

# IMPLEMENTACIÓN DE UN SISTEMA QA



## GRUPO A

Carballal Mato, Adrián: infacm02 : [adriancarballal@gmail.com](mailto:adriancarballal@gmail.com)

(Coordinador) Fernández Pastoriza, José María: infjfp00 : [chemafic@gmail.com](mailto:chemafic@gmail.com)

Luengo Cabanillas, Alberto: infalc00 : [cabiwan@gmail.com](mailto:cabiwan@gmail.com)

Martín Quiroga, Javier: infjmq00 : [javimartinquiroga@yahoo.es](mailto:javimartinquiroga@yahoo.es)



## Contenido

Introducción. ....	1
Arquitectura del sistema. ....	2
Arquitectura detallada del sistema. ....	5
es.udc.Inaturales.practica.ER: ....	5
es.udc.Inaturales.practica.IndexGenerator: ....	5
es.udc.Inaturales.practica.PatternSelector: ....	6
es.udc.Inaturales.practica.Search: ....	6
es.udc.Inaturales.practica.Transformation: ....	7
es.udc.Inaturales.practica.Transformation.util: ....	7
es.udc.Inaturales.practica.util: ....	8
Herramientas. Indexación. ....	10
Apache Lucene. ....	10
TERRIER (TERabyte RetriEver). ....	10
Freeling. ....	12
Descargas. ....	13
Instrucciones de instalación. ....	14
Diario de Trabajo. ....	16
Versión 1.....	17
Versión 2.....	21
Versión 3.....	25
Versión 4.....	28
Versión 5.....	32
Versión 6.....	36
Versión Final.....	39

Tabla de resultados .....	39
Justificación de respuestas.....	42
Documento de Respuestas.....	46
Preguntas erróneas. ¿por qué no funcionan?.....	50
Mejoras: Detalles .....	56
MEJORA DE LOS ADJETIVOS. ....	56
MEJORA DE LAS BUSQUEDAS REDUCIDAS CON NOMBRES PROPIOS. ....	56
MEJORA DE MAYOR NIVEL DE ESPECIFICACIÓN. ....	57
MEJORA DE SINÓNIMOS. ....	57
CAMBIO DE ALGORITMO DE RESPUESTAS. ....	57
MEJORA DE VERBOS. ....	57

## Introducción.

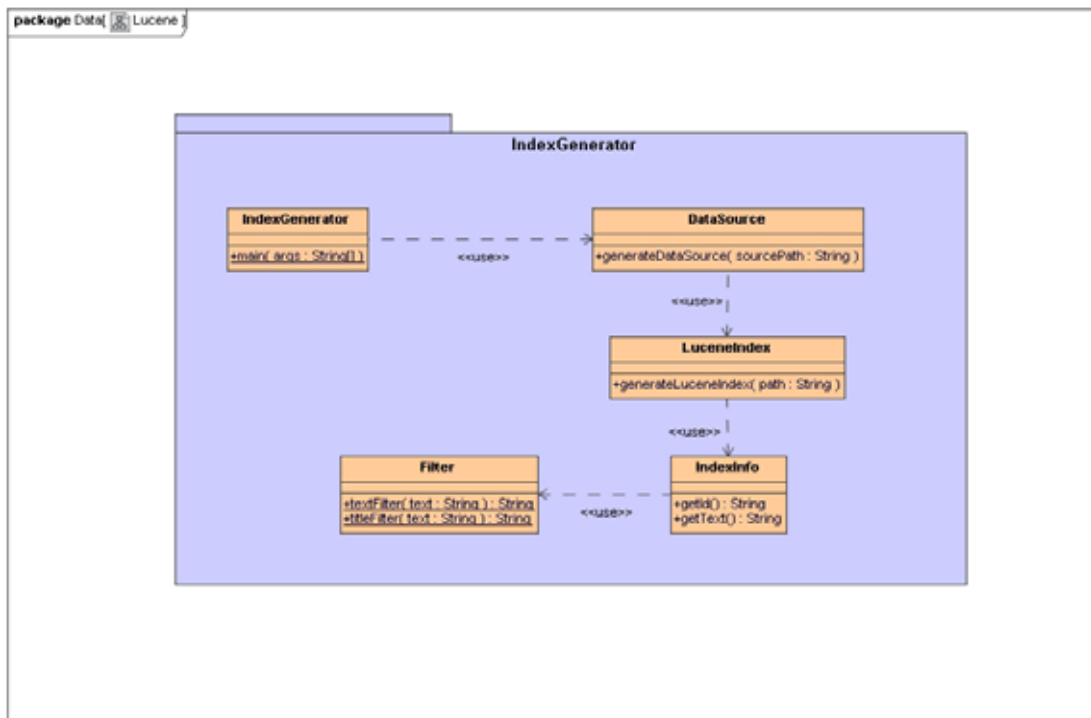
Esta aplicación QA ha sido desarrollada intentando cumplir los objetivos propuestos en la asignatura. Aunque el porcentaje de respuestas no es muy alto, en la memoria viene explicadas las diferentes técnicas que se han utilizado, junto con aquellas otras que han sido finalmente descartadas debido a los malos resultados que nos han dado.

Junto con las explicaciones de por qué se han elegido o rechazado ciertos métodos, incluimos también explicaciones de otras técnicas que podríamos haber utilizado para resolver el resto de preguntas sin responder, y el por qué no las hemos implementado.

También comentamos las mejoras que hemos ido realizando a medida que desarrollamos la aplicación, junto con los resultados correctos que hemos conseguido, de modo que se pueda observar las diferencias entre las distintas versiones de la aplicación.

## Arquitectura del sistema.

La realización del índice, como se ha dicho anteriormente, se ha implementado mediante Apache Lucene. Para la creación del mismo se ha creado un sistema propio independiente tal y como se ve en la siguiente figura.

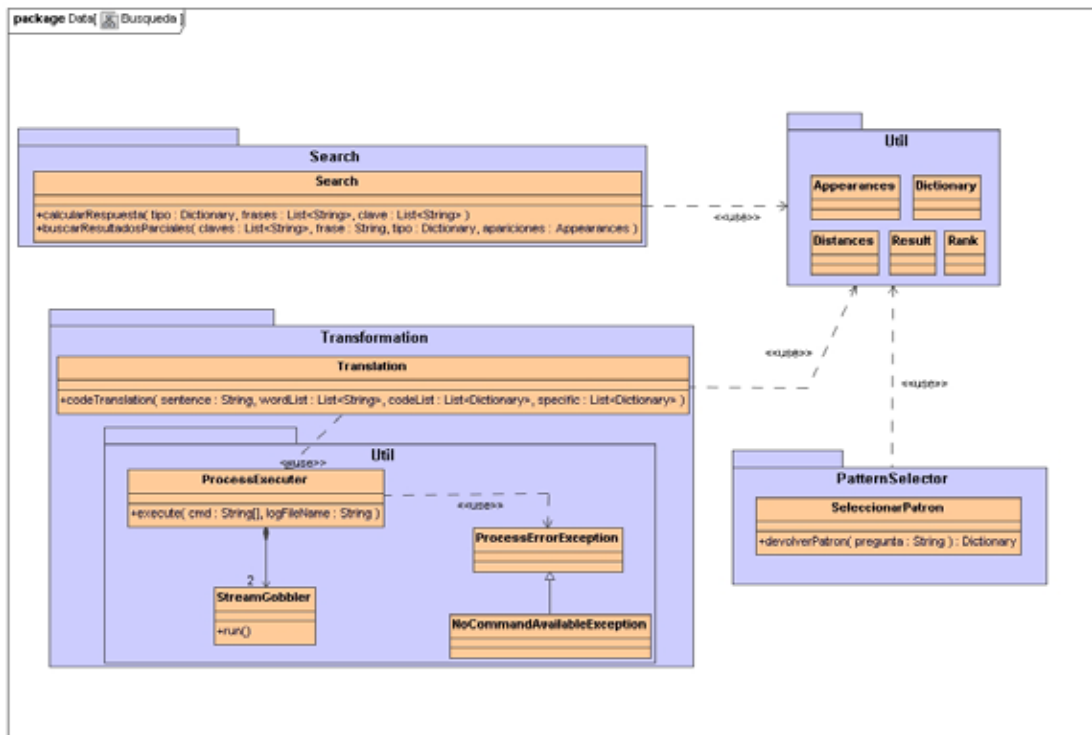


La clase principal IndexGenerator creará un índice completo con el identificador del documento como clave y el título junto con el texto asociado como documento. Se ha escogido que el propio índice guarde de forma nativa los datos sin ningún tipo de codificación, facilitando de esta manera la recuperación tanto de los textos de búsqueda como sobre todo de las claves del mismo, muy útil a la hora de la creación de las respuestas.

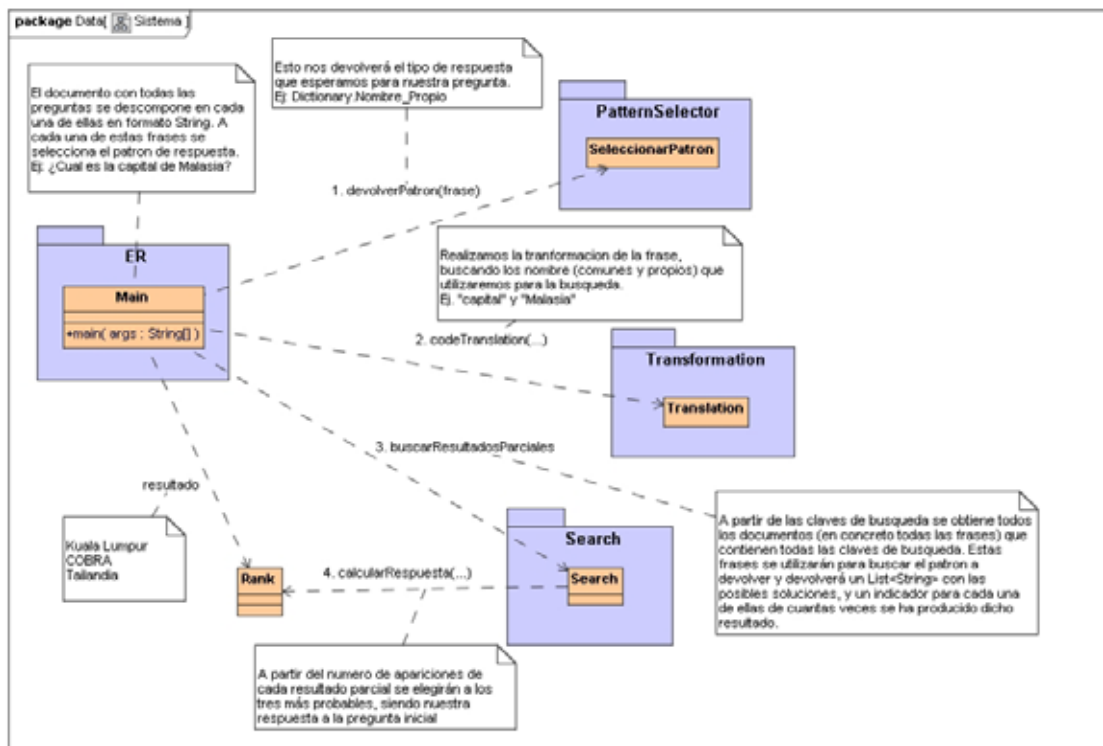
Se han tenido que realizar algunas transformaciones sobre los documentos iniciales para optimizar la creación de los índices. Lucene nos provee de un parseador xml integrado para la creación de índices. Cambiando el árbol del documento a un formato xml genérico óptimo reduciremos notablemente el tiempo de realización, ya que ese proceso ya está optimizado.

