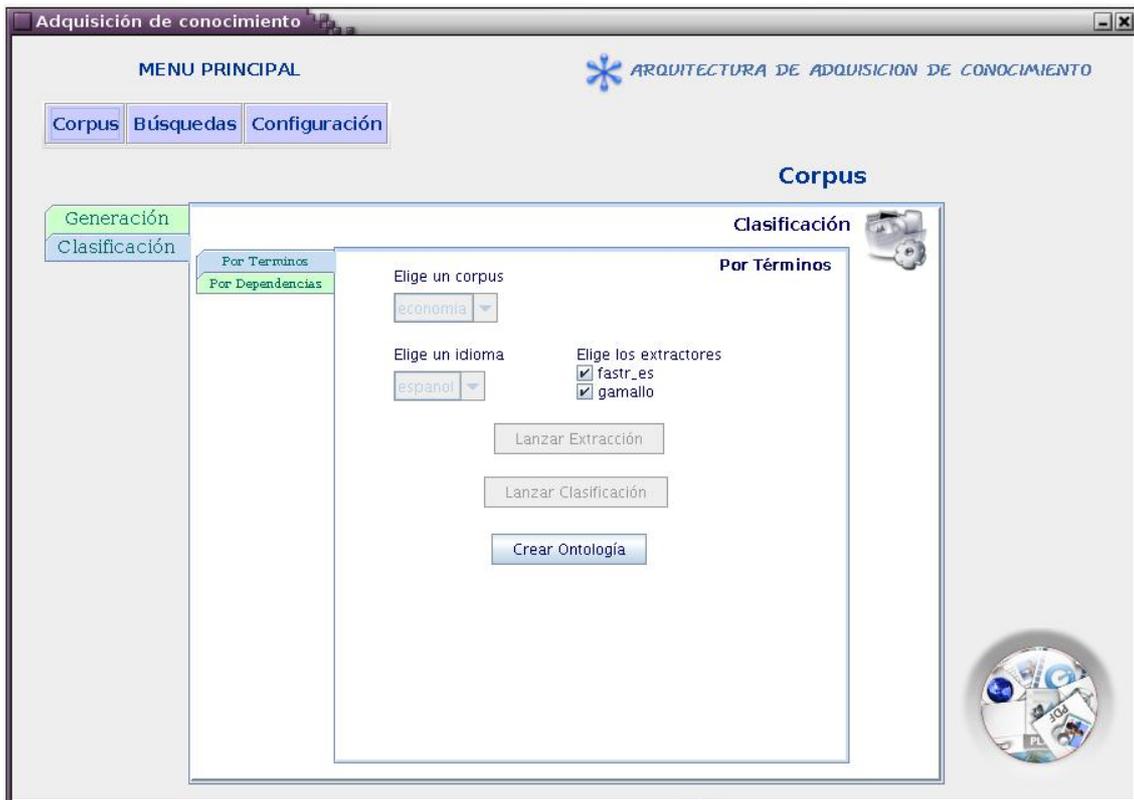


Figura 11 vista clasificación por términos crear la ontología

En esta figura se muestra el último botón activado del proceso de clasificación por términos, el botón de *Crear Ontología*. Una vez se ha creado la ontología, ésta puede recuperarse en el directorio *Ontologia/temaActual/idiomaActual/ontologia*, donde *temaActua* e *idiomaActual* son, respectivamente, el tema e idioma que el usuario ha elegido.

2.3.2. CLASIFICACIÓN POR DEPENDENCIAS

Se selecciona la opción de clasificación y a continuación la pestaña dependencias. El usuario puede, entonces, introducir el tema e idioma. Efectuado esto puede desencadenar dos eventos:

- Si el tema asociado al idioma y al corpus no existe, el sistema mostrará un mensaje como el de la Figura 9.
- Si el corpus asociado existe, se muestra entonces el botón que permite activar la clasificación por dependencias, este botón es el titulado *Lanzar Clasificación*

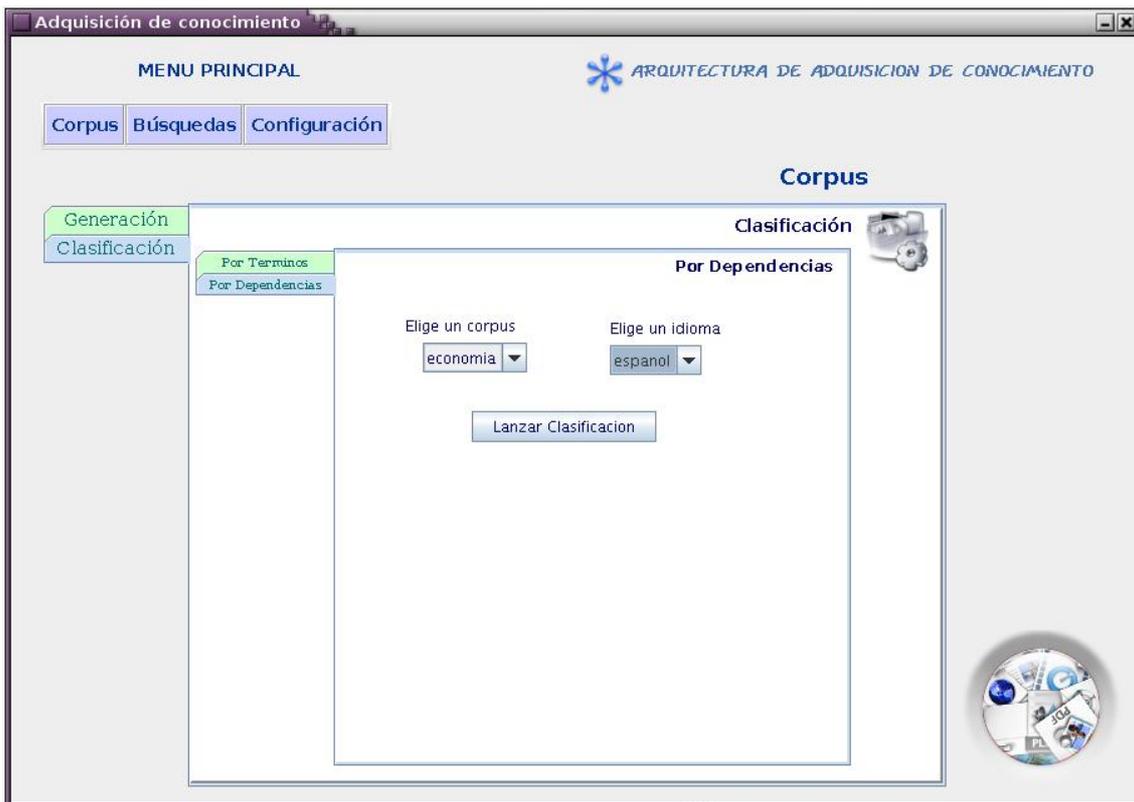


Figura 12 vista principal clasificación por dependencias

Una vez el usuario seleccione el botón *Lanzar Clasificación* se le mostrará un mensaje como el siguiente para que confirme la operación que se va a llevar a cabo:

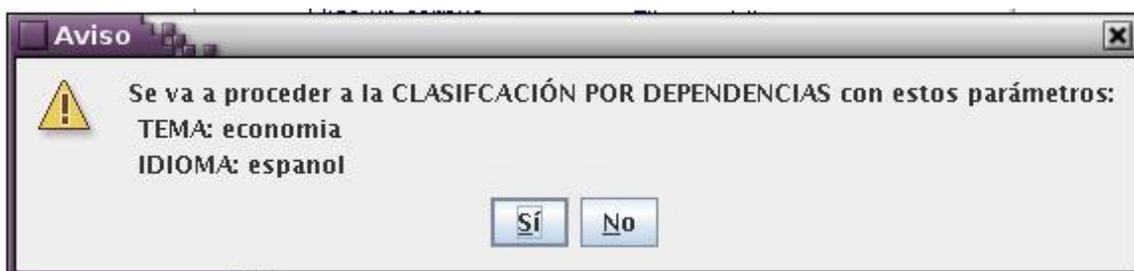


Figura 13 aviso clasificación por dependencias

Si el usuario ha pulsado que sí, se lanza el proceso que habrá finalizado una vez que se muestre el botón *Lanzar Ontología*:

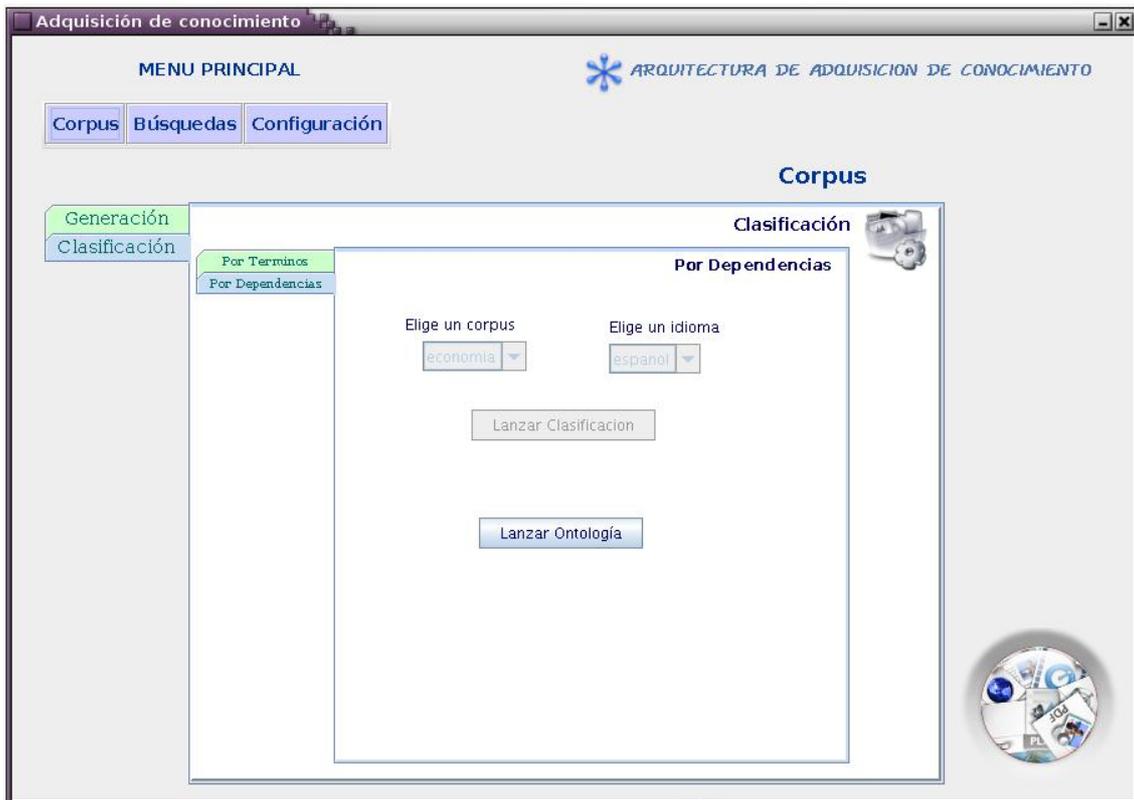


Figura 14 vista clasificación por dependencias crear la ontología

El usuario deberá pulsar el botón *Lanzar Ontología* para generar la ontología por dependencias. Una vez se ha creado la ontología, ésta puede recuperarse en el directorio *Ontologia/temaActual/idiomaActual/ontologia*, donde *temaActual* e *idiomaActual* son, respectivamente, el tema e idioma que el usuario ha elegido.

(A) INTERACCIÓN DEL USUARIO CON LOS FICHEROS

Una vez el proceso se ha ejecutado, el usuario podría volver a lanzarlo para obtener nuevos resultados que sean más acordes a sus necesidades. Para ello puede modificar un conjunto de ficheros (estos ficheros están descritos en el apartado 13.2 del Manual Técnico)

Los ficheros se recuperan del directorio *Ontologia/recuperacionConocimiento/temaActual* dentro del directorio donde esté instalada la aplicación.

FICHERO ELEM.TXT

En el mismo se pueden modificar los nombres asignados a los conceptos y propiedades. Sin embargo deben permanecer MERCADOS y PROPIEDADES:

MERCADOS
ENTIDADES
INDICES
DINEROS
PAISES
PERSONAS
PROPIEDADES
LUGARES
TIPOS
TAMAÑO
PERÍODO
CATEGORÍA

Es necesario añadir una tabulación antes de cada concepto y propiedad.

FICHERO MARCADORES.TXT

En este fichero se definen aquellas palabras que marcan el tipo de relación entre los elementos de una dependencia.

bolsa	MERCADOS	PAISES/INDICES	[2]
bolsa	PROPIEDADES	LUGARES/TIPOS	[2]
mercado	PROPIEDADES	LUGARES/TIPOS	[2]
mercado	MERCADOS	INDICES/DINEROS	[2]
tamaño	PROPIEDADES	TAMAÑO	[2]
gestión	MERCADOS	ENTITES/DINEROS	[1,2]
gestión	PROPIEDADES	TIPO/TAMAÑO/LUGARES	[2]
inicio	PROPIEDADES	PERÍODO	[2]

Cada línea del mismo define un marcador. En la primera columna hay que poner la palabra marcador. A continuación el elemento con él que se puede relacionar, bien un concepto (designado por MERCADOS) o bien una propiedad (designada por PROPIEDADES). En tercer lugar el tipo del concepto o propiedad (según el valor de la columna anterior). Por último se añade un número entre corchetes indicando la posición en la dependencia del elemento que acompaña al marcador. Todos los valores deben separarse por tabulaciones.

