

Definición de un Modelo Semántico aplicado a los Sistemas de Búsqueda de Respuestas

José L. Vicedo y Antonio Ferrández

Departamento de Lenguajes y Sistemas Informáticos

Universidad de Alicante

Apartado 99. 03080 Alicante

{vicedo,antonio}@dlsi.ua.es

Resumen El sistema presentado en este trabajo realiza la tarea de Búsqueda de Respuestas sobre dominios no restringidos (open domain Question Answering - QA) desde una perspectiva semántica. Para ello, se define un modelo semántico de representación de los conceptos referidos en las preguntas y una medida de relevancia que permite la localización y selección de aquellos fragmentos de documentos de los que se extraen respuestas a las preguntas formuladas. La evaluación del sistema demuestra que las medidas del rendimiento del sistema mejoran al aplicar este modelo de representación de conceptos.

1 Introducción

Los sistemas de QA se definen como herramientas capaces de obtener respuestas concretas a necesidades de información muy precisas a partir del análisis de documentos escritos en lenguaje natural. Estos sistemas localizan y extraen la respuesta de aquellas zonas de los documentos de cuyo contenido es posible inferir la información requerida en cada pregunta. Los sistemas de QA están experimentando una mejora continua en sus resultados desde que en las últimas convocatorias del congreso Text Retrieval Conference [11][12] se diseñó una tarea específica para este tipo de sistemas. En estas convocatorias se ha evaluado el rendimiento de los sistemas existentes y se han comparado las diferentes aproximaciones utilizadas.

El análisis de los sistemas actuales más relevantes, [1][4][5][7] permite identificar los componentes principales de un sistema de QA:

1. Análisis de la pregunta.
2. Recuperación de documentos o párrafos.

3. Selección de párrafos relevantes.

4. Extracción de respuestas.

Estos procesos se llevan a cabo generalmente desde una perspectiva basada en la comparación de términos entre la pregunta y los documentos. Sin embargo, dado que cualquier información puede estar expresada de diversas formas (utilizando términos y estructuras diferentes), el rendimiento de esta estrategia está bastante restringido a respuestas que aparecen expresadas utilizando los mismos términos con los que se formulan las preguntas. Esta circunstancia limita, en muchos casos, un mejor funcionamiento de los sistemas actuales.

El objetivo de este artículo consiste en la definición de un modelo semántico que permita representar de forma general los conceptos a los que hace referencia la pregunta y que supere las limitaciones de las aproximaciones basadas en términos. Con la finalidad de evaluar las mejoras de rendimiento propiciadas por el uso de este modelo, esta aproximación se integra en un sistema de QA completo aplicándose, en concreto, al proceso de selección de párrafos. Posteriormente se comparan los resultados obtenidos mediante la utilización de este modelo con los resultados obtenidos por un sistema de referencia que utiliza una aproximación basada en términos. Esta evaluación muestra cómo mejora el rendimiento del sistema con la aplicación del modelo presentado.

La exposición de este trabajo está estructurada de la siguiente forma: la sección 2 describe el estado del arte de los sistemas de QA en dominio no restringido. En la sección 3 se detallan los componentes del sistema de referencia implementado. En la sección 4 se define el modelo de representación semántica de los conceptos referidos en las preguntas y la medida de relevancia que permite valorar su

similitud con los fragmentos de texto de los que se pretende extraer las respuestas. Finalmente, la sección 5 presenta la evaluación efectuada y el análisis y comparación de los resultados obtenidos.

2 Antecedentes

Los sistemas de QA combinan los procesos de búsqueda en grandes bases de datos documentales con los procesos de comprensión de preguntas y documentos hasta un nivel mínimo que permita seleccionar un fragmento de texto como respuesta a una determinada pregunta. Esta actividad está propiciando una relación cada vez mayor entre los campos de Recuperación de Información (RI) y de Procesamiento del Lenguaje Natural (PLN).

En particular, se ha contrastado la importancia de la aplicación de determinadas herramientas de PLN en tareas de búsqueda de respuestas. La mayoría de los sistemas utilizan etiquetadores de categorías gramaticales (POS-tagger), analizadores sintácticos (parser), sistemas de clasificación de entidades (Name-Entity tagger) y clasificadores de preguntas (Question-Type classifier). Generalmente, estos sistemas combinan técnicas de RI, basadas en la coincidencia de términos entre preguntas y respuestas, con clasificadores de entidades para seleccionar aquellas zonas de documentos en donde es más probable encontrar la respuesta esperada.