Otras transformaciones menores dentro de los textos a la hora de crear los índices se basan en la eliminación de caracteres especiales que no facilitan las búsquedas posteriormente, y que en algunos casos provocan errores. En este punto se encontrarán puntos suspensivos, separadores de título y artículo (“-.”), etc.

Una vez que tenemos el índice de artículo operativo, nos centraremos en el sistema de búsqueda de respuestas propiamente dicho. En la siguiente figura se muestran los componentes del mismo.



Veamos ahora una figura explicativa sobre un ejemplo o caso de uso de una pregunta en particular:



En nuestro caso de ejemplo intentamos encontrar respuesta para la pregunta:

M SPA 0060 ¿Cuál es la capital de Malasia?

El sistema inicialmente obtiene qué tipo de respuesta estamos buscando para dicha pregunta en función de los diferentes patrones para preguntas factuales que tenemos recogidos en la clase `SeleccionarPatron`. Para este caso particular se obtiene que nuestra respuesta esperada ha de ser un nombre.

Una vez hecho esto, se realizará una transformación léxica sobre la pregunta, con el objetivo de identificar los nombres que aparecen en la misma. Nótese que sólo se utilizan nombres y no adjetivos o verbos, ya que estos últimos suelen variar con facilidad dentro de los documentos, sobre todo el caso de los tiempos verbales. Los adjetivos por otra parte se explicarán dentro de la versión 5, al tratarse de un aspecto más controvertido a la hora de tomar una decisión al respecto. Estos nombres se utilizarán como claves de búsqueda dentro del indexador Lucene.

Sobre la metodología de búsqueda de resultados parciales (documentos que pueden contener la respuesta buscada) mencionaremos que se realizarán utilizando un operador AND y dejaremos que sea el propio sistema es que nos devuelva los hits o resultados ordenados por orden de relevancia. Esto último lo utilizaremos posteriormente a la hora de la elección del documento representativo de cada una de las respuestas, adoptando el primero en orden de relevancia para ese resultado como el más probable.

Una vez que tenemos los documentos representativos estos se descomponen en frases, y se eliminan de la búsqueda aquellas que no posean todas las claves de búsqueda en sí mismas. A cada una de estas frases se le realizará una transformación léxica, intentando identificar aquellas palabras que coincidan con el patrón o tipo buscado, en este caso nombre. Tener en cuenta que las claves no pueden formar parte de la respuesta, y por lo tanto son eliminadas como posibles. Una vez que tenemos todas las palabras resultantes se realizará una contabilización de apariciones de cada una de ellas, siguiendo la filosofía de “a cuantas más apariciones de una respuesta mayor probabilidad de ser la correcta”. Hemos optado por esta filosofía atendiendo a que nuestro QA podría trabajar sobre otros sistemas de información, como un buscador web, o similar, donde la gran cantidad de información nos proporcionaría mejores estadísticas a la hora de buscar una respuesta correcta.

A partir de las apariciones realizaremos un ranking con los tres mejores, junto con el documento asociado (anteriormente hemos dicho que se utilizará el primero por orden de relevancia devuelto por Lucene. Para nuestro caso se tratará de:

```
0060 plnast03lms 1 EFE19941114-07734 7 Kuala Lumpur
0060 plnaex03lms 2 EFE19940122-11393 2 España
0060 plnaex03lms 3 EFE19940122-11393 2 COBRA
```



## Arquitectura detallada del sistema.

### es.udc.lnaturales.practica.ER:

Contiene el Main de la aplicación.

- Main:
  - **private static** String *questionFile\_PATH* especifica la ruta al fichero que debe contener las preguntas.
  - **private static int** *maxHits* = 100 especifica el numero máximo de documentos preliminares obtenidos por el buscador de Lucene.
  - **public static void** *main*(String[] args)

### es.udc.lnaturales.practica.IndexGenerator:

En este paquete se incluyen ficheros que trabajan con los textos realizando pequeños cambios (aplicando un filtro para corregir algunos caracteres) y utiliza el Lucene, para obtener el conjunto de textos para poder realizar las búsquedas.

- DataSource:
  - **private static** String *efe\_path* especifica la ruta de la carpeta donde se encuentran los archivos de la Agencia EFE.
  - **public static** String *index\_path* especifica la ruta de la carpeta donde queremos almacenar el nuevo índice.
  - **public void** *generateDataSource*() procedimiento que realiza la creación del índice completo.
- Filter:
  - **public static** String *textFilter*(String text) elimina caracteres especiales y puntos suspensivos dentro del contenido de los tags <TEXT> y <TITLE> de los documentos.
  - **public static** String *titleFilter*(String id) elimina caracteres especiales y espacios innecesarios dentro del contenido del tag <ID> de los documentos.

- IndexGenerator:
  - **public static void** *main*(String[] args) realiza la creación del índice.
- IndexInfo: representación de cada uno de los nodo o índices de Lucene.
  - **public** String *getId()* devuelve el indentificador del nodo.
  - **public** String *getText()* devuelve el texto del nodo.
- LuceneIndex:
  - **public void** *generateLuceneIndex*(String path) crea o incrementa un índice en el path que se proporciona a partir de los datos que se pasan en el constructor de la clase.

### **es.udc.lnaturales.practica.PatternSelector:**

A partir de la pregunta realizada, obtiene el tipo de respuesta que habrá que devolver. Dependiendo del pronombre interrogativo utilizado, mediante una tabla hash, se devolverá un tipo de respuesta (nombre, numero,...).

- SeleccionarPatron:
  - **private static** HashMap<String, Dictionary> *preguntas* contiene los patrones de preguntas factuales que es capaz de reconocer el sistema.
  - **public static** Dictionary *devolverPatron*(String pregunta) devuelve un tipo específico de Dictionary obtenido a partir de preguntas y que concuerda con el tipo de respuesta esperado a partir de los patrones conocidos por el sistema.

### **es.udc.lnaturales.practica.Search:**

Se encarga de la búsqueda de respuestas. Una vez que se tiene la pregunta se selecciona, junto con el tipo de respuesta, las palabras clave que se utilizarán. Devolverá todos los resultados que se hayan encontrado en los textos, de acuerdo con las reglas establecidas (AND de todas las palabras clave, palabras más veces repetidas,...).

- Search:
  - **public static** *Appearances calcularRespuesta()* calculará el numero de apariciones de una palabra dentro de los posible resultados proporcionados. Parámetros:
    - Dictionary tipo: especifica el tipo de palabra buscado.
    - List<String> frases: conjunto de frases potenciales que pueden contener la respuesta correcta a nuestra pregunta.
  - List<String> claves: conjunto de claves utilizadas en la búsqueda de las frases pertenecientes al parámetro anterior.

### **es.udc.lnaturales.practica.Transformation:**

Se encarga de la transformación sintáctica, utilizando la aplicación FreeLing. Para cada tipo de palabra devolverá un código.

- Translation:
  - **private static final** String *FILE\_FORMAT* especifica el formato del documento temporal. Es nuestro caso será ISO8859-1.
  - **private static final** String *EXECUTE\_PATH* ruta y comando a la aplicación externa para la realización de la transformación morfológica.
  - **private static final** String *LOGFILE\_PATH* ruta al fichero temporal donde se realiza la transformación morfológica.

### **es.udc.lnaturales.practica.Transformation.util:**

Contiene clases específicas, como las encargadas de recoger los errores de la aplicación.

- ProcessExecuter:
  - **private static final** String *FILE\_FORMAT* especifica el formato del documento temporal. Es nuestro caso será ISO8859-1.
  - **public static void** *execute()* ejecuta un comando que recibe como parámetro.
    - String[] cmd: especifica el comando a realizar

- `String logFileName`: especifica la ruta al fichero de salida de la ejecución del comando.
- `ProcessErrorException`: representa los diferentes errores que se pueden producir durante una ejecución de uno de los comandos.
  - **public** `String getError()` nos devuelve el tipo de error producido
- `NoCommandAvailableException`: representa un tipo de error específico referente a la inexistencia o imposibilidad de ejecución de un determinado comando.
- `StreamGlobber`: esta clase se encarga de realizar la recuperación de los errores y salida durante la ejecución de un programa, ya que lo usamos para recoger la salida del programa FreeLing
  - **public void** `run()`

### **es.udc.lnaturales.practica.util:**

Contiene los objetos de la aplicación, tales como la respuesta, el ranking de respuestas, las distancias entre palabras, los tipos de respuestas,...

- `Appearances`: recoge el número de apariciones de cada una de las posibles respuestas, utilizando las tres que han sido encontradas en mayor número de ocasiones.
  - **public void** `addAppearance()` utilizado para añadir una aparición de un posible respuesta.
    - `String key`: representa la clave o respuesta probable.
    - `String docid`: representa el documento que contiene la posible respuesta
- `Dictionary`: contiene un tipo enumerado del mismo nombre en el que se recogen todos los tipos morfológicos posibles detectados por nuestro sistema, entre los que se encuentran: nombre, nombres propio y comunes, numeral, fecha, adjetivo o desconocido.
- `Distances`: clase utilizada como método de selección de las respuestas más probables a partir de las respuesta preliminares.
  - **public int** `distanceBetweenWords`(`String sentence`, `String wordKey`, `String reference`) {

- String sentence: representa la frase en la que se calcularán las distancias de aquellas respuestas potenciales.
  - String wordKey: cadena que representa a una posible respuesta y de la que queremos conocer la distancia a la clave de la pregunta.
  - String reference: cadena que representa a alguna de las claves de búsqueda que hemos utilizado y que se obtienen a partir de la pregunta que se proporciona.
- Rank: representa un ranking con las tres posibles respuesta que se manejan como salida del sistema.
    - **public void** *setFirst*(String first) establece la respuesta más probable.
    - **public void** *setSecond*(String second) establece la segunda respuesta más probable.
    - **public void** *setThird*(String third) establece la tercera respuesta más probable.

## Herramientas. Indexación.

### Apache Lucene.

#### Características:

- <http://lucene.apache.org/>
- API de desarrollo para indexación y búsqueda.
- Escrito en Java
- Se distribuye bajo licencia Apache (Open Source)
- Eficiente, versátil y robusto
- No incluye Spider (Apache Nutch, OmniFind)
- No es una aplicación completa sino un API
- Disponible también en C++, php y otros lenguajes

#### Ventajas:

- No necesitamos instalar ningún programa.
- Realización de índices dinámicos, con posibilidad de determinar el formato de los nodos del mismo.
- Optimización de almacenamiento y búsqueda.
- Búsquedas específicas capaces de devolver los resultados por orden de preferencia.

#### Desventajas:

- Solo permite almacenar texto plano o comprimido.
- No permite la transformación sintáctica o morfológica.
- No realiza búsquedas mediante raíces o sinónimos.
- No indexa documentos de forma eficiente sin realizar cambios específicos en los mismo, como ocurre con alguno documento en formato XML+DTD.

Se ha tenido en consideración:

### TERRIER (TERabyte RetrIEver).

#### Características:

- Desarrollado en java.
- Disponible en versión bash o con interfaz gráfica.

#### Ventajas:

- Indexación en varios niveles.
- Permite la utilización de varios sistemas de stemming, como por ejemplo Snowball, incluyendo el castellano.
- Capaz de indexar documentos en formatos específicos de forma nativa.

**Desventajas:**

- Enfocado sobre todo al ámbito de la investigación.
- No realiza búsquedas mediante sinónimos.
- Tiempos de análisis muy elevado debido a la gran cantidad de cálculos que realiza con las *queries* realizadas.

Se ha optado por el uso de Apache Lucene con respecto a Terrier debido en gran medida al tiempo de espera que se necesitaría para el análisis y recuperación de cada consulta. Al realizar un sistema preliminar de uso educativo hemos intentado que nuestro sistema no consumiese demasiado tiempo, de modo que la realización de las pruebas con conjunto amplios de preguntas no supusiese un hándicap. Como se explicará en la última versión el gran problema a la hora de responder algunas preguntas radica en la imposibilidad de realizar búsquedas mediante sinónimos, pero ambos sistemas no lo implementan.

