

La Búsqueda de Respuestas: Estado Actual y Perspectivas de Futuro*

José Luis Vicedo
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Carretera San Vicente s/n
03080 Alicante
vicedo@dlsi.ua.es

Resumen

La creciente demanda de sistemas que respondan de forma precisa y escueta a las necesidades de información de los usuarios ha potenciado la aparición de un nuevo campo de investigación: la *Búsqueda de Respuestas* (BR). El objetivo de la investigación en este campo va mucho más allá de la simple localización de documentos relevantes realizadas por los ya tradicionales sistemas de *Recuperación de Información* (RI). Los sistemas de BR afrontan el problema de localizar, extraer y presentar al usuario única y exclusivamente aquella información que desea conocer, evitando así la ardua tarea de recopilar información a través de la lectura de documentos relevantes. Este trabajo realiza un amplio estudio del problema de la BR y presenta el estado actual de las investigaciones a través de su evolución histórica y el análisis de las diferentes tendencias.

Abstract

The increasing demand of systems that respond in a precise way users information needs has fostered a new investigation field: *Question Answering* (QA). The objective of this investigation goes beyond the retrieval of relevant documents carried out by traditional *Information Retrieval* systems (IR). QA systems tackle the problem of locating, extracting and presenting to the user just only the required information avoiding, this way, the arduous task of reading relevant documents. This work carries out a wide study of QA problem and presents the state of the art in this field through the revision of current approaches.

1 ¿Qué es un sistema de Búsqueda de Respuestas?

Los últimos veinte años hemos asistido a un crecimiento exponencial de la cantidad de in-

formación digital disponible y a la explosión de las comunicaciones entre ordenadores como vía principal de transmisión de información entre usuarios. La gran cantidad de información disponible, unido al creciente número de usuarios finales (no especialistas en tratamiento de datos ni en computadores) que disponen de acceso directo a dicha información a través de ordenadores personales, ha impulsado la investigación en sistemas de información textual que faciliten

* *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. No. 20 (2003), pp. 34-52. (ISSN: 1137-3601). (c) AEPIA (<http://www.aepia.org/>).*

la localización, acceso y tratamiento de toda esta ingente cantidad de datos.

Actualmente, los sistemas más conocidos que nos permiten realizar estas tareas son los motores de búsqueda de documentos en Internet. Sin embargo, el tiempo empleado por los usuarios en navegar a través de la red en busca de aquella información que les es útil viene generalmente acompañado de un esfuerzo aún mayor. Una vez recuperados los documentos, el usuario necesita seleccionar aquellos cuyo contenido está relacionado con la información solicitada y posteriormente, indagar en ellos hasta localizar la información deseada, lo que puede suponer una carga de tiempo y trabajo muy importante en algunas ocasiones.

Este problema ha provocado una creciente necesidad de sistemas que permitan a un usuario formular sus necesidades de información empleando un lenguaje cotidiano y recibir, como contestación, una respuesta rápida, precisa y escueta (junto al contexto necesario que permita su validación) que satisfaga dichas necesidades. Precisamente éste es el objetivo de los sistemas de Búsqueda de Respuestas (BR) o *Question Answering*.

Las investigaciones en sistemas Búsqueda de Respuestas se están desarrollando a una velocidad vertiginosa gracias a la combinación de dos factores principales: la creciente demanda de este tipo de sistemas y la organización de una tarea para la evaluación de este tipo de sistemas en el ámbito de las conferencias *Text REtrieval Conference* (TREC¹) en cuyas actas queda patente tanto el progreso de las investigaciones en este campo como los resultados alcanzados por estos sistemas. Una medida de este rápido progreso se fundamenta en el hecho de que actualmente, los mejores sistemas son capaces de contestar correctamente más de dos tercios de las preguntas formuladas en estas evaluaciones.

Este trabajo se ha organizado del siguiente modo. En primer lugar se presenta el contexto histórico en el que desarrolla la investigación en sistemas de BR y se introducen las principales características de este tipo de sistemas. A continuación, se realiza un análisis detallado del problema y de sus diferentes dimensiones que desemboca en un análisis detallado de la

situación actual de las investigaciones. A continuación se analiza la problemática referente a la evaluación de este tipo de sistemas y se revisan los progresos conseguidos en el ámbito de las evaluaciones desarrolladas en las conferencias TREC. Para finalizar, presentaremos las principales líneas de investigación abiertas y las perspectivas de futuro en este campo.

2 Contexto histórico

Generalmente, cuando un usuario emplea un ordenador para buscar una información determinada, lo que realmente está intentando es encontrar respuesta a sus necesidades de información. Para facilitar esta tarea, se necesitaría disponer de sistemas -llamémosles "ideales"- que fuesen capaces de localizar la información requerida, procesarla, integrarla y generar una respuesta acorde a los requerimientos expresados por el usuario. Además, estos sistemas deberían ser capaces de comprender preguntas y documentos escritos en lenguaje natural en dominios no restringidos permitiendo así, una interacción cómoda y adecuada a aquellos usuarios inexpertos en el manejo de computadores. Sin embargo, y aunque las investigaciones avanzan en buena dirección, todavía no existe hoy ningún sistema operacional que cumpla todos estos requisitos.

