AIETI 6

Congreso internacional

TRADUCIMOS DESDE EL SUR

Actas del VI Congreso Internacional
de la Asociación Ibérica de Estudios de Traducción e Interpretación

Las Palmas de Gran Canaria, 23-25 de enero de 2013

José Jorge Amigo Extremera
(coordinador)

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Servicio de Publicaciones y Difusión Científica

Nuevas tendencias en Recuperación de Información: la Búsqueda de Respuestas desde la perspectiva de la traducción

[New Trends in Information Retrieval: Question Answering from Translation Perspective]

María-Dolores Olvera-Lobo (1, 2)
Juncal Gutiérrez-Artacho (3)
juncalgutierrez@ugr.es
(1) CSIC, Unidad Asociada Grupo SCImago
(2) Departamento de Información y Comunicación, Universidad de Granada
(3) Departamento de Traducción e Interpretación, Universidad de Granada

Resumen
En el entorno de la Web la sobrecarga de información se deja sentir aún más que en otros contextos. De esta forma, en demasiadas ocasiones, al plantear una determinada consulta en las herramientas de búsqueda de información Web el número de páginas Web recuperadas resulta excesivo y no todas ellas son relevantes ni útiles para los objetivos del usuario. Un paso en la evolución hacia la mejora de la Recuperación de Información (RI) son los sistemas de búsqueda de respuestas (SBR) multilingües, que se presentan como una alternativa a los tradicionales sistemas de RI tratando de ofrecer respuestas precisas y comprensibles a preguntas factuales. Se han estudiado, analizado y evaluado los diferentes SBR disponibles en la Web, tanto monolingües como multilingües, de dominio general y del especializado. Además, se ha procedido al análisis de los recursos lingüísticos utilizados en los SBR multilingües, y a la evaluación de su uso real y eficacia mediante técnicas cuantitativas y cualitativas.

Palabras clave
sistemas de búsqueda de respuestas; herramientas para la traducción; evaluación; recuperación de información

Abstract
Information overload is felt more strongly on the Web than elsewhere. All too often a query made with a web search tool (search engine or meta-search engine) results in the retrieval of too many pages – many of which are useless or irrelevant to the user. Question Answering Systems are an evolutionary improvement in Information Retrieval systems. As an alternative to traditional IR systems, they give correct and understandable answers to factual questions – rather than just offering a list of documents related to the search. Our first aim is to analyze and evaluate the different Question Answering Systems, monolingual and multilingual, open and restricted-domain, available in Web. The second general aim is to identify the linguistic resources and tools found in these systems, and establishing how much use is made of these tools by multi-lingual QA systems.

Keywords
Question Answering Systems; translator tools; evaluation; Information Retrieval

1. Introducción

En el entorno de la Web la sobrecarga de información se deja sentir aún más que en otros contextos. De esta forma, en demasiadas ocasiones, al plantear una determinada consulta en las herramientas de búsqueda de información web (buscadores, directorios o metabuscadores) el número de páginas web recuperadas resulta excesivo y no todas ellas son relevantes ni útiles para los objetivos del usuario. Los sistemas de búsqueda de respuestas (SBR) se presentan como una alternativa a los tradicionales sistemas de recuperación de información tratando de ofrecer respuestas precisas y comprensibles a preguntas factuales, en lugar de mostrar al usuario una lista de documentos relacionados con su búsqueda (Jackson y Schilder 2005). El funcionamiento de los SBR se basa en los modelos de respuestas cortas (Blair-Goldensohn et al. 2004), y la ventaja principal que ofrece al usuario es que éste no ha de consultar documentos completos para obtener la información requerida puesto que el sistema ofrece la respuesta correcta en forma de un número, un sustantivo, una frase corta o un fragmento breve de texto (Pérez-Coutiño et al. 2004).

Puesto que la búsqueda de respuestas se presenta como un avance destacado en la mejora de la recuperación de información (Kolomiyets y Moens 2011) se hace necesario determinar su eficacia para el usuario final. Con este objetivo se ha realizado un estudio donde se evalúa el rendimiento y la calidad de las respuestas de los principales SBR de dominio general disponibles en la Web (QuALiM, SEMOTE, START y TrueKnowledge) ante preguntas de diversos tipos (de definición, factuales y de lista) y temas (Arte y Literatura, Biología, Personajes, Historia, Economía o Deportes, entre otros), para lo que se aplican diferentes medidas de evaluación. A continuación se detalla el análisis realizado. Los objetivos del trabajo son comparar y evaluar las respuestas ofrecidas por cada SBR de dominio general ante 500 preguntas factuales, de definición y de lista, de modo que podamos confirmar su relevancia y eficacia.

2. Sistemas de búsqueda de respuestas

Desde el punto de vista de la recuperación de información, el uso del lenguaje natural favorece el acceso a los contenidos al permitirle al usuario recurrir a su forma habitual de expresión. Los SBR normalmente presentan una sencilla interfaz con un motor de búsqueda mediante el que los usuarios pueden formular su pregunta, e incluso algunos proporcionan un listado de las últimas cuestiones introducidas para ayudarles a entender cómo han de plantearlas. Ciertamente, estos sistemas intentan emular el comportamiento del lenguaje humano por lo que tratan de entender la pregunta formulada en lenguaje natural y proporcionar respuestas adecuadas. En otras palabras, la interpretación del lenguaje natural por el sistema es un proceso esencial en el desarrollo de los SBR (Belkin y Vickery 1985; Sultan 2006). Tanto es así que el análisis de la pregunta, así como la búsqueda y la extracción de las respuestas son tres importantes tareas llevadas a cabo por los SBR, las cuáles implican, al menos, el procesamiento de las preguntas, el procesamiento de los documentos y el procesamiento de las respuestas (Kangavari, Ghandchi y Golpour 2008).

Los primeros sistemas de búsqueda de respuestas surgieron en los años 60 y utilizaban bases de datos de dominio restringido con información estructurada. Ejemplos clásicos son Baseball (Green et al. 1961), una base de datos de partidos de béisbol —*How many games did the Yankees play in July?,* Lunar (Woods 1972)—, una base de datos de análisis químicos de las misiones lunares de Apollo —*What is the average concentration of aluminium in high alkali rocks?* — o Chat-80 (Warren 1981), una base de datos geográficos —*Which is the largest African country?*— con una versión moderna que convierte la pregunta en SQL. Otro tipo de SBR son los sistemas de diálogo como el clásico Eliza (Weizenbaum 1966). Este sistema simulaba un psicoanalista y puede considerarse precursor de los actuales *chatterbot* (software diseñado para simular una conversación inteligente con uno o más humanos por medio de texto

y/o audio). Por último, los antecesores más inmediatos de los sistemas web de búsqueda de respuestas, en los que aquí nos centramos, son los sistemas de búsqueda en documentos de texto, los cuáles tomaron un importante impulso a partir de la conferencia TREC-8 (Text REtrieval Conference, Voorhees 1999).