Sin embargo, la aplicación de herramientas semánticas en los sistemas de QA no está generalizada. Mayoritariamente se utilizan las relaciones de sinonimia, hiponimia e hiperonimia en procesos relacionados con clasificadores de tipos de preguntas y de entidades así como, para expandir de forma heurística, algunos términos de las preguntas cuando no se localizan párrafos lo suficientemente relevantes [10][4] [2] [9]. El único sistema que aplica un modelo semántico general a la tarea de selección de párrafos es el descrito en Woods [15]. Este sistema aplica un modelo de indexación conceptual basado en conocimiento morfológico, sintáctico y semántico apoyado además, en técnicas de subsunción taxonómica. La selección de párrafos relevantes se realiza mediante la transformación de la pregunta al modelo de indexación y recuperando los párrafos más relevantes sobre la base de dicho modelo.

A diferencia del modelo de Woods, la pro-

puesta de este trabajo plantea un modelo de representación semántica de los conceptos que aparecen en las preguntas que es independiente del proceso de indexación. A tal efecto nuestro sistema utiliza un sistema de RI convencional que realiza una primera selección de documentos. Como se podrá comprobar, el modelo semántico propuesto en este artículo mejora el rendimiento alcanzado por el sistema de Woods.

3 Descripción del sistema

En esta sección se detallan las características principales de los componentes del sistema de QA utilizado como referencia y en el que se integrará posteriormente el nuevo modelo de representación desarrollado. Este sistema está ampliamente descrito en [13].

El sistema está compuesto por los cuatro módulos básicos de un sistema de QA enunciados anteriormente: *análisis de la pregunta, recuperación de documentos, selección de párrafos y extracción de respuestas*. El primer módulo procesa las preguntas expresadas en lenguaje natural y su resultado sirve de entrada a los siguientes módulos del sistema. El módulo de recuperación de documentos realiza una primera selección de documentos utilizando a tal efecto un sistema de RI standard. Posteriormente, estos documentos son analizados por el módulo de selección de párrafos para detectar aquellos fragmentos de texto en los que es más probable encontrar la respuesta. Finalmente, dichos fragmentos son procesados por el módulo de extracción de respuestas con la finalidad de localizar la respuesta correcta.

Se aplican diversas herramientas de PLN tanto a las preguntas como a los documentos. Estas herramientas conforman el sistema SUPAR que está ampliamente detallado en [3] y [14]. Este sistema está integrado por tres módulos independientes que interactúan entre sí: análisis léxico, análisis sintáctico y un último módulo de resolución de diversos problemas de PLN como la resolución de la anáfora. La figura 1 muestra la arquitectura del sistema.

3.1 Análisis de la pregunta

El análisis de la pregunta supone la realización de diversos procesos. En primer lugar, el sistema SUPAR efectúa el etiquetado y el análisis sintáctico de las preguntas. Posteriormente, se obtiene el *tipo de pregunta* y

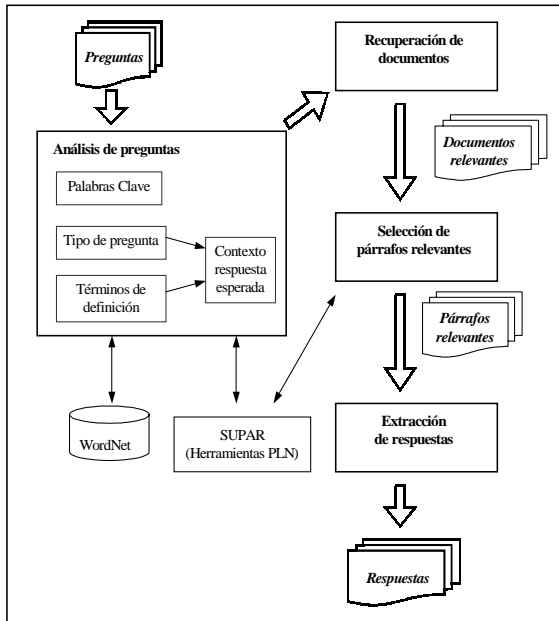


Figura 1: Arquitectura del sistema

se clasifican los restantes términos de la misma en dos categorías: *palabras clave* (keywords) y *términos de definición de respuesta* (definition terms).

