

Evaluación del algoritmo de fusión documental 2stepRSV en un sistema de Búsqueda de Respuestas Multilingüe*

Evaluation of the document fusion algorithm 2stepRSV in a Multilingual Question Answering system

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: En este artículo se evalúa el rendimiento del algoritmo de fusión documental 2stepRSV como parte de un sistema de Búsqueda de Respuestas multilingüe. En estos sistemas el usuario puede formular su consulta en diversos idiomas, y puede obtener la respuesta a partir de colecciones documentales escritas igualmente en diversos idiomas. En concreto, el sistema propuesto denominado BRUJA soporta los idiomas inglés, español y francés tanto en la formulación de la consulta como en la base documental sobre la que se realiza la búsqueda. Se ha evaluado el rendimiento de 2stepRSV atendiendo a dos criterios: por un lado se ha comparado 2stepRSV con dos de los algoritmos de fusión documental más conocidos (*Round-Robin* y *Raw-Scoring*), y por otro se ha comparado el sistema multilingüe con el rendimiento que se habría obtenido en caso de que el idioma de la consulta y el de la respuesta fuera el mismo (búsqueda de respuestas monolingüe). En ambos casos 2stepRSV obtiene mejoras consistentes, entre un 15 % y un 35 %, respecto al rendimiento global de BRUJA.

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

Abstract: This paper presents the evaluation of the performance of the document fusion algorithm 2stepRSV as a part of a multilingual Question Answering system. The user makes questions in several languages and obtains answers extracted from document collections in different languages. The proposed system, called BRUJA, supports English, Spanish and French languages, for questions and document collections. We have evaluated the performance of the 2stepRSV approach according to the following motivations: on the one hand, we have compared the 2stepRSV algorithm with two algorithms related to the documents merging *Round-Robin* and *Raw-Scoring*; on the other hand, we have compared the results obtained with this multilingual system vs a monolingual one. In both cases, the 2stepRSV algorithm obtains consistent improvements, between 15 % and 35 %, with regard to the overall system performance.

Keywords: Multilingual Question Answering, Multilingual Information Retrieval, fusion method

1. Introducción

Los Sistemas de Búsqueda de Respuesta 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 (Q&A) 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. En la literatura se describen dos modelos para implementar uno de estos sistemas Aceves-Pérez et al. (2008): o bien se cuenta con un sistema Q&A para cada idioma, y luego de al-

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

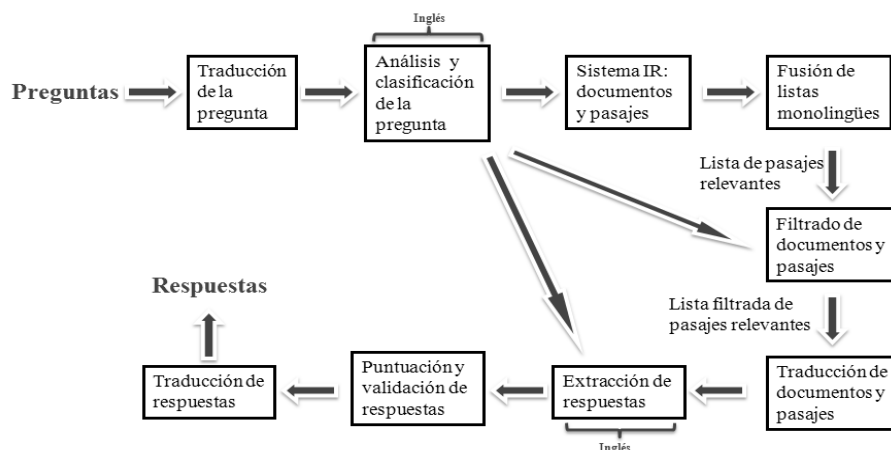


Figura 1: Arquitectura global del sistema de Búsqueda de Respuestas multilingüe BRUJA

guna manera se pondera el peso de cada respuesta con independencia del idioma, o bien el sistema utiliza internamente un único idioma, generalmente el inglés, y traduce tanto consultas como pasajes seleccionados a ese único idioma. Al primero de estos enfoques se le denomina orientado a la fusión de respuestas. El segundo enfoque está orientado a la fusión de pasajes. Mientras el primer enfoque requiere tantos sistemas de búsquedas de respuestas monolingües como idiomas se soporten, el segundo requiere de recursos de traducción automática además de un módulo de recuperación de información multilingüe¹.