En el tratamiento y la gestión de las preguntas, los SBR aplican algoritmos y métodos de análisis lingüístico y de procesamiento del lenguaje natural para identificar sus componentes y determinar la clase de respuesta esperada (Zweigenbaum 2005). El tipo de preguntas que suelen permitir son las denominadas preguntas factuales, de definición y de lista. Las preguntas factuales son las relacionadas con datos o hechos concretos, nombres propios, etc., se expresan mediante partículas interrogativas *(who, what, where, when, how)* y persiguen una respuesta concreta y rápida (un nombre, una fecha, un lugar, una cantidad). Este tipo de preguntas constituye la mayoría de las consultas *(who won the Nobel Prize for Literature in 1994?)*. Las preguntas de definición, como su nombre indica, persiguen obtener la definición de un término, organización, etc., y están formuladas como *what is..? (what is angiotensin?)*. En estos casos, las respuestas más relevantes serán las que ofrezcan información de manera eficiente, con el menor número de palabras, pero de construcción similar a las entradas de una enciclopedia (Greenwood y Saggion 2004; Olvera-Lobo y Gutiérrez-Artacho 2010; Olvera-Lobo y Gutiérrez-Artacho 2011a). Finalmente, las preguntas de lista son aquellas que solicitan un cierto número de respuestas de un mismo tipo y suelen plantearse de forma imperativa *(tell me…, name all of London's airports)*.

Como parte de la arquitectura de los SBR, el módulo de procesamiento de documentos se encarga de realizar una primera selección de los documentos o párrafos que se pueden considerar como relevantes para la pregunta planteada (véase figura 1). Las fuentes de información que utilizan los sistemas para seleccionar estos documentos son de lo más variadas y van desde la omnipresente Wikipedia hasta enciclopedias, diccionarios o bases de datos especializadas de gran prestigio como Medline (Olvera-Lobo y Gutiérrez-Artacho 2011b). La elección de las fuentes de información es una decisión habitualmente condicionada por el hecho de que se trate de un SBR de dominio general —y, por tanto, capaz de atender consultas de temas muy diversos, como START Natural Language Question Answering[1] (Olvera-Lobo y Gutiérrez-Artacho 2010) o NSIR Question Answering System[2]— o de dominio específico —si se centran en un ámbito temático determinado, como HONQA Health On the Net Foundation[3] (Crouch et al. 2005; Olvera-Lobo y Gutiérrez-Artacho 2011c) o EAGLi Engine for question-Answering in Genomics Literature[4] (Abdou, Savoy y Ruch 2006).

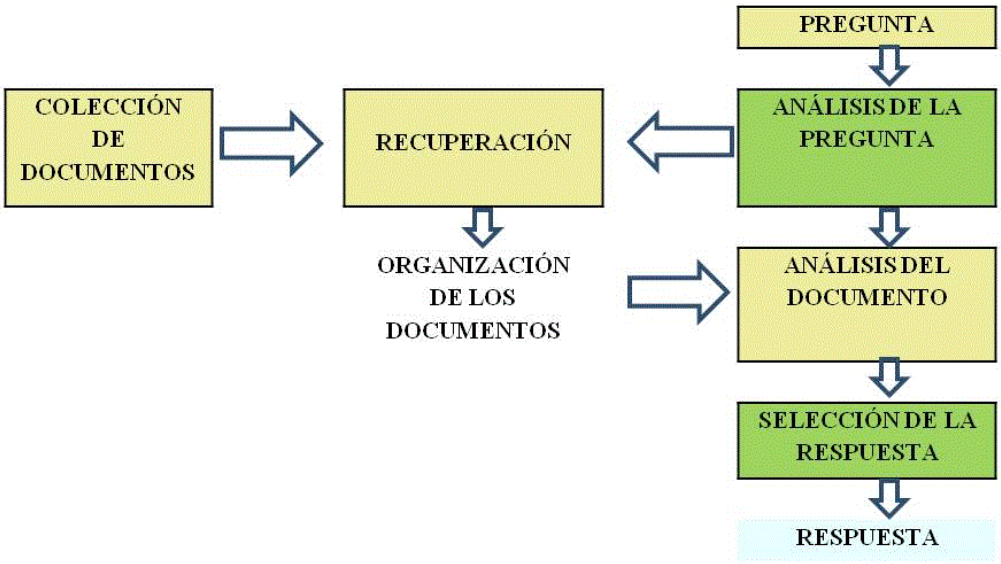

Figura 1. Modelo clásico de sistema de búsqueda de respuesta

Por último, el módulo de procesamiento de la respuesta lo forman dos importantes componentes, el destinado a su extracción y el dedicado a la validación. Las respuestas candidatas se extraen de los documentos que son recuperados por el motor de búsqueda del SBR. Tras ello, mediante el filtrado y la ordenación de las respuestas candidatas, se validan las respuestas que serán las que finalmente se muestren al usuario (Kangavari, Ghandchi y Golpour 2008). El objetivo de esta etapa es eliminar cualquier pasaje incorrecto o redundante que se encuentre en la lista recuperada por el SBR. No siempre la respuesta será única y el sistema puede proveer varias respuestas correctas que satisfagan la necesidad el usuario (Cui, Kan, Cua y Xin 2004; Rodrigo et al. 2010; Tsur 2003).

3. Objetivos de la investigación

Actualmente el nivel de desarrollo y de eficacia en el funcionamiento de los SBR puede calificarse de bastante desigual. Mientras que algunos sistemas despliegan un comportamiento más que aceptable, otros funcionan con dificultad, presentan un deficiente funcionamiento o, incluso, han dejado de estar operativos. Por esta razón, se hace necesario someterlos a diferentes pruebas que permitan el testeo y la evaluación rigurosa de los mismos considerando diferentes aspectos. Sólo así se puede determinar objetivamente cuáles son sus puntos fuertes y sus flaquezas y, por tanto, establecer dónde necesitan introducir mejoras.

Como se ha indicado, los SBR se diferencian de otros sistemas de RI por ofrecer respuestas cortas, palabras, o números, a las consultas planteadas. Esta peculiaridad los obliga a evaluar con nuevas medidas que se adapten de forma más adecuada a su arquitectura y a las expectativas de los usuarios.

El objetivo primordial de este trabajo ha consistido en diseñar y desarrollar una propuesta metodológica válida para la evaluación de la búsqueda de respuestas en el entorno de la W3. Este objetivo se encamina a la constatación de que algunas de las técnicas, medidas y herramientas de evaluación usadas tradicionalmente son adaptables y aplicables en el contexto de los SBR, aunque su arquitectura y tipología documental sea muy diferente a la de sistemas de RI clásicos. Sin embargo, y por otra parte, sus peculiaridades también exigen que dichas propuestas metodológicas las tengan en cuenta con el fin de llevar a cabo evaluaciones rigurosas y que respondan a las expectativas de los usuarios. Por tanto, debido a que en Internet tanto los sistemas RI disponibles como la propia información de la que se nutren son muy dinámicos, la finalidad principal de este estudio no es la de determinar cuál es el mejor SBR actualmente presente en la red, sino desarrollar un método que pueda aplicarse a la evaluación de los resultados por éstos ofrecidos. También pretendemos medir la satisfacción de los usuarios para realizar estudios periódicos que determinen y analicen la evolución de estas herramientas de búsqueda tan apreciadas por los usuarios.