De todas formas, ante la creciente necesidad de aplicaciones que facilitaran -al menos en parte- el acceso y tratamiento de grandes cantidades de información, la comunidad científica concentró sus esfuerzos en la resolución de problemas más especializados y por ello, más fácilmente abordables. Esta circunstancia propició el desarrollo de campos de investigación que afrontaron el problema desde diferentes puntos de vista: la *recuperación de información* (RI), la *extracción de información* (EI) y, posteriormente, la *búsqueda de respuestas* (BR).

2.1 La recuperación de información

Los sistemas de recuperación de información (RI) realizan las tareas de seleccionar y recuperar aquellos documentos que son relevantes a necesidades de información arbitrarias formula-

¹ <http://trec.nist.gov>

das por los usuarios. Como resultado, estos sistemas devuelven una lista de documentos que suele presentarse ordenada, en función de valores que intentan reflejar en qué medida cada documento contiene la información que responde a las necesidades expresadas por el usuario.

Los sistemas de RI más conocidos son aquellos que permiten -con mayor o menor éxito- localizar información a través de Internet. Sirvan como ejemplo algunos de los motores de búsqueda más utilizados actualmente como Google², Alta Vista³ o Yahoo⁴.

Una de las características de estos sistemas reside en la necesidad de procesar grandes cantidades de texto en un tiempo muy corto (del orden de milisegundos para búsquedas en Internet). Esta limitación impone una severa restricción en cuanto a la complejidad de modelos y técnicas de análisis de documentos que pueden emplearse.

Dentro del ámbito de la RI podemos destacar la aparición de dos líneas de investigación orientadas a mejorar el rendimiento de estos sistemas: La recuperación de pasajes (RP) (Callan, 1994; Kaszkiel *et al.*, 1999) y la aplicación de técnicas de procesamiento del lenguaje natural (PLN) al proceso de RI (Strzalkowski *et al.*, 1998).

Uno de los principales foros de investigación en sistemas de RI lo constituye la serie anual de conferencias TREC cuyas actas reflejan con detalle la evolución de las investigaciones desarrolladas en este campo.

2.2 La extracción de información

Los sistemas de extracción de información (EI) realizan la tarea de buscar información muy concreta en colecciones de documentos. Su finalidad consiste en detectar, extraer y presentar dicha información en un formato que sea susceptible de ser tratado posteriormente de forma automática.

Estos sistemas se diseñan y construyen de forma específica para la realización de una tarea determinada, y en consecuencia, dispondremos de un sistema diferente en función del tipo de

información a extraer en cada caso. Un ejemplo podría ser un sistema diseñado para extraer el nombre, DNI y las direcciones de las personas contratantes que aparecen en documentos notariales. Este sistema operaría de forma que cada vez que detectara uno de estos datos, lo extraería y lo incorporaría en el campo correspondiente de una base de datos creada a tal efecto. Como puede deducirse, estos sistemas necesitan aplicar técnicas complejas de PLN dada la gran precisión que se requiere para la detección y extracción del tipo de información que les es relevante.

La investigación en este campo ha sido muy intensa. En particular, la serie de conferencias Message Understanding Conference (MUC) han constituido uno de sus principales foros de promoción (Cinchor, 1998). Estas conferencias han permitido la evaluación y comparación de diversos sistemas, realizando para la EI la misma función que las conferencias TREC en el ámbito de la recuperación de información.

2.3 La búsqueda de respuestas

La investigación en sistemas de RI y EI facilitó el tratamiento de grandes cantidades de información, sin embargo, las características que definieron estas líneas de investigación presentaban serios inconvenientes a la hora de facilitar la obtención de respuestas concretas a preguntas muy precisas formuladas arbitrariamente por los usuarios.

Por una parte, los sistemas de RI se vieron incapaces por sí solos de afrontar tareas de este tipo. De hecho, una vez que el usuario recibía la lista de documentos relevantes a su pregunta, debía revisarlos manualmente para localizar en su interior la información puntual deseada.

Por otra parte, y aunque los sistemas de EI son mucho más precisos en la tarea de encontrar información concreta en documentos, estos sistemas no permiten el tratamiento de preguntas arbitrarias sino que el tipo de información requerida necesita ser definida de forma previa a la implementación del sistema.

Todos estos inconvenientes y principalmente, un creciente interés en sistemas que afrontaran con éxito la tarea de localizar datos precisos en grandes volúmenes de información textual,

² <http://www.google.com/>

³ <http://www.altavista.com/>

⁴ <http://www.yahoo.com/>

dejaron la puerta abierta a la aparición de un nuevo campo de investigación: la búsqueda de respuestas.

3 Los sistemas de búsqueda de respuestas

Se puede definir la búsqueda de respuestas (BR) como aquella tarea automática realizada por ordenadores que tiene como finalidad la de encontrar respuestas concretas a necesidades precisas y arbitrarias de información formuladas por los usuarios. Los sistemas de BR son especialmente útiles en situaciones en las que el usuario final necesita conocer un dato muy específico y no dispone de tiempo -o no necesita- leer toda la documentación referente al tema de la búsqueda para solucionar su problema. A modo de ejemplo, algunas aplicaciones prácticas podrían ser las siguientes:

- Sistemas de ayuda en línea de software.
- Sistemas de consulta de procedimientos y datos en grandes organizaciones.
- Interfaces de consulta de manuales técnicos.
- Sistemas de consulta de bases de datos textuales de todo tipo (financieras, legales, de noticias, ...).
- etc.