Es necesario pues evaluar qué ruido introducen estos dos recursos. La experimentación aportada en Aceves-Pérez et al. (2008) demuestra que los sistemas orientados a la fusión de respuestas superan por muy poco a aquellos orientados a la fusión de pasajes, pero en cualquier caso no se consigue superar el rendimiento de un sistema de búsqueda de respuestas monolingüe. En este artículo demostraremos que no tiene que ser necesariamente así, siempre que se cuente con buenos recursos de traducción y, especialmente, un módulo CLIR eficiente en términos de precisión y cobertura.

El resto de este artículo se organiza como sigue: en la siguiente sección se describe la arquitectura del sistema BRUJA. En la sección 3 se detalla el problema de la fusión documental, y los algoritmos utilizados, 2stepRSV incluido. En la sección 4 se presenta el marco de experimentación utilizado. En la sección 5

¹módulo de recuperación de información multilingüe, de aquí en adelante módulo o sistema CLIR, del inglés *Cross-Lingual Information Retrieval*

se analizan los resultados de los experimentos realizados, tanto sobre colecciones monolingües como multilingües. Finalmente, en la sección 6 se exponen algunas conclusiones relevantes y se esbozan los siguientes pasos previstos en nuestra experimentación.

2. Breve descripción de la arquitectura de BRUJA

BRUJA es un sistema de Búsqueda de Respuestas multilingüe, orientado a la traducción del pasaje, y que usa el idioma inglés como idioma pivote. En la Figura 1 se ilustra la arquitectura global de BRUJA. Nótese que si bien BRUJA usa el inglés como idioma pivote, eso no conlleva que se traduzcan las colecciones completamente a ese idioma, lo que sería excesivamente caro computacionalmente. El módulo de recuperación multilingüe (CLIR) mantiene cada documento en su idioma original. Luego, BRUJA traduce al inglés sólo los pasajes mejor posicionados para incluir la respuesta. Debido a que el módulo CLIR mantiene cada colección en su idioma original, cada idioma cuenta con su propio índice monolingüe y por lo tanto, dada una consulta, se obtienen tantas listas de documentos relevantes como idiomas hay en la colección documental. Este problema es conocido como el problema de la fusión documental Voorhees, Gupta, y Johnson-Laird (1995), problema que no aparece por primera vez en los sistemas CLIR que no traducen los documentos, sino que es heredado de los sistemas de recuperación de información distribuida Callan, Lu, y Croft (1995).

Por otro lado, 2stepRSV es un algoritmo de fusión documental que ha obtenido buenos

resultados tanto en recuperación de información multilingüe Martínez-Santiago, Martín-Valdivia, y Ureña-López (2006) como distribuida Martínez-Santiago, García-Cumbreras, y Ureña-López (2006). Los experimentos y resultados obtenidos demuestran que este algoritmo 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 Aceves-Pérez et al. (2008).

3. El problema de la fusión documental

El problema es el siguiente: es necesario obtener una lista de documentos relevantes única a partir de I listas diferentes, cada una de las cuales es obtenida independientemente de las demás. Si la probabilidad estimada de que un documento sea relevante estuviera correctamente calibrada y fuera independiente de la colección a la cual pertenece ese documento, entonces un enfoque inmediato es ordenar los documentos atendiendo a la puntuación o RSV alcanzado por cada uno (*Raw scoring*) Kwok, Grunfeld, y Lewis (1995; Moffat y Zobel (1994).

Un algoritmo tan inmediato como el anterior es aplicar un algoritmo estilo *Round-Robin*. En este caso no se atiende al RSV obtenido, sino a la posición relativa alcanzada por cada documento en su colección. Se obtiene la lista única de documentos anotando el documento n -ésimo de cada colección, para la posición n -ésima de la lista. Así, si por ejemplo se tienen cinco listas de documentos, cinco idiomas, los cinco primeros documentos de la lista única coincidirán con el primer documento de cada lista, los cinco siguientes, con el segundo documento de cada lista, y así sucesivamente.