Se ha presentado el estado de la cuestión respecto a los SBR monolingües y multilingües, y se han descrito en detalle todos los recursos lingüísticos y herramientas de los que se valen estos sistemas. Se describen las colecciones de preguntas y las medidas empleadas en las evaluaciones orientadas al sistema, la herramienta aplicada para la evaluación orientada al usuario así como los SBR que se han utilizado para llevar a cabo las distintas pruebas de evaluación. Asimismo, se detallan las distintas evaluaciones y análisis llevados a cabo los cuáles se han organizado de la siguiente forma (véase Tabla 1):

- El primer análisis *(A1)* consta de dos estudios sobre los recursos y herramientas lingüísticos utilizados por los SBR multilingües para solucionar el problema de la traducción. El primer estudio *(E1)* se centra exclusivamente en las publicaciones presentadas en *Cross-Language Evaluation Forum*, la conferencia que más ha apostado por este tipo de sistemas y que cada año celebra un *track* exclusivamente dedicado a este ámbito. El segundo estudio *(E2)* amplía el número de contribuciones realizadas en este campo analizando todas las conferencias, congresos y foros existentes en el terreno de la recuperación de información.

- El segundo análisis *(A2)*, denominado «Evaluación orientada al sistema», consta de tres test (*T1, T2, T3*) donde se analizan, evalúan y comparan los SBR monolingües y multilingües, generales y especializados, disponibles en la Web.
- Por último, se ha realizado una evaluación final (*A3*) que aúna todos los objetivos anteriores. Ésta consta de la denominada «evaluación objetiva», que analiza y evalúa las respuestas ofrecidas por los SBR multilingües, y una evaluación subjetiva y analiza las fuentes de información de los que los sistemas multilingües extraen las respuestas.

A1: Análisis orientado a los recursos y herramientas lingüísticos en los SBR multilingües		
E1: Análisis de los recursos lingüísticos utilizados en CLEF	E2: Análisis general de recursos lingüísticos usados en los SBR	
A2: Evaluación orientada al sistema		
T1: Evaluación de los SBR como eficientes fuentes de información terminológica	T2: Evaluación de los SBR de dominio general frente a los de dominio especializado en el ámbito biomédico	T3: Evaluación de la eficacia del funcionamiento de los SBR de dominio general
A4: Evaluación final: La evaluación objetiva y subjetiva de la búsqueda multilingüe de respuestas		

Tabla 1. Evaluaciones realizadas en el estudio

Esta serie de pruebas, con una perspectiva tanto cuantitativa como cualitativa, ha permitido evaluar el funcionamiento y la calidad de los SBR monolingües y multilingües disponibles en la Web así como aquellos que, aun habiendo sido desarrollados, no se han puesto todavía a disposición de los usuarios finales. En los siguientes apartados se describen los materiales y pautas metodológicas de los

4. Método

Este estudio pretende contribuir a establecer una metodología para la evaluación de la búsqueda de respuestas ofrecidas por los SBR en el entorno de la *World Wide Web*. A continuación se detalla el método diseñado —que adapta las técnicas tradicionales de evaluación a las particularidades de los SBR— y aplicado con éxito en los distintos test realizados. Además de medir la satisfacción de los usuarios mediante herramientas específicas, el funcionamiento de los sistemas se evalúa a partir de los resultados ofrecidos recurriendo a diversas medidas de evaluación. El método se ha aplicado al análisis y a la evaluación de SBR monolingües y multilingües, de ámbito general y especializado puesto que, sin investigar de forma rigurosa el funcionamiento de estos sistemas difícilmente es posible determinar con exactitud su eficacia real, ni la cobertura que ofrecen.

La investigación llevada a cabo se organiza en torno a 4 grupos de análisis. El primer análisis (A1) denominado *Análisis de los recursos lingüísticos en los SBR multilingües* se centra en el estudio de las publicaciones sobre SBR presentadas en los más destacados congresos y conferencias internacionales sobre recuperación de información, lo que permite determinar las principales y más novedosas propuestas en este sentido a través de sus propios diseñadores y desarrolladores. La ventaja de esta aproximación es que, si bien muchas de las ideas y propuestas no están disponibles para el usuario final, el análisis basado en la observación documental permite identificar cuáles son los recursos lingüísticos utilizados por los SBR y las tendencias actualmente en auge.

El segundo análisis (A2) denominado *Evaluación orientada al sistema* consiste en una serie de evaluaciones de la eficacia de la recuperación realizadas mediante SBR disponibles en la Web. Aquí se incluyen los test 1, 2 y 3 (T1, T2, T3) en los que se han testado y evaluado un total de seis sistemas, monolingües y multilingües, de dominio general y de dominio especializado.

Por último, el análisis final (A3) intenta aglutinar las diferentes perspectivas de evaluación contempladas en los tres análisis precedentes. Las pruebas realizadas consideran tanto la calidad de las respuestas del sistema, como la opinión de los usuarios respecto a los SBR multilingües, y los recursos lingüísticos que se utilizan para la recuperación de información multilingüe. Para esta prueba se ha utilizado el único representante de SBR multilingüe (inglés, francés e italiano) disponible en la Web que ofrece una cobertura y recuperación aceptable, HONqa.

4.1. Colección de las preguntas

Hoy en día, la recuperación de información en la Web va más allá de la recuperación de documentos mediante buscadores web. Los nuevos sistemas desarrollados en este ámbito persiguen tener una relación real directa con sus usuarios. De este modo, ya no se trata exclusivamente de recuperar documentos que probablemente satisfagan las necesidades de los usuarios, sino de extraer conocimiento de los mismos y elaborar respuestas que interactúen con ellos.

Para ello, es necesaria la implementación de sistemas de RI capaces de procesar el lenguaje natural y de «comprender» tanto las consultas que plantea el usuario como la información almacenada en su base de datos. Entre los diversos tipos de sistemas que responden a esta filosofía se encuentran los SBR.

Estos sistemas tratan de ofrecer una solución a preguntas concretas de los usuarios formuladas en lenguaje natural. Las preguntas que principalmente gestionan los SBR son de tipo factual (en las que se solicitan datos referidos a: persona, tiempo, localización, organización, medida, recuento, objeto, entre otros), preguntas donde se solicita una breve descripción o definición *(what is)* o bien una enumeración o lista de datos (cuáles son los partidos políticos con representación en el parlamento español). Para llevar a cabo esta tarea, los sistemas actuales aplican técnicas cada vez más sofisticadas con el fin de extraer respuestas del conjunto de los documentos que constituyen su base de datos.

Para llevar a cabo el análisis y la evaluación de los procesos de búsqueda de respuesta nos hemos servido tanto de los diferentes SBR disponibles en la Web, como de aquellos SBR presentados por los desarrolladores en los diferentes congresos y foros internacionales dedicados a este tema. Como paso previo, para poder analizar estos sistemas, ha sido necesario construir dos colecciones de preguntas de evaluación (véase Tabla 2) que permitieran analizar el funcionamiento real de los mismos.

c1	Test 3 (A2)
c2	Test 1 y 2 (A2), A4

Tabla 2. Colecciones de preguntas de evaluación

En los siguientes apartados se indican los procedimientos que se han seguido para la recopilar las preguntas de cada test.

4.2. C1: la colección de preguntas factuales y de definición

La colección 1, que incluye preguntas factuales y de definición en lengua inglesa se creó a partir de las colecciones de preguntas de evaluación generadas por las principales conferencias sobre recuperación de información *(Text Retrieval Conference, Cross-Language Evaluation Forum)*. Estas colecciones de preguntas se crean para que los participantes en estos foros las utilicen al llevar a cabo la evaluación de sus sistemas, y los resultados obtenidos puedan compararse con los de los demás. A partir de las listas de preguntas disponibles en estos dos foros durante los

años 2000 a 2004 se obtuvo una serie de casi 2000 preguntas de definición, factuales y de lista, y que versaban sobre diferentes especialidades (véanse Tablas 3 a 5).

