

# Búsqueda de Respuestas multilingüe: ¿es buena idea buscar respuestas en otros idiomas distintos a los de la pregunta?\*

*Multilingual Question Answering: is it a good idea to search answers in different languages from the question one?*

Miguel Angel García Cumbreiras  
L. Alfonso Ureña López

Fernando Martínez Santiago  
José Manuel Perea Ortega

Universidad de Jaén, Campus Las Lagunillas  
Edificio A3. E-23071  
{magc,dofer,laurena,jmperea}@ujaen.es

**Resumen:** Actualmente existen algunas iniciativas dentro de la tarea de Búsqueda de Respuestas orientadas al desarrollo de sistemas capacitados para encontrar respuestas sobre colecciones multilingües. Esto es, el usuario puede formular su consulta en diversos idiomas, y puede obtener la respuesta a partir de colecciones documentales escritas igualmente en diversos idiomas. Por un lado, para una pregunta dada, buscar en más documentos conlleva más documentos relevantes y por lo tanto mayor probabilidad de encontrar una respuesta satisfactoria. Por otro lado, la multilingüidad requiere necesariamente de un proceso de traducción automática, bien de la consulta, bien de los pasajes relevantes o bien de ambos, con la consiguiente introducción de ruido ¿Hasta qué punto ese ruido distorsiona los resultados? ¿Aportan todas las colecciones monolingües al rendimiento global del sistema multilingüe o por el contrario depende de cómo de bien representado esté ese idioma dentro del sistema multilingüe? En este trabajo hemos encontrado que un sistema de búsquedas de respuestas multilingüe puede superar a uno monolingüe, y que la aportación de cada idioma es desigual, pero en todos los casos es significativa.

**Palabras clave:** Búsqueda de Respuestas multilingüe, Recuperación de Información multilingüe, fusión de listas

**Abstract:** Nowadays there are some works in the Question Answering (QA) task that have a multilingual orientation. The user can make questions in several languages, he will obtain the answers in the same language and these answers could be in collections in several languages. On the one hand, a multilingual system requires an automatic translation module to translate the questions, the relevant documents or passages or both, and this translation introduces some noise in the system. How important is the noise? Are all the collections significant to the performance of the QA system? In this work we conclude that a multilingual QA system improves a monolingual one, and the performance of each language is not similar, but it is significant.

**Keywords:** Multilingual Question Answering, Multilingual Information Retrieval,

## 1. Introducción

Los Sistemas de Búsqueda de Respuesta (Q&A) son sistemas que procesan preguntas concretas y devuelven las respuestas específicas tras consultar colecciones de documentos de texto no estructurado. Los sistemas

de búsquedas de respuestas multilingües son aquellos que tienen la capacidad de buscar la respuesta en más de un idioma, con independencia del idioma usado para formular la consulta. Si se pretende desarrollar un sistema Q&A multilingüe escalable con cierta facilidad a otros idiomas es necesario descartar aquellas arquitecturas que implementan un sistema de búsqueda de respuestas monolingüe para cada idioma (Aceves-Pérez et al.,

\* Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología a través del proyecto TIMOM (TIN2006-15265-C06-03)

2008), debido a la complejidad y la cantidad de recursos necesarios para el desarrollo de cada uno de los sistemas monolingües. Un enfoque más fácilmente escalable es el propuesto en (García Cumbresas et al., 2009). Aquí se propone un idioma pivote, en este caso el idioma inglés, de tal modo que la consulta y los pasajes relevantes son traducidos a ese idioma, consiguiendo así que la mayor parte de los recursos usuales en Q&A sólo son necesarios para el idioma pivote, como la clasificación de la pregunta o extracción de la respuesta. Nótese que no es necesario traducir documentos ni colecciones completas, sólo los pasajes relevantes. Una descripción detallada se encuentra en (García Cumbresas et al., 2009), si bien en el siguiente apartado de este mismo artículo se describe tal arquitectura.

