# ¿Cómo influye la resolución de la anáfora pronominal en los Sistemas de Búsqueda de Respuestas?

**Jose Luis Vicedo González**, **Antonio Ferrández Rodríguez** y **Jesús Peral Cortés**
Grupo de Procesamiento del Lenguaje y Sistemas de Información.
Departamento de Lenguajes y Sistemas Informáticos. Universidad de Alicante.
Apdo. Correos 99 - 03080 Alicante, España.
{vicedo, antonio, jperal}@dlsi.ua.es
Teléfono: +34-96-590-3400 ext. 2514,2445,3385
Fax: +34-96-590-93-26

### *Resumen*

La finalidad de este artículo consiste en analizar los efectos de aplicar técnicas de resolución de la anáfora pronominal en Sistemas de Búsqueda de Respuestas sobre dominios no restringidos (open domain Question Answering Systems - QA). Para ello, se ha implementado un sistema completo de QA y se analiza cómo varía su comportamiento al resolver la anáfora pronominal. Las pruebas se han se han efectuado sobre una colección de textos escritos en inglés. La evaluación del sistema demuestra que las medidas del rendimiento del sistema mejoran cuando se tiene en cuenta la información referida anafóricamente en los documentos.

## *Introducción*

Los sistemas QA se definen como herramientas capaces de obtener respuestas concretas (directamente de documentos escritos en lenguaje natural en dominios no restringidos) a necesidades de información precisas de los usuarios. O por lo menos, pretenden localizar y extraer aquellas zonas de los documentos de cuyo contenido es posible inferir la respuesta a una pregunta específica. En definitiva, estos sistemas intentan reducir la cantidad de tiempo que los usuarios emplean para buscar una información concreta.

Este trabajo pretende alcanzar dos objetivos principales. Primero, se analizan varias colecciones de documentos para determinar el nivel de información referida mediante el uso de pronombres. Este estudio nos proporciona una visión general acerca de la cantidad de información que se descarta cuando estas referencias no son analizadas. Como segundo objetivo, intentamos medir las mejoras obtenidas en sistemas QA cuando se resuelven estas clases de referencias. Con este objetivo en mente, se ha implementado un sistema QA completo. Las mejoras obtenidas con la resolución de las referencias pronominales se miden comparando el rendimiento del sistema "con" y "sin" tener en cuenta la información referida pronominalmente. La evaluación efectuada nos muestra cómo mejora el rendimiento del sistema al resolver este tipo de referencias.

En la sección 1 se describe el estado del arte de los sistemas QA en dominios no restringidos. En la sección 2 se analiza la importancia de las referencias pronominales en varias colecciones de documentos. Posteriormente, en la sección 3 se detalla nuestra aproximación y los componentes del sistema implementado. Por último, la sección 4 presenta los resultados comparativos obtenidos al evaluar el sistema.

## *1 Antecedentes*

El interés sobre sistemas QA en dominios no restringidos es bastante reciente. Se disponía de poca información acerca de esta clase de sistemas hasta que en el último congreso TREC (TREC-8, 1999) se diseñó una tarea específica para evaluar este tipo de sistemas (The First Question Answering Track). En este congreso se evaluaron cerca de veinte sistemas diferentes obteniendo porcentajes de éxito muy diferentes. Podemos clasificar las aproximaciones actuales en dos grupos: sistemas *de extracción de zonas de documentos (text-snippet extraction systems)* y sistemas de *extracción de sintagmas nominales (Noun-Phrase extraction systems).*

Las aproximaciones *text-snippet* tienen como objetivo la localización y extracción de las oraciones o párrafos más relevantes a la pregunta suponiendo que este texto contendrá la respuesta correcta a la misma. Esta aproximación ha sido la más empleada por los participantes en el referido congreso. Algunos ejemplos de estos sistemas son (Moldovan *et al*., 1999) (Singhal *et al*., 1999) (Prager *et al*., 1999) (Takaki, 1999) (Hull, 1999) (Cormack *et al*., 1999).

