

ACTUACIÓN DEL USUARIO EN UN SISTEMA INTERACTIVO MULTILINGÜE DE BÚSQUEDA DE RESPUESTAS

Ángel F. ZAZO RODRÍGUEZ, afzazo@usal.es
José Luis ALONSO BERROCAL, berrocal@usal.es
Carlos G. FIGUEROLA, figue@usal.es

Universidad de Salamanca (Salamanca, España). Departamento de Informática y Automática
Grupo de Investigación en Recuperación de Información Avanzada (REINA)

Resumen

Se describe una investigación experimental llevada a cabo para evaluar la actuación del usuario en un sistema interactivo multilingüe de búsqueda de respuestas. La búsqueda de respuestas es una de las facetas más avanzadas en recuperación de la información, que quiere hallar soluciones precisas a necesidades de información concretas. Se persigue encontrar un fragmento de texto mínimo que responda a la pregunta, empleando para ello una extensa colección de documentos. La tarea se complica si se utilizan documentos escritos en un idioma del que el usuario sólo tiene un conocimiento pasivo. En esta situación, la interacción con el mismo puede mejorar los resultados del sistema. El objetivo final es desarrollar mejores sistemas para ayudar al usuario a resolver sus necesidades informativas. Se han explorado dos formas de interacción. Por una parte, haciendo que el sistema sugiera términos relacionados con su búsqueda, para ser incorporados a la pregunta en el idioma de los documentos. Por otra, permitiendo que el usuario visualice no sólo pasajes de texto, sino también, si lo desea, los documentos completos. Los resultados muestran que la posibilidad de añadir términos relacionados fue poco valorada por los usuarios, que la consideraron poco útil. Sin embargo, la posibilidad de obtener y visualizar los documentos completos, partiendo de los pasajes recuperados, mejoró el número de aciertos, así como el tiempo requerido para conseguir una respuesta correcta.

Palabras clave

búsqueda de respuestas, traducción multilingüe de consultas, interactividad en el web, multilingüismo en la información, uso de la información

1 INTRODUCCIÓN

Una de las tareas más avanzadas en recuperación de información (RI) es aquella que busca, no documentos relevantes ante una necesidad informativa dada, sino respuestas precisas a necesidades de información concretas, empleando una gran colección de documentos. Esto es lo que se

conoce como búsqueda de respuestas, *question answering* en inglés, cuyas siglas, QA, suelen utilizarse sin traducir. Los sistemas de QA deben extraer la respuesta después de localizar la zona de los documentos que pueda estar relacionada con la afirmación: «Se busca encontrar un fragmento de texto mínimo que responda la pregunta». Por eso, algunas aproximaciones a este problema están basadas en la recuperación de pasajes o fragmentos de texto (VICEDO 2003; WOODS 2000). La respuesta puede ser extraída después mediante algún proceso automático o, si el sistema no es capaz de proporcionar una respuesta válida de una manera totalmente autónoma, debe disponerse algún proceso de realimentación con el usuario, para que pueda satisfacer su necesidad informativa. La interacción con el usuario ayuda al sistema de QA a encontrar mejores respuestas. A su vez, el sistema de QA ayuda al usuario a encontrar la respuesta más rápidamente.

El proceso interactivo de QA posee hoy día un amplio interés, debido al aumento de la web y a la necesidad cada vez más urgente de obtener información precisa. A ello unimos los avances en RI y en el procesamiento del lenguaje natural (PLN), que han hecho resurgir de nuevo el interés en los sistemas de QA claramente enfocados a la web. La enorme cantidad de información en la web hace de QA una poderosa herramienta para encontrar información ante una necesidad informativa dada. Ahora bien, en contraste con los motores de búsqueda en internet, los sistemas de QA deben proporcionar respuestas breves y concisas.

Un problema adicional de la búsqueda de respuestas en la web es el entorno multilingüe de este medio. Efectivamente, los usuarios que buscan en la web normalmente encuentran documentos en varios idiomas, habitualmente en el idioma en el que se realiza la consulta, y también en el idioma más utilizado en internet, el inglés. Por este motivo, para resolver sus necesidades informativas, el usuario debe tener al menos un conocimiento pasivo de esta lengua, así como algún tipo de herramienta lingüística, como diccionarios bilingües o programas de traducción automática, que le ayuden a traducir su pregunta al inglés o los documentos recuperados a su propio idioma. Muchas de esas herramientas están disponibles gratuitamente en internet.

En esta comunicación, nuestro trabajo se ha concretado en medir cómo los usuarios interactúan con nuestro sistema multilingüe de búsqueda de respuestas. Con él se pretende ayudar al usuario en la localización de respuestas en pasajes de texto escritos en un idioma, el inglés, diferente al utilizado para plasmar la necesidad informativa, el español. Se han explorado los efectos de dos formas de interacción con el usuario. Por un lado, haciendo que el sistema le sugiera un conjunto de términos traducidos al idioma de los documentos. Por otro, permitiendo que el sistema trabaje no sólo con pasajes de texto, sino, a petición del usuario, también con los documentos completos. En ambos casos, el objetivo es evaluar no sólo el número de respuestas correctas, sino también la valoración subjetiva que el usuario hace de esta clase de ayudas.

Cabe señalar que el trabajo que aquí presentamos es un complemento al que nuestro grupo de investigación llevó a cabo en la conferencia

CLEF¹ (Cross-Language Evaluation Forum) del año 2004 (FIGUEROLA 2004), en particular en la tarea interactiva de búsqueda de respuestas (iCLEF). Estas conferencias se desarrollan actualmente bajo el programa *Information Society Technologies* (IST) de la Comisión Europea y son un foro donde diversos grupos de investigación que evalúan sus sistemas de recuperación de información multilingüe. Las conferencias CLEF se desarrollan en colaboración con las conocidas conferencias TREC² (*Text REtrieval Conferences*).