Por otra parte, aun tratándose de una colección multilingüe, raramente todas las colecciones van ser igual de relevantes dentro del área o entorno donde ese sistema Q&A multilingüe opera. Por ejemplo, un sistema Q&A multilingüe que esté orientado a un público alemán es probable que encuentre más consultas en ese idioma, simplemente porque algunas consultas tendrán una naturaleza local y, en consecuencia, recibirán mayor atención en los idiomas de aquel lugar. A ese idioma donde se espera encontrar más respuestas lo denominamos idioma dominante. Uno de los objetivos marcados en este estudio es determinar en qué medida aportan respuestas idiomas no dominantes.

Por lo tanto, atendiendo tanto a las peculiaridades de la arquitectura implementada como a las de la colección es posible distinguir tres escenarios según el idioma dónde se encuentre la respuesta:

- el mismo idioma que el idioma pivote.
- el mismo idioma que el idioma dominante.
- otro idioma. El idioma del documento no coincide ni con el idioma pivote ni con el idioma dominante.

Los experimentos realizados muestran que, incluso aquellas colecciones peor colocadas por no estar escritas ni en el idioma pivote ni en el dominante, aportan respuestas al rendimiento global del sistema. Esto es, respuestas que se han encontrado exclusivamente en ese idioma. Nótese que, usualmente, un sistema Q&A no encuentra la respuesta en

un único documento, sino en varios. En este caso, al tratarse de una colección multilingüe es usual encontrar la misma respuesta en más de un idioma. Lo que aquí se afirma es que todas las colecciones suman, a pesar del ruido que inevitablemente introducen debido principalmente al proceso de traducción, incluso en aquellos casos en que la respuesta se encuentra exclusivamente en los idiomas peor situados.

El resto de este artículo se organiza describiendo en la sección segunda la arquitectura del sistema BRUJA. En la sección 3 se presenta el marco de experimentación utilizado. En la sección 4 se muestran y analizan los resultados de los experimentos realizados. Finalmente, en la sección 5 se exponen las conclusiones y trabajo futuro relacionado con esta investigación.

## 2. *Arquitectura*

BRUJA es un sistema de Búsqueda de Respuestas multilingüe orientado a la traducción de pasajes y fusión a nivel de pasajes. El idioma inglés actúa como pivote en las fases de análisis de las preguntas y de extracción final de respuestas, siendo el idioma dominante en el sistema, aunque siempre la entrada de las preguntas y la salida de las respuestas será en el idioma del usuario. La Figura 1 muestra la arquitectura global del sistema BRUJA. El módulo de recuperación multilingüe (CLIR) trabaja sobre varias colecciones monolingües (español, inglés y francés), realizando una posterior fusión de pasajes relevantes, aplicando el algoritmo de fusión documental 2stepRSV (Martínez-Santiago, Martín-Valdivia, y Ureña-López, 2006). Este método trabaja partiendo de la siguiente idea: dado un término de la consulta, la frecuencia documental de ese término en todas las colecciones debe agruparse. Este método requiere calcular la puntuación obtenida por cada documento cambiando la frecuencia documental de cada término, donde su nueva frecuencia documental será el resultado de sumar a su frecuencia documental original, la frecuencia documental alcanzada por tal término en el resto de las subcolecciones seleccionadas. Una vez realizada esta fusión de pasajes monolingües relevantes, BRUJA traduce al inglés sólo los pasajes mejor posicionados para incluir la respuesta.

Los experimentos y resultados obtenidos en experimentos anteriores demuestran que el