Tras revisar estas aproximaciones, hemos observado que hay un acuerdo tácito general acerca de la importancia de varias técnicas de Procesamiento del Lenguaje Natural (PLN) para tareas de QA. El Etiquetado de las categorías gramaticales de las palabras (POS-tagging), el análisis sintáctico (parsing) y los sistemas de clasificación de entidades (Name-Entity tagging) son utilizados por la mayoría de los sistemas. Sin embargo, pocos sistemas aplican otras técnicas de PLN. Particularmente, sólo cuatro sistemas modelan algún tipo de relación de correferencia (Morton, 1999) (Breck *et al*., 1999) (Oard *et al*., 1999) (Humphreys *et al*., 1999). Como ejemplo, la aproximación presentada por Morton (1999) modela las relaciones de identidad, los sintagmas nominales definidos y los pronombres no posesivos de tercera persona. No obstante, los efectos de la aplicación de estas técnicas no han sido analizadas y evaluadas de forma separada.

El segundo grupo incluye los sistemas *Noun-Phrase*. Estas aproximaciones son más ambiciosas ya que intentan encontrar la contestación precisa requerida por preguntas cuya respuesta está definida típicamente por un Sintagma Nominal (SN).

MURAX es uno de estos sistemas (Kupiec, 1999). Puede usar información de diferentes oraciones, párrafos e incluso de documentos diferentes para determinar la respuesta (el SN más relevante) a la pregunta. Sin embargo, este sistema no tiene en cuenta la información referida mediante el uso de pronombres en los documentos, simplemente la ignora.

Nuestro sistema, pretende evaluar las mejoras obtenidas tras aplicar técnicas de resolución de la anáfora pronominal en sistemas de QA. Para resolver las anáforas pronominales, se utiliza el sistema computacional Slot Unification Parser for Anaphora Resolution – SUPAR – (Ferrández *et al*., 1999). La arquitectura de SUPAR está formada por tres módulos independientes: análisis léxico, análisis sintáctico y un módulo de resolución de problemas lingüísticos (anáfora pronominal, extraposición de elementos, elipsis, etc.).

Para evaluar nuestra aproximación se han implementado un sistema de Recuperación de Información (RI) y un sistema QA de extracción de oraciones. Ambos están basados en la aproximación de Salton (1989). Una vez que el sistema RI recupera los documentos relevantes, nuestro sistema QA procesa estos documentos "con" y "sin" resolver las referencias pronominales con el objetivo de comparar el rendimiento de ambas aproximaciones.

Como se demostrará, la resolución de la anáfora pronominal mejora notablemente el rendimiento de los sistemas de QA. Por esta razón, creemos que esta técnica debería ser incluida en el grupo de técnicas de PLN importantes para tareas de QA.

## *2 Importancia de la información pronominal en los documentos*

Para medir la importancia de la información que aparece en los documentos en forma de pronombres, hemos analizado varias colecciones de textos que se han utilizado para la tarea de QA en el TREC-8 así como otras usadas frecuentemente para la evaluación de sistemas de RI. Estas colecciones son las siguientes: Los Angeles Times (LAT), Federal Register (FR), Financial Times (FT), Federal Bureau Information Service (FBIS), TIME, CRANFIELD, CISI, CACM, MED y LISA. Este análisis consiste en determinar la cantidad y tipo de pronombres que aparecen en cada una de ellas. Como medida de la cantidad de información que se oculta tras las referencias pronominales, hemos utilizado la razón entre la cantidad de pronombres y el número total de oraciones que aparecen en cada colección. Esta medida (a la que denominamos Nivel de Referencia Pronominal - NRP) proporciona el nivel de información que se ignora si estas referencias no son resueltas. La Figura 1 muestra los resultados obtenidos en este análisis.