2 EL EXPERIMENTO

El experimento siguió las indicaciones de la organización de la tarea iCLEF³ (GONZALO 2004), que seguidamente se resumen. Se dispuso de una colección de dieciséis preguntas en español, elegidas de tal modo que tuviesen respuesta en la colección documental en inglés, proporcionada por la organización y formada por documentos provenientes de noticias de los periódicos *Los Angeles Times* y *Glasgow herald*, de los años 1994 y 1995, respectivamente. Nos referiremos a estas colecciones como LA94 y GH95. Este tipo de documentos condiciona bastante cualquier actuación, dado que en general son documentos muy cortos y monotemáticos. La tipología de las preguntas se restringió a buscar nombres de personas y de organizaciones, fechas, o cantidades medibles. Para resolver una pregunta, se disponía de un tiempo límite de cinco minutos. Las preguntas debían resolverse por uno o más grupos de ocho usuarios.

El objetivo del experimento era contrastar dos formas diferentes de interacción, para lo cual los grupos participantes debían disponer de dos sistemas. Cada usuario resolvía la mitad de las preguntas en un sistema, y la otra mitad en el otro. Todos los usuarios eran instruidos previamente en ambos sistemas, disponiéndose para ello de dos preguntas de prueba en cada uno de ellos. Los usuarios rellenaron un cuestionario previo al experimento, otro al finalizar la sesión de preguntas con cada sistema y uno más al final. Para reducir el efecto del tipo de pregunta, su orden y el sistema utilizado, cada usuario debía seguir un esquema diferente (HERSH 2000; OARD 2003). Finalmente, cada respuesta correcta se clasificaba en dos categorías, dependiendo de si coincidía con una respuesta prefijada de antemano (respuesta «exacta») o bien no coincidía (respuesta «relajada»).

2.1 Nuestros sistemas

El primer sistema, Sistema A, consistió en presentar al usuario pasajes de texto en los que posiblemente se encontrase la respuesta a la pregunta, sin más información adicional (es decir, sin incluir el contexto con que

1. <http://clef.iei.pi.cnr.it/>

2. <http://trec.nist.gov/>

3. <http://nlp.uned.es/iCLEF>

TABLA 1. Preguntas de la tarea i CLEF

N.	Pregunta
1	¿En qué año le fue concedido el premio Nobel a Thomas Mann?
2	¿Cuántos genes humanos hay?
3	¿Quién es el ministro alemán de economía?
4	¿Quién cometió el atentado en el metro de Tokio?
5	¿Cuánto costó el túnel del Canal de la Mancha?
6	¿Cuándo ganó Letonia la independencia?
7	¿Cuántos desaparecidos causó en Filipinas el tifón Ángela?
8	¿Quién es el director gerente del Fondo Monetario Internacional?
9	¿Cuándo murió Lenin?
10	¿Cuántos muertos por asfixia hubo en el metro de Bakú?
11	¿Cómo se llama el presidente de Burundi?
12	¿De qué grupo político es presidente Charles Millon?
13	¿De qué club es entrenador Bobby Robson?
14	¿Cuándo ha tenido lugar el atentado en la estación de metro de Saint-Michel en París?
15	¿Cuánta gente vive en Bombay?
16	¿Quién ganó el Premio Nobel de Literatura en 1994?

ayudarse). El segundo sistema, Sistema B, además de mostrar estos pasajes, incluía la posibilidad de, a petición del usuario, ver el documento completo para ayudarse en la búsqueda de la respuesta. En ambos casos, el número de pasajes de texto mostrado fue 30. Se eligió este número para que fuese superior a lo habitual en la actuación de los usuarios que utilizan motores de búsqueda en internet, que no ven más allá de dos pantallas de resultados (es decir, 20 documentos) (SPINK 2001).

En ambos sistemas las preguntas se lanzaban en español y los pasajes de texto se mostraban en inglés. Si el usuario no encontraba la respuesta a la pregunta, podía replantearla mediante un mecanismo de sugerencia de términos (en inglés), que podía incorporar a su gusto. Este mecanismo mostraba términos relacionados con los de la pregunta original en inglés y estaba basado en una técnica automática de expansión de términos. La razón de haber incluido este mecanismo de sugerencia de términos fue comprobar si el usuario se comportaba de forma similar a como lo hacía en otros sistemas de búsqueda sobre colecciones documentales bien definidas, como catálogos de bibliotecas (JONES 2000) o servicios de información en línea (ZAZO 2004).

En nuestro experimento no hemos empleado un típico sistema de QA (VICEDO 2003). Tampoco hemos adaptado uno a la tarea interactiva. Por el contrario, utilizamos un sistema convencional de recuperación basado en el modelo vectorial (SALTON 1968), pero con algunas adaptaciones. El proceso fue el siguiente (véase la Figura 1):

1. Traducir la pregunta al idioma de los documentos por medio de un programa de traducción automática.
2. Llevar a cabo una recuperación convencional a partir de la pregunta traducida. Se utilizaron pasajes de texto, no documentos completos.
3. Presentación de los pasajes recuperados al usuario, en los cuales podría encontrar la respuesta a la pregunta. Para el sistema B, se podía ver el documento completo.