La primera discusión acerca de las características de un sistema de BR y la primera aproximación a un sistema funcional (QUALM) fueron introducidos por Wendy Lehnert a finales de los 70 (Lehnert, 1977; Lehnert, 1980). En estos trabajos se definieron las características ideales de un sistema de BR. Estos sistemas deberían entender la pregunta del usuario, buscar la respuesta en una base de datos de conocimiento y posteriormente, componer la respuesta para presentarla al usuario. En consecuencia, estos sistemas deberían integrar técnicas relacionadas con el Entendimiento del Lenguaje Natural, la Búsqueda de Conocimiento (incluyendo posiblemente técnicas de inferencia) y la Generación de Lenguaje Natural.

La investigación en sistemas de BR tuvo sus inicios en la comunidad científica relacionada con la Inteligencia Artificial (IA). Desde esta perspectiva, la investigación desarrollada consideró requisito indispensable que los sistemas de BR tenían que satisfacer todas y cada una de las características ideales anteriormente citadas. Sin embargo, hasta la fecha únicamente se han podido obtener algunos resultados a costa de restringir mucho los dominios sobre los que se realizan las consultas. El trabajo de Levine muestra con detalle este tipo de aproximaciones (Levine & Fedder, 1989).

Recientemente, la investigación en sistemas de BR también se ha afrontado desde el punto de vista de la comunidad especializada en sistemas de RI. Sin embargo, desde esta perspectiva, el poder desarrollar la tarea sobre dominios no restringidos constituye el requisito básico e innegociable a cumplir. Partiendo de este requerimiento inicial, las investigaciones se han orientado hacia el desarrollo de sistemas que van incorporando progresivamente herramientas más complejas que permiten la convergencia de estos sistemas hacia las propuestas de Lehner.

El interés en sistemas de BR por parte de la comunidad científica tradicionalmente dedicada a la RI es bastante reciente. A mitad de los 90 se presentó MURAX (Kupiec, 1993), el primer sistema de BR que combinó técnicas tradicionales de RI con técnicas superficiales de PLN para conseguir lo que su autor denominó una recuperación de información de alta precisión.

La investigación en sistemas de BR en dominios no restringidos vive actualmente momentos de gran auge. De hecho, ya existen algunos sistemas de estas características que son accesibles a través de Internet, como por ejemplo START⁵ o IO⁶.

3.1 Arquitectura Básica de un sistema de BR

El análisis de algunas de las aproximaciones actuales más relevantes (Moldovan *et al.*, 2002; Soubbotin & Soubbotin, 2002; Yang & Chua, 2002; Magnini *et al.*, 2002) permite identificar

⁵<http://www.ai.mit.edu/projects/infolab/globe.html>

⁶<http://www.ionaut.com:8400/>

los componentes principales de un sistema de BR:

1. Análisis de la pregunta.
2. Selección de documentos o pasajes.
3. Extracción de respuestas.

Estos componentes se relacionan entre sí procesando preguntas y documentos en diferentes niveles hasta obtener la respuesta. La figura 1 muestra gráficamente la secuencia de ejecución de estos procesos.

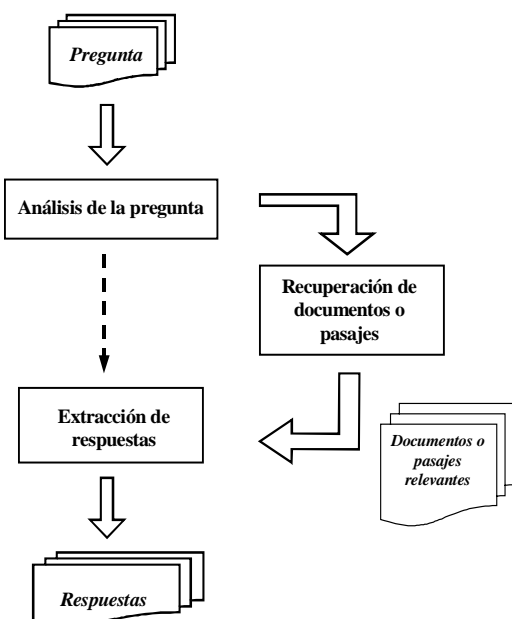


Figura 1: **Arquitectura básica de un sistema de BR**

Las preguntas formuladas al sistema son procesadas inicialmente por el módulo de *análisis de la pregunta*. Este proceso realiza dos tareas principales: Detectar el tipo de información que la pregunta espera como respuesta (una fecha, una cantidad, etc) y seleccionar aquellos elementos de la pregunta que van a permitir la localización de los documentos susceptibles de contener la respuesta. Este proceso es de vital importancia puesto que de la calidad de la información extraída en este análisis dependerá en gran medida el rendimiento de los restantes módulos y por ende, el rendimiento del sistema.