## ***Herramientas. Análisis Morfo-sintáctico.***

### **Freeling.**

Para el desarrollo de esta práctica se ha utilizado este analizador morfológico principalmente por su soporte para el lenguaje castellano que es el que nos ocupa en esta práctica y su facilidad de uso y adaptación a nuestra práctica.

Freeling utiliza un conjunto de etiquetas para representar la información morfológica de las palabras. Este conjunto de etiquetas se basa en las etiquetas propuestas por el grupo EAGLES para la anotación morfosintáctica de lexicones y corpus para todas las lenguas europeas. Así pues está previsto que recojan los accidentes gramaticales existentes en las lenguas europeas. Es por esto que hay atributos que pueden no especificarse. Si un atributo no se especifica significa que o bien expresa un tipo de información que no existe en la lengua o que la información no se considera relevante. Esto ha limitado en cierta medida la búsqueda de algunas respuestas debido a que Freeling no siempre es capaz de especificar nombres propios de personas o nombres propios de lugares (topónimos) tal y como se intentó en una de las mejoras propuestas que se comentan más adelante.

Las palabras reconocidas y codificadas por Freeling son adjetivos, adverbios, determinantes, nombres, verbos, pronombres, conjunciones, interjecciones, preposiciones, signos de puntuación, numerales y fechas. Sin embargo se han tenido en cuenta solamente nombres (propios y comunes), numerales y fechas puesto que se manejan únicamente preguntas factuales.

En la presente práctica se realiza una llamada a través de la clase “Translation.java” a Freeling a través de la línea de comandos ejecutando la aplicación a la que se le pasa un fichero de texto con las palabras a analizar y almacenando los resultados del análisis en un archivo temporal especificado en dicha clase. A partir del mencionado archivo temporal nuestra aplicación lee los códigos y realiza las posteriores búsquedas.

Otro analizador contemplado fue NooJ que incluye diccionarios y gramáticas. Los diccionarios y gramáticas son aplicados a textos para localizar el modelo morfológico, léxico y sintáctico etiquetando palabras simples y compuestas. Es capaz de construir concordancias complejas, en lo que concierne a todos los tipos de modelo de estados finitos y patrones de contexto libre. Finalmente se descartó su uso puesto que Freeling permitió cubrir en gran medida las necesidades planteadas.



## Descargas.

- Apache Lucene:
  - <http://ftp.udc.es/apache-dist/lucene/java/>
- Terrier:
  - <http://ir.dcs.gla.ac.uk/terrier/Download.html>
- FreeLing:
  - [https://lafarga.cpl.upc.edu/frs/?group\\_id=32](https://lafarga.cpl.upc.edu/frs/?group_id=32)

## Instrucciones de instalación.

La primera y más importante consideración que se debe hacer a la hora de instalar el sistema es que es dependiente de la plataforma, i.e. se ha construido y probado en Windows (tanto XP como Vista), por lo que la presencia de este sistema operativo es fundamental para su funcionamiento.

Otra consideración que hay que tener en cuenta debido a su implementación en Java es que si se desea modificar la ruta de los ficheros que se emplean, siempre se debe introducir una barra adicional para que funcione correctamente, es decir las rutas serán del estilo “C:\\fichero.txt” en vez de “C:\fichero.txt”. Asimismo es indiferente emplear la notación “C:\\fichero.txt” que “C://fichero.txt”.

Por último, se comentarán ciertos aspectos que se pueden modificar en los ficheros fuente entregados para garantizar la portabilidad del sistema. Los valores que se muestran son los que contemplan una instalación por defecto:

- Dentro del paquete `es.udc.lnaturales.practica.Transformation` se declaran cuatro Strings referentes a la instalación y ejecución de Freeling. Individualmente:
  - ***FILE\_FORMAT***: Esta es la codificación que va a tener el fichero de salida, una vez se haya realizado el análisis con Freeling. No es necesario cambiarlo.
  - ***EXECUTE\_PATH***: Contiene la sentencia de ejecución del analizador Freeling con dos consideraciones: que Freeling debe estar instalado en el directorio `C:\Freeling` y que el fichero que recibe el analizador se encuentra en el fichero `naturales.txt` ubicado en la partición `D:`. Si se desea modificar estas rutas sólo hará falta cambiar el valor de dicho String, así como es libre la modificación de dicho fichero.
  - ***LOGFILE\_PATH***: Ruta al fichero de log que crea Freeling después de realizar el análisis sintáctico del fichero `naturales.txt`. Esta variable también se puede modificar libremente.
  - Dentro del método `executeAnalyzer()` de la clase `Translation.java` presente en dicho paquete, se va a ir construyendo un fichero frase por frase que después analizará Freeling. Para ello, se ha establecido un nombre y una ruta por defecto (`d:\naturales.txt`) que es modificable, siempre que concuerde con el fichero presente en la variable ***EXECUTE\_PATH***.
- Dentro del paquete `es.udc.lnaturales.practica.IndexGenerator`, y más concretamente en la clase `DataSource.java`, están declarados dos Strings relacionados con la generación del árbol de índices de Lucene. Concretamente:

- *efe\_path*: Declara la ruta al fichero de preguntas que se utilizará posteriormente –tras su descomposición en elementos- para la generación de índices con Lucene. Por defecto se encuentra ubicado en C: pudiendo ser modificado libremente tanto su nombre como su ubicación.
- *index\_path*: En esta variable se almacenarán los índices que se van generando con Lucene a partir de los elementos del documento. Su ubicación por defecto también es el disco C: pudiendo ser asimismo modificado su nombre y ubicación.
- Dentro del paquete `es.udc.lnaturales.practica.ER`, en la única clase presente en dicho paquete, `Main.java`, se declara un `String` referente a la ubicación del fichero de preguntas sobre el que se trabajará. Entrando más en detalle:
  - *questionFile\_PATH*: Declara la ruta en la que se encuentra el fichero de preguntas que se va utilizar (el cual no contiene todas las preguntas por motivo de pruebas). Por defecto se encuentra en C: y es modificable.

## Diario de Trabajo.

Ahora veremos las diferentes versiones que hemos realizado sobre nuestro sistema, así como los problemas que se han detectado con cada ejecución de las mismas y los resultados que ofrecen sobre las 50 preguntas que se han utilizado como entrada. Con cada versión se muestran además de los problemas detectados posibles soluciones para los mismos, aunque no todas se han utilizado, o mejor dicho referenciado, a modo de versión, ya que no ofrecían mejora de resultados sobre la anterior.

## Versión 1

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Nil	Nil	Nil
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	IRA	paz
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Swissair	abales	Abasolo
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Nil	Nil	Nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Actualidad	Antártida	agujero
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	Fuentes	Tailandia
M SPA 0061 ¿Cuál es la capital de Irán?		
Oposición	Analistas	Autoridad
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
República	Ashjabad	repúblicas
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
Nil	Nil	Nil
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Nil	Nil	Nil
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Integrantes	edificio
M SPA 0068 ¿Quién escribió "Star Trek"?		
Star Trek	Seria	generación
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Pamplona	Universidad de Navarra	Francesco Cossiga
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Radio_Pyongyang	país	Kim_Jong_Il
M SPA 0071 ¿Qué significan las siglas ETA?		
Bilbao	Lucha	mexico
M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		

Nil	Nil	Nil
M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Nil	Nil	Nil
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Nil	Nil	Nil
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Nil	Nil	Nil
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Nil	Nil	Nil
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Fuentes	marinos
M SPA 0083 ¿Quién es el presidente de Rusia?		
Nil	Nil	Nil
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Nil	Nil	Nil
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Nil	Nil	Nil
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Fundación_Carlos_...	Palacio Real	sedes
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Corea de el Norte	Pyongyang	Pekín
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Moratoria	Japón
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Nil	Nil	Nil
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
WWF	OMS	control
M SPA 0094 ¿Quién es el director de la CIA?		
James_Woolsey	Espionaje	presidente
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin? 17docs		
Premio	Nobel de Literatura	Rusia

M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Sede	Alcalde	presidente
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
Nil	Nil	Nil
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
Nil	Nil	Nil
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1º opción	12	12
2º opción	4	2
3º opción	1	.333
MRR Total		15.333

El sistema realiza búsquedas únicamente para preguntas cuyas respuesta esperada entra en el tipo de nombre (los cuales pueden ser indistintamente propios o comunes, ya que no los distingue). Las preguntas correspondientes a fechas y numerales no están implementadas, por lo que su respuesta por defecto siempre será nil. Algunas preguntas en la fase inicial de búsqueda de documentos relacionados devuelven gran cantidad de documentación relacionada con las claves. En estos casos, y debido a que el tiempo de transformación de frases crece de forma exponencial, estas preguntas se han omitido dando como resultados también nil. Este es el caso de las siguientes preguntas:

M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?  
(800 documentos preliminares encontrados)

M SPA 0083 ¿Quién es el presidente de Rusia?  
(908 documentos preliminares encontrados)

M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?  
(430 documentos preliminares encontrados)

Problemas de esta versión:

1. Los documentos anteriormente mencionados no devuelven ningún resultado debido que encuentran demasiados probables y el tiempo de búsqueda sería desproporcionado.
2. En numerosas preguntas cuya respuesta tiene que ser un nombre propio existen 2 clases de problemas fundamentales.
  - a. La respuesta correcta no entra entre las 3 mejores, siendo estos nombres comunes.
  - b. La respuesta correcta está entre las 3 mejores pero por encima de ella existen resultados de tipo nombre común
3. No se tienen en cuenta preguntas cuya salida tiene que ser una fecha o cantidades.

A partir de estos 2 primeros problemas, intentaremos solucionarlos en las siguientes versiones. Primero atajaremos las preguntas que no se implementa su resolución, para fechas y cantidades. Después realizaremos una especificación para poder realizar búsquedas más exhaustivas distinguiendo nombre propios de nombre comunes.



## Versión 2

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Nil	Nil	Nil
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	Ira	paz
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Gastos	Vuelos	meses
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Diez	Nil	nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Regiones	Actualidad	Chile
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	Fuentes	Tailandia
M SPA 0061 ¿Cuál es la capital de Irán?		
Oposición	Analistas	autoridad
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
república	Ashjabad	repúblicas
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
Opep	Colombia	Cuba
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Integrantes	edificios
M SPA 0068 ¿Quién escribió "Star Trek"?		
Seria	Generación	premios
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Pamplona	Universidad de Navarra	Francesco Cossiga
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Radio_Pyongyang	País	Kim_Jong_II
M SPA 0071 ¿Qué significan las siglas ETA?		
Bilbao	Lucha	México
M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		
Petróleo	Accidente	suelo

M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Invierno	Juegos Olímpicos de Atlanta	COI
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	sábado
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	Ocasión	fotografía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	15 de junio de 1989
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Fuentes	marino
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Agencia	Ministros
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Ministro de Asuntos Exteriores	Ministro	Javier Solana
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Delegación	Yaser Arafat
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Nil	Nil	Nil
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Nil	Nil	Nil
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Moratoria	Japón
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Selección	Daniel Passarella	seleccionador
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil
M SPA 0094 ¿Quién es el director de la CIA?		
James Woolsey	Espionaje	presidente
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin? 17docs		
Nobel de Literatura	Rusia	escritor

M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Lillehammer	COI	candidatura
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
Personas	China	fuentes
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	2,7 millones	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1ª opción	16	16
2ª opción	2	1
3ª opción	1	.333
MRR Total		17.333

Soluciones que hemos intentado arreglar:

1. Para las preguntas que daban un número desproporcionado de respuestas se ha optado por una decisión salomónica. Dado que Lucene devuelve los hits o resultados por orden de preferencia se ha optado por reducir a un máximo de 100 documentos preliminares, en cuyo caso el tiempo de respuesta del sistema es de 12 minutos por pregunta.
2. Se ha implementado la resolución de las preguntas cuya salida tiene que ser del tipo fecha o cantidad.