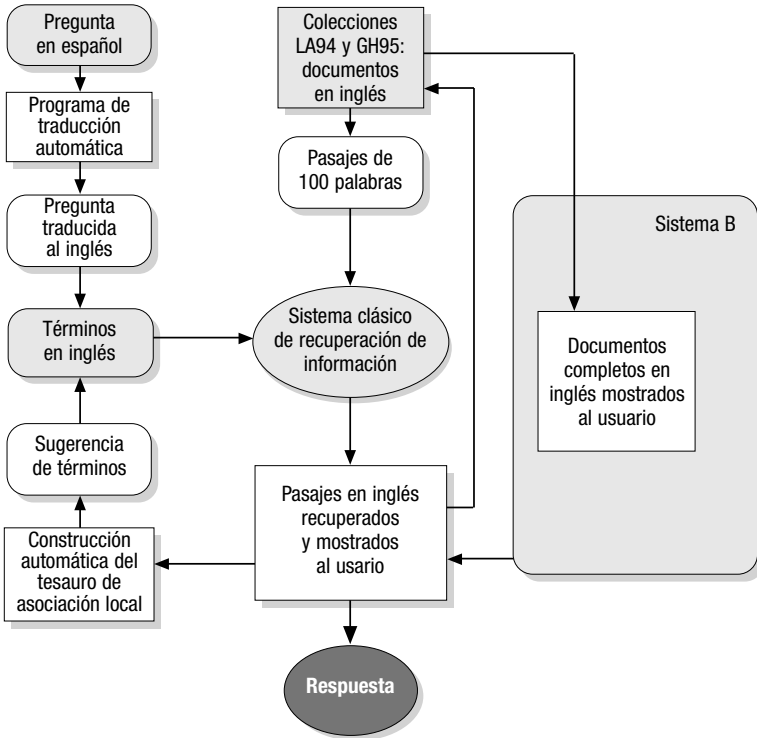


FIGURA 1. Esquema del sistema interactivo multilingüe de búsqueda de respuestas

El sistema de recuperación ha sido el mismo que hemos utilizado en anteriores participaciones de nuestro grupo de investigación en las conferencias CLEF (ZAZO 2002). Se dotó este sistema de algunas facilidades adicionales para la tarea interactiva: traducción de las preguntas, división de los documentos en pasajes de texto, sugerencia de términos para la expansión de la pregunta y, por supuesto, una interfaz de usuario adecuada. A continuación, se detallan estas facilidades.

— *Traducción en línea.* Las pruebas realizadas con gran número de traductores automáticos disponibles gratuitamente en internet reflejan que la calidad de sus traducciones deja mucho que desear. Por este motivo, elegimos el que consideramos posee el menor índice de error en los idiomas origen y destino de nuestro experimento. Se trata del

programa de traducción *Reverso*, con licencia de Softissimo.⁴ Este programa tiene la característica adicional de incluir traducciones alternativas de términos polisémicos, lo cual sin duda mejora los efectos de expansión de las consultas sobre el sistema de recuperación.

- *División de los documentos en pasajes de texto.* La unidad básica de indexación no ha sido el documento completo, sino pasajes de texto formados por ventanas de 100 palabras, incluidas las palabras vacías. A la lista de palabras vacías de SMART (SALTON 1971), se le han añadido las palabras que aparecen en más del 15% de los documentos. Dada esta división en pasajes, hay que considerar la situación habitual de encontrar al principio de los mismos frases ya iniciadas y al final frases inacabadas, que pueden oscurecer el sentido de las mismas. El promedio final de palabras por pasaje es menor que 100 debido a los fragmentos finales de los documentos. El motor de recuperación que hemos utilizado ha trabajado con los pasajes de texto y la expansión de términos también ha considerado estos pasajes. En la Tabla 2 pueden verse las características de la colección documental.

TABLA 2. Características de la colección documental

	LA94 & GH95	Pasajes
Nº de documentos	169.477	915.283
Nº de términos índice	302.241	302.241
Longitud media (incluidas palabras vacías)	229,94	42,71
Nº medio de términos índice	163,77	38,01

- *Expansión de términos.* La expansión de términos es un mecanismo ampliamente utilizado en los sistemas de QA (MAGNINI 2002), sobre todo si tenemos en cuenta que las preguntas suelen contener muy pocos términos. También son muy usados en experimentos interactivos (BELKIN 2001). Nuestro grupo de investigación posee una gran experiencia en expansión automática de consultas aplicada a tareas clásicas de RI (ZAZO 2003). Por ello, hemos utilizado la técnica de expansión que ofrece los mejores resultados con bajo coste computacional: la utilización de tesauros de asociación local. Las relaciones de coocurrencias de los términos en los primeros pasajes recuperados se utiliza para construir el tesoro, en base a la hipótesis de asociación (RIJSBERGEN 1979, p. 104). Utilizando esta técnica de expansión se han obtenido los mejores términos relacionados con todos los términos de la pregunta original, pero la expansión no se ha aplicado automáticamente al sistema. Para cada pregunta, se obtuvieron los mejores 30 términos relacionados, y se presentaron al usuario para que los incorporase, si lo deseaba, a la pregunta original (Figura 2). Es muy importante señalar que la única manera de reformular la pregunta fue dada mediante este mecanismo. Es decir, el usuario no tenía libertad para

4. Puede probarse gratuitamente en la página web <http://www.reverso.com/>

utilizar los términos que quisiera; por el contrario, la reformulación de la pregunta sólo podía hacerse añadiendo o quitando términos de una lista.

- *Interfaz de usuario.* Dado que el proceso interactivo con el usuario ayuda al motor de QA a encontrar mejores respuestas, el diseño de la interfaz de usuario es un aspecto muy importante. Este se ha desarrollado con típicos formularios de páginas web, puesto que este entorno es muy conocido por los usuarios que han realizado los experimentos. La característica más importante es la posibilidad de efectuar varias iteraciones (reformular la pregunta traducida, examinar los pasajes recuperados, etc.), así como, en el segundo experimento, la posibilidad de obtener y leer el documento completo a partir de un pasaje recuperado. Cuando el usuario encontraba la respuesta a la pregunta, debía marcar el pasaje correspondiente y escribir la respuesta la final de la página (véanse las Figuras 3, 4 y 5). Otro aspecto destacado de la interfaz de usuario era que mostraba resaltados en color rojo aquellos términos no vacíos de la pregunta que aparecían en los pasajes de texto o en el documento completo. De esta manera, el usuario podía fácilmente localizar la respuesta a las preguntas.