Raw scoring parte de la hipótesis de que la descripción hecha de cada colección es similar a la descripción del resto de las colecciones, además de que se está usando exactamente el mismo modelo de Recuperación de Información en todas las colecciones. Por su parte, *Round-Robin* presupone que los documentos relevantes están más o menos repartidos equitativamente entre todas las colecciones. En este trabajo se han utilizado las colecciones CLEF², descritas en el siguiente apartado, y dada la forma artificial en que se crean estas

colecciones se explica que ambos algoritmos rindan razonablemente bien, a pesar de su simplicidad.

3.1. El algoritmo de fusión documental 2stepRSV

El cálculo de la relevancia documental en dos pasos o 2stepRSV parte de la siguiente idea: dado un término de la consulta, la frecuencia documental de ese término en todas las colecciones debe agruparse. De esta manera, el 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. Por ejemplo, si las colecciones I_1 e I_2 son seleccionadas, y la consulta contiene el término "government", entonces la nueva frecuencia global será $df_{I_1}(\text{government}) + df_{I_2}(\text{government})$.

De esta manera, dada una consulta los dos pasos son:

1. La fase de preselección de documentos se corresponde con el lanzamiento de la consulta sobre cada subcolección seleccionada I_j . Resultado de unir los documentos más relevantes recuperados para cada idioma, obtenemos una única colección de documentos preseleccionados I' .
2. La fase de reordenamiento consiste en reindexar la colección I' , pero considerando tan sólo el vocabulario de la consulta. Se elabora así una nueva consulta formada por los términos anotados. Finalmente, se lanza tal consulta sobre el nuevo índice.

La hipótesis de este método es la siguiente: dados dos documentos, la puntuación de ambos documentos será comparable si la frecuencia documental es la misma para cada término de la consulta. Agrupando la frecuencia documental para cada término aseguramos la hipótesis.

Posiblemente, la restricción más fuerte impuesta por este método es que cada término de la consulta debe ser alineado con el resto de sus traducciones. Pero esta información no está siempre disponible, por ejemplo si se utiliza traducción automática. Por ello es necesario utilizar un algoritmo que

²<http://www.clef-campaign.org>

permita alinear los términos con sus traducciones. En este trabajo se ha utilizado el algoritmo de alineación de términos descrito en Martínez-Santiago, Martín-Valdivia, y Ureña-López (2006), que consigue alinear correctamente alrededor del 90% de palabras no vacías (Tabla 1).

Español	Francés
91 %	87 %

Tabla 1: Porcentaje de palabras no vacías alineadas correctamente a partir de las consultas expresadas en el idioma inglés (200 consultas)

Una manera de tratar con estas consultas alineadas parcialmente es calcular la puntuación del documento mediante la combinación lineal de dos valores: la parte alineada dará una puntuación o RSV según la aplicación del algoritmo 2stepRSV original. Por contra, la parte no alineada dará una segunda puntuación calculada localmente, como si de una consulta adicional se tratara, que es resuelta mediante el sistema de RI de ese idioma. Finalmente ambos valores se combinan linealmente, como podemos ver en la fórmula 1.

$$RSV_i' = \alpha \cdot RSV_i^{align} + (1 - \alpha) \cdot RSV_i^{nonalign} \quad (1)$$

donde RSV_i^{align} es la puntuación obtenida por la parte alineada, y $RSV_i^{nonalign}$ es calculado localmente. α es una constante (usualmente fijada a $\alpha = 0,75$).

4. 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 2 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.

El sistema BRUJA multilingüe se ha probado con el conjunto de 200 preguntas de la tarea CLEF@QA del año 2006, concretamente 200 preguntas de las tareas mono y bilingües con origen en el idioma español.