Una parte de la información resultado del análisis de la pregunta es utilizado por el

módulo de *recuperación de documentos o pasajes* para realizar una primera selección de textos. Dado el gran volumen de documentos a tratar por estos sistemas y las limitaciones de tiempo de respuesta con las que trabajan, esta tarea se realiza empleando sistemas de recuperación de información, generalmente orientada a la detección de extractos de texto más reducidos que el documento completo (RP). El resultado obtenido es un subconjunto muy reducido de la base de datos documental sobre el que se afrontará la extracción de la respuesta.

Finalmente, el módulo de *extracción de respuestas* se encarga de realizar un análisis más detallado del subconjunto de textos relevantes resultado del proceso anterior, con la finalidad de localizar y extraer la respuesta buscada.

Las primeras investigaciones en este campo utilizaron, como base de desarrollo, la aplicación de técnicas de RI adecuadas al proceso de BR (Cormack *et al.*, 1999; Fuller *et al.*, 1999; Allan *et al.*, 2000). Sin embargo, estas aproximaciones presentaron un pobre rendimiento en tareas en las que se requería una respuesta escueta y precisa como contestación a la pregunta.

Inmediatamente se empezó a experimentar con la aplicación de técnicas de PLN. Aunque había quedado demostrado que el uso de estas técnicas no introducían mejoras de rendimiento significativas en los sistemas de RI, esos mismos estudios habían dejado entrever que su uso podría tener mucha importancia en tareas en las que las unidades de información a analizar fuesen mucho menores al tamaño de un documento.

Este problema condujo a los investigadores a la aplicación de técnicas de PLN cada vez más complejas que permitiesen, sobre todo, mejorar la precisión de los sistemas a la hora de localizar y extraer la respuestas exacta buscada, evolucionando hacia sistemas que realizan un uso intensivo de técnicas de análisis del lenguaje natural. Desde etiquetadores léxicos, lematizadores y etiquetadores de entidades, pasando por herramientas de nivel sintáctico, como analizadores sintácticos parciales y completos, hasta llegar a complejas técnicas de análisis semántico y contextual. Este proceso se desarrolló de forma vertiginosa en el periodo 2000-2001.

Algunas de estas técnicas demostraron sobradamente su efectividad. Sobre todo, aquellas que realizan tareas encuadradas en los primeros niveles del análisis del lenguaje natural como son el análisis léxico y sintáctico. Sin embargo, los resultados obtenidos al aplicar técnicas de mayor complejidad fueron contradictorios. Aunque son varios los sistemas que aplican técnicas enmarcadas en estos niveles, solo el trabajo descrito en (Moldovan *et al.*, 2002) consigue un nivel de satisfacción que justifica el esfuerzo empleado en su aplicación.

Estas diferencias de resultados provocaron un intenso debate en torno a la aplicación eficiente de técnicas de PLN a los sistemas de BR. De hecho, aunque la comunidad científica estaba de acuerdo en la conveniencia de su aplicación, también asumía que la mejora del rendimiento del sistema no dependía directamente de la complejidad de las herramientas empleadas, sino de su correcta aplicación e integración.

“...for future, question-answering in general depends on linguistic analysis, even if this may sometimes be done by linguistically-shallow means.”

Spark-Jones, 1999

En consecuencia, y siguiendo la afirmación de Karen Spark-Jones en 1999 los últimos años hemos asistido a un cambio de tendencia orientada a la aplicación de técnicas superficiales de PLN en detrimento del uso de técnicas más complejas (Soubbotin & Soubbotin, 2002; Hermjacob *et al.*, 2002; Magnini *et al.*, 2002; Yang & Chua, 2002).

3.2 Otros campos relacionados

Además de los ya citados, se han desarrollado investigaciones en otros campos también cercanos a la búsqueda de respuestas: la Búsqueda de Preguntas Frecuentes (*Frequently Asked Questions Finding*) y el Proceso de Tests de Lectura y Comprensión de Textos (*Reading Comprehension Tests*).

3.2.1 Búsqueda de preguntas frecuentes

Los sistemas de búsqueda de preguntas frecuentes tienen como objetivo localizar y devolver pasajes de texto como respuesta a preguntas de los usuarios. Las principales diferencias entre estos sistemas y los de BR radican en las características de la base de datos documental sobre la que realizan el proceso y la forma de búsqueda de la respuesta (Burke *et al.*, 1997; Berger *et al.*, 2000).

Estos sistemas utilizan bases documentales formadas por conjuntos de preguntas que tienen asociadas sus correspondientes respuestas. Ejemplos de estas bases de datos pueden ser los conjuntos de *preguntas más frecuentes* (Frequently Asked Questions - FAQs) disponibles en Internet y que versan sobre temas muy diversos.

Realmente, estos sistemas no realizan una búsqueda de respuestas tal y como se ha definido previamente. Simplemente localizan aquellas preguntas incluidas en la base documental que son similares a la realizada por el usuario y como resultado, presentan sus correspondientes respuestas asociadas. Sistemas como FAQ Finder⁷ o Askjeeves⁸ constituyen ejemplos de algunas implementaciones que están actualmente disponibles en Internet.