| COLECCIONES | LAT | FBIS | TIME | CISI | FT | LISA | MED | FR | CACM | CRANFIELD |
|---|---|---|---|---|---|---|---|---|---|---|
| HE, SHE, THEY | 34,51% | 26,51% | 27,28% | 14,82% | 24,38% | 12,22% | 13,69% | 11,59% | 8,73% | 6,20% |
| IT, ITS | 24,07% | 36,00% | 19,63% | 49,92% | 43,43% | 45,72% | 49,72% | 53,26% | 66,17% | 78,80% |
| HIS, HER, THEIR | 23,11% | 19,58% | 29,90% | 21,88% | 19,09% | 29,05% | 18,06% | 18,03% | 15,20% | 9,07% |
| HIM, THEM | 6,30% | 6,44% | 9,12% | 4,17% | 4,13% | 5,15% | 4,10% | 3,00% | 2,81% | 1,35% |
| HIM, HER,IT(SELF), THEMSELVES | 1,44% | 2,39% | 2,81% | 1,98% | 2,02% | 1,65% | 1,30% | 1,30% | 2,34% | 1,35% |
| WHO, WHOSE, WHOM | 10,58% | 9,07% | 11,26% | 7,23% | 6,95% | 6,21% | 13,13% | 12,82% | 4,75% | 3,23% |
| **Número de Pronombres** | **2.433.732** | **1.563.035** | **7.874** | **2.422** | **2.003.718** | **4.541** | **1.074** | **246.249** | **1.283** | **1.113** |
| Número de Frases | 3.354.108 | 2.623.506 | 13.757 | 4.875 | 4.073.700 | 14.447 | 5.602 | 1.376.422 | 7.916 | 11.200 |
| **Razón N.Pronombres / N.Frases** | **72,56%** | **59,58%** | **57,24%** | **49,68%** | **49,19%** | **31,43%** | **19,17%** | **17,89%** | **16,21%** | **9,94%** |

Figura 1. Referencias pronominales en las colecciones de textos.

Como se puede observar, la cantidad y tipo de pronombres que aparecen en las colecciones analizadas varía dependiendo del tema del que tratan los documentos. LAT, FT, FBIS y TIME son colecciones compuestas de artículos publicados en diferentes periódicos. La cantidad de pronombres que utilizan estas colecciones en relación al número de frases que las componen es muy elevada (desde 49'19% hasta 72'56%). Estos documentos contienen un gran número de referencias pronominales en tercera persona – los pronombres en castellano: *él* (*he*), *ella* (*she*), *ellos/ellas* (*they*) y *su* (*his/her/their*) – cuyos antecedentes son principalmente nombres de personas. En este tipo de documentos la resolución de la anáfora pronominal parece ser muy necesaria para una correcta modelización de las relaciones entre las entidades. La colección CISI presenta un NRP similar al de las colecciones anteriores. Este conjunto de documentos está compuesto por extractos y comentarios generales acerca de gestión, clasificación, indexación, etc. de documentos. Aunque el NRP presentado por esta colección (48'68%) es también muy alto, el grupo más importante de referencias pronominales está formado por los pronombres *it* e *its*. En este caso, los antecedentes de estas referencias son principalmente conceptos representados básicamente por SN. De nuevo parece importante resolver estas referencias para modelar correctamente las relaciones entre los conceptos expresados por SN. El NRP más bajo aparece al analizar la colección CRANFIELD con un 9'94%. Este bajo nivel es debido al contenido de los textos. Esta colección está compuesta por extractos de temas muy técnicos. Entre los porcentajes descritos encontramos las colecciones FR, CACM, MED y LISA. Estas colecciones están formadas por resúmenes y documentos extraídos del Registro Federal, de la revista CACM, de revistas médicas y de "Library and Information Science Abstracts" respectivamente. Como regla general podemos observar que a medida que los contenidos de los documentos se vuelven más técnicos, los pronombres *it* e *its* se convierten en los más utilizados y el NRP decrece.