Por cada parámetro que variemos en el sistema surgen nuevos experimentos y resultados de forma exponencial (variando los métodos de traducción, los sistemas de recuperación de información, los esquemas de pesado, el uso de realimentación, distintas expansiones de la consulta). Algunos de estos parámetros, basándonos en los objetivos de este trabajo de investigación y en experimentos y resultados previos y paralelos obtenidos, han quedado fijados, y son los siguientes:

- El módulo de traducción utilizado ha sido SINTRAM García Cumbreiras et al. (2007). SINTRAM selecciona para cada idioma el traductor online que, empíricamente hemos encontrado que mejor funciona en tareas CLIR y QA. Además, estas traducciones son enriquecidas mediante el uso de algunas sencillas heurísticas. Se lista a continuación el traductor online que usa SINTRAM para cada par de idiomas:
 - Español o Francés a Inglés: Systran³.
 - Inglés o Francés a Español: Prompt⁴.
 - Español o Inglés a Francés: Reverso⁵.
- Los sistemas de recuperación de información utilizados han sido LEMUR⁶ (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 X primeros (X es un parámetro del sistema), teniendo en

³ disponible en <http://www.systransoft.com>

⁴ disponible en <http://translation2.paralink.com>

⁵ disponible en <http://www.reverso.net>

⁶ 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

Tabla 2: Colecciones y características utilizadas en el sistema de QA BRUJA

cuenta que el mismo identificador de documento no se encuentre en las dos listas, para no procesar dos veces el mismo documento. Si esto ocurre el peso de ese documento relevante se ve incrementado.

- El esquema de pesado utilizado ha sido OkapiRobertson y S.Walker (1999).
- En todos los experimentos se ha aplicado realimentación por relevancia o PRF-Martínez-Santiago, García-Cumbreras, y Ureña-López (2006).
- No se ha aplicado ningún método externo de expansión de la consulta (ni Wordnet ni Google).
- La evaluación en términos de MRR, Accuracy y estadísticas obtenidas con estos experimentos se ha realizado tomando las cinco primeras respuestas y entre ellas la primera respuesta correcta encontrada.

Para la evaluación del sistema de Búsqueda de Respuestas se utilizan las medidas MRR (*Mean Reciprocal Rank*) Voorhees, Gupta, y Johnson-Laird (1995) y *Accuracy*. El MRR mide la cercanía entre la respuesta correcta y las primeras opciones del conjunto de respuestas candidatas devueltas, y se calcula mediante la fórmula 2.

$$mrr(Q, p, n) = \frac{\sum_{\forall q \in Q} rr(R(P, q, n))}{|Q|} \quad (2)$$

donde Q es el conjunto de preguntas, q es cada pregunta del conjunto Q , P es la colección de documentos o pasajes, $R(P, q, n)$ es el conjunto de los n primeros documentos devueltos por el sistema de RI para una cierta pregunta q y $rr(R(P, q, n))$ es una función llamada *Reciprocal Rank* (RR) que depende

de la posición del primer documento o pasaje. Este valor es 0 si no se ha respondido correctamente en los n primeros documentos o pasajes.

Se asignan valores de RR iguales a 1, 0.5, 0.33, 0.25, 0.2 ó 0, en función de la posición de la respuesta correcta, lo que indica que respuestas acertadas en posiciones bajas toman valores bajos.

El *Accuracy* es una medida de la proporción entre el número de respuestas correctas respecto al número total de respuestas que el sistema devuelve. Se calcula mediante la fórmula 3.

$$Accuracy = \frac{NumRespuestasCorrectas}{NumRespuestasDevueltas} \quad (3)$$

5. Experimentación y Análisis de resultados

Con la finalidad de identificar con mayor sencillez cada experimento se ha seguido el convenio que se define en el siguiente patrón:

```
EXP ::= TIPOEXP _ IDIOMAPREGUNTAS _
      IDIOMACOL [ _ ALGORITMOFUSION ]
TIPO_EXP ::= MONO | MULTI
IDIOMAPREGUNTAS ::= EN | FR | ES
IDIOMACOL ::= EN | FR | ES | ALL
ALGORITMOFUSION ::= RR | RS | 2STEP
```

De forma esquemática, los experimentos **monolingües** realizados con el sistema BRUJA son los siguientes:

- **MONO_ES_ES**: realizado con todo el conjunto de preguntas para el caso *monolingüe español* contra la colección en español.
- **MONO_EN_EN**: realizado con todo el conjunto de preguntas para el caso *monolingüe inglés* contra la colección en inglés.