	CLEF	TREC	Total
Nº	597	1383	1980

Tabla 3. Lugar de extracción

	2000	2001	2002	2003	2004	Total
Nº	730	475	100	475	200	1980

Tabla 4. Años

Animales	23
Arte y literatura	193
Biología	44
Ciencia	184
Cine	41
Demografía	43
Deportes	91
Economía	156
General	324
Geografía	160
Historia	255
Medicina	86
Música	35
Orografía	18
Personajes	219
Política	108
Total	1980

Tabla 5. Temas

Esta colección ha sido la utilizada en el test 3 (T3) del Análisis 2 (A2), *Evaluación orientada a sistemas*.

4.3. C2: la colección de preguntas de dominio especializado

La segunda colección de preguntas de evaluación (C2) es de dominio especializado. Se creó una muestra de preguntas médicas que permitieran evaluar el funcionamiento y las fuentes de información utilizadas por los SBR analizados. Las preguntas se obtuvieron del sitio web WebMD[5], un portal estadounidense creado por especialistas médicos para dar respuesta a las incertidumbres de los pacientes y en el que se ofrece información, en forma de preguntas y respuestas, sobre una amplia lista de temas médicos de diferente grado de especialización de los que se explican brevemente, en su caso, las características de esa enfermedad, tratamiento o síndrome, entre otros. La colección de preguntas se ha obtenido tras plantear la expresión *What is* en el motor interno de búsqueda del sitio web, ante la cual WebMD proporcionó más de 6 000 preguntas.

Esta colección ha sido utilizada en los test 1, 2 y en el *Análisis final* (A3). La elección de preguntas de ámbito biomédico se debe al hecho de han proliferado los SBR especializados en esta área del saber; por tanto, en las evaluaciones se ha querido comprobar la eficacia de los sistemas especializados en ese dominio temático.

5. Evaluación de la experiencia

En el apartado § 4 se describe el método seguido durante el proceso metodológico para obtener de manera objetiva los resultados de la investigación. La meta de este apartado es componer una detallada imagen de lo realizado en los diferentes test de la evaluación para poder cotejar y comparar los resultados posteriormente. Por ello, se presentan los datos, en primer lugar, desde el análisis de los recursos lingüísticos, con la intención de establecer qué recursos o herramientas son los más usados y los más efectivos para llevar a cabo la traducción en los SBR multilingües. Los resultados se han obtenido mediante dos análisis independientes de las publicaciones presentadas en los foros, congresos y conferencias relativas a esta área durante la última década. El análisis ha sido exhaustivo y se ha tratado un número diferente de años en cada análisis para poder mostrar datos reales.

En segundo lugar, se presentan los datos obtenidos de la evaluación orientada al sistema. Se han realizado tres test con una metodología similar para extraer datos que puedan ser comparados. No obstante, cada test persigue una serie de objetivos individuales que permiten ofrecer una visión real de la cobertura y eficacia de los SBR disponibles actualmente en la Web.

El test 3 aúna todos los objetivos individuales propuestos en cada una de las evaluaciones anteriores en un único estudio. Para ello, primero se ha planteado 120 preguntas de definición en los tres idiomas disponibles por el SBR multilingüe de dominio especializado HONqa (inglés, francés y alemán). Después se han neutralizado todas fuentes o recursos lingüísticos, en su mayoría páginas o sitios web especializados, de las cuales el sistema extrae las respuestas.

5.1. A.1. Resultados del análisis de los recursos lingüísticos en los SBR

El auge en la creación de nuevos prototipos que superen las barreras lingüísticas muestra la importancia creciente de los SBR multilingües en la RI, y la preocupación de los desarrolladores, investigadores y empresas por llevar a cabo programas que tengan una buena acogida en el mercado. Cualquier avance que se realice en la resolución de problemas de la comunicación multilingüe podrá ser incorporado a los sistemas de recuperación existentes.

Teniendo en cuenta la situación y la naturaleza de esta investigación, se han planteado las siguientes premisas que han guiado a los dos análisis realizados:

a) El uso de los sistemas de BR multilingües, que permiten la búsqueda en varios idiomas, disminuye el tiempo de obtención de la respuesta y diversifica las posibilidades de que el usuario obtenga la respuesta adecuada.

b) En el novedoso ámbito de los sistemas multilingües de BR es necesario determinar cuáles son las herramientas lingüísticas que más afectan a la eficacia de su funcionamiento.

En las evaluaciones de este apartado se plantea el análisis e incorporación de la disciplina de la traducción mediante el estudio de los sistemas multilingües de BR. El objetivo general es proceder al análisis y evaluación de las herramientas y los recursos lingüísticos utilizados en estos sistemas, divididos en dos análisis complementarios en donde en el primero se ha centrado exclusivamente en las publicaciones presentadas en CLEF desde el 2000 al 2011, y en la segunda se han incluido también el resto de conferencias dedicadas a este ámbito. Estos

objetivos confieren una nueva perspectiva al problema de la recuperación de información multilingüe.

Además, como objetivos específicos se ha planteado identificar los principales tipos de herramientas y recursos lingüísticos útiles en los procesos de RI multilingüe, concretamente en el caso de los SBR multilingües, y determinar el grado de utilización real que hacen los SBR multilingües de cada uno de los recursos y herramientas analizados mediante estos análisis.

Nuestro trabajo ha consistido en el estudio de los recursos y herramientas utilizados por los sistemas multilingües de BR a partir del análisis de las publicaciones realizadas en estos últimos años. En total, hemos analizado 170 artículos que se han presentado en las principales conferencias que tratan estos sistemas.

Al analizar todas las publicaciones, hemos comprobado que el recurso que más se utiliza por los SBR multilingües son los traductores automáticos, seguidos de los corpus, traducción y diccionarios (véase Figura 2). Según Nguyen et al. (2009), los tres recursos más utilizados hasta hace relativamente poco habían sido los traductores automáticos, los corpus (principalmente los paralelos) y los diccionarios. Sin embargo, hemos podido comprobar que el panorama está cambiando y un número mayor de sistemas utiliza la traducción en lugar de los diccionarios. No obstante, debemos afirmar que nuestros datos no son concluyentes ya que solamente hemos analizado una parte del material publicado, y por tanto, será necesario realizar más investigaciones y análisis para conseguir resultados finales.