Cabe destacar la mejora que ha sufrido el sistema con la actualización, pero ésta no está exenta de inconvenientes. Vemos qué preguntas que antes se resolvían en la primera versión.

M SPA 0087 ¿En qué ciudad está el Museo del Prado?

M SPA 0088 ¿Cuál es la capital de Corea del Norte?

M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?

El problema se debe a que el analizador morfológico realiza una transformación sobre las contracciones gramaticales “del” y “al”. En lugar de trabajar con estas utiliza “de el” y “a el”, y lógicamente, al realizar la búsqueda dentro de Lucene no devuelve ningún resultado.

Viendo que este problema es a causa del analizador morfológico, más que un problema del diseño del sistema, nos encargaremos de él en la siguiente versión, antes de realizar la transformación específica de tipos para la distinción de tipos de nombre.

## Versión 3

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Nil	Nil	Nil
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	Ira	paz
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Gastos	Vuelos	meses
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Diez	Nil	nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Regiones	Actualidad	Chile
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	Fuentes	Tailandia
M SPA 0061 ¿Cuál es la capital de Irán?		
Oposición	Analistas	Autoridad
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
República	Ashjabad	Repúblicas
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
Opep	Colombia	Cuba
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	Nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Integrantes	Edificio
M SPA 0068 ¿Quién escribió "Star Trek"?		
Serie	Generación	Premios
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Pamplona	Universidad de Navarra	Francesco Cossiga
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Radio_Pyongyang	País	Kim_Jong_Il
M SPA 0071 ¿Qué significan las siglas ETA?		
Bilbao	Lucha	México
M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		
Petróleo	Accidente	Suelo

M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Lillehammer	Juegos Olímpicos de Barcelona	Juegos Olímpicos de Atlanta
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	Sábado
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	Ocasión	Fotografía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	15 de junio de 1989
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Fuentes	marinos
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Agencia	Ministro
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Ministro de Asuntos Exteriores	Ministro	Javier Solana
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Delegación	Yaser Arafat
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Fundación Carlos de Amberes	Palacio Real	sedes
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Corea de el Norte	Pyongyang	Pekín
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Moratoria	Japón
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Selección	Daniel Passarella	seleccionador
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil
M SPA 0094 ¿Quién es el director de la CIA?		
James Woolsey	Espionaje	presidente
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin?		
Nobel de Literatura	Rusia	escritor

M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Lillehammer	COI	candidatura
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
Personas	China	fuentes
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	2,7 millones	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1º opción	16	16
2º opción	2	1
3º opción	2	.666
MRR Total		17.666

Soluciones que hemos intentado arreglar:

1. Para solucionar el problema dado en la versión anterior con las contracciones gramaticales, hemos realizado un “pequeño parche”. Ahora, si una de las claves utilizadas en la búsqueda contiene una contracción del tipo “de\_el” o “a\_el”, se realiza una conversión directa a “del” y “al” respectivamente.

Como podemos observar, de las 3 preguntas anteriores que fallaban por culpa de este problema, ahora se responde a 2 de ellas. La pregunta de Corea del Norte se responde correctamente en 2º posición, y la pregunta del Museo del Prado que devuelve de momento 3 respuestas incorrectas.

Para la pregunta del Fondo Mundial, no encuentra respuesta ya que “Fondo Mundial para la Protección de la Naturaleza” lo interpreta como un nombre y realiza la búsqueda de todas las palabras que lo forman (Fondo AND Mundial AND para AND la AND Protección AND de...).

## Versión 4

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Nil	Nil	Nil
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	Ira	Irlanda_de_el_Norte
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Gastos	Vuelos	Meses
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Diez	Nil	nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Antártida	Argentina	Chile
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	COBRA	Tailandia
M SPA 0061 ¿Cuál es la capital de Irán?		
Arafat	EFE	ONU
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
Ashjabad	Rusia	Uzbekistán
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
Opep	Colombia	Cuba
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	Nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Fundación	Manhattan
M SPA 0068 ¿Quién escribió "Star Trek"?		
The_Next_Generation	Voyager	GENEVIEVE_BUJOLD
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Pamplona	Universidad_de_Navarra	Francesco_Cossiga
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Kim_Song_Ae	Kyodo	Radio_Pyongyang
M SPA 0071 ¿Qué significan las siglas ETA?		
Bilbao	Estado_de_Chiapas	México
M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		



Nil	Nil	Nil
M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Lillehammer	Juegos Olímpicos de Barcelona	Juegos Olímpicos de Atlanta
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	Sábado
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	EN_CANAL	Doña_Sofía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	15 de junio de 1989
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Francia	Tolón
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Moscú	Kremlin
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Javier Solana	Gobierno	España
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Yaser Arafat	Israel
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Nil	Nil	Nil
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Corea_de_el_Norte	Pyongyang	Pekín
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Japón	Antártida
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Daniel Passarella	Alfio Basile	Diego Armando Maradona
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil
M SPA 0094 ¿Quién es el director de la CIA?		
James Woolsey	Bill Clinton	Servicio Central de

		Información
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin?		
Nobel de Literatura	Rusia	Alexandr_Solzhenitsin
M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Juegos Olímpicos de Invierno	Lillehammer	COI
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
China	Ministerio de Seguridad Pública	Xinjiang
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	2,7 millones	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1º opción	17	17
2º opción	2	1
3º opción	1	.333
MRR Total		18.333

Soluciones que hemos implementado:

1. Uno de los mayores problemas que hemos tenido hasta ahora era que simplemente se realizaban búsquedas de sustantivos. Ahora se diferencia entre nombres propios y nombres comunes.

Como podemos observar en casi todas las preguntas en las que había que buscar un nombre propio, ahora devuelve siempre como respuestas nombres propios. Ahora tenemos varias preguntas (0056, 0073) que antes se fallaban porque todas las respuestas que daban eran nombres comunes, y ahora se aciertan. Otras, que aunque se acertaban, la respuesta correcta no estaba en primer lugar y ahora sí (0062). Y otras que antes se respondían, ahora se fallan (0072).

Introducir las diferencias entre nombres propios y nombres comunes ha mejorado en cierta medida los resultados, porque aunque estadísticamente sólo se ha observado un

pequeño cambio, podemos ver que las respuestas de nombres propios no son tan incongruentes como lo eran antes.

Los fallos que hemos cometido en la pregunta 0072 son debidos a que antes las respuestas dadas eran nombres comunes, además de erróneas, por lo que no afectan en absoluto.

## Versión 5

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Festival de Cine de Venecia	Festival	Nagisa Oshima
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	Ira	Irlanda_de_el_Norte
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Swissair	Abasolo	Sabena
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Diez	Nil	nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Antártida	Argentina	Chile
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	COBRA	Tailandia
M SPA 0061 ¿Cuál es la capital de Irán?		
Arafat	CiU	Sáenz
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
Ashjabad	Rusia	Uzbekistán
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
Opep	Colombia	Cuba
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	Nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Fundación	Manhattan
M SPA 0068 ¿Quién escribió "Star Trek"?		
The_Next_Generation	Voyager	GENEVIEVE_BUJOLD
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Pamplona	Universidad_de_Navarra	Francesco_Cossiga
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Kim_Song_Ae	Kyodo	Radio_Pyongyang
M SPA 0071 ¿Qué significan las siglas ETA?		
Bilbao	Estado_de_Chiapas	México

M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		
Komi Artic Oil	Comité	Tambov
M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Lillehammer	Juegos Olímpicos de Barcelona	Juegos Olímpicos de Atlanta
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	Sábado
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	EN_CANAL	Doña_Sofía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	15 de junio de 1989
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Francia	Tolón
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Moscú	Kremlin
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Javier Solana	Gobierno	España
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Yaser Arafat	Israel
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Fundación Carlos Ambers	Palacio Real	Museo_de_el_Prado
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Corea_de_el_Norte	Pyongyang	Pekín
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Puerto Vallarta	Japón
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Daniel Passarella	Alfio Basile	Consejo Ejecutivo
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil
M SPA 0094 ¿Quién es el director de la CIA?		

James Woolsey	Bill Clinton	Servicio Central de Información
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin?		
Nobel de Literatura	Rusia	Alexandr Solzhenitsin
M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Juegos Olímpicos de Invierno	Lillehammer	COI
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
China	Ministerio de Seguridad Pública	Xinjiang
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	2,7 millones	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1º opción	19	19
2º opción	3	1.5
3º opción	2	.666
MRR Total		21.166

Soluciones que hemos implementado:

1. Hemos mejorado el comportamiento de las búsquedas. Ahora no hace distinción con letras capitales, es decir, no distingue las mayúsculas de las minúsculas y por lo tanto las interpreta igualmente. El motivo de este cambio es porque había palabras que al iniciar textos o frases no se tenían en cuenta (al llevar mayúsculas eran descartadas).
2. Otra mejora implementada para esta versión es la de “el definido no entra en la definición”. Muchas veces mostraba como una de las respuestas palabras de la propia pregunta, por lo que se ha incluido una cláusula que las respuestas que coincidan con las palabras de la pregunta, queden descartadas del grupo final. Desgraciadamente esta regla no se cumple para las

contracciones gramaticales “del” y “al”, debido a la transformación que realizamos en la versión 3.

De esta manera hemos observado un aumento de aciertos, provocado tal vez por lo que hemos dicho anteriormente, que muchas de esas respuestas iniciaban frases o textos.

## Versión 6

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
Festival de Cine de Venecia	Festival	Bienal
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
Gerry Adams	Ira	Irlanda_de_el_Norte
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
Swissair	Sabena	Abasolo
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
Diez	Nil	Nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Chile	Argentina	Antártida
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
Nil	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
Nil	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
Nil	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
Kuala Lumpur	España	COBRA
M SPA 0061 ¿Cuál es la capital de Irán?		
Arafat	Argentina	EFE
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
Ashjabad	Rusia	Uzbekistán
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
OPEP	Colombia	Gobierno
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	Nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
Beatles	Manhattan	Nureyev
M SPA 0068 ¿Quién escribió "Star Trek"?		
The_Next_Generation	NYPD	Northen Exposure
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Francesco_Cossiga	Pamplona	Universidad_de_Navarra
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Kyodo	Kim_Song_Ae	Radio_Pyongyang
M SPA 0071 ¿Qué significan las siglas ETA?		
Estado_de_Chiapas	México	Alfredo Harp Helú



M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		
EFE	Comité	Situaciones de Emergencia
M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Lillehammer	Juegos Olímpicos de Barcelona	Juegos Olímpicos de Atlanta
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	Viernes
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	Trueba	Doña_Sofía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	26 de diciembre de 1992
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Francia	Defensa
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Kremlin	Moscú
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Javier Solana	Gobierno	España
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Yaser Arafat	Israel
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Academia de Bellas Artes de San Fernando	Calcografía Nacional	Palacio Real
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Pyongyang	Pekín	OEIA
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Puerto Vallarta	Antártida
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Daniel Passarella	Alfio Basile	Benito Floro
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil

M SPA 0094 ¿Quién es el director de la CIA?		
James Woolsey	Bill Clinton	Servicio Central de Información
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin?		
Nobel de Literatura	Rusia	Alexandr_Solzhenitsin
M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Lillehammer	Juegos Olímpicos de Invierno	COI
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
China	EFE	Ministerio de Seguridad Pública
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	Un millón y medio	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados		
	Acertadas	MRR parcial
1º opción	20	20
2º opción	1	0.5
3º opción	2	.666
MRR Total		21.166

Soluciones que hemos implementado:

1. En la versión anterior, se implementó la mejora de “el definido no entra en la definición” y habíamos indicado que no se cumplía para aquellas claves que incluyesen una contracción gramatical. En esta nueva versión se ha corregido este problema.