Exp	Acertadas	MRR	Accuracy
MULTI_ES_ALL_2STEP	77	0,356	0,385
MULTI_ES_ALL_RR	65	0,306	0,325
MULTI_ES_ALL_RS	61	0,229	0,305

Tabla 3: Experimentos multilingües atendiendo al algoritmo de fusión utilizado

Exp	Acertadas	MRR	Accuracy
MULTI_ES_ALL_2STEP	77	0,356	0,385
MULTI_EN_ALL_2STEP	74	0,300	0,370
MULTI_FR_ALL_2STEP	71	0,297	0,375

Tabla 4: Experimentos multilingües atendiendo al idioma

- **MONO_FR_FR**: realizado con todo el conjunto de preguntas para el caso *monolingüe francés* contra la colección en francés.

Los experimentos **multilingües** realizados quedan enumerados en la siguiente lista:

- **MULTI_ES_ALL_RR**: realizado con el conjunto de preguntas en español y traducidas al resto de idiomas, y con todas las colecciones, utilizando el esquema de fusión *“Round Robin”*.
- **MULTI_ES_ALL_RS**: realizado con el conjunto de preguntas en español y traducidas al resto de idiomas, y con todas las colecciones, utilizando el esquema de fusión *“Raw Scoring”*.
- **MULTI_ES_ALL_2STEP**: realizado con el conjunto de preguntas en español y traducidas al resto de idiomas, y con todas las colecciones, utilizando el esquema de fusión *“2-step RSV”*.
- **MULTI_EN_ALL_2STEP**: realizado con el conjunto de preguntas en inglés y traducidas al resto de idiomas, y con todas las colecciones, utilizando el esquema de fusión *“2-step RSV”*.
- **MULTI_FR_ALL_2STEP**: realizado con el conjunto de preguntas en francés y traducidas al resto de idiomas, y con todas las colecciones, utilizando el esquema de fusión *“2-step RSV”*.

Con estos resultados obtenidos podemos analizar el rendimiento general del sistema de búsqueda de respuestas multilingüe BRUJA. Un primer análisis nos lleva a comparar los resultados obtenidos en los casos multilingües con los distintos métodos de fusión

de listas empleados (Tabla 3). El algoritmo de fusión que mejores resultados consigue es 2stepRSV, que alcanza una mejora superior al 15 %, en términos de MRR para el caso de *“Round Robin”*, y más de un 35 % para *“Raw Scoring”*.

De los experimentos realizados sobre la colección multilingüe fijando 2stepRSV como algoritmo de fusión y variando el idioma de la consulta (Tabla 4) obtenemos el siguiente análisis de los resultados: las preguntas acertadas varían entre 77 y 71 preguntas, lo que demuestra que BRUJA es estable respecto del idioma utilizado por el usuario. Atendiendo al MRR, el rendimiento varía entre 0,356 si el idioma base es el español y el 0,297 para el caso del francés. Esto es consecuencia de que, para el caso del español, las respuestas se encuentran en los primeros documentos recuperados. La explicación hay que buscarla en el hecho de que las 200 consultas utilizadas fueron definidas por el CLEF pensando justamente en ese idioma, por lo que es de esperar que el módulo CLIR funcione sensiblemente mejor cuando recupera documentos a partir de consultas expresadas nativamente en ese idioma.

Si nos centramos en los resultados monolingües (Tabla 5) se puede comprobar que el experimento monolingüe español (MONO_ES_ES) es el que mejor resultado ha obtenido, tanto en número de respuestas acertadas como en MRR y Accuracy. Los experimentos monolingües en inglés y en francés obtienen un resultado similar, tomando como medida el MRR, con un decremento de rendimiento de un 11.6 %. En el caso del francés, además, se produce un menor número de respuestas acertadas, como consecuencia de ser el caso más complejo: un idioma distinto al de las preguntas originales y dis-

Exp	Acertadas	MRR	Accuracy
MONO_ES_ES	68	0,308	0,340
MONO_EN_EN	62	0,272	0,310
MONO_FR_FR	59	0,272	0,295

Tabla 5: Experimentos monolingües

tinto al idioma pivote del sistema BRUJA.