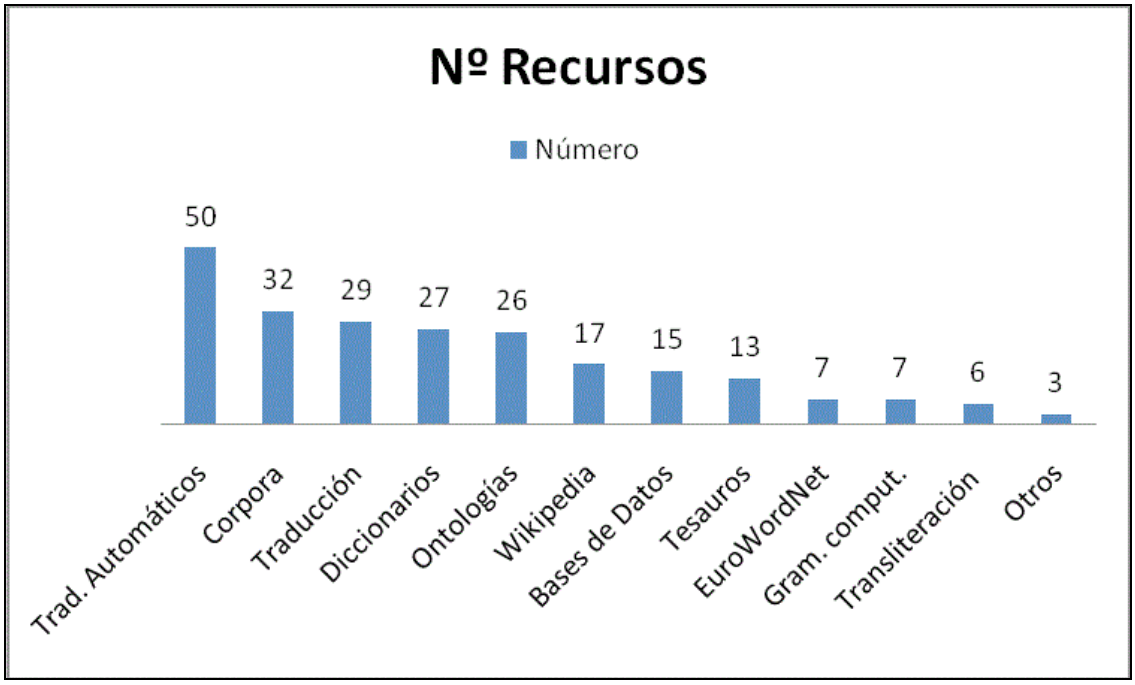

Figura 2. Número de recursos lingüísticos utilizados

Gracias a este análisis hemos comprobado el uso de cada uno de los recursos y herramientas, así como de sus características, ventajas y desventajas. Con todos estos datos, ya podemos decidir qué recursos son los más útiles y cuáles pueden ofrecer un producto final mejor. Además, este análisis nos ha dado una visión general de cómo se encuentra el problema de la traducción en los SBR multilingües y las posibles pautas que debieran seguirse en un futuro.

5.2. A.2. Resultados de la evaluación orientada al sistema: ámbito biomédico

Tras plantear las 200 preguntas en los dos SBR se identificaron las fuentes utilizadas por éstos para obtener las respuestas. En START, se extrajeron las respuestas para las preguntas médicas de seis fuentes de información: *Wikipedia, American Medical Association* (la única especializada en medicina que utilizó START), *The Internet Movie Data Base*, *Webopedia.com, Yahoo* y *Merriam Webster Dictionary*.

La fuente de información que más respuestas ofreció fue *Wikipedia,* con un total de 182, seguida por *Merriam Webster Dictionary,* que proporcionó 84 (si bien 31 de esas respuestas fueron repeticiones y, por lo tanto, se desestimaron). La única fuente especializada, *American Medical Association,* aportó 36 respuestas y las que contribuyeron en menor medida fueron *The Internet Movie Data Base (*IMDB), *Yahoo* y *Webopedia.com,* con un total de 5,2 y 1 respuestas recuperadas, respectivamente.

Fuente	Respuestas
Wikipedia	182
Merriam Webster Dictionary	*84
American Medical Association	36
IMDB	5
Yahoo	2
Webopedia.com	1
Total	**310**

**31 repetidos*

Tabla 6. Fuentes utilizadas

En relación a la calidad de los resultados ofrecidos por estas fuentes que fueron seleccionados por START como respuestas a las preguntas formuladas, (véase tabla 7) *Wikipedia* fue la fuente que más respuestas correctas ofreció (104), con 42 de ellas inexactas y 36 incorrectas. Parte de las respuestas inexactas fueron aquellas que no ofrecían directamente la respuesta sino una ventana intermedia que le ofrecía varias opciones.

El diccionario general *Merriam-Webster Dictionary* mostró un total de 45 respuestas correctas, 7 inexactas y sólo una incorrecta. La *American Medical Association* ofreció una única respuesta correcta y 35 inexactas. *Webopedia.com* ofreció una respuesta correcta mientras que las también escasas respuestas recuperadas por las fuentes *IMDB* y *Yahoo* fueron incorrectas (Tabla 7).

Fuente	incorrectas	inexactas	correctas
Wikipedia	36	42	104
American Medical Association	0	35	1
IMDB	5	0	0
Yahoo	2	0	0
Webopedia.com	0	0	1
Merriam- Webster Dictionary	36	42	104
Total	0	35	1

Tabla 7. Fuentes usadas por START

El número de respuestas obtenidas fue superior en MedQA y, en su mayor parte, fueron extraídas de fuentes médicas especializadas. La que proporcionó más respuestas fue *Medline* con 200 resultados. Se trata de una base de datos bibliográfica no gratuita creada por la Biblioteca Nacional de Medicina de los Estados Unidos que recoge referencias bibliográficas y artículos especializados publicados en casi cinco mil revistas médicas desde 1966.

Fuente	Respuestas
Medline	200
Dictionary of Cancer Terms	192
Wikipedia	191
Google	174
Dorland's Illustrated Medical Dictionary	143
Medline Plus	105
Technical and Popular Medical Terms	29
National Immunization Program Glossary	3
Total	**1037**

**34 repetidos*

Tabla 8. Fuentes usadas por MedQA

Fuente	incorrectas	inexactas	correctas
Medline Plus	9	1	95
Wikipedia	43	31	117
Google	26	26	122
Dictionary of Cancer Terms	140	0	51
Technical and Popular Medical Terms	5	3	21
Dorland's Illustrated Medical Dictionary	35	94	14
National Immunization Program Glossary	1	0	2
Medline	127	61	12
Total	**386**	**216**	**434**

Tabla 9.

Basándonos en las medidas de evaluación utilizadas en el trabajo de Raved (2001), se analizó la medida de evaluación MRR *(Mean Reciprocal Rank)*, que analiza el valor inverso de la posición en el que la respuesta se mostró, dando 1 o 0 si se encontró o no. La media de respuestas correctas en MedQA fue superior al de START. Los dos sistemas ofrecieron una media favorable en sus resultados, manteniéndose la línea general de los resultados por encima de la media.

Al comprobar que los sistemas no ofrecían una única respuesta correcta, se tuvo en cuenta la medida *Total Reciprocal Rank* (TRR). Tanto la media de respuestas correctas como la media de respuestas devueltas por el sistema fue superior en MedQA. Cuando nos encontramos con una respuesta del tipo *How, When, Who* si el sistema devuelve varias veces la misma respuesta correcta que nos ayudará para cerciorarnos de su fiabilidad.

No ocurre exactamente lo mismo con las preguntas definicionales, puesto que la devolución de respuestas repetidas distorsiona la calidad de los sistemas. Ambos sistemas ofrecieron una media comparable de respuestas repetidas, pero mientras que START devolvía la misma respuesta de la misma fuente *(Merriam-Webster Dictionary)*, MedQA ofrecía la misma respuesta de diferentes fuentes *(Wikipedia, Google)*, en ocasiones con ciertos cambios.

También se evaluó la medida de evaluación FHS *(First Hit Success)* que mide si la primera respuesta recuperado fue o no correcta. Esta medida es muy relevante, puesto que los profesionales tienden a centrarse en la primera respuesta recuperada obviando el resto. Los dos sistemas mostraron un porcentaje superior a la media, siendo superior en MedQA (75 %) que en START (61 %).