FICHERO PROPIEDADES.TXT

En este fichero se tienen inicialmente un conjunto de propiedades base para iniciar el proceso. Además en las sucesivas ejecuciones este fichero se va modificando incorporando nuevas propiedades. En la primera columna se tiene la palabra que marca la propiedad y en la segunda el tipo de propiedad. El usuario puede modificar el tipo de propiedad así como, eliminar aquellas que no considere oportunas. Todo ello con cuidado de separar las palabras de tipo de propiedad con una tabulación.

europeo	LUGARES
baja	TAMAÑO
caída	TAMAÑO
krach	TAMAÑO
junio	PERÍODO
año	PERÍODO
día	PERÍODO
abril	PERÍODO
mitad	TAMAÑO
americano	LUGARES
parisino	LUGARES
alto	TAMAÑO
internacional	LUGARES
francés	LUGARES
general	TIPO
único	TIPO

FICHERO MERCADOS.TXT

La interpretación de este fichero y su manipulación es la misma que la del fichero PROPIEDADES.txt explicado arriba. Se diferencia en que MERCADOS.txt se refiere a los conceptos hallados en el corpus.

ministro	PERSONAS
fluctuación	DINEROS
banco	ENTIDADES
divisa	DINEROS
grupo	ENTIDADES
inversor	PERSONNES
Dow&Jones	INDICES
Nikkei	INDICES

2.4.VISTA BÚSQUEDA

Se describe en este apartado como se procede a la recuperación de documentos relevantes a la consulta del usuario. Antes de describir la operación para cada una de las posibles búsquedas, se indica que las consultas del usuario pueden incluir las palabras clave AND, NOT y OR y pueden combinar los términos con paréntesis.

2.4.1.BÚSQUEDA POR TÉRMINOS

En primer lugar será necesario elegir un tema y un idioma. Esto puede desencadenar tres tipos de eventos:

- Si no existe un corpus asociado al tema e idioma elegidos, el sistema muestra un diálogo de error como el de la Figura 9.
- Si no existe una ontología creada para ese tema y ese idioma, el sistema muestra un diálogo como el siguiente:

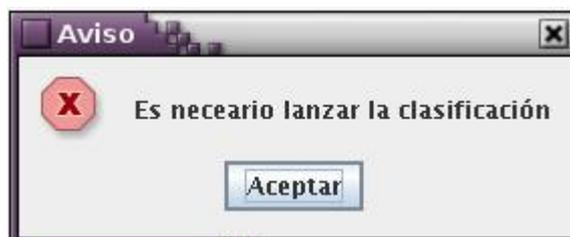


Figura 15 error lanzar la clasificación

- Si todos los datos son correctos y se muestra el conjunto de ontologías que existen para ese tema e idioma¹, el usuario deberá elegir una de ellas. Una vez elegida se activa el botón de *Lanzar Consulta*; y el usuario puede introducir una consulta.

¹ Estas ontologías serán las que se han generado anteriormente en la clasificación por términos.

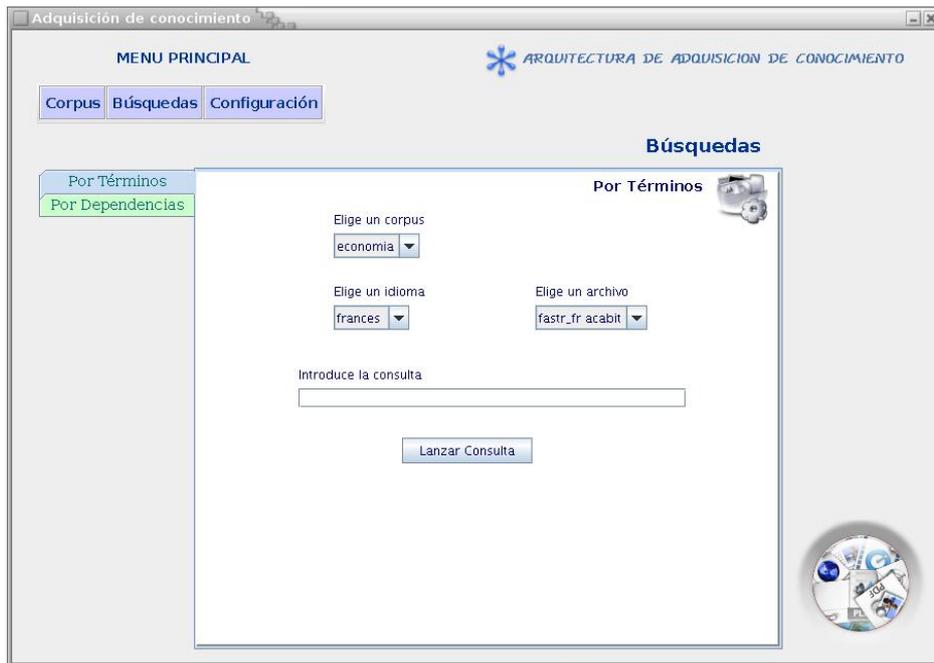


Figura 16 búsqueda por términos

RESULTADOS DE UNA CONSULTA

El sistema muestra los documentos devueltos en una ventana como la siguiente:

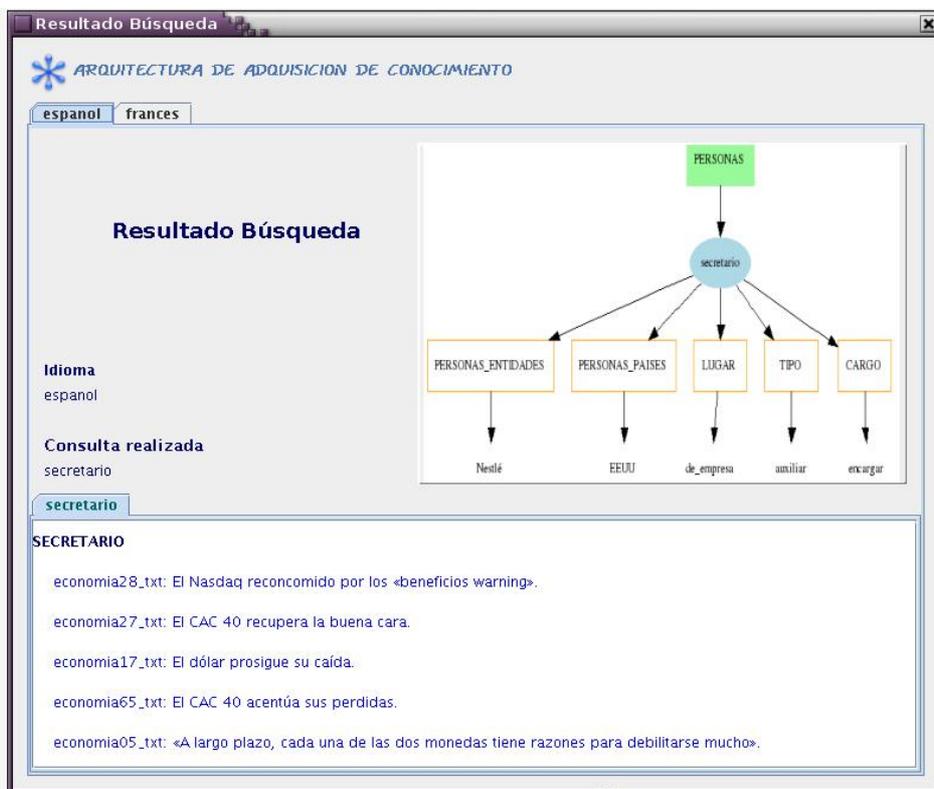


Figura 17 resultado de los documentos

En la parte superior pueden verse dos pestañas con el nombre de un idioma cada una. Marcan el conjunto de documentos devueltos para cada uno de ellos. Para cada idioma se muestra un gráfico con la sección de la ontología implicada en la consulta. Es el gráfico que se puede ver en la parte superior derecha. También se muestra la consulta introducida por el usuario. En la parte inferior se muestran los documentos agrupados según el término que los ha devuelto y organizados por su relevancia. El nombre de estos documentos es en realidad un enlace que, al pulsarlo, muestra el contenido del documento, tal y como se puede ver en la Figura 18.

Es necesario indicar, que si bien en la recuperación de documentos por términos no se realiza la búsqueda en más de un idioma, la vista es la misma pero sólo incluiría una pestaña de idioma. A modo de ejemplo, en la Figura 17 para el caso de la búsqueda por términos, si la búsqueda ha sido para el español sólo se mostraría la pestaña correspondiente a ese idioma.

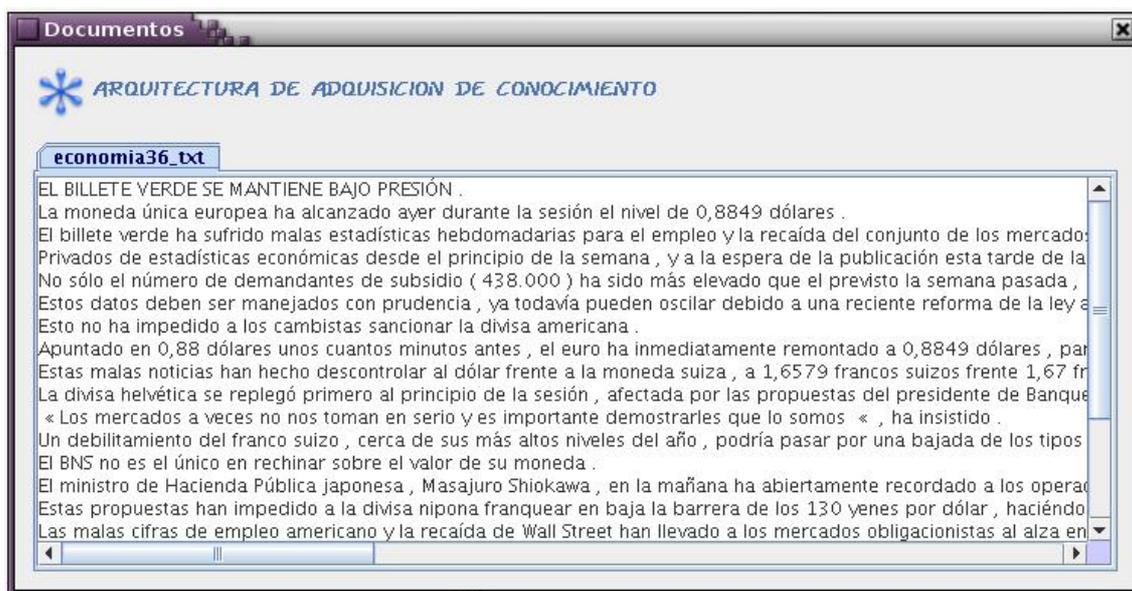


Figura 18 vista de un documento

2.4.2. BÚSQUEDA POR DEPENDENCIAS

En primer lugar será necesario elegir un tema y un idioma. Esto puede desencadenar tres tipos de eventos:

- Si no existe un corpus asociado al tema e idioma elegidos, el sistema muestra un diálogo de error como el de la Figura 9.
- Si no existe una ontología creada para ese tema y ese idioma, el sistema muestra un diálogo de error como el de la Figura 15.

No se vuelve a detallar cómo se muestran los documentos pues esa operación ya está explicada en el apartado anterior. Sólo añadir que en este caso sí se devuelven los resultados para más de un idioma.

2.5. VISTA CONFIGURACIÓN

El usuario deberá elegir en el menú principal la opción de *Configuración*. En ella se muestra un contenedor principal con dos pestañas, *declarar corpus* y *herramientas*. En el primer caso se trata de manipular los temas e idiomas y en el segundo caso se trabaja con los analizadores sintácticos y extractores.

2.5.1.GESTIÓN DE TEMAS

La gestión de temas se lleva a cabo a través de la siguiente vista:

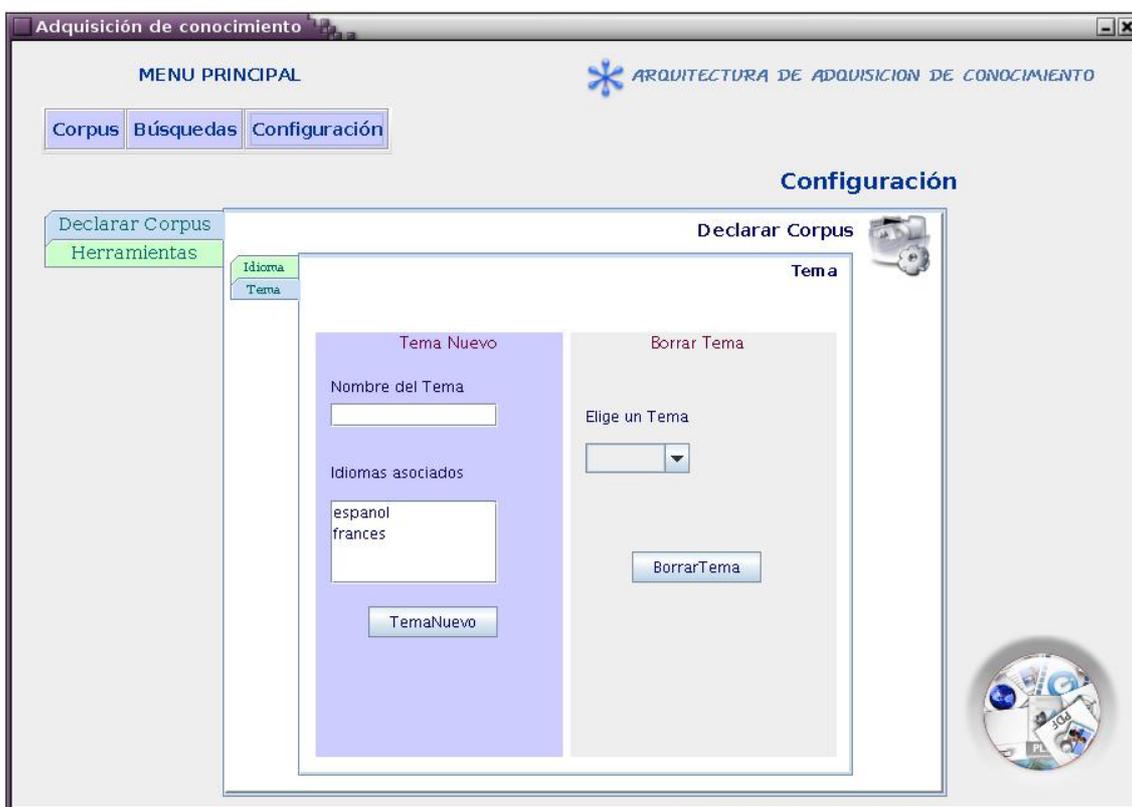


Figura 19 vista gestión tema

En la figura anterior, se muestran dos formularios. El de la izquierda permite añadir un nuevo tema y el de la derecha borrarlo.

AÑADIR UN TEMA

En el formulario de alta, es necesario que el usuario introduzca el nombre del tema así como el conjunto de idiomas asociados a ese tema. Aquí una importante aclaración, para añadir un idioma a un tema, es necesario que el primero esté creado. De este modo, al crear un idioma, mientras no se le asocie un tema, su estado permanece desactivado y no puede participar en ninguna operación.

- Si el usuario introduce como nombre del tema un valor que ya existe en el fichero de configuración, entonces se presenta un mensaje como éste:

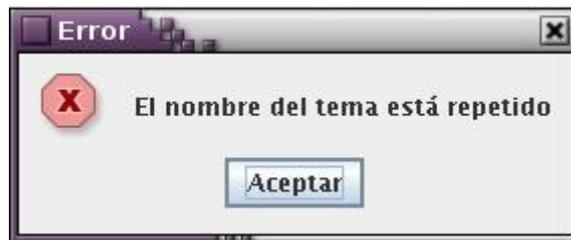


Figura 20 error tema repetido

- Si la operación se realiza correctamente entonces se muestra una frase indicando que hubo éxito. El sistema añade el nuevo tema que ya estará disponible para todas las operaciones oportunas.

BORRAR UN TEMA

En el caso de que el usuario desee borrar un tema, entonces deberá seleccionar el formulario de la derecha. Elegirá un tema para borrar, entonces el sistema mostrará un diálogo de aviso solicitando la confirmación del usuario. Si confirma el diálogo se borrarán todas las referencias del tema en el fichero de configuración, así como todas las estructuras de directorios asociadas.

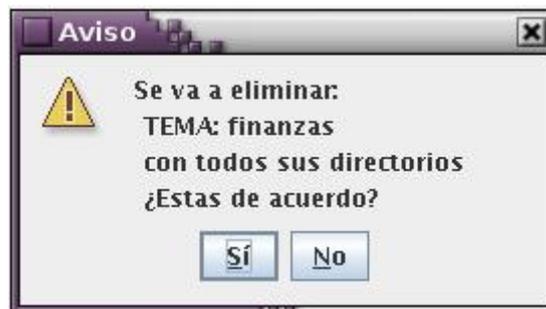


Figura 21 aviso eliminar un elemento

2.5.2.GESTIÓN DE IDIOMAS

La gestión de los idiomas se lleva a cabo a través de la siguiente vista:

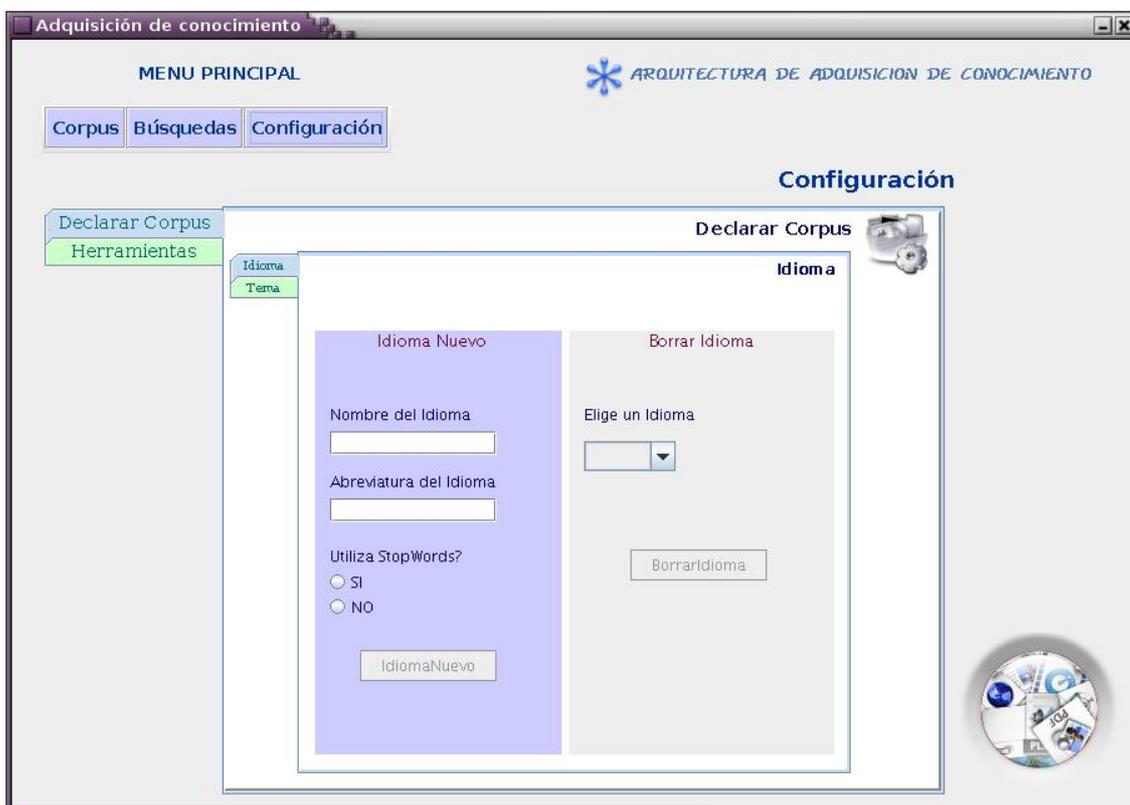


Figura 22 vista gestión idioma

En la figura anterior, se muestran dos formularios. El de la izquierda permite añadir un nuevo idioma y el de la derecha borrar un idioma.