## Versión Final

## Tabla de resultados

M SPA 0051 ¿En qué festival se entregan los premios "León de Oro"?		
<b>Festival de Cine de Venecia</b>	Festival	Bienal
M SPA 0052 ¿Quién es el líder del Sinn Fein?		
<b>Gerry Adams</b>	Ira	Irlanda_de_el_Norte
M SPA 0053 ¿Cómo se llama la compañía aérea nacional de Suiza?		
<b>Swissair</b>	Sabena	Abasolo
M SPA 0054 ¿Cuántos tripulantes murieron en el submarino Emeraude?		
<b>Diez</b>	Nil	Nil
M SPA 0055 ¿En qué tipo de procesador se descubrió un error en la unidad aritmética?		
Nil	Nil	Nil
M SPA 0056 ¿Sobre qué continente se detectó el agujero de ozono?		
Chile	Argentina	<b>Antártida</b>
M SPA 0057 ¿Quién es el mayor exportador europeo de aceite de oliva?		
<b>Nil</b>	Nil	Nil
M SPA 0058 ¿Cuándo se constituyó la República de Sudáfrica?		
<b>Nil</b>	Nil	Nil
M SPA 0059 ¿Qué porcentaje del comercio mundial de drogas está controlado por el Cartel de Cali?		
<b>Nil</b>	Nil	Nil
M SPA 0060 ¿Cuál es la capital de Malasia?		
<b>Kuala Lumpur</b>	España	COBRA
M SPA 0061 ¿Cuál es la capital de Irán?		
Arafat	Argentina	EFE
M SPA 0062 ¿Cuál es la capital de Turkmenistán?		
<b>Ashjabad</b>	Rusia	Uzbekistán
M SPA 0063 ¿Cuál es el principal país productor de petróleo en el mundo?		
OPEP	Colombia	Gobierno
M SPA 0064 ¿Cuántos países son miembros de la Unión Europea?		
Dos	Tres	Nueve
M SPA 0065 ¿Cuándo se firmó el Acta Única Europea?		
Nil	Nil	Nil
M SPA 0066 ¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?		
Nil	Nil	Nil
M SPA 0067 ¿A qué grupo pertenecía John Lennon?		
<b>Beatles</b>	Manhattan	Nureyev
M SPA 0068 ¿Quién escribió "Star Trek"?		
The_Next_Generation	NYPD	Northern Exposure
M SPA 0069 ¿Quién es el presidente de la República de Italia?		
Francesco_Cossiga	Pamplona	Universidad_de_Navarra
M SPA 0070 ¿Quién ostenta el poder en Pyongyang?		
Kyodo	Kim_Song_Ae	Radio_Pyongyang
M SPA 0071 ¿Qué significan las siglas ETA?		
Estado_de_Chiapas	México	Alfredo Harp Helú
M SPA 0072 ¿En qué parte de Rusia se rompió un oleoducto?		

EFE	Comité	Situaciones de Emergencia
M SPA 0073 ¿Dónde se celebraron los Juegos Olímpicos de 1996?		
Lillehammer	Juegos Olímpicos de Barcelona	Juegos Olímpicos de Atlanta
M SPA 0074 ¿Cuántos hijos tiene Anthony Quinn?		
Nil	Nil	Nil
M SPA 0075 ¿Cuál es la profesión de Renzo Piano?		
Nil	Nil	Nil
M SPA 0076 ¿En qué año se creó el Fondo Monetario Internacional?		
Miércoles	Jueves	Viernes
M SPA 0077 ¿Quién dirigió "Con la muerte en los talones"?		
Alfred Hitchcock	Trueba	Doña Sofía
M SPA 0078 ¿Cuántas personas murieron en el juzgado de Euskirchen?		
Nil	Nil	Nil
M SPA 0079 ¿Cuándo se fundó la CEE?		
Jueves	Noviembre	26 de diciembre de 1992
M SPA 0080 ¿En qué ciudad europea está la Torre Eiffel?		
Nil	Nil	Nil
M SPA 0081 ¿A qué país pertenece el agente inmobiliario Schneider?		
Nil	Nil	Nil
M SPA 0082 ¿Qué submarino nuclear francés sufrió un accidente?		
Emeraude	Francia	Defensa
M SPA 0083 ¿Quién es el presidente de Rusia?		
Borís Yeltsin	Kremlin	Moscú
M SPA 0084 ¿Quién es el presidente italiano de Asuntos Exteriores?		
Javier Solana	Gobierno	España
M SPA 0085 ¿Cuál es el nombre de pila de la mujer de Nelson Mandela?		
Nil	Nil	Nil
M SPA 0086 ¿Qué significa OLP?		
Organización para la Liberación de Palestina	Yaser Arafat	Israel
M SPA 0087 ¿En qué ciudad está el Museo del Prado?		
Academia de Bellas Artes de San Fernando	Calcografía Nacional	Palacio Real
M SPA 0088 ¿Cuál es la capital de Corea del Norte?		
Pyongyang	Pekín	OEIA
M SPA 0089 ¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?		
CBI	Puerto Vallarta	Antártida
M SPA 0090 ¿Quién es el entrenador del equipo nacional de fútbol noruego?		
Daniel Passarella	Alfio Basile	Benito Floro
M SPA 0091 ¿Cuál es la causa más frecuente de los accidentes de coche?		
Nil	Nil	Nil
M SPA 0092 ¿Qué país de África ha adoptado una nueva constitución?		
Nil	Nil	Nil
M SPA 0093 ¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?		
Nil	Nil	Nil
M SPA 0094 ¿Quién es el director de la CIA?		
James Woolsey	Bill Clinton	Servicio Central de

		Información
M SPA 0095 ¿Qué premio Nobel ganó Solzhenitsin?		
Nobel de Literatura	Rusia	Alexandr_Solzhenitsin
M SPA 0096 ¿En qué ciudad se celebraron los Juegos Olímpicos de invierno?		
Lillehammer	Juegos Olímpicos de Invierno	COI
M SPA 0097 ¿Cuándo tomará China la posesión de Hong Kong?		
Nil	Nil	Nil
M SPA 0098 ¿Qué causó el incendio en un cine en la ciudad china de Karamai?		
China	EFE	Ministerio de Seguridad Pública
M SPA 0099 ¿Cuántos habitantes hay en Moscú?		
1954	Un millón y medio	000
M SPA 0100 ¿En qué mes se produjo el naufragio del Estonia?		
Nil	Nil	Nil

Resultados (Permisivo)		
	Acertadas	MRR parcial
1º opción	20	20
2º opción	1	.5
3º opción	2	.666
MRR Total		21.166

Resultados (Estricto)		
	Acertadas	MRR parcial
1º opción	18	18
2º opción	1	.5
3º opción	2	.666
MRR Total		19.166

## Justificación de respuestas

En algunos casos los documentos relacionados con las respuestas que ofrece el sistema final no concuerdan con los proporcionados como válidos. En nuestro caso, aquellas en las que el documento asociado no concuerda se proporciona como parte de la respuesta un documento a partir del cual se puede inferir la respuesta buscada. Veamos ahora cada uno de esos casos individualmente.

0051|¿En qué festival se entregan los premios "León de Oro"?|(EFE19940902-00973;En el Festival de Cine de Venecia)

```
<DOCID>EFE19941125-16022</DOCID>
<TITLE>    ITALIA-CINE
          SOLO UN "LEON DE ORO" PARA FESTIVAL VENEZIA
</TITLE>
<TEXT>    Venecia (Italia), 25 nov (EFE).- En la próxima edición del
Festival de Cine de Venecia se otorgará un solo León de Oro, en lugar
de los cinco premios máximos concedidos este año, dijo hoy, viernes,
el director de la Bienal, Gillo Pontecorvo
    Precisó que el Consejo Directivo había aprobado su propuesta de
reducir el número de premios en la 52 edición del Festival, que se
celebrará en septiembre de 1995, para valorizarlos más.
    Los tres Leones de Oro a la carrera quedaron convertidos en tres
"osellas" de oro (la "osella" es una antigua moneda de la República
Veneciana), mientras se prohibió la posibilidad del "ex aequo" a la
mejor película.
    En la 51 edición del Festival, el León de Oro al mejor filme fue
compartido por el macedonio "Antes de la lluvia" y el chino "Viva el
amor".
    En 1995, se entregarán un León de Oro al mejor filme y un Gran
Premio del Jurado, cuatro Copas Volpi (mejor actor y mejor actriz,
protagonistas y no protagonistas) y las "osellas" a la carrera. EFE
</TEXT>
```

0052|¿Quién es el líder del Sinn Fein?|(EFE19940105-01866;Gerry Adams),(EFE19941013-07212;Gerry Adams)

```
<DOCID>EFE19940104-01096</DOCID>
<TITLE>    ULSTER-ATENTADO
          EXPLOSION EN SEDE CENTRAL SINN FEIN
</TITLE>
<TEXT>    Dublín, 4 ene (EFE).- Una leve explosión, causada al parecer
por
una carta bomba, se registró hoy, martes, en la sede central del Sinn
Fein, sin que se registraran víctimas, informó la Policía.
    Expertos de la Policía irlandesa examinan un segundo paquete
sospechoso, mientras las oficinas del Sinn Fein, rama política del
Ejército Republicano Irlandés (IRA), y las del periódico de dicho
partido, "An Phoblacht", han sido evacuadas.
    La sede central del Sinn Fein se encuentra en la plaza de "Parnell
Square", en Dublín, y muy cerca están las oficinas del "An
Phoblacht".
    La explosión ocurre un día después de las manifestaciones del
líder del Sinn Fein, Gerry Adams, en el sentido de que la declaración
angloirlandesa, firmada el pasado 15 de diciembre, tiene un potencial
```

significativo ya que contempla el reconocimiento, por parte de Londres, de que los irlandeses tienen derecho a la autodeterminación.