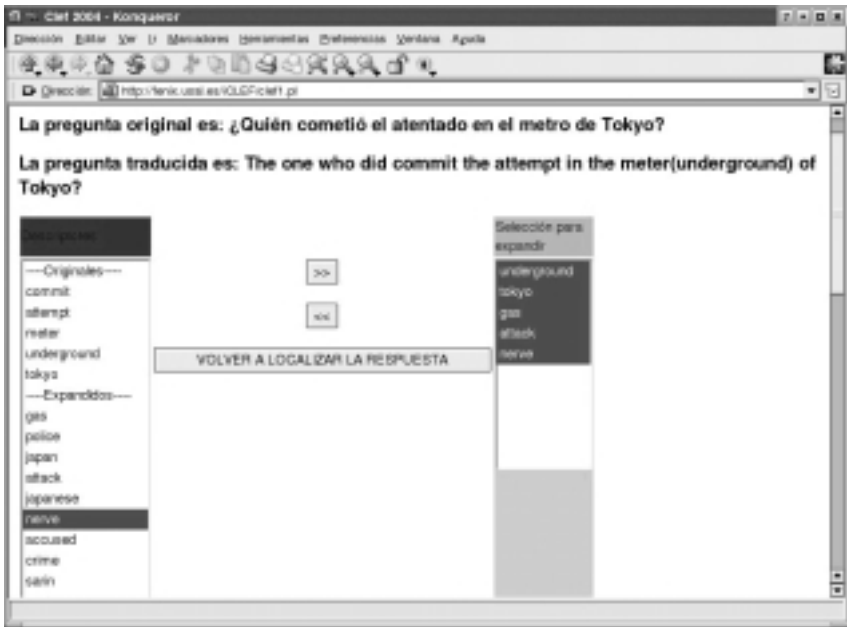


FIGURA 2. Mecanismo de sugerencia de términos

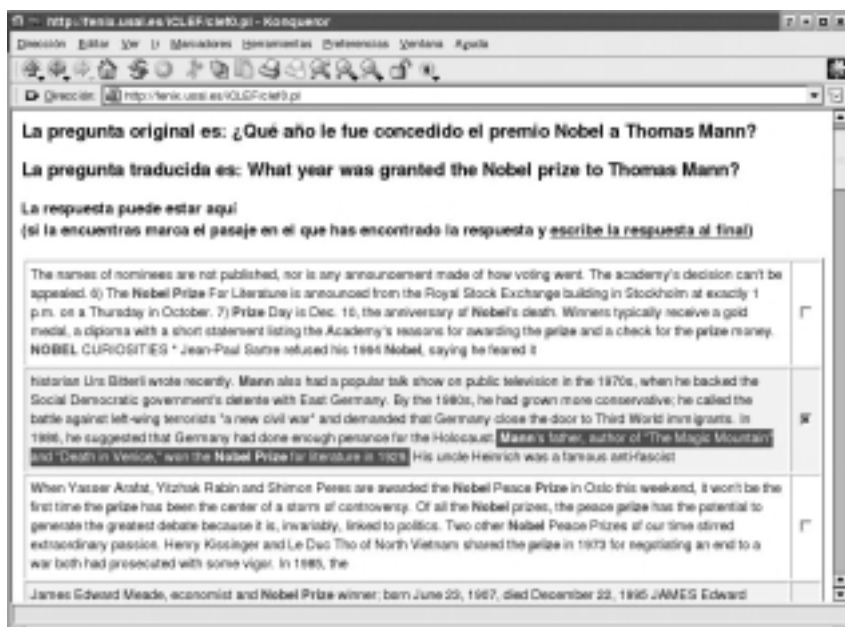


FIGURA 3. Presentación de pasajes de texto al usuario en el sistema A

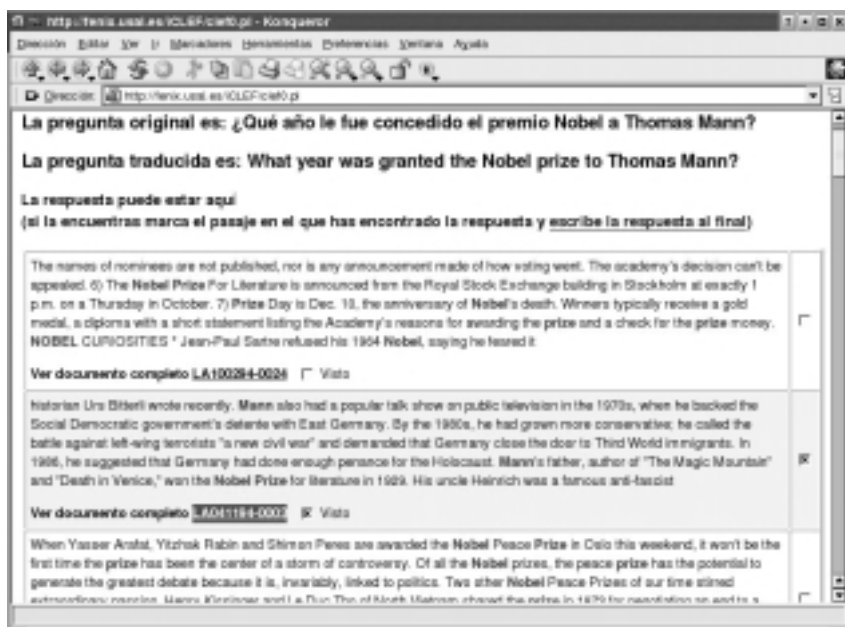


FIGURA 4. Presentación de pasajes de texto al usuario en el sistema B

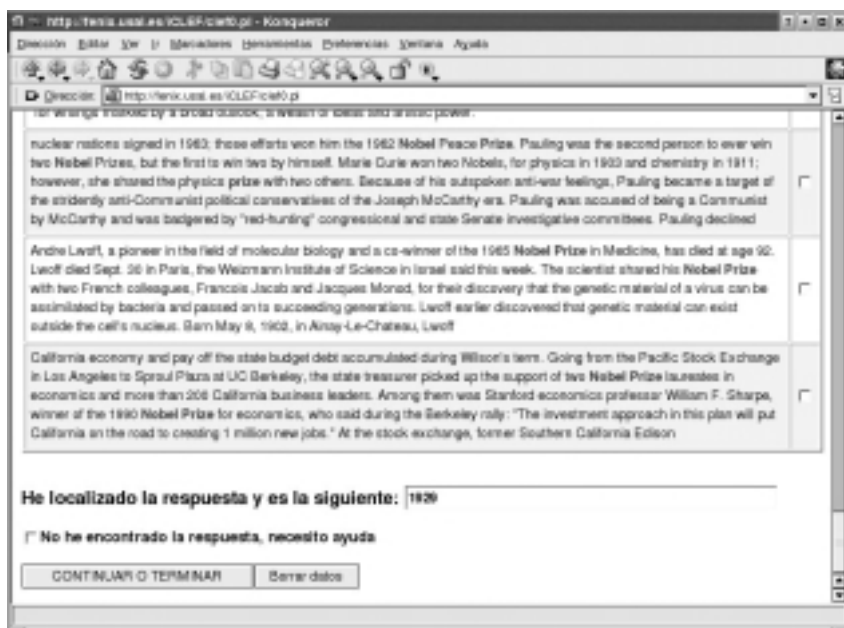


FIGURA 5. Introducción de la respuesta

2.2 Los usuarios

Se formó un grupo de ocho usuarios, con una edad media de 24,13 años, todos ellos con amplia experiencia en sistemas clásicos de recuperación de información (es decir, aquellos que utilizan los términos de búsqueda como términos índice). Eran alumnos de la Licenciatura en Documentación de la Universidad de Salamanca, con un extenso bagaje en búsquedas sobre bases de datos bibliográficas y búsquedas en internet. Tenían el español como lengua materna y poseían solamente un conocimiento pasivo del inglés, que les permitía comprender, aunque de manera incompleta, textos en inglés o incluir palabras en esa lengua para la expansión de consultas. En cuanto a su conocimiento de programas de traducción automática, era bastante escaso. En la Tabla 3 se pueden ver los resultados del cuestionario previo al experimento. Se indican valores medios en una escala de 1 a 5. Destacamos que la frecuencia de búsqueda se sitúa en 4,25; es decir, prácticamente realizaban búsquedas a diario.

TABLA 3. Resultado del cuestionario previo al experimento

Utilización de interfaces gráficas	4,38
Experiencia de búsquedas en OPACs	4,26
Experiencia de búsqueda en sistemas comerciales	3,25
Experiencia en buscadores de internet	4,50
Utilización de programas de traducción automática	2,38
Frecuencia de búsqueda	4,25
Conocimiento de inglés escrito	3,13