Antes de realizar la evaluación, se pensaba que los sistemas mostraban los resultados según lo recurrentes que fueran a la respuesta planteada. Sin embargo, se comprobó que no es exactamente así, sino que los sistemas poseen una lista previa de las fuentes de documentación utilizadas ordenada según su eficacia. El orden de las fuentes siempre es el mismo en todas las respuestas, excepto en una ocasión donde START ofreció primero el resultado de *American Medical Association* y luego el de *Wikipedia.*

Por último, se evaluó la precisión que muestran estos dos sistemas al recuperar y mostrar las respuestas seleccionadas. El porcentaje de cada uno de los sistemas fue relativamente parecido, mostrando ambos una media de 40. Sin embargo, START mostró una media ligeramente superior al otro sistema.

5.3. A2: Comparativa de búsquedas especializadas en el ámbito biomédico en sistemas de búsqueda de respuestas generales y específicos

Tras plantear las 150 preguntas en los cuatro SBR se analizaron las cinco primeras respuestas de cada uno de estos sistemas, al ser el promedio de respuestas recuperadas por la mayoría de ellos y debido a que los usuarios, pretendidamente, utilizarían este tipo de sistemas para la recuperación rápida de información, centrando su atención en las primeras. Para ciertas preguntas sin embargo, algunos SBR ofrecieron un número superior y otros no llegaron a ofrecer las cinco respuestas.

El volumen de respuestas recuperadas por los sistemas de dominio abierto es inferior a los de dominio restringido, siendo el menor el de START (con 1,6 respuestas como media) seguido de cerca de QuALiM (con 3 respuestas). En los SBR de dominio especializado los resultados aumentan sustancialmente, sobre todo en el caso de HONqa (con 44,23 respuestas), mientras que en MedQA es ligeramente superior a la de los SBR de dominio abierto (5,34 respuestas a cada pregunta de promedio).

Las respuestas correctas están presentes en mayor medida en START (70,08 %), en los dos SBR de dominio especializado este promedio desciende (MedQA, con 46,67 % y HONqa, con 47,25 %); QuALiM se presenta como el más deficiente con el 40,89% de respuestas correctas. La figura 3 muestra el número de respuestas correctas, inexactas e incorrectas de los cuatro SBR analizados.

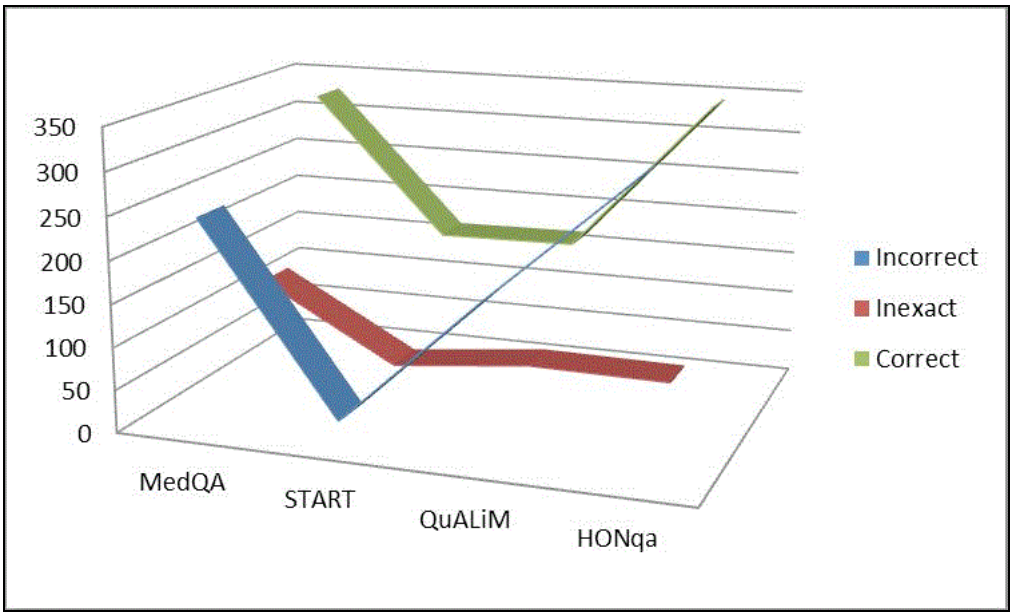

Figura 3. Promedio de respuestas correctas en los diferentes sistemas analizados

En relación a las respuestas inexactas, SBR con el promedio inferior es HONqa (7,97 %), ofreciendo el resto de los SBR un promedio similar (MedQA, 18,58 %; START, 18,38 %; QuALiM, 14,78 %).

Las respuestas incorrectas en HONqa (44,78 %) y QuALiM (44,33 %) alcanzan una cifra similar. En MedQA es algo inferior (34,75 %) aunque continúa siendo elevada siendo START (11,54 %) el sistema que presenta menos respuestas inexactas.

El valor obtenido al aplicar la medida de evaluación MRR y FHS indica que MedQA ordena mejor las respuestas, de manera que la primera respuesta correcta aparece en los primeros lugares de la lista de resultados. Esto es significativo puesto que no emplea ningún algoritmo de ranking para llevar a cabo este proceso sino que siempre recurre a la misma ordenación de las respuestas, según de la fuente de la que provengan. FHS resulta una medida muy relevante puesto que los usuarios, en muchas ocasiones, tienden a centrarse en la primera respuesta recuperada obviando el resto.

	MRR	TRR	FHS	Precisión	Exhaustividad
HONqa	0,75	1,15	0,55	47,24 %	45,73 %
QuALiM	0,65	0,77	0,59	40,88 %	22,13 %
MedQA	0,87	1,29	0,76	46,66 %	43,87 %
START	0,67	0,81	0,64	70,08 %	21,87 %

Tabla 10.

En la situación contraria nos encontramos con el otro SBR de dominio especializado, HONqa, ya que ofrece el FHS con valor más bajo, aún cuando se trata del segundo sistema con un promedio superior en MRR. El comportamiento de los dos SBR de dominio abierto es muy similar y no hay una gran diferencia entre ambas medidas, lo que se explica si se tiene en cuenta que, para cada pregunta, las respuestas recuperadas suelen oscilar entre 1 y 3.

No obstante, la medida TRR es superior en los sistemas de dominio especializado puesto que pondera el valor de cada respuesta correcta en función del lugar que ocupa en la lista de resultados. Necesariamente, este valor aumenta al ofrecer un mayor número de resultados. El

valor de la precisión en los SBR de dominio abierto —principalmente en START— ha sido superior al de los SBR de dominio especializado ya que éstos últimos presentan una alta tasa de ruido documental.

Como se puede observar, ninguna de las medidas aplicadas presenta valores muy altos. Esta circunstancia se ha visto claramente influida por el hecho de que las condiciones requeridas para evaluar una respuesta como correcta tenían un alto nivel de exigencia. En muchas ocasiones, por ejemplo en MedQA y en START se encontraron respuestas correctas con más de 100 caracteres, lo que provocó que fueran consideradas como inexactas, como se indicó en el apartado Metodología. En HonQA, por ejemplo, se encontraron respuestas correctas que, al no responder a la pregunta de forma totalmente concreta y precisa, fueron consideradas también inexactas.

5.3. A3: Análisis de búsquedas mediante preguntas factuales y de definición

Tras plantear las 500 preguntas en los cuatro SBR se analizaron las respuestas ofrecidas por cada uno de ellos. Las medidas de evaluación aplicadas tuvieron en cuenta tanto si las respuestas eran correctas para las preguntas planteadas como el orden en las que éstas aparecían en la lista ofrecida por el sistema. A continuación se muestran los resultados obtenidos (Tabla 11).