3.2.2 Proceso de tests de lectura y comprensión de textos

Los tests de lectura y comprensión de textos conforman una herramienta tradicionalmente utilizada para evaluar el nivel de comprensión que un lector adquiere al leer un documento.

Un test de lectura y comprensión está formado por dos elementos: un texto en el que se narra una historia o noticia y un conjunto de preguntas de test relativas a dicha narración. La complejidad de estas preguntas suele ser creciente, lo que permite evaluar el nivel y la capacidad de comprensión del texto alcanzado por el lector mediante la comprobación de sus respuestas.

El proceso automático de este tipo de tests presenta varias vertientes de interés. La primera de ellas reside en el uso de estos tests como

⁷<http://faqfinder.cs.uchicago.edu:8001/>

⁸<http://www.askjeeves.com>

material de evaluación de sistemas automáticos de comprensión del lenguaje natural (Hirschman *et al.*, 1999; Charniak *et al.*, 2000). Su uso se está considerando como alternativa a los sistemas actuales utilizados para evaluar técnicas avanzadas de PLN. En particular, pueden utilizarse como banco de pruebas para medir el rendimiento de los sistemas de BR si bien, cabría tener en cuenta el escaso volumen de información que contienen.

Quizás el ámbito de aplicación más interesante se basa en la construcción de sistemas que permitan evaluar de forma automática el nivel de comprensión que un lector o bien, un sistema automático, alcanzan al leer un documento. Este proceso se realizaría mediante la comparación de las respuestas correctas incluidas en el test con las que suministra el lector o el sistema automático que procesa dicho test. La eficiencia de este método radica básicamente en la obtención de medidas de similitud que permitan determinar, de forma fiable, cuando ambas respuestas (la correcta y la suministrada) son equivalentes.

4 Dimensiones del problema

Para contestar una pregunta, un sistema debe analizar la pregunta (probablemente en un contexto determinado y a través de un proceso de interacción con el usuario), debe consultar una serie de recursos para localizar la respuesta y debe presentar dicha respuesta de forma adecuada a las necesidades del usuario (generalmente asociada a datos que que permitan su validación).

El análisis de cada uno de estos procesos descubre una serie de aspectos o dimensiones del problema que resulta conveniente abordar de forma previa a su resolución:

1. Usuarios
2. Preguntas
3. Respuestas
4. Nivel de conocimiento requerido

4.1 Usuarios

Seguramente, el grado de satisfacción de diferentes usuarios, ante el mejor sistema de BR disponible en la actualidad, será totalmente variable en función de las expectativas de cada uno de ellos.

Podemos encontrar un amplio espectro de usuarios que requieren diferentes capacidades del sistema para satisfacer sus necesidades de información. Estas necesidades pueden variar entre las solicitadas por un usuario casual, que interroga al sistema para la obtención de datos puntuales, y las que puede necesitar un analista profesional. Estos tipos representan los extremos de esa amplia tipología de usuarios potenciales de un sistema de BR.

Podemos clasificar los diferentes usuarios de un sistema de BR en cuatro tipos generales en función de la complejidad de sus requerimientos.

El usuario casual. Este tipo de usuario necesita información puntual acerca de hechos concretos. Realiza preguntas cuya contestación puede encontrarse en un documento expresada, generalmente, de forma simple: “¿En qué año nació Nixon?” o “¿Cuántos habitantes tiene China?”.

El recopilador de información. Este usuario realiza preguntas cuya contestación necesita de un proceso de recopilación de varias instancias de información indicadas en la pregunta. Veamos algún ejemplo: “¿Qué países visitó el Papa en 1998?” o “Dime los principales datos biográficos de Nelson Mandela”. Como puede observarse, este tipo de preguntas requiere la localización de varias informaciones (probablemente en diferentes documentos) y su posterior combinación como respuesta final.

El periodista. Imaginemos un periodista al que se le encarga la redacción de un artículo relacionado con un evento determinado, por ejemplo, un terremoto en la ciudad de Shanghai. Para ello, el reportero necesitará recabar tanto datos concretos del suceso (intensidad del terremoto, lugar del epicentro, daños materiales, ...) como informaciones anteriores más o menos relacionadas que permitan enmarcar el suceso en un contexto adecuado (terremotos anteriores en la zona, estudios sismológicos previos, predicciones, ...). En ambos casos, el sistema de BR necesitaría tener en cuenta el contexto de la se-

rie de preguntas que el usuario interpondrá al sistema. Este contexto permitiría al sistema determinar la amplitud de la búsqueda y la necesidad de profundizar en determinados aspectos del mismo.

El analista profesional. El perfil de este usuario corresponde con el de un consumidor profesional de información experto en temas concretos. Por ejemplo, analistas financieros, personal de organismos estatales especializados en política internacional, tráfico de drogas, etc.