El bagaje cultural de los usuarios puede llevar a distorsiones en los resultados de un experimento interactivo de estas características, no por el conocimiento del inglés, sino por la propia temática de las preguntas. En este sentido, un cuestionario previo sobre la familiaridad del usuario con las preguntas mostró que para el 77% de las mismas los usuarios no tenían conocimiento alguno de cuáles podrían ser sus respuestas. Para el 23% restante, los usuarios indicaron que el tema les sonaba, pero que no conocían la respuesta. En resumen, ningún usuario conocía de antemano la respuesta exacta a ninguna de las preguntas.

3 RESULTADOS Y COMENTARIOS

3.1 Precisión de los sistemas

Todas las operaciones de los usuarios fueron registradas por el sistema. En la Tabla 4 se puede ver la precisión de nuestros sistemas A y B, tanto en precisión «exacta» como «relajada». La precisión «relajada» es más favorable y también es más realista. Por ejemplo, ante la pregunta: «¿De qué grupo político es presidente Charles Millon?», algunos usuarios respondieron «Union for French Democracy» y otros simplemente «UDF». Ambas respuestas fueron correctas en sentido relajado y solamente la segunda en sentido estricto.

TABLA 4. Precisión de las respuestas

Sistema	Precisión	
	Estricta	Relajada
A (pasajes)	0,49	0,55
B (pasajes + docs)	0,55	0,70

Podemos apreciar que en ambos casos el sistema B es superior, lo que significa que la posibilidad de acceder al documento completo, a partir de los pasajes recuperados, produce más éxitos (una mejora del 27,27%). Se esperaban mejores resultados con el sistema B, pero no tan abismales. Este resultado es general para todos los usuarios (Figura 6), salvo para el primer usuario, que obtiene un número de respuestas inferior con el sistema B.