	Total de respuestas	Promedio de respuestas	Respuestas correctas	Respuestas inexactas	Respuestas incorrectas
START	744	1,49	627 *(84,27 %)*	55 *(7,39 %)*	62 *(8,34 %)*
QuaLiM	1871	3,74	892 (47,68 %)	302 *(16,14 %)*	677 *(36,18 %)*
SEMOTE	5000	10	1588 *(31,76 %)*	538 *(10,76 %)*	2874 *(57,48 %)*
TrueKnowledge	766	1,53	516 (67,37 %)	149 (19,45 %)	101 (13,18 %)

Tabla 11

El número total de respuestas recuperadas en SEMOTE (5000 y un promedio de 10 respuestas para cada pregunta) ha sido bastante superior al resto, seguido por QuALiM con menos de la mitad de respuestas recuperadas (1871 y 1,49 de promedio), mientras que en los otros dos sistemas el total de respuestas recuperadas fue similar (744 respuestas para START y 766 en TrueKnowledge, con un promedio de 1,49 y 1,53 respectivamente).

SEMOTE, como era de esperar, fue también el sistema que más respuestas correctas totales ofreció (1588), seguido de QuaLiM con 892 respuestas correctas, START con 627 y, finalmente, por TrueKnowledge con 516. Sin embargo si se tiene en cuenta la ratio de respuestas correctas respecto al número total de respuestas recuperadas por cada sistema se observa que es START el de funcionamiento más eficaz, con el 84,27 % de respuestas correctas, mientras que SEMOTE, el sistema con más respuestas totales recuperadas, ha sido el que presenta un porcentaje inferior (31,74 %). Es decir, los sistemas que menos respuestas recuperaron (START, QuaLiM y TrueKnowledge) fueron finalmente los más eficaces, los que constata que una larga lista de respuestas no garantiza que éstas sean mejores ni más precisas.

En lo que a respuestas incorrectas se refiere SEMOTE (con un 57,48 %) es el que mayor índice presenta seguido de QuaLiM con una proporción también bastante considerable (36,18 %). Frente a éstos, tanto START como TrueKnowledge (8,34 % y 13,18 %, respectivamente) presentan una ratio de respectas incorrectas que podríamos calificar de «aceptable». Por último, y en relación al tipo de respuestas que aquí se han considerado como inexactas en los cuatro

sistemas el porcentaje de las mismas ha sido bastante bajo, oscilando entre casi el 20 hasta el 7 %. La figura 4 muestra el porcentaje de respuestas correctas, inexactas e incorrectas para las preguntas formuladas en los cuatros SBR online.

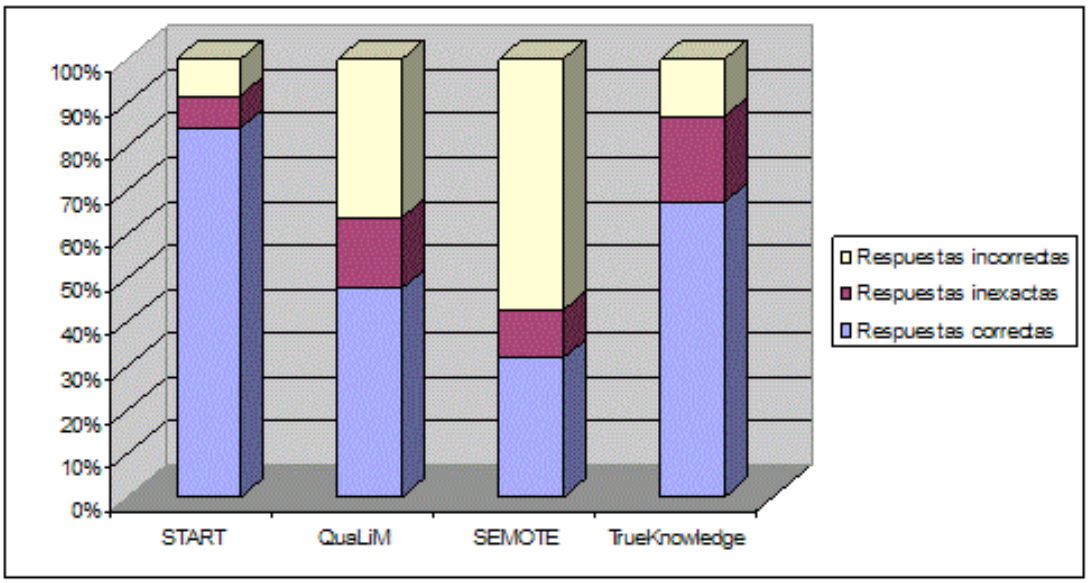

Figura 4. Promedio de respuestas recuperadas por los diferentes sistemas analizados

Los datos arrojados por las medidas de evaluación utilizadas ilustran igualmente el comportamiento de estos sistemas, considerando además el orden de las respuestas. Los resultados en general pueden considerarse bastante buenos y por tanto, demuestran que los sistemas son potencialmente útiles para recuperar información de diferentes tipos y dominios temáticos. El valor de MRR para las respuestas recuperadas es bastante elevado en todos los sistemas analizados excepto en SEMOTE (0,38). Sin embargo, aunque SEMOTE ha ofrecido un valor de MRR muy bajo, el valor para TRR ha sido similar al obtenido por los otros sistemas, lo que indica que el ranking de las respuestas ofrecidas no es el más adecuado.

FHS es una medida muy relevante puesto que los usuarios, en muchas ocasiones, tienden a centrarse en la primera respuesta recuperada obviando el resto. Se observa que más del cincuenta por ciento de las primeras respuestas ofrecidas en todos los sistemas han sido correctas. El sistema que presenta un valor superior en FHS ha sido START (0,89) siguiendo el mismo patrón que en las otras medidas aplicadas, frente a SEMOTE que, como era de esperar y al igual que en las otras medidas, es el que obtuvo un valor inferior (0,55).

6. Reflexiones finales

El usuario actual confía en recuperar información específica y de calidad que responda a sus necesidades. Los SBR presentan una interesante alternativa a la recuperación de información en Internet intentando satisfacer sus exigencias y demandas. Sin embargo, a pesar del aumento de esta clase de sistemas y del avance que supone el poder contar con herramientas de búsqueda de información de este tipo, los SBR disponibles en la Web son escasos y no todos proporcionan una cobertura adecuada. De hecho, las investigaciones que se vienen realizando y que culminan en interesantes propuestas plasmadas en diferentes publicaciones, foros y congresos, salvo contadas excepciones —y ya sea porque su utilidad se limita a contextos muy concretos, o bien por sus dificultades de implementación—no se desarrollan para el usuario final.

En este análisis se han evaluado de diferentes modos los SBR de dominio general y específicos accesibles desde la Web mediante dos colecciones de preguntas cuyas respuestas, conforme a la metodología TREC, fueron juzgadas como correctas, incorrectas o inexactas por estudiantes y

especialistas en diferentes campos temáticos. En base a estas valoraciones se aplicaron diferentes medidas de evaluación mediante las que se ilustra claramente la eficacia del funcionamiento de los sistemas analizados. Los estudios realizados revelan resultados alentadores debido a que presentan este tipo de herramienta como una nueva posibilidad para obtener información precisa y fiable en un corto período de tiempo.