Un ejemplo del tipo de preguntas que el sistema de BR debería de soportar a este nivel sería el siguiente. Un analista de la policía intuye que puede haber cierta conexión entre las actividades de dos grupos terroristas e intenta investigar la existencia de dicha conexión. Para ello, el analista podría realizar al sistema las siguientes preguntas: “¿Hay alguna evidencia de conexión, comunicación o contacto entre estos dos grupos terroristas o sus miembros conocidos?”, “¿Hay alguna evidencia de que estos grupos estén planeando alguna acción conjunta?”, “En caso afirmativo, ¿Cuándo y dónde?”. Un sistema de BR que trabaje a este nivel debe poder aceptar preguntas muy complejas cuyas respuestas pueden basarse en conclusiones y decisiones realizadas por el propio sistema. Estas respuestas necesitarán de la recopilación y síntesis de información obtenida en diferentes fuentes y deberá ser presentada al usuario de una forma adecuada a su forma de trabajo. Además, este sistema debería de disponer de potentes herramientas de navegación multimedia que permitieran no sólo revisar la respuesta propuesta por el sistema a través de todo el proceso de su obtención (revisión de la información de soporte, interpretaciones, conclusiones y decisiones realizadas) sino también facilitar la interacción con el usuario en cada uno de esos procesos. Esta interactividad daría como resultado una respuesta conjunta entre el sistema de BR y el analista.

4.2 Las preguntas

La experiencia nos ha demostrado que resulta difícil determinar cuales son las características que hacen que unas preguntas resulten más difíciles de contestar que otras. Esta circunstancia hace inviable una clasificación general de las mismas desde este punto de vista. Por ello,

la clasificación generalmente aceptada se fundamenta en el tipo de respuesta que requieren:

- *De hechos concretos.* Estas preguntas requieren como respuesta uno o varios datos muy específicos como fechas, nombre de entidades, cantidades, etc: “¿Cuál es la capital de Brasil?”, “¿Cuándo murió Bob Marley?”.
- *De resumen.* Estas preguntas necesitan localizar instancias de información relacionadas con la pregunta y resumirlas para su presentación final. Por ejemplo, “¿Qué asuntos trató George Bush en su última visita a Alemania?”.
- *De opinión.* Corresponden a preguntas muy complejas que requieren de la recopilación de datos y de la aplicación de técnicas de deducción en base a ellos: “¿Qué podría pasar en Marruecos si el rey Hassan II es asesinado?”.

Las investigaciones en BR se están centrando actualmente en el tratamiento de preguntas *de hechos concretos*. A su vez, este tipo de preguntas se pueden organizar, en cuatro subtipos: (1) *preguntas si/no* que esperan la confirmación positiva o negativa de la pregunta (“¿Es el Everest el monte más alto del mundo?”), (2) *preguntas con términos interrogativos* que incluyen términos como “quién” o “dónde” que claramente indican el tipo de respuesta que esperan, (3) *preguntas indirectas* como “Me gustaría saber el nombre del inventor de la radio” y finalmente, (4) *preguntas de requerimiento*, por ejemplo “Dime los países que visitó el Papa en 1996”.

Esta clasificación resulta muy importante desde el punto de vista de la efectividad del proceso de BR dado que, un correcto análisis de la pregunta permitirá reducir considerablemente el espacio de respuestas posibles a considerar por el sistema en el proceso de localización de la misma. En (Light *et al.*, 2001) se puede consultar un buen estudio al respecto.

4.3 Las respuestas

La forma de las respuestas suministradas por un sistema de BR puede ser variada y está

íntimamente relacionada tanto con el tipo de pregunta como con el tipo de usuario del sistema.

Desde el punto de vista de su extensión, éstas pueden ser *cortas*, como respuesta a preguntas de hechos concretos (el nombre de una ciudad, una fecha, una cantidad, una lista de entidades, ...) o *largas*, en respuesta a preguntas de opinión y resumen.

Desde la perspectiva del usuario, estas respuestas pueden ir acompañadas de los documentos y de los criterios de selección empleados que justifiquen la respuesta. De esta forma, el usuario puede validar la corrección de las respuestas suministradas por el sistema.

Por otra parte, existen diferentes técnicas relacionadas con la forma de construcción de la respuesta: la extracción y la generación. La *extracción* consiste en seleccionar uno o varios extractos de los documentos analizados y presentarlos como respuesta tal y como aparecen en los documentos originales. Por otra parte, la *generación* consiste en elaborar una presentación coherente de la respuesta a partir de la información original localizada en los documentos. En este caso necesitaríamos la aplicación de técnicas de generación de lenguaje.

4.4 Nivel de conocimiento requerido

El poder contemplar con éxito el desarrollo de sistemas de BR que soporten los diferentes niveles de la tipología de usuarios enumerada previamente, necesita inexcusablemente de un incremento progresivo del *nivel de conocimiento* utilizado por estos sistemas.

Podemos estructurar este conocimiento en cuatro niveles en función de la necesidad de su participación para afrontar preguntas de creciente complejidad. Cada nivel incluiría el conocimiento de los niveles anteriores:

De hechos concretos. Corresponde al nivel mínimo exigido en un sistema de BR. Este conocimiento permite la contestación de preguntas cuya respuesta es un hecho concreto que bien puede ser el nombre de una persona u organización, una cantidad, un lugar o una fecha. Las bases de conocimiento utilizadas pueden estar

formadas por diccionarios o enciclopedias.