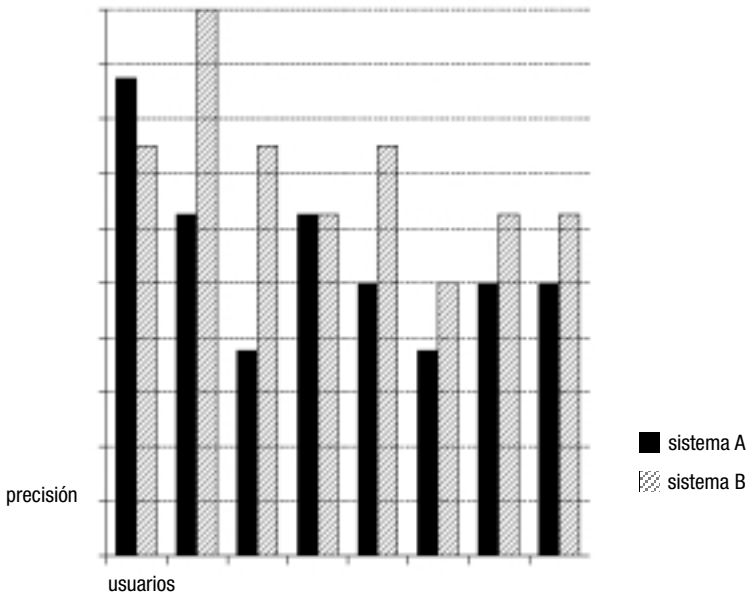


FIGURA 6. Precisión en sentido «relajado» por usuario en ambos sistemas

Es interesante ver también el tiempo medio empleado para resolver las preguntas. Cada sesión con una pregunta tuvo un tiempo máximo de cinco minutos; las que sobrepasaron este tiempo fueron consideradas no válidas (el 19,53%). En la Tabla 5 se han computado solamente las respuestas en sentido «relajado» y que no superaban los cinco minutos. Podemos ver que con el sistema B se emplea menos tiempo. Ello puede ser debido a que con el sistema A se requieren más reformulaciones. El mecanismo de sugerencia de términos para realizar la reformulación de preguntas se estudia más adelante.

TABLA 5. Tiempo medio por pregunta (excluyendo las que superaron cinco minutos)

Sistema	Tiempo medio por pregunta
A (pasajes)	146,6
B (pasajes + docs)	115,9

3.2 Dificultad de las preguntas

El número de preguntas que superaron los cinco minutos coincide, claro está, con las preguntas consideradas más difíciles de resolver. La Figura 7 muestra la precisión en sentido «relajado» de las preguntas, independientemente del sistema utilizado. Podemos ver que hay varias preguntas que no alcanzan el 50% de respuestas correctas (preguntas 1, 8, 14 y 15). Existen varios motivos para ello. Por un lado, la diferencia de criterios entre usuarios y asesores (personas de la organización iCLEF que deciden si una respuesta es correcta o no, en ambos sentidos, «exacto» y «relajado»). Por ejemplo, a la pregunta catorce: «¿Cuándo ha tenido lugar

el atentado en la estación de metro de Saint-Michel en París?», dos usuarios respondieron «*July, 25*», que fue considerado incorrecto por no incluir el año. Esta situación fue asumida como un error en la interfaz de consulta, ya que no mostraba la fecha de las noticias, lo cual provocaba respuestas incorrectas. Algo parecido sucedió con la pregunta cinco, «¿Cuánto costó el túnel del Canal de la Mancha?», a la que dos usuarios respondieron «*15 billion*», omitiendo la moneda y juzgándose como incorrectas.

En cuanto a la pregunta ocho: «¿Quién es el director gerente del Fondo Monetario Internacional?», el principal motivo por el que se obtuvieron malos resultados fue que en una gran parte de los documentos aparecían las siglas de tal organismo en inglés, IMF, despistando a los usuarios.

La explicación a los malos resultados para la pregunta quince: «¿Cuánta gente vive en Bombay?», hay que buscarla en el poco conocimiento del idioma inglés que tenían los usuarios, ya que tres de ellos respondieron «*16 millions*», siendo en realidad una expectativa de crecimiento para el año 2001 y no la cantidad actual en aquellos años, doce millones.

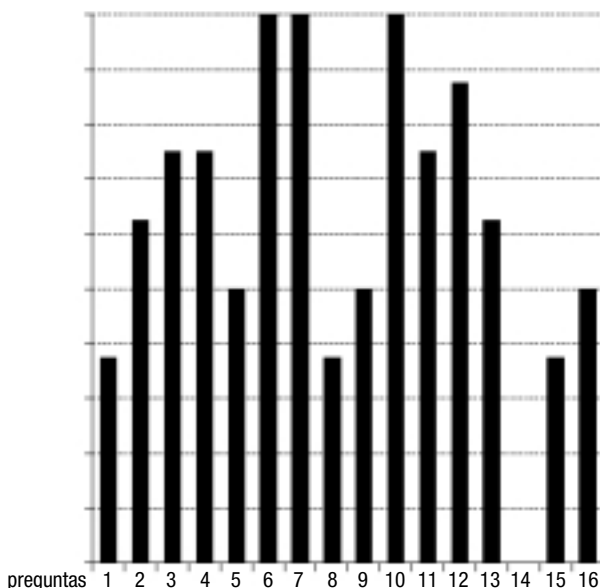


FIGURA 7. Precisión de las respuestas en sentido «relajado»

3.3 Reformulación de las preguntas

Uno de los objetivos del experimento fue analizar el comportamiento del usuario ante el mecanismo de sugerencia de términos, el cual permitía reformular la pregunta en inglés, incluyendo términos relacionados para obtener nuevos pasajes de texto. La duración de cada sesión parece estar

relacionada con el número de reformulaciones realizadas. Por este motivo, el sistema B, que consigue respuestas correctas más rápidamente (véase al Tabla 5), necesita de menos de estas reformulaciones.

De las 128 sesiones de búsqueda realizadas, en 91 de ellas no se añadió ningún término, en 35 solamente dos, y en las dos restantes seis términos. De las 80 preguntas respondidas correctamente con ambos sistemas, sólo en nueve se añadieron términos. De las respuestas incorrectas, sin embargo, en más de la mitad se añadieron términos. La Tabla 6 muestra los resultados.