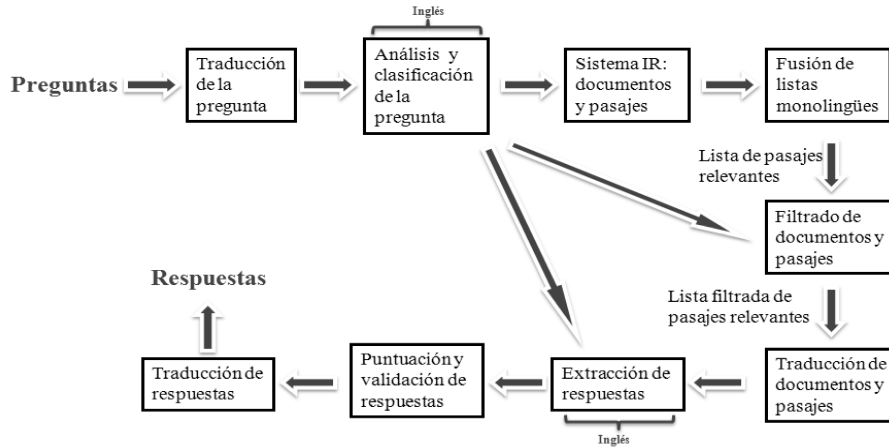


Figura 1: Arquitectura global del sistema de Búsqueda de Respuestas multilingüe BRUJA

algoritmo de fusión documental 2stepRV no sólo supera a otros usuales en la literatura, sino que permite que BRUJA supere ampliamente el rendimiento global de los sistemas de búsqueda de respuestas monolingüe.

### 3. Marco de experimentación

El marco de experimentación utilizado es el proporcionado por el foro de competición CLEF, y más concretamente la tarea de Búsqueda de Respuestas CLEF@QA del año 2006. En la Tabla 1 resumimos las colecciones utilizadas para cada uno de los tres idiomas con los que trabaja BRUJA (español, inglés y francés) junto con las características más relevantes de cada colección. Se muestra la siguiente información:

- Colección: idioma y nombre de la colección.
- Año: año en el que se añadió la colección.
- Tam: tamaño en megabytes de la colección.
- Docs: número de documentos que la componen.
- TamDoc: tamaño medio por documento.
- PalDoc: número medio de palabras por documento.

Como conjunto de preguntas se han utilizado 200 preguntas de la tarea CLEF@QA del año 2006 (tareas mono y bilingües con origen en el idioma español). Dentro de estas 200 preguntas, tras hacer una clasificación automática de las mismas, encontramos mayoritariamente preguntas factuales, y en menor

medida preguntas de definición. La Tabla 2 muestra esta distribución de preguntas.

Algunos de los parámetros variables del sistema (módulo de traducción, sistemas de Recuperación de Información, esquemas de pesado, etc.) han sido los siguientes:

- Como módulo de traducción se ha utilizado SINTRAM (García Cumberas et al., 2007). Se trata de un módulo desarrollado en nuestro grupo de investigación que selecciona para cada idioma el traductor online que, empíricamente obtiene mejores resultados en varias tareas, y aplica varias heurísticas con el fin de mejorar las traducciones. Se lista a continuación el traductor online que usa SINTRAM para cada par de idiomas:
  - Español o Francés a Inglés: Systran<sup>1</sup>.
  - Inglés o Francés a Español: Prompt<sup>2</sup>.
  - Español o Inglés a Francés: Reverso<sup>3</sup>.
- Los sistemas de recuperación de información utilizados han sido LEMUR<sup>4</sup> (basado en recuperación de documentos) y JIRS (Gómez-Soriano et al., 2005) (basado en recuperación de pasajes). De cada lista de documentos o pasajes el sistema procesa los n primeros (n es un parámetro del sistema), sin procesar dos veces el mismo documento.

<sup>1</sup> disponible en <http://www.systransoft.com>

<sup>2</sup> disponible en <http://translation2.paralink.com>

<sup>3</sup> disponible en <http://www.reverso.net>

<sup>4</sup> disponible en <http://www.lemurproject.org>

Colección	Año	Tam	Docs	TamDoc	PalDoc
Español: EFE 94	2001	511	215738	2172	290
Español: EFE 95	2003	577	238307	2221	299
Inglés: LA Times 94	2000	425	113005	2204	421
Inglés: Glasgow Herald 95	2003	154	56472	2219	343
Francés: Le Monde 94	2000	158	44013	1994	361
Francés: Le Monde 95	2001	156	47646	ND (No disponible)	ND
Francés: ATS 94	2001	86	43178	1683	227
Francés: ATS 95	2003	88	42615	1715	234