Después de analizar estos resultados nos podemos preguntar: ¿Merece la pena resolver las referencias pronominales en los documentos? Parece razonable pensar que la resolución de la anáfora pronominal sólo se debería llevar a cabo cuando el NRP de los textos a tratar excediera un nivel mínimo. Sin embargo, hemos de tener en cuenta que el coste de resolver estas referencias es proporcional al número de pronombres analizados y consecuentemente, proporcional a la cantidad de información que el sistema ignorará si no se resuelven. De cualquier forma, la evaluación realizada y su posterior análisis contribuyen con datos empíricos a recomendar la aplicación de técnicas de resolución de la anáfora pronominal en tareas de QA.

## 3 *Nuestra aproximación*

Nuestro sistema se compone de tres módulos. El primero es un sistema RI que recupera, en una lista ordenada, aquellos documentos que son relevantes a cada pregunta. El segundo módulo tratará la resolución de la anáfora en las preguntas y los documentos recuperados. Para este propósito hemos utilizado el sistema SUPAR (sección 3.1). El tercer módulo es un sistema de QA basado en la extracción de oraciones que utiliza el resultado de SUPAR y ordena las oraciones de los documentos recuperados con el fin de localizar la respuesta correcta (sección 3.2).

Para la evaluación, se ha implementado un sistema de RI. Este sistema está basado en la aproximación de recuperación de información documentos descrita en Salton (1989). La misma aproximación se ha empleado como sistema de referencia (baseline) para la tarea de QA, pero utilizando las oraciones como unidades de texto. Los términos son procesados previo a su indexación y comparación mediante la aplicación de una versión del *stemmer* de Porter. Cada término de las preguntas y los documentos se pondera asignándoles el valor inverso de la frecuencia del término en los documentos de la colección (*idf – inverse document frequency*). Este valor se calcula de la siguiente forma:

$$idf(t) = \log(\frac{N}{df(t)})$$

donde $N$ es el número total de documentos de la colección y $df(t)$ es el número de documentos que contienen el término $t$. La similitud entre los documentos (para RI) y las frases (para QA) a cada pregunta se calcula aplicando la medida de similitud del coseno. Se ha seleccionado el corpus LAT como colección para realizar las pruebas debido a su alto nivel de referencias pronominales.

## 3.1 Resolución de la anáfora pronominal

En esta sección se describe, a grandes rasgos, el sistema de PLN Slot Unification Parser for Anaphora Resolution – SUPAR – (Ferrández *et al.*, 1999). La arquitectura de SUPAR está formada por tres módulos independientes que interactúan entre sí. Estos módulos son: análisis léxico, análisis sintáctico y un módulo de resolución de problemas lingüísticos.

**Módulo de análisis léxico.** Este módulo toma como entrada cada oración a analizar junto con una herramienta que proporciona al sistema toda la información léxica de cada una de las palabras de la oración. Esta herramienta puede ser un diccionario o un etiquetador de categorías gramaticales. Como salida, este módulo devuelve una lista con toda la información necesaria para los módulos restantes. SUPAR procesa el texto de entrada "oración por oración", pero almacena información de las oraciones anteriores que será usada en otros módulos (por ejemplo, la lista de antecedentes de las oraciones anteriores para la resolución de la anáfora).

**Módulo de análisis sintáctico.** Este módulo toma como entrada la salida del módulo de análisis léxico y la información sintáctica representada por medio del formalismo gramatical Slot Unification Grammar – SUG – (Ferrández *et al.*, 1997). Como salida, devuelve lo que se denomina una Estructura de Huecos (Slot Structure - SS) que almacena toda la información necesaria para los módulos siguientes. Una de las principales ventajas de este sistema es que permite realizar un análisis parcial o completo del texto.

**Módulo de resolución de problemas de PLN.** En este módulo se tratan los problemas de PLN (anáfora, extraposición de elementos, elipsis, etc.). Toma como entrada la SS que corresponde a la oración analizada. La salida es una nueva SS en la que se han resuelto todas las anáforas. Las clases de conocimiento que se han usado en la resolución de la anáfora son: etiquetador, análisis parcial, conocimiento estadístico, restricciones de c-dominio y de concordancia morfológica y varias heurísticas tales como paralelismo sintáctico o preferencia por los SN en la misma oración que el pronombre.