0082|¿Qué submarino nuclear francés sufrió un accidente?|(EFE19940330-19318;El "Emeraude"),(EFE19940330-19191;Emeraude)

```
<DOCID>EFE19940330-19078</DOCID>
<TITLE>    FRANCIA-ACCIDENTE
            MUEREN DIEZ MARINOS EN ACCIDENTE SUBMARINO NUCLEAR
</TITLE>
<TEXT>    París, 30 mar (EFE).- Diez marinos han muerto en un
accidente
ocurrido a bordo del submarino nuclear francés de ataque "Emeraude",
informaron hoy, miércoles, fuentes militares.
    El accidente se produjo en el compartimento de turbo-alternador
del submarino, que no se ha visto afectado en su instalación nuclear,
precisaron las fuentes.
    El navío, que se encontraba sumergido cerca de la costa de Tolón
(sureste de Francia) en el momento del accidente, salió a la
superficie y regresó a la base militar de esa ciudad.
</TEXT>
```

0083|¿Quién es el presidente de Rusia?|(EFE19940710-05944;Boris Yeltsin),(EFE19940408-04084;Boris Yeltsin),(EFE19940709-05653;Boris Yeltsin)

```
<DOCID>EFE19941215-08415</DOCID>
<TITLE>    URGENTE
            RUSIA-CHECHENIA
</TITLE>
<TEXT>
    Moscú, 15 dic (EFE).- El presidente de Rusia, Borís Yeltsin,
emitirá en las próximas horas una declaración pública sobre la
situación en Chechenia, informó hoy, jueves, la agencia oficial
ITAR-TASS.
</TEXT>
```

0086|¿Qué significa OLP?|(EFE19940115-07057;Organización para la Liberación de Palestina),(EFE19940816-07711;Organización para la Liberación de Palestina),(EFE19940312-07775;Organización para la Liberación de Palestina)

```
<DOCID>EFE19940127-14347</DOCID>
<TITLE>    TUNEZ-OLP
            REGRESO A TUNEZ DEL JEFE DE LA OLP
</TITLE>
<TEXT>    Túnez, 27 ene (EFE).- El jefe de la Organización para la
Liberación de Palestina (OLP), Yaser Arafat regresó hoy, jueves, a
Túnez tras un viaje en el que visitó Noruega, Siria, A. Saudí y en
tres ocasiones Egipto, según informa la agencia palestina Wafa.
    Arafat emprenderá de nuevo viaje el próximo sábado con destino a
Davos (Suiza) donde se entrevistará con Simon Peres, ministro de
Asuntos Exteriores israelí.
</TEXT>
```

0089|¿Dónde se celebró la asamblea anual de la Comisión Ballenera Internacional?|(EFE19941110-05535;En Puerto Vallarta, México)

<DOCID>EFE19940524-13903</DOCID>

<TITLE> BALLENAS-CBI

FONDO MUNDIAL SOLICITA APOYO PARA RESERVA DE BALLENAS

</TITLE>

<TEXT> Puerto Vallarta (México), 23 may (EFE).- El Fondo Mundial para la

Naturaleza (WWF) pidió a los países que no han fijado su posición sobre la intención de crear un refugio austral para las ballenas, que imiten a México y a Dominica, que apoyan esta iniciativa, informó hoy, lunes, un comunicado de la institución.

Científicos y expertos de los 38 países miembros de la Comisión Ballenera Internacional (CBI) comenzaron hoy su 46 asamblea anual en Puerto Vallarta, con dos cuestiones fundamentales: la creación de una reserva para las ballenas en la Antártida y el mantenimiento de la moratoria sobre su captura.

El documento del WWF señala que la creación de un refugio en los mares del Sur significaría un complemento de la moratoria internacional vigente sobre la pesca comercial de ballenas, que debe ser ratificada en la reunión de la CBI.

El Fondo Mundial para la Naturaleza advierte que al menos 20 países de los 30 participantes en la reunión pueden votar a favor de la creación de la reserva ballenera en la Antártida, con lo que se lograría el mínimo de las dos terceras partes de los votos para ser ratificado.

Hasta ahora son cinco los países que se han mostrado contrarios a la aprobación del refugio, y desconocen la posición que adoptarán las naciones restantes, por lo que convocan a estas naciones a seguir el ejemplo de los que se han pronunciado en favor de la protección de los cetáceos.

Los principales países que mantienen una postura contraria a la reserva ballenera son Japón y Noruega, que mantienen una fuerte presión económica sobre Santa Lucía, Saint Vincent y Granada, afirma el documento.

El WWF advierte que aquellos países cuyas "intenciones no están definidas eviten manejos sucios y procuren no caer en las componendas de último minuto que han caracterizado a las decisiones del CBI en los últimos años".

La seguridad de las ballenas es muy importante para que se someta a las "tácticas de chequera" del gobierno japonés, afirma la WWF.

La propuesta de crear un refugio ballenero en esa zona la presentó el gobierno francés en 1992 y es apoyada por varios países de la CBI, entre ellos Francia, Alemania, España, Gran Bretaña, Holanda, Estados Unidos, Argentina, Brasil, Nueva Zelanda y Australia.

La institución expresa su preocupación ante la creciente actividad pesquera de la industria ballenera noruega, a pesar de la oposición de la opinión pública internacional y de la moratoria vigente desde 1982.

Una investigación del comité científico de la CBI pone en duda las cifras de las comunidades de ballenas Minke en el Atlántico que han sido esgrimidos por la industria ballenera noruega para justificar la captura de los cetáceos, indica el comunicado.

Este estudio demuestra que Noruega "no tiene ninguna razón valedera para matar una sola ballena, mucho menos las 226 ballenas Minke cazadas y sacrificadas el año pasado", señala.

La organización ecologista Greenpeace denunció la pretensión del gobierno japonés al proponer excluir de la prohibición a la ballena Minke, con lo que pretende contrarrestar el intento de crear una reserva ballenera.



A principios del presente siglo, en la zona austral, donde actualmente se intenta establecer un refugio ballenero, la población de especies de grandes cetáceos fue muy abundante, pero ha sido precisamente en esa región donde se ha producido la mayor caza de grandes ballenas.

Cerca de un millón y medio de ballenas fueron cazadas en la zona austral por las flotas pesqueras de muchos países durante la primera parte de este siglo, lo que originó que la población de grandes cetáceos haya disminuido en un 95 por ciento y que siete de las ocho especies de grandes ballenas se encuentren sobreexplotadas.

Esta protección a las ballenas deberá durar no menos de 50 años, para que se pueda recuperar la población de especies de ballenas casi extinguidas y luego reanudar la caza comercial controlada.

</TEXT>

0094|¿Quién es el director de la CIA?|(EFE19941229-15677;James Woolsey),(EFE19941228-15229;James Woolsey)

<DOCID>EFE19941228-15209</DOCID>

<TITLE> Urgente  
EEUU-CIA

</TITLE>

<TEXT>

Washington, 28 dic (EFE).- El director de la Agencia Central de Información de Estados Unidos (CIA), James Woolsey, presentó hoy, miércoles, su dimisión al presidente Bill Clinton, informó la Casa Blanca.

La inesperada dimisión de Woolsey se produce después de que sendos informes del Congreso criticaran severamente el manejo del caso del espía Aldrich Ames, un veterano de la agencia que espío durante siete años para la URSS.

</TEXT>

0095|¿Qué premio Nobel ganó Solzhenitsin?|(EFE19940721-12710;Literatura)

<DOCID>EFE19940825-12415</DOCID>

<TITLE> RUSIA-DISIDENTE  
SOLZHENITSIN RECIBE PREMIO DE LITERATURA "LEON TOLSTOI"

</TITLE>

<TEXT> Moscú, 25 ago (EFE).- El antiguo disidente soviético Alexandr

Solzhenitsin, que regresó recientemente a su patria tras dos décadas de exilio, fue galardonado con el premio de Literatura "León Tolstoi", que le será entregado en una ceremonia el próximo día 8.

El jurado del premio, que comenzó a otorgarse en 1992, decidió agradecer así a Solzhenitsin su enorme contribución a la Literatura rusa, informó hoy, jueves, la agencia Interfax.

Solzhenitsin, premio Nobel de Literatura, es mundialmente conocido por su obra "Archipiélago GULAG", en la que relató los horrores, sufridos en carne propia, de un campo de trabajos forzados de la Unión Soviética.

Expulsado de la URSS, el disidente pasó la mayor parte de su exilio en Vermont (EEUU), de donde regresó a Rusia en la primavera pasada.

</TEXT>

## Documento de Respuestas

0051 plnast03lms 1 EFE19941125-16022 2 Festival de Cine de Venecia  
 0051 plnaex03lms 2 EFE19941125-16022 2 Festival  
 0051 plnaex03lms 3 EFE19941125-16022 1 Bienal  
  
 0052 plnast03lms 1 EFE19940104-01096 40 Gerry Adams  
 0052 plnaex03lms 2 EFE19940105-01866 40 IRA  
 0052 plnast03lms 3 EFE19940105-01866 17 Irlanda de el Norte  
  
 0053 plnaex03lms 1 EFE19940904-01821 3 Swissair  
 0053 plnaex03lms 2 EFE19941215-08626 3 Sabena  
 0053 plnaex03lms 3 EFE19940607-03862 3 Abasolo  
  
 0054 plnaex03lms 1 EFE19940330-19318 1 diez  
 0054 plnaex03lms 2 NIL  
 0054 plnaex03lms 3 NIL  
  
 0055 plnaex03lms 1 NIL  
 0055 plnaex03lms 2 NIL  
 0055 plnaex03lms 3 NIL  
  
 0056 plnaex03lms 1 EFE19941111-06535 1 Chile  
 0056 plnaex03lms 2 EFE19941111-06535 1 Argentina  
 0056 plnaex03lms 3 EFE19940225-14832 1 Antartida  
  
 0057 plnaex03lms 1 NIL  
 0057 plnaex03lms 2 NIL  
 0057 plnaex03lms 3 NIL  
  
 0058 plnaex03lms 1 NIL  
 0058 plnaex03lms 2 NIL  
 0058 plnaex03lms 3 NIL  
  
 0059 plnaex03lms 1 NIL  
 0059 plnaex03lms 2 NIL  
 0059 plnaex03lms 3 NIL  
  
 0060 plnast03lms 1 EFE19941114-07734 7 Kuala Lumpur  
 0060 plnaex03lms 2 EFE19940122-11393 2 España  
 0060 plnaex03lms 3 EFE19940122-11393 2 COBRA  
  
 0061 plnaex03lms 1 EFE19941125-15752 3 Arafat  
 0061 plnaex03lms 2 EFE19940816-08057 2 Argentina  
 0061 plnaex03lms 3 EFE19940816-08057 2 EFE  
  
 0062 plnaex03lms 1 EFE19940912-06480 3 Ashjabad  
 0062 plnaex03lms 2 EFE19941021-12278 2 Rusia  
 0062 plnaex03lms 3 EFE19941021-12278 2 Uzbekistán  
  
 0063 plnaex03lms 1 EFE19940128-14783 3 OPEP  
 0063 plnaex03lms 2 EFE19940324-15928 2 Colombia  
 0063 plnaex03lms 3 EFE19940128-14783 1 Gobierno  
  
 0064 plnaex03lms 1 EFE19940324-15316 7 dos  
 0064 plnaex03lms 2 EFE19940718-10401 6 tres  
 0064 plnaex03lms 3 EFE19940512-06814 4 nueve  
  
 0065 plnaex03lms 1 NIL

0065 plnaex03lms 2 NIL  
 0065 plnaex03lms 3 NIL  
  
 0066 plnaex03lms 1 NIL  
 0066 plnaex03lms 2 NIL  
 0066 plnaex03lms 3 NIL  
  
 0067 plnaex03lms 1 EFE19940121-10455 2 Beatles  
 0067 plnaex03lms 2 EFE19941025-15078 1 Manhattan  
 0067 plnaex03lms 3 EFE19941025-15078 1 Nureyev  
  
 0068 plnast03lms 1 EFE19940721-12600 2 The Next Generation  
 0068 plnaex03lms 2 EFE19940721-12600 1 NYPD  
 0068 plnast03lms 3 EFE19940721-12600 1 Northen Exposure  
  
 0069 plnast03lms 1 EFE19940128-14834 1 Francesco Cossiga  
 0069 plnaex03lms 2 EFE19940128-14834 1 Pamplona  
 0069 plnast03lms 3 EFE19940128-14834 1 Universidad de Navarra  
  
 0070 plnaex03lms 1 EFE19940720-11505 4 Kyodo  
 0070 plnast03lms 2 EFE19940720-11505 4 Kim Song Ae  
 0070 plnast03lms 3 EFE19940720-11505 4 Radio Pyongyang  
  
 0071 plnast03lms 1 EFE19940416-09096 1 Estado de Chiapas  
 0071 plnaex03lms 2 EFE19940416-09096 1 México  
 0071 plnast03lms 3 EFE19940416-09096 1 Alfredo Harp Helú  
  
 0072 plnaex03lms 1 EFE19940210-05693 1 EFE  
 0072 plnaex03lms 2 EFE19940210-05693 1 Comité  
 0072 plnast03lms 3 EFE19940210-05693 1 Situaciones de Emergencia  
  