AÑADIR UN IDIOMA

El usuario puede solicitar añadir un nuevo idioma, para ello accede al formulario de la izquierda. En primer lugar deberá escribir el nombre del idioma así como el nombre de la abreviatura asociada. En último lugar elegirá si va a trabajar con un fichero con *palabras de parada*. Una vez proporcionados todos los parámetros y pulsado el botón *IdiomaNuevo*, pueden desencadenarse tres eventos:

- Si el nombre del idioma introducido ya existe en el sistema, el sistema informa de ello con una caja de diálogo como la de la Figura 20 en la que se indicará que el idioma está repetido.

- Si el nombre de la abreviatura ya existe en el sistema, el sistema informa de ello de la misma manera que en el punto anterior.
- Si todos los parámetros son correctos. El sistema añade el nuevo idioma que ya estará disponible para todas las operaciones oportunas. Se muestra una frase indicando el éxito de la operación.

Si el usuario desea que la búsqueda de documentos se produzca en más de un idioma debe tener en cuenta cuales son los idiomas disponibles en el traductor de Google, así como las abreviaturas asociadas a esos idiomas. Para ello puede consultar el documento *abreviaturasGoogle.txt* en el directorio *distribución/configuracion*.

El fichero de *palabras de parada* debe contener una palabra por línea y el nombre del mismo será *nombre_del_idiomaStopWords.txt* y deberá guardarse en el directorio *Ontologia/stopWords*.

BORRAR UN IDIOMA

En el caso de que el usuario desee borrar un idioma, entonces deberá seleccionar el formulario de la derecha. Elegirá un idioma para borrar, entonces el sistema mostrará un diálogo de aviso como el de la Figura 21 solicitando la confirmación del usuario sobre la operación que se va a realizar. Si confirma el diálogo se borrarán todas las referencias del idioma en el fichero de configuración así como todas las estructuras de directorios asociadas.

2.5.3.GESTIÓN DE EXTRACTORES

La gestión de los extractores se lleva a cabo a través de la siguiente vista:

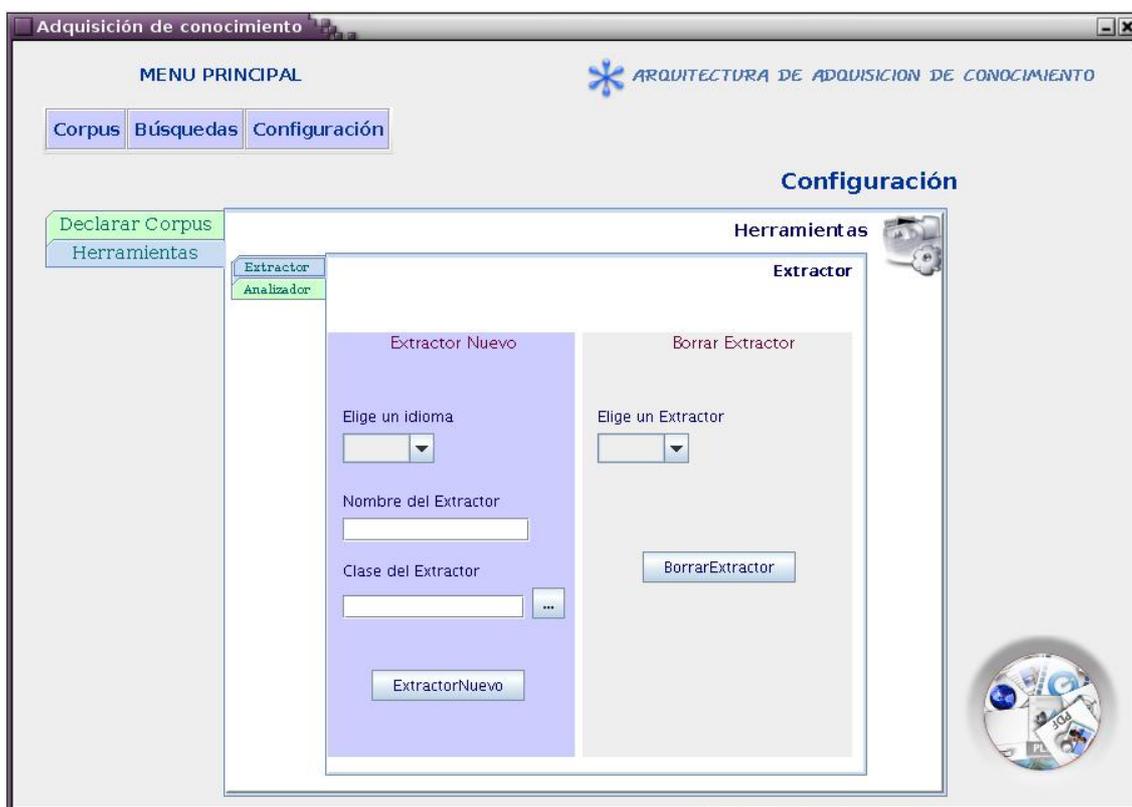


Figura 23 vista gestión extractor

En la figura anterior, se muestran dos formularios. El de la izquierda permite añadir un nuevo extractor y el de la derecha borrar un extractor.

AÑADIR UN EXTRACTOR

El usuario elige añadir un nuevo extractor a través del formulario de la izquierda. Para ello debe elegir el idioma del extractor, introducir el nombre del extractor y el nombre del recurso asociado, esto es, la clase que lo implementa. Esto puede hacerlo escribiendo la ruta completa o bien pulsando el botón “...” que desplegará un diálogo para seleccionar archivos con extensión *.java* o *.class* (si la clase ya está compilada). Una vez proporcionados estos parámetros y pulsado el botón de *ExtractorNuevo* se pueden desencadenar tres eventos:

- Si el nombre del extractor ya existe en el sistema, se muestra una caja de diálogo similar a la de la Figura 20, en la que se indicará que el nombre del extractor está repetido.
- Si la clase introducida contiene errores, se muestra el siguiente diálogo:



Figura 24: error en la clase introducida

- Si todos los valores son correctos, el sistema añade el nuevo extractor y se muestra un mensaje de éxito. A partir de ese momento el extractor está disponible para las tareas oportunas.

BORRAR UN EXTRACTOR

En el caso de que el usuario desee borrar un extractor, deberá seleccionar en el formulario de la derecha el extractor que desea eliminar. Al igual de como sucedía con los idiomas y los temas, se muestra una ventana de confirmación pero en este caso se informará de que se van a eliminar todas las referencias del extractor en el fichero de configuración. Se ha optado por no eliminar los directorios que contienen a los extractores, caso de que estos se implementen como un recurso externo.