El *tipo de pregunta* se obtiene mediante el análisis de los términos "Wh" (ej. Who, What, Which, etc). Este proceso relaciona los términos Wh con una de las categorías relacionadas en la figura 2. Cada una de estas categorías está relacionada con conceptos base definidos en WordNet [6]. Cuando este análisis no consigue asignar una categoría en concreto, se utiliza la categoría NONE (ej. preguntas del tipo "What"). El resultado de este análisis proporciona al sistema la siguiente información: (1) restricciones léxicas que debe cumplir la posible respuesta (ej. nombre propio), (2) la forma de detectar los términos de definición en la pregunta (si existen) y (3) conceptos base de WordNet relacionados con el tipo de respuesta esperada.

PERSON	GROUP	LOCATION	TIME
QUANTITY	REASON	MANNER	NONE

Figura 2: Categorías de tipo de pregunta

Una vez obtenido el tipo de pregunta, el sistema selecciona los *términos de definición*. Un término de la pregunta es considerado término de definición, si expresa características semánticas del tipo de respuesta esperada. Estos términos no se utilizan para lo-

calizar fragmentos de texto en los que puede aparecer la respuesta buscada, sino que definen el tipo de información requerido en la pregunta. En función del tipo de pregunta, se emplearán diferentes patrones para localizar estos términos. Por ejemplo, para las preguntas del tipo "What", "Which" o "How", los términos de definición estarán formados por aquellos sintagmas nominales que aparecen a continuación del término Wh. En el caso de preguntas del tipo "Find the number of whales..." o "Name a flying mammal..." se considerarán los sintagmas nominales que aparecen tras el verbo.

El tipo de la pregunta y los términos de definición se utilizan para la obtención del *contexto de la respuesta esperada (CRE)*. Dicho contexto define las características léxicas que la respuesta esperada debe cumplir para ser considerada como tal y las características semánticas con las que la respuesta esperada es compatible. Las características semánticas se representan mediante un vector ponderado formado por aquellos términos que están relacionados semánticamente con los términos de definición. El valor asignado a cada término mide el grado de relación entre ambos. Los términos relacionados semánticamente con los términos de definición se obtienen valorando los sinónimos, hipónimos de primer nivel y todos los hiperónimos de los términos de definición extraídos de la base de datos léxica WordNet [6]. Este proceso se describe de forma detallada en [13].

El resto de términos de la pregunta son clasificados como *palabras clave* (keywords). Las palabras clave son utilizadas por el módulo de recuperación de documentos y de selección de párrafos para detectar aquellos fragmentos de texto susceptibles de contener la respuesta a la pregunta.

A modo de ejemplo, el análisis de la pregunta "What is the name of the company that manufactures the American Girl doll collection?" produciría el siguiente resultado:

- Términos de definición: *company*.
- Keywords: *manufactures, American, Girl, doll, collection*.
- Contexto de respuesta esperada: Nombre propio relacionado principalmente con los conceptos *group, person, organization* y *social-work*.

3.2 Recuperación de documentos

Los términos clave seleccionados en el análisis de la pregunta se utilizan para recuperar los documentos relevantes a la pregunta. El sistema de RI utilizado está basado el sistema de recuperación de documentos SMART descrito en [8]. Previo a su indexación y comparación, los términos son procesados mediante la aplicación de una versión del *stemmer* de Porter.

3.3 Selección de párrafos

Los cincuenta documentos más relevantes recuperados en la etapa anterior se procesan para seleccionar aquellos párrafos más relevantes a la pregunta. El algoritmo de puntuación y selección es el siguiente:

1. Los documentos relevantes se dividen en frases.
2. Cada frase define un párrafo junto con las frases anterior y posterior a ella.
3. Las frases se puntúan con un valor que mide la similitud entre el párrafo que define y la pregunta.
4. Las frases se ordenan en función de la puntuación obtenida.

El valor asignado a cada frase (*valor-frase*) sera la suma de los valores *idf* [8] de los términos clave que aparecen en la pregunta y en el párrafo que define cada frase. En el caso de que un término de la pregunta aparezca más de una vez en el párrafo analizado, su *idf* se suma una sola vez. De esta forma, el *valor-frase* mide el nivel de aparición de los conceptos de la pregunta en cada uno de los párrafos de los documentos.