Hay que destacar que cuando trabajamos con textos de dominios no restringidos (como ocurre en este artículo) no se utiliza conocimiento semántico (aunque en el futuro se pretende utilizar herramientas como WordNet). Actualmente, SUPAR resuelve la anáfora pronominal para español e inglés con un porcentaje de éxito del 84% y 87% respectivamente.

## 3.2 Resolución de la anáfora y QA

Nuestra aproximación de QA proporciona un segundo nivel de procesamiento para los textos relevantes recuperados con el sistema de RI: Análisis de documentos relevantes y ordenación de las oraciones.

**Análisis de documentos relevantes.** Este paso se aplica sobre los mejores documentos recuperados del sistema RI. Estos documentos son analizados por el módulo SUPAR que resuelve las anáforas pronominales. Como resultado, a cada pronombre que aparece en los documentos, se le asocia el SN al que hace referencia. Posteriormente, los documentos son divididos en oraciones como unidades de texto básicas para la tarea de QA. Este conjunto de oraciones se procesa en la etapa de ordenación de oraciones.

**Ordenación de oraciones.** A cada término de la pregunta se le asigna un peso. Este peso es el valor *idf* descrito anteriormente que se ha calculado sobre la base de su ocurrencia en la colección LAT. Los términos de las frases de los documentos se valoran de la misma forma. La única diferencia con respecto al sistema de referencia (baseline) es que a los pronombres se les asigna el peso de la entidad a la que se refieren. Como queremos analizar los efectos de la resolución de la anáfora pronominal no se introduce ningún cambio más en el esquema de asignación de valores. Para medir la similitud entre cada frase y la pregunta se ha utilizado la medida del coseno.

## 4 *Evaluación*

Para evaluar nuestra aproximación, personas no implicadas en este trabajo propusieron 150 preguntas cuyas respuestas correctas aparecían al menos una vez en la colección analizada. El proceso de evaluación fue el siguiente:

En primer lugar, se recuperaron los documentos relevantes para cada pregunta usando el sistema de RI descrito anteriormente. Únicamente los 50 documentos más relevantes fueron seleccionados para la evaluación de QA. El documento que contenía la respuesta correcta apareció dentro de los 50 más relevantes sólo para 93 preguntas (un 62% de las preguntas propuestas). Por ello, las restantes 57 preguntas se descartaron y no se tuvieron en cuenta para la evaluación del sistema de QA.

Una vez recuperados los documentos relevantes para cada pregunta, el sistema SUPAR efectúa la resolución de la anáfora en estos documentos. Finalmente se procede con la ordenación de las frases según se describe en la sección 3.2. Como resultado el sistema presenta una lista ordenada que contiene las 10 oraciones más relevantes a cada pregunta.

Con el fin de facilitar el análisis de los resultados de la evaluación, las preguntas se clasificaron en tres grupos dependiendo de las siguientes características:

- Grupo A. No aparecen pronombres en la oración objetivo (oración que contiene la respuesta correcta).
- Grupo B. La información requerida como respuesta se encuentra indicada mediante una anáfora pronominal en la oración objetivo.
- Grupo C. Algunos términos de la pregunta aparecen referidos mediante pronombres en la oración objetivo.

El Grupo A constaba de 37 preguntas y los Grupos B y C contenían 25 y 31 preguntas respectivamente.