Explicativo. Este nivel de conocimiento ha de permitir que el sistema responda a preguntas más complejas en las que la respuesta constituye la explicación, justificación o causa de un suceso. En este caso, las bases de conocimiento utilizadas pueden estar formadas por ontologías y bases de conocimiento léxico-semánticas como WordNet.

Modal. Un mayor nivel de conocimiento es necesario para que un sistema pueda afrontar preguntas de opinión y resumen: “¿Qué podría pasar en Marruecos si el rey Hassan II es asesinado?” La respuesta a esta pregunta se obtendría dentro de un dominio específico, por ejemplo, la evaluación de las posibles consecuencias políticas, económicas y militares de dicho suceso. El tipo de conocimiento requerido para realizar este análisis vendría representado por lo que se conoce como *bases de conocimiento de alto rendimiento* (High-Performance Knowledge Bases - HPKB). Estas bases de conocimiento estarían formadas por ontologías restringidas al dominio de la pregunta junto con axiomas particulares y estrategias genéricas de solución de problemas asociados a dicho dominio.

General del mundo. Un amplio conocimiento general del mundo permitiría al sistema procesar preguntas del tipo anterior pero sin limitar el dominio de aplicación. De hecho el sistema podría ser capaz de “descubrir” nuevo conocimiento relacionado con la pregunta, “aconsejar” y “justificar” los motivos de dicha relación e incluso facilitar al usuario la posibilidad de interactuar con el sistema para dirigir el proceso de generación de la respuesta en función del descubrimiento de información relacionada.

Llegados a este punto, el análisis de las diferentes dimensiones del problema de la BR nos ha permitido acercarnos al problema desde una perspectiva general facilitando así, el acotar el ámbito general del problema y definir una base que permite situar el estado actual de las investigaciones en este campo.

5 Clasificación de los sistemas de BR

Los sistemas de BR actualmente operacionales, afrontan la tarea desde la perspectiva del *usuario casual*. Un usuario que realiza preguntas simples que requieren un hecho, situación o dato concreto como contestación. Estos sistemas utilizan como fuente de información una base de datos textual compuesta por documentos escritos en un único lenguaje. El conocimiento utilizado en estos sistemas corresponde también con el nivel mínimo detallado anteriormente (*de hechos concretos*). En algunos casos se ha avanzado un poco más mediante el uso de bases de datos léxico-semánticas (principalmente WordNet) y la integración de algún tipo particular de ontología como SENSUS (Hovy *et al.*, 2000) o Mikrokosmos (Odgen *et al.*, 1999).

Clasificar los sistemas existentes resulta una tarea bastante complicada. Esta dificultad radica principalmente en la selección de la perspectiva desde la que se desea realizar dicha clasificación y en la gran variedad de aproximaciones existentes.

En primer lugar se presenta una clasificación que tiene en cuenta la situación actual de los sistemas de BR en el ámbito de una perspectiva general. En segundo lugar, y con la intención de profundizar en las diferentes aproximaciones existentes, este trabajo propone una clasificación más detallada en función del nivel de análisis del lenguaje natural que estos sistemas utilizan.

5.1 Según una perspectiva general.

La taxonomía presentada en (Moldovan *et al.*, 1999) propone una clasificación de los sistemas de BR desde una perspectiva general. Esta taxonomía clasifica los sistemas de BR en cinco clases en función de tres criterios:

1. Las bases de conocimiento empleadas.
2. El nivel de razonamiento requerido.
3. Las técnicas de indexación y de PLN utilizadas.

Las bases de conocimiento y los sistemas de razonamiento proporcionan el medio que facilita la construcción del contexto de la pregunta y la búsqueda de la respuesta en los documentos. Por otra parte, las técnicas de indexación permiten localizar aquellos extractos de documentos en los que pueden aparecer las respuestas. Finalmente, las técnicas de PLN proporcionan el entorno general que permite la localización y extracción de dichas respuestas de forma precisa.

La figura 2 reproduce esta clasificación junto con sus características principales y algunos ejemplos de preguntas, respuestas y comentarios aclaratorios. Esta clasificación asume que los requisitos de una clase inferior están incluidos en los de la clase superior.

Según esta clasificación, los sistemas actuales de BR estarían enmarcados en las clases 1 y 2 confirmando de esta forma, el estado incipiente de la investigación en este tipo de sistemas.

5.2 Según el nivel de análisis del lenguaje empleado.

Con el objetivo de poder analizar las diferentes soluciones existentes, este trabajo propone su clasificación en función del nivel de análisis del lenguaje que utilizan. Las clases propuestas son las siguientes:

- Clase 0. Sistemas que no utilizan técnicas de PLN.
- Clase 1. Sistemas que emplean técnicas de análisis superficial.
- Clase 2. Sistemas que utilizan técnicas de análisis profundo.

Al igual que en caso anterior, esta clasificación asume que las características enmarcadas en una clase inferior están incluidas en las de clases superiores. A continuación se describen las características principales de cada una de las clases. Se presentan las aproximaciones más relevantes y se destacan las diferencias básicas que caracterizan las propuestas enmarcadas en una misma clase.