Cuadro 1: Colecciones y características utilizadas en el sistema de QA BRUJA

Clase	Total	%
ABBR	6	3 %
DESC	37	18.5 %
ENTY	46	23 %
HUM	31	15.5 %
NUM	51	25.5 %
LOC	29	14.5 %

Cuadro 2: Distribución de preguntas por clasificación genetal en el sistema de QA BRUJA

- El esquema de pesado utilizado ha sido Okapi (Robertson y S.Walker, 1999).
- En todos los experimentos se ha aplicado realimentación por relevancia o PRF (Martínez-Santiago, García-Cumberas, y Ureña-López, 2006).

#### 4. Experimentación y análisis de resultados

En trabajos previos ((García Cumberas et al., 2009) y en la Tesis Doctoral (García Cumberas, 2009)) se muestra el rendimiento global del sistema, y se demuestra que efectivamente un sistema multilingüe puede superar a uno monolingüe, pero queda por resolver la cuestión de lo que aporta cada idioma, a qué se debe esa mejora. En estos mismos trabajos previos se mostraron que sobre un total de 52 preguntas con respuesta nula en las colecciones, el sistema encontró globalmente 48 de las mismas, no habiendo prácticamente diferencias de rendimiento en este sentido entre experimentos mono y multilingües.

Una vez analizados estos resultados previos se obtuvieron algunos valores estadísticos de respuestas acertadas y no acertadas, comparando los experimentos mono y multilingües que tenían como origen el mismo idioma. En este punto nuestro principal objetivo es comparar cuántas respuestas correctas introduce el sistema multilingüe (utilizando nuestro método de fusión de pasajes 2stepRSV), y cuántas respuestas co-

rrectas obtenidas con el sistema monolingüe no aparecen con el sistema multilingüe, esto es, calcular la ganancia y la pérdida de respuestas correctas obtenidas. En la tabla 3 podemos observar este análisis comparativo de respuestas acertadas y no acertadas entre los sistemas mono y multilingües tomando como origen el mismo conjunto de 200 preguntas.

La primera columna muestra el idioma de origen de las preguntas de entrada al sistema. A continuación aparecen dos columnas con valores lógicos que indican si la respuesta se ha acertado en el experimento monolingüe, en el multilingüe o en ambos (una respuesta se considera acertada cuando está entre las cinco primeras devueltas). La última columna muestra el total de coincidentes en cada caso.

Como podemos comprobar en la tabla, la proporción entre respuestas acertadas con los sistemas multilingües y no con los monolingües y viceversa, demuestra una mejora significativa del rendimiento con el aspecto multilingüe, con una pérdida de respuestas inferior a la mitad de las nuevas respuestas encontradas (por ejemplo, 27 contra 18 con español como idioma origen). Por lo tanto, una primera conclusión que se deriva de estos resultados es que, si bien la multilingüalidad introduce ruido (respuestas perdidas), éste es muy inferior a la ganancia que supone disponer de más documentos relevantes donde bus-

Idioma Origen	Acertadas MONO	Acertadas MULTI	Total
ES	Sí	Sí	50
ES	Sí	No	18
ES	No	Sí	27
EN	Sí	Sí	46
EN	Sí	No	16
EN	No	Sí	28
FR	Sí	Sí	48
FR	Sí	No	11
FR	No	Sí	27

Cuadro 3: Análisis comparativo de respuestas acertadas y no acertadas entre experimentos mono y multilingües con el mismo conjunto de preguntas

car respuestas (respuestas ganadas).