La Figura 2 muestra los resultados de la evaluación. En ella aparecen el número de oraciones objetivo que aparecieron entre las 10 oraciones más relevantes devueltas por el sistema para cada pregunta y por otra parte, el número de estas oraciones que podían considerarse respuestas correctas. Una respuesta se considera correcta si el usuario puede conocerla simplemente leyendo la oración objetivo. Los resultados se presentan clasificados según el tipo de pregunta antes mencionado. La segunda columna de la tabla indica el número de preguntas de cada grupo. La tercera y cuarta columnas muestran los resultados del sistema de referencia (sin resolver la anáfora). La quinta y sexta columnas muestran los resultados obtenidos cuando las referencias pronominales se resuelven.

Los resultados muestran varios aspectos a tener en cuenta. Las mejoras obtenidas tras aplicar la resolución de la anáfora pronominal varían dependiendo del tipo de pregunta. Los resultados para las preguntas de los Grupos A y B indican que la relevancia a la pregunta es la misma que la obtenida por el sistema de referencia. Por ello, parece que la resolución de la anáfora pronominal no aporta ninguna mejora al sistema. Esto es cierto sólo para las preguntas del Grupo A. Aunque las oraciones objetivo se ordenan de un modo similar, para las preguntas del Grupo B las oraciones devueltas por el sistema de referencia no se pueden considerar correctas porque la respuesta no aparece textualmente en ellas. La respuesta correcta aparece en la oración objetivo sólo si se resuelve la anáfora pronominal y se sustituye por el SN al que hace referencia. Sólo de esta forma el usuario no necesitará leer más texto para conocer la respuesta correcta.

| Tipo de Pregunta | Cantidad | | Sistema de Referencia Objetivo Incluida | | Respuesta Correcta | | Anáfora resuelta Objetivo Incluida | | Respuesta Correcta | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 37 | (39,78%) | 18 | (48,65%) | 18 | (48,65%) | 18 | (48,65%) | 18 | (48,65%) |
| B | 25 | (26,88%) | 12 | (48,00%) | 0 | (0,00%) | 12 | (48,00%) | 12 | (48,00%) |
| C | 31 | (33,33%) | 9 | (29,03%) | 9 | (29,03%) | 21 | (67,74%) | 21 | (67,74%) |
| **A+B+C** | **93** | **(100,00%)** | **39** | **(41,94%)** | **27** | **(29,03%)** | **51** | **(54,84%)** | **51** | **(54,84%)** |

Figura 2. Resultados de la evaluación.

Analizando los resultados de las preguntas del Grupo C, se observa que el rendimiento mejora de nuevo. Estas preguntas tienen las siguientes características: algunos de los términos de las preguntas aparecen referidos mediante pronombres en la oración objetivo. Cuando esta situación se produce, las oraciones objetivo aparecen en posiciones más elevadas – en la lista final de oraciones ordenadas si resolvemos las referencias pronominales – que en el sistema de referencia (donde no se han resuelto). Esta mejora es debida a que la similitud entre la pregunta y la oración objetivo aumenta cuando los pronombres se valoran igual que los términos a los que se refieren. El porcentaje de las oraciones objetivo obtenidas aumenta 38’71 puntos (de 29’03% a 67’74%).

Por último, en la última fila de la Figura 2 se presentan las mejoras obtenidas para el conjunto del sistema (todas las preguntas). El porcentaje general de las oraciones objetivo localizadas aumenta en 12’9 puntos (del 41’94% al 54’84%) y el nivel de respuestas correctas devuelto por el sistema aumenta en 25’81 puntos (del 29’03% al 54’84%).

Según se desprende del análisis de los resultados, podemos concluir que la resolución de la anáfora pronominal mejora el rendimiento de los sistemas de QA en dos aspectos principales. En primer lugar, la precisión aumenta cuando los términos de la pregunta están referidos anafóricamente en la oración objetivo. En segundo lugar, la resolución de la anáfora pronominal reduce la cantidad de texto que un usuario tiene que leer cuando en la oración objetivo se han sustituido las referencias pronominales por los SN a los que se refieren.