Un resultado que consideramos muy significativo se hace evidente al comparar el rendimiento de cada experimento monolingüe (Tabla 5) con su respectivo multilingüe (Tabla 4). En los tres idiomas hay una sensible mejora cuando se realiza la búsqueda sobre una colección multilingüe. Por ejemplo, comparando el resultado monolingüe español (MONO_EN_EN) con su contrapartida multilingüe con 2stepRSV (MULTILES_ALL_2STEP) se produce una mejora en más de un 15 % en MRR. Además, encuentra más respuestas correctas que su contrapartida monolingüe (77 contra 68). Estas dos métricas indican que se están recuperando respuestas no sólo en el idioma español, sino también extraídas de las colecciones francesa e inglesa. Este mismo razonamiento puede extrapolarse para los casos francés e inglés.

De forma resumida se demuestra que utilizar un base documental multilingüe compensa el ruido introducido durante las fases de traducción de la consultas y de los pasajes relevantes, siempre que el módulo CLIR cuente con un algoritmo de fusión de colecciones suficientemente eficiente. Este hecho ha sido cuantificado y se muestra en la Tesis García Cumbreras (2009)

6. Conclusiones y trabajo futuro

La principal conclusión a la que se puede llegar a través de la experimentación aportada es que es posible que un sistema de Búsqueda de Respuestas multilingüe basado en la traducción de pasajes y la fusión de listas consiga mejores resultados que el mismo sistema aplicado sobre una única colección multilingüe. Esto es, el hecho de contar con una mayor cantidad de documentos candidatos donde buscar la respuesta correcta compensa el ruido introducido por el sistema durante la traducción en los módulos de clasificación de la pregunta y extracción de la respuesta.

El modelo propuesto, BRUJA, resuelve el problema de la multilingüedad documental mediante un módulo de Recuperación de Información multilingüe basado en la traducción

de la consulta. Esto es, los documentos no se traducen, sino que, para una consulta dada, se obtiene una lista de documentos relevantes para cada idioma. Así pues, una pieza clave en la arquitectura de BRUJA es la elección del algoritmo que debe fusionar todas esas listas en una única lista de documentos relevantes, lista que será usada por BRUJA para seleccionar los pasajes más prometedores. En este sentido, el algoritmo 2stepRSV se muestra como un algoritmo con un excelente rendimiento para la fusión de colecciones multilingües en el ámbito de la búsqueda de respuestas multilingüe.

Como trabajo futuro se abre un amplio abanico de posibilidades, tratándose los sistemas de Búsqueda de Respuestas de procesos complejos formados por muchos módulos. Algunos de estos futuros desarrollos pasan por añadir más idiomas al sistema, siendo sólo necesario disponer de buenos traductores automáticos, trabajar con otros tipos de preguntas (preguntas de listado, preguntas con restricciones temporales, preguntas agrupadas por contexto), y por supuesto la mejora y optimización de los módulos que conforman BRUJA (traducción automática, clasificación automática de preguntas, recuperación de información, extracción de respuestas).

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).
- Callan, J. P., Z. Lu, y W. B. Croft. 1995. Searching distributed collections with inference networks. En *Proceedings of the 18th International Conference of the ACM SIGIR '95*, páginas 21–28, New York. The ACM Press.

García Cumbreras, M.A. 2009. *BRUJA: Un*

- sistema de Búsqueda de Respuestas Multilingüe*. Ph.D. tesis, Universidad de Jaén.
- 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)*.
- Kwok, K. L., L. Grunfeld, y D. D. Lewis. 1995. TREC-3 ad-hoc, routing retrieval and thresholding experiments using PIRCS. En *Proceedings of TREC'3*, volumen 500-215, páginas 247-255, Gaithersburg. NIST.
- 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.
- Moffat, A. y J. Zobel. 1994. Information retrieval systems for large document collections. En *Proc. Third Text REtrieval Conference (TREC-3)*, páginas 85-93, Gaithersburg. NIST.
- Robertson, S. E. y S.Walker. 1999. Okapi-keenbow at trec-8. En *Proceedings of Text Retrieval Conference (TREC-8)*.
- Voorhees, E., N. K. Gupta, y B. Johnson-Laird. 1995. The collection fusion problem. En D. K. Harman, editor, *Proceedings of the 3th Text Retrieval Conference TREC-3*, volumen 500-225, páginas 95-104, Gaithersburg. National Institute of Standards and Technology, Special Publication.