Referencias bibliográficas

Abdou, Samir, Jacques Savoy y Patrick Ruch. 2006. Dépister efficacement de l'information dans une banque documentaire: L'exemple de MEDLINE. @ *Actes du XXIVème Congrès INFORSID*, 129–143.

Belkin, Nicholas John y Alina Vickery. 1985. *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-based systems*. LIR Report No 35. Londres: The British Library.

Blair-Goldensohn, Sasha, Kathleen McKeown y Andrew Hazen Schlaikjer. 2004. Answering Definitional Questions: A Hybrid Approach. @ M.T. Maybury, ed. *New Directions in Question Answering*. Palo Alto: AAAI Press, pp. 47–58.

Cui, Hang, Kan, Min-Yen, Chua, Tat-Seng y Jing Xiao. 2004. A Comparative Study on Sentence Retrieval for Definitional Question Answering. @ *SIGIR Workshop on Information retrieval for Question Answering (IR4QA)*, Sheffield.

Crouch, Dick; Saurí, Roser y Abraham Fowler. 2005. AQUAINT Pilot Knowledge-Based Evaluation: Annotation Guidelines. @ *Palo Alto Research Center*. Disponible en: <http://www2.parc.com/isl/groups/nltt/papers/aquaint_kb_pilot_evaluation_guide.pdf> Consultado el 28–I–2014.

Green Jr., Bert F., Wolf, Alice K, Chomsky, Carol y Kenneth Laughery. 1961. Baseball: An Automatic Question Answerer. @ *Proceedings of the Western Joint Computer Conference* 19: 219–224.

Greenwood, Mark A. y Horacio Saggion. 2004. A Pattern Based Approach to Answering Factoid, List and Definition Questions. @ *Proceedings of the 7th RIAO Conference (RIAO 2004):* 232–243.

Jackson, Peter y Frank Schilder. 2005. Natural Language Processing: Overview. @ K. Brown, ed. *Encyclopedia of Language & Linguistics* 2. Amsterdam: Elsevier Press, pp. 503–518.

Kangavari, Mohammad Reza, Ghandchi, Samira y Manak Golpour. 2008. A New Model for Question Answering Systems. @ *World Academy of Science, Engineering and Technology* 42: 506–513.

Kolomiyets, Oleksander y Marie-Francine Moens. 2011. A survey on question answering technology from an information retrieval perspective. @ *Journal of Information Sciences* 181: 5412–5434. DOI: 10. 1016/j. ins. 2011. 07. 047. Elsevier.

Nguyen, Dong, Overwijk, Arnold, Hauff, Claudia, Trieschnigg, Dolf, Hiemstra, Djoerd y Fransiska M. G. De Jong.2 008. WikiTranslate: Query Translation for Cross-lingual Information Retrieval using only Wikipedia @ *Lecture notes in computer science*, 5706, (CLEF 2008): 58–65.

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2013. Evaluación del rendimiento de los sistemas de búsqueda de respuestas de dominio general. *Revista Española de Documentación Científica* 36/2: 1–10. ISSN-L 0210-0614. DOI 10.3989/redc.2013.2.921. Disponible en <http://redc.revistas.csic.es/index.php/redc/article/view/791> Consultado el 28–I–2014.

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2012. Language Resources for Translation in Multi-lingual Question Answering Systems. @ *Translation Journal* 16/2. ISSN 1536-7207. Disponible en <http://www.bokorlang.com/journal/60qa.htm> Consultado el 28–I–2014.

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2011a. Language resources used in multi-lingual Question Answering Systems. @ *Online Information Preview* 35/4: 543–557. DOI 10.1108/14684521111161927.

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2011b. Evaluation of Open- vs. Restricted-Domain Question Answering Systems in the Biomedical Field. @ *Journal of Information Science*. 37/2:152–162

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2011c. Multilingual Question-Answering System in biomedical domain on the Web: an evaluation. @ P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas y M. de Rijke, eds. *Multilingual and Multimodal Information Access Evaluation. Second International Conference of the Cross-Language Evaluation Forum, CLEF 2011*. Amsterdam: Springer. ISBN 9783642237072. DOI 10.1007/978-3-642-23708-9, pp 83–89.

Olvera-Lobo, María-Dolores y Juncal Gutiérrez-Artacho. 2010. Question-Answering Systems as Efficient Source of Terminological Information: Evaluation. @ *Health Information and Library Journal* 27/4: 268–276. ISSN 1471-1842 / 1471-1834. Disponible en <http://hdl.handle.net/10481/23990>. Consultado el 28–I–2014.

Pérez-Coutiño, Manuel Alberto, Solorio, Tomás, Montes y Gómez, Manuel, López López, Antonio y Luis Villaseñor Pineda. 2004. The Use of Lexical Context in Question Answering for Spanish. @ *Workshop of the Cross-Language Evaluation Forum (CLEF 2004):* 377–384. Disponible en <http://clef.isti.cnr.it/2004/working_notes/WorkingNotes2004/46.pdf> Consultado el 28–I–2014.

Radev, Dragomir R., Qi, Hong, Wu, Harris y Weiguo Fan. 2001. *Evaluating Web-based Question Answering Systems*. Informe técnico, University of Michigan.

Rodrigo, Álvaro, Pérez-Iglesias, Joaquín, Peñas, Anselmo, Garrido, Guillermo y Lourdes Araujo. 2010. A Question Answering System based on Information Retrieval and Validation. @ *Proceedings of CLEF 2010 LABs and Workshops*, Notebook Papers, 22–23 September 2010, Padua, Italy

Sultan, Meshael. 2006. *Multiple Choice Question Answering*. Tesis doctoral. Sheffield: University of Sheffield.

Tsur, Oren. 2003. *Definitional Question-Answering Using Trainable Text Classifiers*. Tesis doctoral. Ámsterdam: University of Amsterdam.

Voorhees, Ellen M. y Dawn Tice. 1999. The TREC-8 question answering track evaluation. @ E. Voorhees y D. Harman. *Proceedings of the Eighth Text Retrieval Conference*. Gaithersburg, MD: NIST Publicación Especial. Disponible en <http://comminfo.rutgers.edu/~muresan/IR/TREC/Proceedings/t8_proceedings/t8_proceedings.html> Consultado el 28–I–2014.

Warren, David. 1981. Efficient Processing of Interactive Relational Database Queries Expressed in Logic. @ *Proceedings Seventh International Conference on Very Large Data Bases*, v.7. Cannes, VLDB Endowment, v.7., pp. 272–283.

Weizenbaum, Joseph. 1966. Eliza: A computer program for the study of natural language communication between man and machine. @ *Communications of the ACM*. 9/1: 36–45.

Woods, William A. 1972. The Lunar Sciences Natural Language Information System. @ *BBN Final Report 2378*. Cambridge: Bolt, Beranek and Newman.

Zweigenbaum, Paul. 2005. Question answering in biomedicine. @ *Proceedings Workshop on Natural Language Processing for answering*. Budapest: ACL, EACL 2003, pp. 1–4.

Notas

[1] Disponible en <http://start.csail.mit.edu/>. Consultado el 15–XII–2013

[2] Disponible en <http://tangra.si.umich.edu/clair/NSIR/html/nsir.cgi>. Consultado el 15–XII–2013

[3] Disponible en <services.hon.ch/cgi-bin/QA10/qa.pl>. Consultado el 15–XII–2013

[4] Disponible en <http://eagl.unige.ch/EAGLi/>. Consultado el 15–XII–2013

[5] Disponible en <http://www.webmd.com>. Consultado el 15–XII–2013