## *Conclusiones y trabajos futuros*

En este trabajo se han analizado y medido los efectos producidos en los sistemas de QA la resolución de la anáfora pronominal. La aplicación de esta técnica ha mejorado considerablemente el rendimiento de estos sistemas y en algunos casos, ha sido esencial para obtener fácilmente la información deseada. Los resultados obtenidos nos inducen a afirmar que, el análisis de la información referida mediante pronombres es importante sobre todo, cuando se realizan tareas en las que se requiere un alto nivel de precisión.

Mientras se desarrollaba la investigación han aparecido tres áreas importantes de trabajo futuro: 1) El sistema RI usado para la recuperación de documentos tiene que ser adaptado para tareas de QA. El sistema actual de RI sólo proporciona el documento que contenía la oración objetivo para 93 de las 150 preguntas propuestas. Por lo tanto, su precisión debe ser mejorada. 2) Debemos extender el algoritmo de resolución de la anáfora para el tratamiento de otros tipos de anáfora, tales como descripciones definidas, anáfora superficial numérica, anáfora adjetiva, etc. 3) Se debe reconsiderar la medida de similitud entre preguntas y frases con el fin de maximizar el porcentaje de las oraciones objetivo incluidas en las 10 oraciones respuesta proporcionadas por el sistema.

## *Bibliografía*

Breck E., Burger J., Ferro L., House D., Light M., Mani. I.(1999) A Sys Called Quanda, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8).* http://trec.nist.gov/pubs/ trec8/ t8_proceedings.html.

Cormack G. V., Clarke C. L. A., Palmer C. R., Kisman D. I. E.(1999). Fast Automatic Passage Ranking (MultiText Experiments for TREC-8), In *Proceedings of the Eighth Text REtrieval Conference*

*(TREC-8)*. http://trec.nist.gov/pubs/trec8/ t8_proceedings.html.

Ferrández A., Palomar M. and Moreno L. (1997) Slot Unification Grammar. In *Proceedings of the Joint Conference on Declarative Programming. APPIA-GULP-PRODE*. Grado (Italy).

Ferrández, A., Palomar, M., Moreno, L. (1999) An empirical approach to Spanish anaphora resolution. To appear in *Machine Translation*. 26 pp.

Hull D. A. (1999). Xerox TREC-8 Question Answering Track Report, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist.gov/pubs/trec8/ t8_proceedings.html.

Humphreys K., Gaizauskas R., Hepple M., Sanderson M.(1999). University of Sheffield TREC-8 Q&A System, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist.gov/pubs/trec8/ t8_proceedings.html.

Kupiec, J. (1999). MURAX: Finding and Organising answers from Text Search. In *Natural Language Information Retrieval*. Kluwer Academic Publishers. T.Strzalkowski (ed). pp. 311-331.

Moldovan D., Harabagiu S., Pasca M., Mihalcea R., Goodrum R., Gîrju R., Rus V.(1999). LASSO: A Tool for Surfing the Answer Net, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist.gov/pubs/trec8/ t8_proceedings.html.

Morton T.(1999) Using coreference in Question Answering, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist. gov/pubs/trec8/ t8_proceedings.html.

Oard D., Wang J., Lin D., Soboroff I.(1999). TREC-8 Experiments at Maryland: CLIR, QA and Routing, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist. gov/pubs/trec8/ t8_proceedings.html.

Prager J., Radev D., Brown E., Coden A., Samn V.(1999). The Use of Predictive Annotation for Question Answering, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec. nist.gov/pubs/trec8/ t8_proceedings.html.

Salton G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley Publishing, New York.

Singhal A., Abney S., Bacchiani M., Collins M., Hindle D., Pereira F.(1999). ATT at TREC-8, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist.gov/pubs/ trec8/t8_proceedings.html.

Takaki T.(1999). NTT DATA: Overview of system approach at TREC-8 ad-hoc and question answering, In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist. gov/pubs/trec8/t8_proceedings.html.

TREC-8.(1999). *Eighth Text REtrieval Conference (TREC-8)*. http://trec.nist.gov/pubs/trec8/ t8_proceedings.html.