Un análisis más detallado de los resultados lo realizamos de cara a comprobar, para los casos multilingües, la colección o el idioma de la colección origen de la primera respuesta acertada. El objetivo principal de este análisis es demostrar que el sistema BRUJA multilingüe está encontrando respuestas en colecciones expresadas en un idioma distinto al de la consulta, y dando un paso más allá, que esto ocurre independientemente del idioma de origen de las preguntas. Para demostrar este objetivo, se han realizado dos experimentos:

- Para el caso de consultas expresadas en inglés, averiguar cuántas respuestas no se encuentran en documentos escritos en inglés. Nótese que el inglés es el idioma que usa internamente BRUJA, por lo que se podría pensar que el sistema tiende a encontrar siempre las respuestas en este idioma, quedando en entredicho las bondades del enfoque multilingüe. Por eso es especialmente relevante conocer en tal caso qué aportan los otros dos idiomas por sí solos.
- En el otro extremo, nos preguntamos cuántas respuestas se responden usando exclusivamente la colección en francés, sea el idioma de la consulta el que sea. Se ha seleccionado este idioma por tratarse del idioma más difícil: recuérdese que ni BRUJA usa el francés internamente, ni las preguntas fueron diseñadas para esa colección. Este dato representa, en definitiva, el peor caso posible, la mínima aportación de una colección al rendimiento global del sistema.

Destacamos que en ambos experimentos

se han buscado respuestas exclusivas: de las colecciones francesa y española en el primer caso, y de la colección francesa en el segundo. Esto es, son respuestas que únicamente se han encontrado en esas colecciones. Esto en general no es así, ya que es usual que BRUJA encuentre la misma respuesta en diversos documentos que pueden estar escritos en uno o más de entre los tres idiomas contemplados. En definitiva, se han buscado aquí casos en los que las respuestas se obtienen gracias a la multilingüalidad del sistema. La Tabla 4 muestra los resultados estadísticos de los idiomas donde se han obtenidos respuestas acertadas, para el caso multilingüe MULTI\_EN\_ALL\_2STEP. La última columna muestra el porcentaje de respuestas acertadas en cada idioma o grupo de idiomas, respecto del total.

Idioma/s	Respuestas Acertadas	%
EN	40	53,96%
<i>ES + FR</i>	26	34,91%

Cuadro 4: Respuestas acertadas por idioma con el experimento multilingüe MULTI\_EN\_ALL\_2STEP

Según estos datos obtenidos, y tomando el inglés como idioma origen de las preguntas, más de la mitad de las respuestas acertadas en este experimento, se han obtenido, al menos, en las colecciones del inglés (esto no significa que la respuesta se encuentre, además, en documentos escritos en otros idiomas). Lo más interesante aquí es que casi el 35% de las mismas no se han encontrado en las colecciones en inglés, sino en las colecciones del español y el francés, respuestas que no se habrían acertado con un sistema mono o bilingüe.

Siguiendo con este análisis, la Tabla 5 muestra para cada uno de los experimentos multilingües que hacen uso del método “2-step RSV”, variando el idioma de origen de las preguntas, las respuestas obtenidas únicamente en las colecciones del francés, idioma que, como ya hemos comentado, supone el caso más problemático al no ser el original de las preguntas ni el idioma pivote del sistema de QA. La última columna muestra, de nuevo, el porcentaje de respuestas acertadas en cada idioma o grupo de idiomas, respecto del total.

Estos valores por sí solos suponen que con sistema multilingüe BRUJA, en el caso del inglés como idioma origen de las preguntas, más del 15 % de las respuestas acertadas se obtuvieron exclusivamente del francés. Esto es, puede que otras preguntas también obtuvieran respuesta en este idioma, pero un 15 % se encontraba exclusivamente en la colección escrita en francés. Esta es la ganancia neta que aporta a BRUJA las colecciones en francés, cuando el idioma de origen es el inglés. En cualquier caso, con independencia del idioma usado en la consulta, el número de respuestas encontrado en la colección francesa es similar, variando entre 10 y 12 respuestas.

## 5. Conclusiones y trabajo futuro

Este artículo muestra los experimentos realizados sobre un sistema de Búsqueda de Respuestas Multilingüe con el fin de comprobar si la introducción de colecciones en diversos idiomas produce una mejora en los resultados finales, en el número de respuestas acertadas.