Una vez ordenadas las frases relevantes, se seleccionan las primeras 100 que incluyen *respuestas posibles* en los párrafos que definen. Estos párrafos se analizan en la fase de extracción de la respuesta para localizar aquellas respuestas consideradas correctas. Un término se considera respuesta posible si satisface las restricciones léxicas determinadas en el análisis de la pregunta.

Continuando con la pregunta planteada anteriormente, los siguiente párrafos resultan los mejor puntuados puesto que contienen (1) la mayoría de las palabras clave buscadas y (2) nombres propios considerados respuestas posibles:

1. "That meant more baby *dolls* for *girls*, wrestling action figures for boys and games and toys spun off from TV shows and comic books. The Toy *Manufacturers of America*, the trade group that sponsors the big show, said sales of traditional toys."
2. "Its *American Girls Collection* is centered on four 18-inch historical *dolls* – Felicity (1774) , Kirsten Larson (1854) , Samantha Parkington (1904) and Molly McIntire (1944) – aimed at *girls* aged seven to 12 . Each *doll* is the heroine of a serie."

Sin embargo, ninguno de estos párrafos contiene la respuesta correcta. En el primer párrafo aparecen los términos clave sin ninguna conexión entre ellos. En el segundo, no existe ninguna relación entre el significado del párrafo y la fabricación de la colección de muñecas. Como puede comprobarse, cualquier respuesta extraída de estos dos párrafos será del todo incorrecta.

3.4 Extracción de respuestas

Este proceso analiza los párrafos relevantes con la finalidad de extraer las respuestas más probables. El sistema selecciona una ventana de la longitud deseada como cadena respuesta que estará centrada en cada una de las respuestas posibles que aparecen en los párrafos analizados. A cada ventana se le asigna un valor (*valor-ventana*) calculado de la siguiente forma:

$$\text{Valor-ventana} = \text{Valor-frase} * (1 + \text{coseno}(\text{CRE}, \text{CPR}))$$

donde *CRE* es el vector de contexto de la respuesta esperada y *CPR* el vector de contexto de la respuesta posible que define la ventana analizada. El CPR se calcula de la misma forma que el CRE pero a partir de los términos de la ventana que no son términos clave [13].

De forma intuitiva, el *valor-ventana* combina la compatibilidad semántica entre cada respuesta posible y la respuesta esperada (*coseno(CRE, CPR)*) con el nivel de aparición de conceptos de la pregunta en el párrafo que contiene dicha respuesta (*valor-frase*).

Finalmente el sistema ordena las ventanas en función de la puntuación obtenida y devuelve las primeras cinco como respuestas a la pregunta.

4 Modelo semántico de representación de conceptos

El objetivo principal de este artículo es la definición de un modelo que permita la representación de los conceptos referenciados en las preguntas que un sistema de QA ha de resolver y que supere las limitaciones impuestas por los modelos basados en términos clave. Para lograr esta finalidad, dicho modelo ha de cumplir dos requisitos básicos:

1. La correcta separación de los diferentes conceptos expresados en una pregunta.
2. La representación de las diferentes formas de expresar cada uno de los conceptos.

El primero de los requisitos se afronta mediante la realización del análisis sintáctico de las preguntas. Una vez descartadas las estructuras sintácticas que formarán parte del conjunto de términos de definición, cada una de las restantes estructuras (sintagmas nominales y verbales) definirá un *concepto* cuyo elemento dominante estará formado por el núcleo y el resto de términos modificarán o precisarán el concepto básico representado por el núcleo.

El segundo requisito impone la obtención y representación de las diferentes formas de expresar cada uno de los conceptos obtenidos previamente. Para ello, cada uno de los términos que componen un concepto se relacionan con sus sinónimos y con sus hipónimos e hiperónimos de primer nivel. Estas relaciones se obtienen a partir de la base de datos léxica WordNet. La representación de cada término y sus relaciones se realiza mediante un vector ponderado que contiene el término que aparece en la pregunta y los términos con él relacionados extraídos de WordNet. El término que aparece en la pregunta se valora con su peso *idf*. Los términos obtenidos a partir de las relaciones de sinonimia, hiponimia e hiperonimia se valoran con el 80%, 50% y 50% de dicho valor respectivamente. Estos vectores se denominan *contenido semántico de un término* (CS).