TABLA 6. Reformulación de preguntas

	Términos añadidos	Número de preguntas
Correctas	0	71
	2	9
Incorrectas	0	20
	2	26
	6	2

A primera vista, esto indica que la sugerencia de términos no es eficaz para obtener respuestas correctas. Recordemos que el mecanismo de sugerencia es bastante estricto: solamente permite añadir o quitar términos de una lista y no otros que el usuario pudiera haber utilizado fuera de esa lista. Este es su principal inconveniente, derivado precisamente del propio mecanismo de expansión. Es bien sabido que los mecanismos automáticos de expansión que utilizan en su tarea la información de los primeros documentos recuperados mejoran el resultado medio de los sistemas, pero para algunas consultas provocan resultados muchos peores. Ello se debe a que si los documentos recuperados no son pertinentes para la consulta, poseen términos que al ser incorporados en la reformulación de la misma degradan todavía más los resultados. El resultado de nuestro experimento corrobora esta circunstancia.

Asimismo, hemos observado un aspecto relacionado con el idioma de los términos sugeridos, el inglés. Efectivamente, algunos usuarios manifestaron que la sugerencia de términos no les había aportado nada porque no entendían el significado de los mismos.

3.4 Resultado de los cuestionarios

Cada usuario respondió un cuestionario después de las sesiones en cada uno de los sistemas. El resultado se sintetiza en la Figura 8. En todos los aspectos, el sistema B fue considerado ligeramente mejor que el A. Todos los usuarios indicaron que el sistema B les parecía mejor porque facilitaba la contextualización de la pregunta al poder ver el documento completo; de hecho, este sistema fue elegido como el mejor en su conjunto.

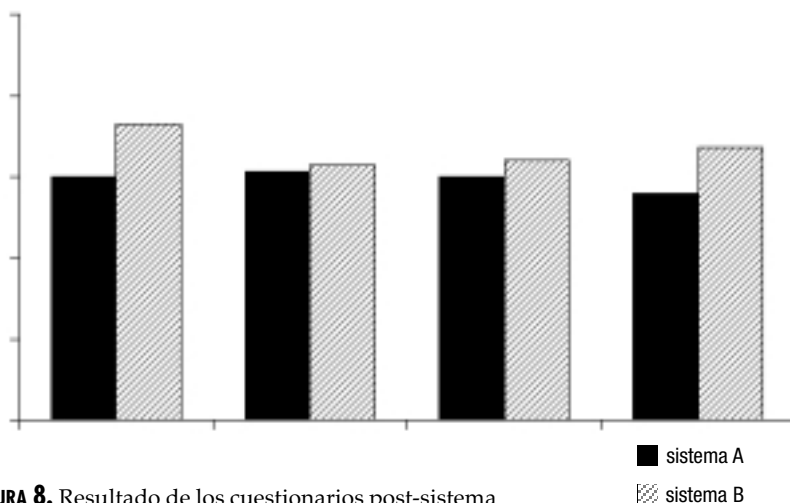


FIGURA 8. Resultado de los cuestionarios post-sistema

4 CONCLUSIONES

Hemos explorado la actuación del usuario en un sistema interactivo multilingüe de búsqueda de respuestas. En este sentido, se ha utilizado un sistema convencional basado en el modelo vectorial para recuperación de información monolingüe, al que se le han añadido ciertas características adicionales para la tarea interactiva: traducción de las preguntas, división de los documentos en pasajes de texto, sugerencia de términos para la expansión de la pregunta y una interfaz de usuario adecuada. Uno de los primeros aspectos que deseamos resaltar en estas conclusiones es el hecho importante que supone el haber obtenido una precisión del 70% en la resolución de las preguntas, que es verdaderamente muy alta si consideramos que el sistema no estaba diseñado para la tarea de búsquedas de respuestas. Esto sugiere que cuando la interacción es posible, un sistema simple proporciona mejores resultados que un sistema completamente automático (GONZALO 2004).

No debemos olvidar que estos resultados deben interpretarse teniendo en cuenta la naturaleza de los documentos y también de las preguntas. Los documentos, al ser noticias de prensa, tienden a ser de corta extensión y monotemáticos. En este sentido, pudiera parecer que la fragmentación en pasajes de texto fuese de poco interés. Sin embargo, está claro que si esta fragmentación se acompaña con la visualización del documento completo, los resultados son realmente muy buenos. La división en pasajes de texto hace que el sistema de recuperación obtenga los mejores fragmentos relacionados con los términos de la pregunta. La posibilidad de obtener y visualizar los documentos completos, a partir de los pasajes recuperados, mejora el número de aciertos, así como el tiempo requerido para obtener una respuesta correcta. Todos los usuarios destacaron el sistema B como el mejor sistema porque les permitía ver el contexto de los pasajes.

En relación con el sistema de traducción de las preguntas, hemos utilizado un sistema gratuito de los varios que existen en internet. Dado que las preguntas son cortas, contienen nombres propios y términos de fácil traducción, comprobamos que la traducción fue lo suficientemente aceptable para obtener los términos más significativos de la misma en el idioma de los documentos, que son los que interesan en un sistema de recuperación vectorial. Ello nos indica que es un método lo suficientemente válido en la búsqueda interactiva de respuestas cuando intervienen dos idiomas distintos y los usuarios tienen al menos un conocimiento pasivo de la lengua de los documentos.

Uno de los aspectos que los usuarios indicaron como más útiles para la localización de la respuesta fue el resaltado de los términos de búsqueda en los pasajes y en los documentos, pues rápidamente les permitía situarse en las zonas más prometedoras para encontrar la solución a las preguntas. La interfaz se diseñó con ese propósito.