2.5.4. GESTIÓN DE ANALIZADORES SINTÁCTICOS

La gestión de los analizadores sintácticos se lleva a cabo a través de la siguiente vista:

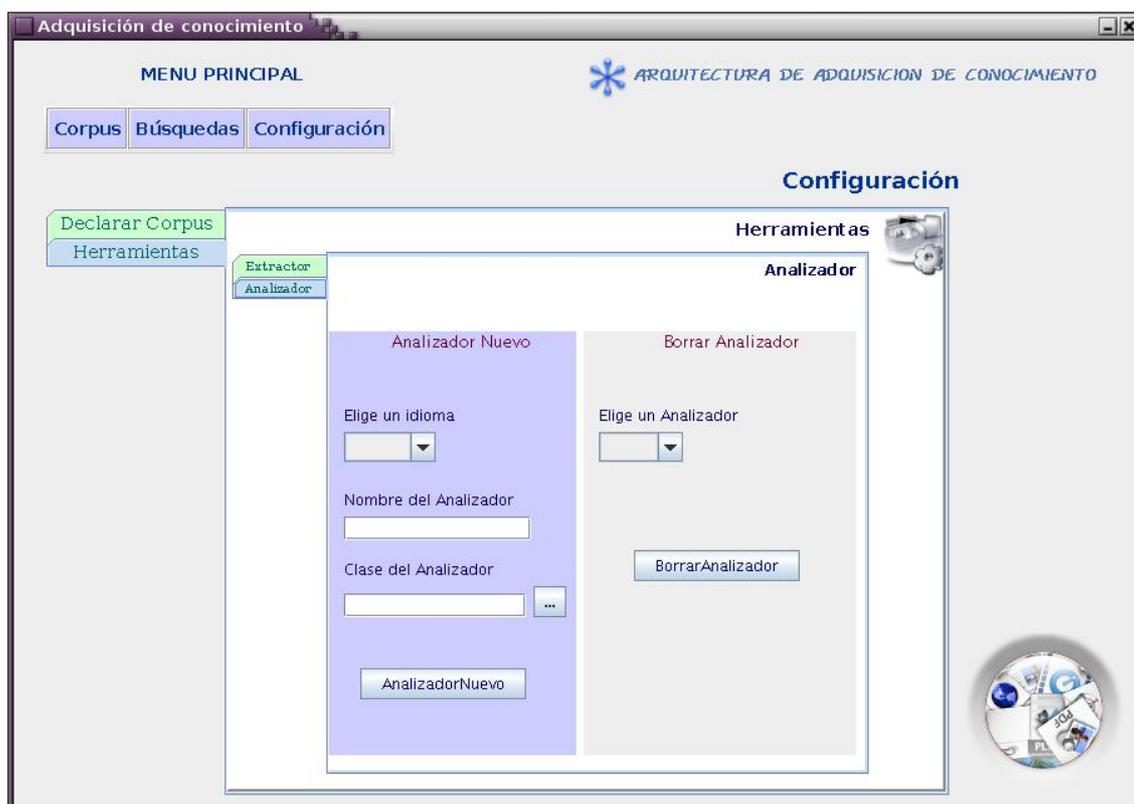


Figura 25 vista gestión analizador

En la figura anterior, se muestran dos formularios. El de la izquierda permite añadir un nuevo analizador sintáctico y el de la derecha borrar un analizador sintáctico.

AÑADIR UN ANALIZADOR SINTÁCTICO

El usuario elige añadir un nuevo analizador a través del formulario de la izquierda. Para ello debe elegir el idioma del analizador sintáctico, introducir el nombre del mismo y el nombre del recurso asociado, esto es, la clase que lo implementa. Esto puede hacerlo escribiendo la ruta completa o bien pulsando el botón "... " que desplegará un diálogo para seleccionar archivos con extensión *.java* o *.class* (si la clase ya está compilada). Una vez proporcionados estos parámetros y pulsado el botón de *AnalizadorNuevo* se pueden desencadenar tres eventos:

- Si el nombre del analizador ya existe en el sistema, se muestra, entonces, una caja de diálogo similar a la de la Figura 20 indicando que el analizador ya existe en el sistema.
- Si la clase introducida contiene errores, se muestra una caja de diálogo similar a la de la Figura 24 en la cual se indica que el nombre del analizador ya existe en el sistema.
- Si todos los valores son correctos, el sistema añade el nuevo analizador y se muestra un mensaje de éxito. A partir de ese momento el analizador está disponible para las tareas oportunas.

BORRAR UN ANALIZADOR SINTÁCTICO

En el caso de que el usuario desee borrar un analizador sintáctico, deberá seleccionar en el formulario de la derecha el analizador que desea eliminar. Al igual de como sucedía con los extractores, se muestra una ventana de confirmación en la que se informará de que se van a eliminar todas las referencias del analizador en el fichero de configuración. Al igual que sucede con los extractores no se eliminarán los directorios asociados.

3. MANUAL DEL USUARIO EXPERTO

Se ha comentado en el Manual Técnico, apartado *3 especificación de requisitos* que la aplicación puede ser utilizada por dos tipos de usuarios. La interfaz gráfica puede ser manejada por cualquiera de ellos, pero si lo que se pretende es modificar la aplicación sin utilizar la misma, son necesarios algunos conocimientos a mayores. Dichos conocimientos se considera que los posee el usuario llamado **experto**.

Se describe aquí cómo el usuario experto debe implementar las clases que realicen bien la extracción bien el análisis sintáctico. Se detalla el comportamiento esperado de los métodos que desarrollen así como la descripción de las modificaciones en el fichero de configuración.

3.1 .AÑADIR UN EXTRACTOR

La interfaz de un extractor solicita que se implementen cuatro métodos:

- **setValores(String tema):** Se le pasa el tema actual. Este método no devuelve ningún valor.
- **etiquetador:** En este método el usuario deberá implementar la etiquetación de las palabras. Se ha comentado a lo largo de la documentación que esta funcionalidad no tiene porque ser necesaria, pues el propio extractor podría traer incorporado una técnica de etiquetación propia. Este método no devuelve ningún valor.
- **execExtract:** En este método se deberá indicar toda la lógica de la extracción. El usuario puede implementar un algoritmo directamente, o realizar una llamada a un proceso externo. En cualquier caso el resultado de la extracción debe almacenarse siguiendo la siguiente nomenclatura.

Utilizando una clase de apoyo como *RdoTerminos*, cuya estructura es la siguiente:

ATRIBUTOS

- termino: Término devuelto por la extracción.
- frecuencia: Frecuencia de aparición del término.
- extractor: Extractor que ha dado lugar al término.

MÉTODOS

- RdoTerminos (String termino, String frecuencia, String extractor)
- getTermino: Devuelve el valor del término actual. El retorno es una cadena.
- getFrecuencia: Devuelve la frecuencia de aparición del término. El retorno es una cadena.
- getExtractor: Devuelve el extractor que ha dado lugar al término. El retorno es una cadena.

El usuario deberá guardar los resultados de la extracción en un *ArrayList* de tipo *RdoTerminos*. Con el constructor se inicializan los valores de los atributos.

- getTerminos: Este método devuelve el vector de tipo *RdoTerminos* anteriormente descrito.