Se define el *contenido semántico de un concepto* (CSC) como el par de vectores ponderados (CSN, CSM) en donde CSN corresponde a la suma de los CS de los términos que conforman el núcleo del concepto y CSM corresponde a la suma de los CS de los términos del mismo concepto que modifican al núcleo.

El conjunto de los CSC que aparecen en una pregunta conforman el *contenido semántico de una pregunta* (CSP). De esta forma, el CSP representa los conceptos indicados en la pregunta y las diversas formas de expresión de los mismos.

La figura 3 muestra la obtención del contenido semántico de la pregunta ejemplo. En primer lugar, se identifican los conceptos "manufactures" y "American Girl doll collection" y posteriormente, se genera el contenido semántico de cada uno de ellos.

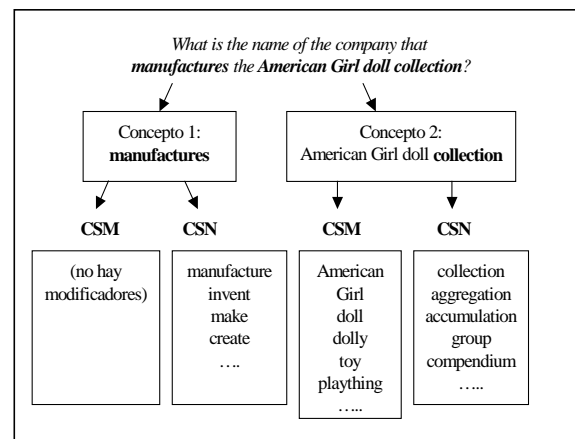


Figura 3: Ejemplo de CSP de una pregunta

4.1 Aplicación a la selección de párrafos relevantes

Con la intención de evaluar las posibilidades de la representación CSP definida en la sección anterior, se aplicará al proceso de selección de párrafos relevantes en un sistema de QA. Para ello, se define una medida que permita valorar la similitud entre los párrafos de los documentos relevantes y el CSP de una pregunta. El algoritmo de valoración de frases a utilizar será el mismo que se describe en la sección 3.3 pero sustituyendo la medida de similitud anterior por la que se define a continuación:

1. Cada CSC de la pregunta se compara con las estructuras sintácticas del mismo tipo (sintagmas nominales o verbales) del párrafo que define la frase analizada y se selecciona el valor máximo obtenido en las comparaciones. Como resultado, para cada CSC de la pregunta se obtiene un valor.
2. El valor de cada frase (*valor-frase*) se obtiene sumando los valores obtenidos para

todos los conceptos de la pregunta según el paso anterior.

3. El valor que mide la similitud entre un CSC y una estructura sintáctica del mismo tipo se calcula sumando los pesos de los términos que aparecen en los vectores del CSC y en la estructura sintáctica analizada. En el caso de que el núcleo de dicha estructura sintáctica no aparezca en el vector del núcleo del CSC, este valor será cero (aunque existan modificadores coincidentes).

Continuando con la pregunta ejemplo, el párrafo que resulta ahora mejor puntuado es el siguiente:

"Wis. firm, Pleasant Co., in September, 1986. As part of her business she *created* an *American Girls Collection* of clothes for girls ages 7 to 10, with matching outfits for dolls. All the clothes and doll costumes are inspired by ..."

Como puede comprobarse, el texto devuelto contiene la respuesta a la pregunta planteada (*Pleasant Co.*). La correcta localización del párrafo que contiene la respuesta es debido a que se han encontrado los dos conceptos referidos en la pregunta:

- El concepto "*create*" está relacionado con el CS de "manufactures"
- El concepto "*American Girls Collection*" está relacionado con el CS de "American Girl doll collection".

Una vez definido el proceso de cálculo de similitud entre la pregunta y los párrafos de los documentos relevantes utilizando el modelo semántico definido, se evaluará su rendimiento en el ámbito del sistema de QA presentado en la sección 3.