 0073 plnaex03lms 1 EFE19940131-16362 20 Lillehammer  
 0073 plnast03lms 2 EFE19940919-10763 9 Juegos Olímpicos de Barcelona  
 0073 plnast03lms 3 EFE19941110-05630 8 Juegos Olímpicos de Atlanta  
  
 0074 plnaex03lms 1 NIL  
 0074 plnaex03lms 2 NIL  
 0074 plnaex03lms 3 NIL  
  
 0075 plnaex03lms 1 NIL  
 0075 plnaex03lms 2 NIL  
 0075 plnaex03lms 3 NIL  
  
 0076 plnaex03lms 1 EFE19940511-06076 3 miércoles  
 0076 plnaex03lms 2 EFE19940929-16574 2 jueves  
 0076 plnaex03lms 3 EFE19940225-15023 1 viernes  
  
 0077 plnast03lms 1 EFE19940512-06920 2 Alfred Hitchcock  
 0077 plnaex03lms 2 EFE19940512-06920 1 Trueba  
 0077 plnast03lms 3 EFE19940512-06920 1 Doña Sofía  
  
 0078 plnaex03lms 1 NIL  
 0078 plnaex03lms 2 NIL  
 0078 plnaex03lms 3 NIL  
  
 0079 plnaex03lms 1 EFE19940512-06697 1 jueves  
 0079 plnaex03lms 2 EFE19940712-07415 1 noviembre  
 0079 plnast03lms 3 EFE19941219-10572 1 26 de diciembre de 1992  
  
 0080 plnaex03lms 1 NIL  
 0080 plnaex03lms 2 NIL

0080 plnaex03lms 3 NIL  
 0081 plnaex03lms 1 NIL  
 0081 plnaex03lms 2 NIL  
 0081 plnaex03lms 3 NIL  
 0082 plnaex03lms 1 EFE19940330-19078 5 Emeraude  
 0082 plnaex03lms 2 EFE19940330-19111 3 Francia  
 0082 plnaex03lms 3 EFE19940330-19111 2 Defensa  
 0083 plnast03lms 1 EFE19941215-08415 43 Borís Yeltsin  
 0083 plnaex03lms 2 EFE19940607-03865 10 Kremlin  
 0083 plnaex03lms 3 EFE19940705-02352 10 Moscú  
 0084 plnast03lms 1 EFE19940210-05835 16 Javier Solana  
 0084 plnaex03lms 2 EFE19940825-12290 15 Gobierno  
 0084 plnaex03lms 3 EFE19940509-04612 10 España  
 0085 plnaex03lms 1 NIL  
 0085 plnaex03lms 2 NIL  
 0085 plnaex03lms 3 NIL  
 0086 plnast03lms 1 EFE19940127-14347 19 Organización para la  
 Liberación de Palestina  
 0086 plnast03lms 2 EFE19940127-14347 17 Yaser Arafat  
 0086 plnaex03lms 3 EFE19940106-02080 16 Israel  
 0087 plnast03lms 1 EFE19940912-06375 1 Academia de Bellas Artes de San  
 Fernando  
 0087 plnast03lms 2 EFE19940912-06375 1 Calcografía Nacional  
 0087 plnast03lms 3 EFE19940912-06375 1 Palacio Real  
 0088 plnaex03lms 1 EFE19940803-01159 5 Pyongyang  
 0088 plnaex03lms 2 EFE19940612-07298 3 Pekín  
 0088 plnaex03lms 3 EFE19940525-14765 2 OEIA  
 0089 plnaex03lms 1 EFE19940502-00816 8 CBI  
 0089 plnast03lms 2 EFE19940524-13903 3 Puerto Vallarta  
 0089 plnaex03lms 3 EFE19940524-13903 3 Antártida  
 0090 plnast03lms 1 EFE19940708-04624 2 Daniel Passarella  
 0090 plnast03lms 2 EFE19940708-04624 2 Alfio Basile  
 0090 plnast03lms 3 EFE19941223-13386 1 Benito Floro  
 0091 plnaex03lms 1 NIL  
 0091 plnaex03lms 2 NIL  
 0091 plnaex03lms 3 NIL  
 0092 plnaex03lms 1 NIL  
 0092 plnaex03lms 2 NIL  
 0092 plnaex03lms 3 NIL  
 0093 plnaex03lms 1 NIL  
 0093 plnaex03lms 2 NIL  
 0093 plnaex03lms 3 NIL  
 0094 plnast03lms 1 EFE19941228-15209 19 James Woolsey  
 0094 plnast03lms 2 EFE19941228-15209 6 Bill Clinton  
 0094 plnast03lms 3 EFE19941229-15677 6 Servicio Central de Información  
 0095 plnast03lms 1 EFE19940825-12415 9 Nobel de Literatura

0095 plnaex031ms 2 EFE19940520-12093 8 Rusia  
0095 plnast031ms 3 EFE19940721-12710 6 Alexandr Solzhenitsin

0096 plnaex031ms 1 EFE19940210-05701 10 Lillehammer  
0096 plnast031ms 2 EFE19940316-10557 10 Juegos Olímpicos de Invierno  
0096 plnaex031ms 3 EFE19941105-02928 6 COI

0097 plnaex031ms 1 NIL  
0097 plnaex031ms 2 NIL  
0097 plnaex031ms 3 NIL

0098 plnaex031ms 1 EFE19941215-08754 3 China  
0098 plnaex031ms 2 EFE19941209-04670 1 EFE  
0098 plnast031ms 3 EFE19941209-04670 1 Ministerio de Seguridad Pública

0099 plnaex031ms 1 EFE19940131-16194 3 1954  
0099 plnast031ms 2 EFE19940803-01099 2 un millón y medio  
0099 plnaex031ms 3 EFE19940131-16194 2 000

0100 plnaex031ms 1 NIL  
0100 plnaex031ms 2 NIL  
0100 plnaex031ms 3 NIL

## Preguntas erróneas. ¿por qué no funcionan?

**0055|¿En qué tipo de procesador se descubrió un error en la unidad aritmética?|(EFE19941209-04635;Pentium)**

Esta pregunta no puede ser respondida actualmente debido al algoritmo utilizado. Tal como está implementado, buscaría “tipo + procesador + error + unidad”. Si buscamos dentro del texto, vemos que sólo la palabra error se encuentra en ese texto. Para poder solucionarlo y que se respondiese a esta pregunta, habría quizás que buscar palabras relacionadas (chips, computadoras, semiconductores, ordenadores...), puede que funcione la opción de utilizar sinónimos.

**0061|¿Cuál es la capital de Irán?|(EFE19941108-04093;Teherán),(EFE19940816-08057;Teherán)**

Esta pregunta no puede ser respondida actualmente debido al algoritmo utilizado. Tal como está implementado, buscaría “capital + Irán”. Si buscamos dentro del texto, vemos que se utiliza la expresión “capital iraní”, en el otro texto apenas hay 1 sola referencia a la ciudad, que se indica de manera que es imposible averiguar si es o no la capital. Para poder solucionarlo y que se respondiese a esta pregunta, habría quizás que buscar topónimos relacionados con el nombre propio utilizado.

**0063|¿Cuál es el principal país productor de petróleo en el mundo?|(EFE19941115-08606;Arabia Saudí),(EFE19940128-14783;Rusia)**

Esta pregunta no puede ser respondida actualmente debido al algoritmo utilizado. Busca “país + productor + petróleo + mundo”, sin embargo en el primer texto viene la frase “principal productor de la Organización de Países Exportadores de Petróleo (OPEP)”, con lo que interpreta que es OPEP. Lógicamente esta es la palabra que más veces se repite en los textos, y es por eso que sale como primera opción. El problema de esta pregunta está relacionado con el algoritmo de selección de respuestas, que da más importancia al número de veces que aparece una palabra, quizás para solucionarlo habría que cambiar ese algoritmo por otro que diese más importancia a la cercanía de las palabras (aunque eso provocaría que fallasen otras preguntas).

**0064|¿Cuántos países son miembros de la Unión Europea?|(EFE19941218-10091;Doce)**

En el documento viene claramente “Los doce países miembros de la Unión Europea”. Creemos que no se responde porque “Los doce” se interpreta como un nombre, mientras

que la pregunta se refiere a un numeral, es por eso por lo que no lo identifica. Desgraciadamente desconocemos cómo se podría solucionar este problema.

**0065|¿Cuándo se firmo el Acta Única Europea?|(EFE19941223-13130;En 1986)**

Esta pregunta no puede ser respondida actualmente debido al algoritmo utilizado. Interpreta Acta Única Europea como una palabra, mientras que en el texto, viene sólo Acta Única. Por eso no localiza ninguna respuesta. La solución pasaría por evitar que realice ese tipo de interpretaciones y tomase cada palabra por separado, pero el hecho de no utilizar adjetivos, provocaría que buscase Acta solamente. Desconocemos si podría funcionar este método, pero sí que podría afectar a muchas otras preguntas.

**0066|¿Qué cargo ostentaba Rabbani al estallar la guerra civil de Afganistán en 1992?|(EFE19940104-01122;Presidente de la república)**

Tampoco devuelve respuesta, debido a que busca “cargo + Rabbani + guerra + Afganistán”. El problema es similar a otros comentados, la palabra cargo no viene indicada en los textos por lo que una solución sería la de obligar a que busque tan sólo los nombres propios: Rabbani y Afganistán, y una vez localizados los documentos, buscar el resto de palabras (cargo, guerra) mediante un OR, en lugar de un AND. Cuando se realizó la implementación de buscar sólo nombres propios, las respuestas dadas no eran correctas, debido al algoritmo de respuestas utilizado. Si quisiésemos que esta pregunta funcionase, deberíamos cambiarlo también. Tantos cambios para una sola pregunta que además perjudicaría a otras no compensa.

**0068|¿Quién escribió "Star Trek"?|(NIL)**

Esta pregunta, tal como está definido nuestro algoritmo, es imposible de no contestar. No porque seamos frikis (o en su caso trekkies), sino porque basta con que haya un solo texto que hable de ello, como para que devuelva alguna respuesta. En nuestro caso, como sólo busca StarTrek y solicita un nombre propio, lo que hace es devolver el nombre propio de esos textos que más veces se repita. Para solucionarlo, será necesario incluir los verbos, de modo que obligase a buscar “StarTrek + escribir”.

**0069|¿Quién es el presidente de la República de Italia?|(EFE19940520-12291;Oscar Luigi Scalfaro),(EFE19940131-16267;Oscar Luigi Scalfaro)**

Esta pregunta tampoco devuelve respuesta, debido a que “República de Italia” no viene en los textos, aunque venga bien claro “presidente italiano”. La solución pasaría por utilizar topónimos.

**0070|¿Quién ostenta el poder en Pyongyang?|(EFE19940720-11505;Kim Il Sung),(EFE19941005-02810;Kim Jong Il),(EFE19940518-10478;Kim Il Sung),(EFE19941019-10732;Kim Jong Il)**

Esta pregunta se responde erróneamente, devolviendo como resultados principales la mujer del anterior dirigente de Corea del Norte, la radio del país y la agencia de noticias japonesa. Esto es debido al algoritmo de respuestas, que premia a las más utilizadas, en lugar de las más cercanas. Cambiando el algoritmo de respuestas quizás se podría contestar.

**0071|¿Qué significan las siglas ETA?|(EFE19940418-10306;Euskadi Ta Askatasuna (Euskadi y Libertad))**

Debido al algoritmo de respuestas, responde erróneamente. La solución estaría en utilizar el verbo significar junto con las claves.

**0075|¿Cuál es la profesión de Renzo Piano?|(EFE19940723-13695;Arquitecto)**

El problema una vez más, está en una de las palabras: “profesión”, que no se encuentra en el documento. La solución más simple pasaría por utilizar sólo el nombre propio. Cuando se implementó esta opción, las respuestas eran incorrectas, debido al algoritmo utilizado.