En cuanto al fichero de configuración, es necesario acceder al nodo correspondiente a los extractores.

```
<extractores>
  <ext lang="es">
    <extractor name="fastr_es">
      <rutaExtract>Extract/espanol/FastrEs/</rutaExtract>
      <clase>universidad.proyecto.recursos.extractor.FastrES</clase>
    </extractor>
  </ext>
  <ext lang="fr">
    <extractor name="fastr_fr">
      <rutaExtract>Extract/frances/FastrFr/</rutaExtract>
      <clase>universidad.proyecto.recursos.extractor.FastrFR</clase>
    </extractor>
    <extractor name="acabit">
      <rutaExtract>Extract/frances/Acabit/</rutaExtract>
      <clase>universidad.proyecto.recursos.extractor.Acabit</clase>
    </extractor>
  </ext>
</extractores>
```

En primer lugar hay que seleccionar el idioma asociado al extractor, para ello se introducirá el nodo correspondiente al extractor bajo el nodo *ext* cuyo atributo *lang* sea la abreviatura del idioma elegido.

Si se supone que el nuevo extractor añadido es "acabit" (señalado en negrita) el usuario deberá acceder al nodo *ext* con *lang="fr"*. A continuación añadirá un nodo de tipo *extractor* con un atributo *name* que sea el nombre del extractor. Este nombre debe ser único en el sistema. En el nodo *rutaExtract* se definirá la ruta donde se ubica el extractor si éste es un recurso externo. En el nodo *clase* es necesario indicar el nombre de la clase que ejecuta el extractor. Es necesario indicar el paquete entero donde se encuentra la clase.

3.2. AÑADIR UN ANALIZADOR SINTÁCTICO

La interfaz de un analizador sintáctico solicita que se implementen cinco métodos. Es importante tener en cuenta que el analizador sintáctico debe analizar los documentos frase a frase.

- **setValores(String tema):** Se le pasa el tema actual. Este método no devuelve ningún valor.
- **startParser (String frase):** Este método debe lanzar el análisis sintáctico de la frase que se le pasa como parámetro. El usuario puede implementar directamente un proceso o indicar una llamada externa. En cualquier caso debe almacenar los resultados de la siguiente forma:

Para cada una de las palabras de la frase se debe guardar su información morfosintáctica y léxica. Para ello se deberá crear un *ArrayList* de tipo *Tagger*. Esta interfaz proporciona los siguientes métodos:

- getLabel: Devuelve la etiqueta asignada a la palabra. El retorno debe ser una cadena.
- getForma: Devuelve la forma de la palabra. El retorno debe ser una cadena.

- getLema: Devuelve el lema de la palabra. El retorno debe ser una cadena.

Existe en el sistema una implementación de esta interfaz, se llama *TaggerData*. Implementa los métodos descritos. Tiene tres atributos que son *label*, *forma* y *lema*. El constructor inicializa esos valores: *TaggerData (String forma, String label, String lema)*.

En cuanto a las dependencias generadas, se debe crear un *ArrayList* de tipo *Dependency*. Esta interfaz proporciona los siguientes métodos:

- getLabel: Devuelve la etiqueta asociada a la dependencia. El retorno debe ser una cadena.
- getLema1: Devuelve el lema de la primera palabra de la dependencia. El retorno debe ser una cadena.
- getPos1: Devuelve la posición en la frase de la primera palabra de la dependencia. El retorno debe ser una cadena.
- getLema2: Devuelve el lema de la segunda palabra de la dependencia. El retorno debe ser una cadena.
- getPos2: Devuelve la posición en la frase de la segunda palabra de la dependencia. El retorno debe ser una cadena.

Por lo tanto las dependencias que devuelve el análisis sintáctico deben ser pares de palabras, con la posición de cada una en la frase actual y la etiqueta asociada a la dependencia.

Existe en el sistema una implementación de la interfaz *Dependency*, es la clase *DependencyData*. Implementa los métodos descritos. Tiene cinco atributos que son, *label*, *lema1*, *lema2*, *pos1*, *pos2* y su constructor inicializa esos valores, *DependencyData(String label, String lema1, String lema2, String pos1, String pos2)*.

- **getVectorDependency**: Este método devuelve el *ArrayList* generado con las dependencias.
- **getVectorTagger**: Este método devuelve el *ArrayList* generado con la información de las palabras.
- **getSentence**: Este método devuelve una cadena que es la frase actual.

En cuanto al fichero de configuración es necesario acceder al nodo correspondiente a los analizadores.

```

<parsers>
  <parser lang="es">
    <analyzer name="erialES">
      <rutaParser>Analyzer/espanol/erialES</rutaParser>
      <claseParser>universidad.proyecto.recursos.parser.ErialES</claseParser>
    </analyzer>
  </parser>
  <parser lang="fr">
    <analyzer name="erialFR">
      <rutaParser>Analyzer/frances/erialFR</rutaParser>
      <claseParser>universidad.proyecto.recursos.parser.ErialFR</claseParser>
    </analyzer>
  </parser>
</parsers>

```

En primer lugar hay que seleccionar el idioma asociado al analizador, para ello se introducirá el nodo correspondiente al nuevo analizador bajo el nodo *parser* cuyo atributo *lang* sea la abreviatura del idioma elegido.

Si se supone que el nuevo analizador añadido es "erialFR" (señalado en negrita) el usuario deberá acceder al nodo *parser* con *lang="fr"*. A continuación añadirá un nodo de tipo *analyzer* con un atributo *name* que sea el nombre del analizador. Este nombre debe ser único en el sistema. En el nodo *rutaParser* se definirá la ruta donde se ubica el analizador si éste es un recurso externo. En el nodo *claseParser* es necesario indicar el nombre de la clase que ejecuta el analizador. Es muy importante indicar el paquete entero donde se encuentra la clase.

ÍNDICE DE FIGURAS

4. ÍNDICE FIGURAS

Figura 1 entrada a la aplicación.....	17
Figura 2 elegir un fichero de configuración	18
Figura 3 error en la estructura del fichero de configuración.....	18
Figura 4 elementos repetidos en el fichero de configuración.....	19
Figura 5: menú principal	19
Figura 6 error idioma sin recurso asociado	20
Figura 7 vista generación.....	21
Figura 8 vista principal de clasificación por términos.....	22
Figura 9 error no existe corpus.....	23
Figura 10 aviso clasificación por términos.....	23
Figura 11 vista clasificación por términos crear la ontología.....	24
Figura 12 vista principal clasificación por dependencias	25
Figura 13 aviso clasificación por dependencias.....	25
Figura 14 vista clasificación por dependencias crear la ontología.....	26
Figura 15 error lanzar la clasificación.....	29
Figura 16 búsqueda por términos	30
Figura 17 resultado de los documentos	30
Figura 18 vista de un documento	31
Figura 19 vista gestión tema.....	33

Figura 20 error tema repetido	34
Figura 21 aviso eliminar un elemento	34
Figura 22 vista gestión idioma	35
Figura 23 vista gestión extractor	37
Figura 24: error en la clase introducida	38
Figura 25 vista gestión analizador	39