5 Evaluación

La evaluación de este tipo de sistemas resulta bastante compleja, sobre todo teniendo en cuenta que debe permitir la comparación de los resultados con otros sistemas del mismo tipo. Con esta finalidad, la evaluación del modelo se realiza utilizando el conjunto de preguntas, la colección de documentos y las medidas de rendimiento empleadas en el TREC-9 [12] para evaluar este tipo de sistemas.

El test está formado por 682 preguntas cuya respuesta ha de extraerse de un total de 978,952 documentos de las siguientes colecciones TIPSTER y TREC:

- AP Newswire.
- Wall Street Journal.
- San Jose Mercury News.
- Financial Times.
- Los Angeles Times.
- Foreign Broadcast Information Service.

Como medida del rendimiento del sistema se utiliza la media recíproca (MRR). La puntuación individual de una pregunta es el valor inverso de la posición en la que se encuentra la primera respuesta correcta, o cero si no aparece la respuesta correcta entre las 5 respuestas devueltas por el sistema:

$$MRR = \left(\sum_{i=1}^Q \frac{1}{far(i)} \right) / Q$$

donde Q corresponde al número de preguntas de test y $far(i)$ indica la posición de la primera respuesta correcta para la pregunta i . Con el fin de facilitar la comparación de resultados el valor MRR se presentará multiplicado por cien.

Para evaluar correctamente las respuestas del sistema y calcular la medida de relevancia, se emplea un programa suministrado por la organización del TREC-9 que realiza este proceso de forma totalmente automática.

5.1 Experimento

Para comparar el rendimiento de la aplicación del modelo definido en este artículo, se han realizado dos pruebas diferentes sobre el mismo conjunto de test. Para la primera prueba se utiliza el sistema de referencia definido en la sección 3. La segunda prueba se efectuará empleando el mismo sistema pero cambiando el proceso de selección de párrafos relevantes por la aproximación semántica presentado en la sección 4. Cada una de estas pruebas se divide a su vez en otras dos en función de la longitud máxima estipulada para la cadena de respuesta. Estas longitudes son de 250 y 50 bytes respectivamente.

5.2 Resultados

La figura 4 muestra los resultados obtenidos en cada una de las dos pruebas realizadas. Como puede comprobarse, la aplicación del modelo semántico de selección de párrafos mejora el rendimiento del sistema en 9,8 puntos para las respuestas de longitud 250 bytes y en 8,6 puntos para las de 50 bytes.

Número de Preguntas			682	
Pruebas	250 bytes	50 bytes		
	MRR	MRR		
Sistema de Referencia	35,6	23,0		
Modelo semántico	45,4	31,6		

Figura 4: *Resultados de la evaluación*

Estos resultados pueden compararse con los obtenidos por algunos de los mejores sistemas de QA existentes. La figura 5 muestra el rendimiento de estos sistemas sobre los mismos datos de prueba. Algunas aproximaciones alcanzan un rendimiento superior al de nuestro sistema. Sin embargo es importante reseñar que todos estos sistemas utilizan sistemas de recuperación de documentos orientados a la tarea de QA y el sistema presentado en este trabajo emplea un sistema standard que sólo recupera el documento que contiene la respuesta correcta para 601 de las 682 preguntas con las que se evalúa el modelo. En cuanto a la comparación con el modelo de Woods, puede comprobarse que el rendimiento de nuestra aproximación es superior en 10.9 puntos para la tarea de 250 bytes. No pueden compararse los resultados para la tarea de 50 bytes puesto que el sistema de Woods no presenta resultados para dicha longitud de respuesta. Esta comparación es la más relevante puesto que ambos modelos aplican una aproximación semántica general al proceso de selección de párrafos.

6 Conclusiones y trabajos futuros

El objetivo perseguido en este trabajo consiste en la definición de un modelo semántico de representación de conceptos que permita afrontar la tarea de QA desde una perspectiva semántica. Como demuestran los resultados, se ha conseguido mejorar el rendimiento del sistema al efectuar el proceso de selección de párrafos desde esta nueva perspectiva.