**0076|¿En qué año se creó el Fondo Monetario Internacional?|(EFE19940509-05165;En 1944)**

El problema está en que busca una fecha para el FMI, usando sólo como clave FMI, por lo que devuelve como resultados las 3 palabras más usadas, que son días de la semana. El hecho de no poder diferenciar las fechas entre sí (meses, años, días,...) es lo que provoca este error. El FreeLing no permite especificar tipos de fecha, por lo que conseguir la respuesta a esta pregunta sería difícil.

**0078|¿Cuántas personas murieron en el juzgado de Euskirchen?|(EFE19940309-06041;Siete)**

Manda buscar “personas + juzgado + Euskirchen”, lógicamente el problema está en la segunda palabra que no se encuentra en el documento. Solucionar ese problema pasaría por utilizar sólo el nombre propio, y al aplicar el cambio, la solución correcta apareció en 2º lugar, pero con el mismo número de resultados que el resto. Si accedemos al documento original, vemos numerosas referencias a distintas cifras que podrían



provocar confusión, ya que sólo se menciona la solución en el titular, por lo que se encontró la solución simplemente por suerte.

**0079|¿Cuándo se fundó la CEE?|(EFE19940228-16727;El 25 de marzo de 1957)**

Al igual que ocurrió con la pregunta 0076, devuelve fechas, pero no son las correctas. La solución, al igual que antes pasaría por filtrar los tipos de fecha, algo de lo que no es capaz el FreeLing. Incluir el verbo podría servir para mejorar la búsqueda.

**0080|¿En qué ciudad europea está la Torre Eiffel?|(EFE19941105-02620;En París),(EFE19940127-14054;París)**

Tampoco responde nada, por culpa de la búsqueda “ciudad + Torre Eiffel”, el problema está en ciudad, que no aparece en ninguno de los dos documentos. La solución puede ser la de utilizar sólo el nombre propio. Cuando se implementó este método, la pregunta se respondió correctamente.

**0081|¿A qué país pertenece el agente inmobiliario Schneider?|(EFE19940416-09172;A Alemania),(EFE19940421-12070;Alemania)**

Igual que la pregunta anterior, las palabras que provocan el fallo son 2 en esta ocasión: país y agente. La solución: utilizar sólo el nombre propio. De todas maneras cuando se implementó, las respuestas eran incorrectas, debido al algoritmo de respuestas.

**0084|¿Quién es el presidente italiano de Asuntos Exteriores?|(EFE19940524-14347;Antonio Martino),(EFE19940204-02527;Beniamino Andreatta)**

El error en esta pregunta es un poco más complejo de lo normal. La búsqueda sería de “presidente + Asuntos Exteriores”, pero lo cierto es que el puesto es ministro, por lo que los textos donde viene, pasan desapercibidos. Devuelve como resultados Javier Solana, Gobierno y España, es decir que salta de Italia a España, simplemente por culpa de la palabra presidente (en Italia son todos ministros, incluido el jefe de estado, que es primer ministro). La solución quizás sirviese simplemente con buscar el nombre propio, pero podría funcionar también la idea de usar sinónimos de presidente.

**0085|¿Cuál es el nombre de pila de la mujer de Nelson Mandela?|(EFE19941218-10151;Winnie)**

Busca “nombre de pila + mujer + Nelson Mandela” y fallan las 2 primeras palabras. Una solución válida podría ser la de usar sólo nombres propios, o bien la de sinónimos (mujer=esposa). La opción de nombres propios tampoco funciona, debido al algoritmo de resultados.

**0087|¿En qué ciudad está el Museo del Prado?|(EFE19940622-13700;En Madrid)**

Pregunta bastante difícil de responder, ya que tan sólo se hace referencia al final del documento sobre la ciudad, por lo que es comprensible que no esté entre las 3 respuestas principales. La idea de utilizar el algoritmo de respuestas por aproximación es posible que tampoco funcionase en este caso.

**0090|¿Quién es el entrenador del equipo nacional de fútbol noruego?|(EFE19940619-12097;Egil Olsen)**

Falla porque busca “entrenador del equipo + futbol”, por lo que ninguna de las respuestas dadas, no es la válida. La solución pasaría por los adjetivos (noruego), que en este caso simplificarían la búsqueda.

**0092|¿Qué país de África ha adoptado una nueva constitución?|(EFE19940422-12669;Sudáfrica),(EFE19941215-08330;República Centroafricana)**

La búsqueda es “país + África + constitución”, y una vez más, la palabra país es la que provoca el fallo. La búsqueda del nombre propio no funcionaría en este caso, ya que buscaría cualquier resultado relacionado con África. Puede que el uso de adjetivos también sirviese para esta pregunta.

**0093|¿Cuáles son las siglas del Fondo Mundial para la Protección de la Naturaleza?|(EFE19940908-04330;WWF)**

Caso importante, ya que la primera de las versiones funcionó, pero a partir de la segunda no volvió a hacerlo. Busca “siglas + Fondo Mundial para la Protección de la Naturaleza”, la palabra “siglas” es la que provoca el error. Una búsqueda sólo del nombre propio solucionaría el problema. Cuando se implementó esta solución se encontró la respuesta en 3º lugar, pero con el mismo número de resultados que las 2 primeras, es decir, que lo encontró por pura suerte.

**0097|¿Cuándo tomará China la posesión de Hong Kong?|(EFE19940426-14917;En 1997)**

El problema está en la palabra posesión, que no se encuentra dentro del documento. La opción más útil puede que sea la de utilizar nombres propios, ya que en este caso al tener 2, puede que la búsqueda se reduzca (el hecho de buscar una fecha, también podría ayudar a que funcionase). Al implementarlo, no se encontró una respuesta válida, ya que fallaba por culpa del FreeLing, que no distingue los tipos de fecha (devolvía días de la semana).

**0098|¿Qué causó el incendio en un cine en la ciudad china de Karamai?|(EFE19941209-04685;Un cortocircuito)**

Esta pregunta no puede ser respondida actualmente debido al algoritmo utilizado. Tal como está implementado, buscaría “incendio + cine + ciudad + Karamai”. Si buscamos dentro del texto, vemos que la palabra ciudad no se menciona, en su lugar utiliza localidad. Para poder solucionarlo y que se respondiese a esta pregunta, habría quizás que buscar palabras sinónimas (ciudad=localidad). Personalmente la opción de utilizar sinónimos es una locura desde el punto de vista computacional, quizás una opción más válida sería la de obligar a que busque tan sólo el nombre propio Karamai. Al implementar esta opción, las respuestas seguían siendo incorrectas.

**0099|¿Cuántos habitantes hay en Moscú?|(EFE19940708-05108;Once millones),(EFE19940708-05108;once millones)**

El problema que tenemos en esta pregunta es que interpreta una fecha (1954) como un numeral, y también el 000, debido a la desaparición de los puntos (5.000 pasa a ser 5 000). Es por eso que 2 respuestas de esta pregunta son erróneas y no deberían estar. No se encuentra la respuesta simplemente porque no es la más mencionada. Cambiar el algoritmo de resultados podría funcionar.

**0100|¿En qué mes se produjo el naufragio del Estonia?|(EFE19941110-05743;En septiembre)**

La pregunta es “mes + naufragio + Estonia”, lo que provoca 0 resultados al no venir la palabra “mes”. Buscar el nombre propio tampoco sirvió aquí, al devolver resultados incorrectos.

## Mejoras: Detalles

### **MEJORA DE LOS ADJETIVOS. MEJORA DESCARTADA:**

Después de implementar las búsquedas sólo con claves formadas por sustantivos, lo siguiente que se intentó fue utilizar adjetivos, con el fin de reducir el tiempo de búsqueda y mejorar los resultados. La idea era que aquellas preguntas que no daban respuesta, seguirían sin tenerla, pero con este sistema conseguiríamos mejorar aquellas preguntas que daban respuestas incorrectas, o en 2º y 3º posición.

El problema surgió con el resto de preguntas. La mayoría de las preguntas correctas, pasaron a obtener 0 resultados, debido a que los adjetivos reducían tanto las posibilidades que no encontraban respuestas.

Esto se debía a que utilizando los adjetivos, se seleccionaban sólo aquellos documentos que contuviesen todos los nombres y todos los adjetivos, y dentro de cada documento, se buscaba las frases que contuviesen también todo. Y es aquí donde el programa cometía la mayoría de los errores, ya que el conseguir que hubiese nombres y adjetivos en la misma frase ya no era tan frecuente y por lo tanto se cometían tantos fallos.

Es por eso que en un principio, utilizar los adjetivos ha quedado descartado. Quizás en otra ocasión.

### **MEJORA DE LAS BUSQUEDAS REDUCIDAS CON NOMBRES PROPIOS. MEJORA DESCARTADA:**

Una de las soluciones más comentadas para los posibles problemas que teníamos era la de buscar sólo con los nombres propios. Esta opción serviría para aquellas preguntas que tenían alguna clave “eliminatória”, que impedía encontrar respuestas.

A la hora de implementarlo se consiguió el primer objetivo: aquellas preguntas sin respuesta pasaban a tener respuestas. El problema es que tenía 2 inconvenientes: El tiempo de búsqueda era demasiado alto (algo que en un principio no debería ser un problema importante) y las soluciones dadas no eran las correctas (al desaparecer el resto de claves de búsqueda, solo tenía información sobre la clave principal y tipo de respuesta buscada, con lo que el número de aciertos era realmente bajo).

Es por esto que en un principio utilizar este método también queda descartado, ya que no ofrece demasiados resultados positivos.

**MEJORA DE MAYOR NIVEL DE ESPECIFICACIÓN. MEJORA DESCARTADA:**

Esta posible mejora tuvo que ser descartada por culpa de la aplicación utilizada FreeLing, ya que el nivel de especificación máximo que pudimos lograr sin cometer demasiados errores fue con los nombres propios y nombres comunes. Las ideas de utilizar topónimos o de especificar los tipos de fechas no hemos podido utilizarlas debido a este problema.

**MEJORA DE SINÓNIMOS. MEJORA DESCARTADA:**

Para poder utilizar esta mejora, era necesario contar con algún programa auxiliar que a partir de un nombre, buscara los sinónimos posibles. Por desgracia no hemos encontrado ningún programa de sinónimos en español.

**CAMBIO DE ALGORITMO DE RESPUESTAS. MEJORA DESCARTADA:**

En vista de los numerosos problemas que había con las preguntas sin contestar con motivo del algoritmo de respuestas utilizado, se intentó otro método. El que llegamos a implementar daba mayor peso a la distancia entre las claves y las respuestas.

Para realizar los cálculos, utilizábamos la suma de los cuadrados de la distancia de esa palabra y cada una de las claves. En un principio da mayor importancia a la cercanía de las palabras a las claves, pero a la hora de analizar los resultados, el número de preguntas correctamente respondidas era prácticamente igual que con el algoritmo actual (con el algoritmo de cercanía obteníamos muchas más respuestas en segundo y tercer lugar mientras que con el de número de respuestas estaban en primer lugar).

En cuanto a la decisión final, optamos por seguir utilizando el de número de respuestas, ya que en conjunto apenas hubo variaciones en el número de resultados.

**MEJORA DE VERBOS. MEJORA DESCARTADA:**

Para poder utilizar esta mejora, era necesario tener la raíz del verbo en el Lucene. No hemos desarrollado esta idea debido a su complejidad.

Para poder hacerlo, tendríamos que guardar las raíces de los verbos con lo que tendríamos que analizar cada documento en tiempo real y el tiempo de computación para miles de documentos sería demasiado largo.