Analizando los resultados obtenidos podemos concluir que el sistema QA multilingüe desarrollado, utilizando nuestro método de fusión de pasajes 2stepRSV, mejora los resultados de los sistemas monolingües, en términos de mayor número de respuestas correctas acertadas y de un mejor posicionamiento de aquellas respuestas acertadas que se obtienen también con los sistemas monolingües. Además, analizando la colección donde se han encontrado las respuestas correctas hemos observado que casi el 35 % de las mismas no se han encontrado en las colecciones en inglés, respuestas que no se habrían acertado con un sistema mono o bilingüe para este idioma. Y no sólo eso sino que más del 15 % de las respuestas acertadas se obtuvieron exclusivamente del francés, siendo el idioma más

complejo para nuestro sistema QA (no es el idioma de entrada de las preguntas ni el idioma pivote).

Como trabajo futuro son varias las líneas de investigación que un sistema tan complejo abarca, algunas de las cuales son las siguientes:

- Trabajo con más idiomas. No supone un gran esfuerzo adaptar el sistema para trabajar con otros idiomas, siendo necesario disponer de buenas herramientas de traducción automática para el nuevo idioma.
- Buscar respuestas en una etapa previa offline. Algunos sistemas QA actuales preprocesan las colecciones extrayendo patrones de definición y respuestas factuales, las cuales son guardadas en una base de datos. De tal forma, ante una nueva pregunta de un usuario primero se buscan respuestas en esta base de datos, no siendo necesario el módulo online de extracción de respuestas. Este hecho proporciona la capacidad de trabajo en tiempo real para estos sistemas.
- Incorporar nuevos tipos de preguntas. Una vez optimizado el trabajo con preguntas factuales y de definición es nuestro objetivo incorporar nuevos tipos de preguntas, tal como los listados, preguntas temporales o preguntas que requieren conocimiento del medio.

## Bibliografía

- Aceves-Pérez, R. M., M. Montes-Gómez, L. Villaseñor, y L. A. Ureña. 2008. Two approaches for multilingual question answering: Merging passages vs. merging answers. *International Journal of Computational Linguistics and Chinese Language Processing Special Issue on Cross-Lingual Information Retrieval and Question Answering*, 13(1).
- García Cumberras, M.A. 2009. *BRUJA: Un sistema de Búsqueda de Respuestas Multilingüe*. Ph.D. tesis, Universidad de Jaén.
- García Cumberras, M.A., F. Martínez Santiago, L.A. Ureña-López, y J.M. Perea Ortega. 2009. Evaluación del algoritmo de fusión documental 2stepsv en un sistema de búsqueda de respuestas multilingüe. En *Revista Sociedad Española para el Procesamiento del Lenguaje Natural*.

Exp	Respuestas FR	%
MULTLES_ALL_2STEP	10	8,10 %
MULTLEN_ALL_2STEP	12	15,87 %
MULTLFR_ALL_2STEP	10	6,75 %

Cuadro 5: Respuestas acertadas únicamente en las colecciones del francés, con experimentos multilingües y “2Step RSV”

García Cumbreras, M.A., L.A. Ureña-López, F. Martínez Santiago, y J.M. Perea Ortega. 2007. Bruja system. the university of jaén at the spanish task of qa@clef 2006. En *Proceedings of the Cross Language Evaluation Forum (CLEF 2006)*.

Gómez-Soriano, J.M., M. Montes y Gómez, E. Sanchis-Arnal, y P. Rosso. 2005. A passage retrieval system for multilingual question answering. En *8th International Conference of Text, Speech and Dialogue 2005 (TSD'05)*.

Martínez-Santiago, Fernando, Miguel Angel García-Cumbreras, y Luis Alfonso Ureña-López. 2006. Does pseudo-relevance feedback improve distributed information retrieval systems? *Inf. Process. Manage.*, 42(5):1151–1162.

Martínez-Santiago, Fernando, Maite Martín-Valdivia, y Luis Alfonso Ureña-López. 2006. A merging strategy proposal: The 2-step retrieval status value method. *Information Retrieval*, 9(1):71–93.

Robertson, S. E. y S.Walker. 1999. Okapi-keenbow at trec-8. En *Proceedings of Text Retrieval Conference (TREC-8)*.