Durante el desarrollo de la investigación se han detectado varias áreas de trabajo futuro entre las que destacaremos como más impor-

Número de Preguntas		682	
Sistema	250 bytes MRR	50 bytes MRR	Sistema
Univ. Methodista	76,0	58,0	Univ. Methodista
Queens College	46,4	32,1	Multitext
Multitext	46,0	31,8	USC
IBM (Yorktown)	45,7	31,5	IBM (Hawthorne)
IBM (Hawthorne)	42,5	29,0	IBM (Yorktown)
LMSI	40,7	28,4	Queens College
Univ. de Korea	40,6	26,0	Univ. de Korea
NTT DATA	39,1	24,9	Syracuse CNLP
Imperial College	38,5	23,7	NTT DATA
Syracuse CNLP	38,5	23,1	Imperial College
Univ. de Montreal	36,6	22,7	Xerox
Xerox	35,3	21,2	Kaist
Univ. de Sheffield	34,5	20,6	Univ. de Sheffield
SUN (Woods)	34,5	19,6	Microsoft
Univ. Fudan	33,9	19,5	Univ. Fudan
U. de Massachusetts	32,9	18,0	Univ. de Alberta
Kaist	32,7	17,9	Univ. de Montreal
Univ. de Taiwan	31,5	17,8	LMSI
CL Research	29,6	13,5	CL Research
Microsoft	26,4	10,1	Univ. de Seoul
Univ. de Pisa	23,8	8,4	Univ. de Pisa
Univ. de Iowa	22,7	6,4	Conexor Oy
Univ. de Seoul	21,7	3,8	Mitre
Mitre	10,2		

Figura 5: *Resultados de otros sistemas*

tante la adaptación del sistema de RI a tareas de QA. El sistema actual de RI sólo proporciona el documento que contenía la respuesta correcta para el 88% de las preguntas. Dado que los sistemas que presentan un mejor rendimiento que el nuestro realizan la recuperación directa de párrafos a partir de la totalidad de la colección, se debe profundizar en la aplicación del modelo definido en este trabajo sobre la totalidad de la colección de documentos.

Referencias

- [1] Charles L. Clarke, G. V. Cormack, Derek I. Kisman, and T. R. Lynam. Question Answering by Passage Selection (MultiText Experiments for TREC-9). In *Nineth Text REtrieval Conference* [12].
- [2] David Elworthy. Question Answering using a large NLP System. In *Nineth Text REtrieval Conference* [12].
- [3] Antonio Ferrández, Manuel Palomar, and Lidia Moreno. An empirical approach to Spanish anaphora resolution. *Machine Translation Special Issue on Anaphora Resolution in Machine Translation. Kluwer Academic Publishers. ISSN 0922-6567, (14(3/4)):191-216, 1999.*

- [4] Sanda Harabagiu, Dan Moldovan, Marius Pasca, Rada Mihalcea, Mihai Surdeanu, Razvan Bunescu, Roxana Girju, Vasile Rus, and Paul Morarescu. FALCON: Boosting Knowledge for Answer Engines. In *Nineth Text REtrieval Conference* [12].
- [5] Abraham Ittycheriah, Martin Franz, Wei-Jing Zu, and Adwait Ratnaparkhi. IBM's Statistical Question Answering System. In *Nineth Text REtrieval Conference* [12].
- [6] G. Miller. Wordnet: A Lexical Database for English. In *Communications of the ACM 38(11)*, pages 39–41, 1995.
- [7] John Prager, Eric Brown, Dragomir Radev, and Krzysztof Czuba. One Search Engine or Two for Question-Answering. In *Nineth Text REtrieval Conference* [12].
- [8] Gerard A. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison Wesley, New York, 1989.
- [9] Sam Scott and Robert Gaizauskas. University of Sheffield TREC-9 Q&A System. In *Nineth Text REtrieval Conference* [12].
- [10] Kim Soo-Min, Baek Dae-Ho, Kim Sang-Beom, and Rim Hae-Chang. Question Answering Considering Semantic Categories and Co-occurrence Density. In *Nineth Text REtrieval Conference* [12].
- [11] TREC-8. *Eighth Text REtrieval Conference*, 1999.
- [12] TREC-9. *Nineth Text REtrieval Conference*, 2000.
- [13] José Luis Vicedo and Antonio Ferrández. A semantic approach to Question Answering systems. In *Nineth Text REtrieval Conference* [12].
- [14] José Luis Vicedo and Antonio Ferrández. Importance of Pronominal Anaphora resolution in Question Answering systems. In *38th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 555–562, 2000.
- [15] W. A. Woods, S. Green, P. Martin, and A. Houston. Halfway to Question Answering. In *Nineth Text REtrieval Conference* [12].