En cuanto al método interactivo de reformulación de la pregunta utilizando los términos de una lista, obtenidos de acuerdo a la construcción automática de un tesoro de asociación local, los resultados muestran que es un sistema muy estricto y en general los usuarios lo consideraron poco útil. Dos de nuestros usuarios indicaron que les hubiese gustado ser ellos mismos quienes introdujesen sus propios términos de búsqueda. En algunos experimentos en los que se deja libertad de actuación (HE 2004), se aprecia que los usuarios utilizaban la reformulación como mecanismo de verificación de respuestas, al incluir los términos de las mismas en la nueva reformulación. Es uno de los aspectos que deseamos analizar en futuras investigaciones.

BIBLIOGRAFÍA CITADA

- (BELKIN 2001) BELKIN, N.J.; COOL, C.; KELLY, D.; LIN, S.-J.; PARK, S.Y.; PÉREZ-CARBALLO, J.; SIKORA, C. «Iterative exploration, design and evaluation of support for query reformulations in interactive information retrieval». *Information Processing & Management*, 37 (3), p. 403-434.
- (CLEF2004) *Working Notes for the CLEF 2004 Workshop, 15-17 September, Bath, UK. Internal Document, IST Project Number IST-2000 31002* [recurso electrónico] <http://clef.iei.pi.cnr.it/2004/working_notes/CLEF2004WN-Contents.html> [consulta: 8-febrero-2005]
- (FIGUEROLA 2004) FIGUEROLA, C.G.; ZAZO, Á.F.; ALONSO BERROCAL, J.L.; RODRÍGUEZ, E. «REINA at the iCLEF 2004». En: (CLEF2004).
- (GONZALO, 2004) GONZALO, J.; OARD, D.W. «iCLEF 2004 track overview: Interactive Cross-Language Question Answering». En: (CLEF2004).
- (HE 2004) HE, D.; WANG, J.; LUO, J.; OARD, D.W. «iCLEF 2004 at Maryland: Summarization design for interactive cross-language question answering». En: (CLEF2004)
- (HERSH 2000) HERSH, W.; OVER, P. «The TREC-9 interactive track report». En: (TREC9 2000), p. 41-49.
- (JONES 2000) JONES, S.; CUNNINGHAM, S.J.; McNAB, R.J.; BODDIE, S. «A transaction log analysis of a digital library». *International Journal on Digital Libraries*, 3(2), p. 152-169.

- (MAGNINI 2002) MAGNINI, B.; NEGRI, M.; PREVETE, R.; TANEV, H. «Mining knowledge from repeated co-occurrences: Diogene at trec 2002». En: *The Eleventh Text REtrieval Conference (TREC 2002)*, (Gaithersburg, Maryland (USA). NIST, 2002)
- (OARD 2004) OARD, D.W.; GONZALO, J. «The CLEF 2003 interactive track». En: PETERS, C.; GONZALO, J.; BRASCHLER, M.; KLUCK, M. (eds.). *Comparative Evaluation of Multilingual Information Access Systems, 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 21-22, 2003, Revised Selected Papers, volume 3237 of Lecture Notes in Computer Science*, p. 425-434.
- (RIJSBERGEN 1979) RIJSBERGEN, C. J. VAN *Information Retrieval*. Glasgow: Dept. of Computer Science, University of Glasgow, 1979.
- (SALTON, 1968) SALTON, G. *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968.
- (SALTON, 1971) SALTON, G. *The SMART Retrieval System. Experiments in Automatic Document Processing*. Englewoods Clifs, N. J.: Prentice Hall, 1971.
- (SPINK 2001) SPINK, A.; WOLFRAM, D.; JANSEN, B.J.; SARACEVIC, T. «Searching the web: The public and their queries». *Journal of the American Society for Information Science and Technology*, 52(3), p. 226-234.
- (TREC9 2000) *The Ninth Text REtrieval Conference (TREC 9)*, Gaithersburg, Maryland (USA). NIST Special Publication 500-249.
- (VICEDO 2003) VICEDO, J.L. *Recuperación de información de alta precisión: los sistemas de búsqueda de respuestas*. Alicante: Sociedad Española para el Procesamiento del Lenguaje Natural, 2003.
- (WOODS 2000] WOODS, W. A.; GREEN, S.; MARTIN, P.; HOUSTON, A. «Halfway to question answering». En: (TREC9 2000), p. 489-500.
- (ZAZO 2002) ZAZO, Á.F.; FIGUEROLA, C.G.; ALONSO BERROCAL, J.L.; GÓMEZ DÍAZ, R. *Recuperación de información utilizando el modelo vectorial. Participación en el taller CLEF-2001* [recurso electrónico]. Informe Técnico DPTOIA-IT-2002-006, Departamento de Informática y Automática - Universidad de Salamanca. [Consulta: 02-febrero-2005] <<http://tejo.usal.es/inftec/2002/DPTOIA-IT-2002-006.pdf>>
- (ZAZO 2003) ZAZO, Á.F.; FIGUEROLA, C.G.; BERROCAL, J.L.A.; RODRÍGUEZ, E.; GÓMEZ, R. «Experiments in term expansion using thesauri in Spanish». En: Peters, C. (ed.) *Advances in Cross-Language Information Retrieval. Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Rome, Italy. September, 2002 Revised Papers, volume 2785 of Lecture Notes in Computer Science*, p. 301-310.
- (ZAZO 2004) ZAZO, Á.F.; BERROCAL, J.L.; FIGUEROLA, C.G.; RODRÍGUEZ, E. «Estudio de usuarios de DATATHÈKE. propuestas de mejora utilizando expansión de consultas». En: CONGRESO INTERNACIONAL SOBRE TECNOLOGÍA DOCUMENTAL Y DEL CONOCIMIENTO (1º: 2004: MADRID) (Madrid: Asociación Hispana de Documentalistas en Internet, 2004). Ed. En CD-ROM.