Class	KB	Reasoning	NLP/Indexing	Examples and Comments
1	dictionaries	simple heuristics, pattern matching	complex noun, apposition, simple semantics, keyword indexing	Q33: <i>What is the largest city in Germany?</i> A: .. <i>Berlin, the largest city in Germany.</i> Answer is: simple datum or list of items found verbatim in a sentence or paragraph.
2	ontologies	low level	verb nominalization, semantics, coherence, discourse	Q198: <i>How did Socrates die?</i> A: .. <i>Socrates poisoned himself.</i> Answer is contained in multiple sentences, scattered throughout a document.
3	very large KB	medium level	advanced nlp, semantic indexing	Q: <i>What are the arguments for and against prayer in school?</i> Answer across several texts.
4	Domain KA and Classification, HPKB	high level		Q: <i>Should Fed raise interest rates at their next meeting?</i> Answer across large number of documents, domain specific knowledge acquired automatically.
5	World Knowledge	very high level, special purpose		Q: <i>What should be the US foreign policy in the Balkans now?</i> Answer is a solution to a complex, possible developing scenario.

Figura 2: Taxonomía de los sistemas de BR (Moldovan *et al.*, 1999)

5.2.1 Clase 0. Sistemas que no utilizan técnicas de PLN.

Estos sistemas emplean técnicas de RI adaptadas a la tarea de BR. La forma general de actuación de estos sistemas se basa en la recuperación de extractos de texto relativamente pequeños con la suposición de que dichos extractos contendrán la respuesta esperada.

Generalmente, el análisis de la pregunta consiste en seleccionar aquellos términos de la pregunta que deben aparecer cerca de la respuesta. Para ello, se eliminan las palabras de parada y se seleccionan aquellos términos con mayor “valor discriminatorio” (palabras clave). Estos términos se utilizan para recuperar directamente fragmentos relevantes de texto que se presentan directamente como respuestas (Cormack *et al.*, 1999) o bien, para recuperar documentos que posteriormente serán analizados. Este análisis consiste en dividir el texto relevante en ventanas de un tamaño inferior o igual a la longitud máxima permitida como cadena respuesta. Cada una de estas ventanas se valora en función de determinadas heurísticas para finalmente presentar como respuestas aquellas ventanas que consiguen la mejor puntuación. Esta valoración suele tener en cuenta aspectos como el valor de discriminación de las palabras clave contenidas en la ventana, el orden de aparición de dichas palabras en comparación con el orden establecido en la pregunta, la distancia a la ventana de aquellas palabras clave que no se

aparecen en la ventana, etc.

Además del sistema de la universidad de Waterloo, citado previamente, se pueden incluir en este grupo los sistemas utilizados por la universidad de Massachusetts (Allan *et al.*, 2000) y los laboratorios RMIT/CSIRO (Fuller *et al.*, 1999).

El rendimiento alcanzado por este tipo de sistemas es relativamente bueno cuando la longitud permitida como respuesta es grande (del orden de 250 caracteres o más), sin embargo, decrece mucho cuando se requiere una respuesta corta y precisa.

5.2.2 Clase 1. Sistemas que emplean técnicas de análisis superficial.

En esta clase se enmarcan la mayoría de las aproximaciones existentes. Estos sistemas se caracterizan, en primer lugar, por la realización de un análisis detallado de la pregunta que permite extraer y representar aquella información que será de utilidad en las sucesivas fases del proceso. De forma general, este proceso permite obtener la siguiente información:

1. El tipo de entidad que cada pregunta espera como respuesta (una fecha, el nombre de una persona, etc).
2. Restricciones y características adicionales relacionadas con el tipo de respuesta espe-

rada:

- Términos de la pregunta que permiten la recuperación de aquellos extractos de texto susceptibles de contener la respuesta.
- Relaciones -sintácticas o semánticas- que deben aparecer entre las entidades de la pregunta y la respuesta a localizar.

La obtención del tipo de respuesta requiere que estas entidades estén organizadas en clases semánticas como por ejemplo, “persona”, “organización”, “tiempo”, “lugar”, etc. La identificación del tipo de respuesta esperada se suele afrontar mediante el análisis de los términos interrogativos de la pregunta (términos *wh*). Por ejemplo, el término “where” indica que la pregunta está buscando como respuesta una expresión de lugar. Sin embargo, en otros casos, se necesita del análisis de algunas estructuras sintácticas de la pregunta para obtener la clase semántica de la respuesta. En el caso de la pregunta “Which is the largest city . . . ?” es el término “city” -núcleo del sintagma nominal “largest city”- el que indica el tipo de respuesta esperado, en este caso, el nombre de una ciudad. La realización de este tipo de análisis suele requerir el uso de etiquetadores léxicos y analizadores sintácticos o bien, la aplicación de patrones léxico-sintácticos.

Del análisis de la pregunta se deriva además, aquella información que permite la generación de las consultas que, procesadas por un sistema de RI, facilitan la selección de los extractos de texto de la colección documental susceptibles de contener la respuesta. La obtención de estas consultas sigue actualmente dos tendencias diferenciadas:

1. La selección de palabras clave (*keywords*).
2. La generación de patrones de respuesta.

La *selección de palabras clave* consiste en seleccionar aquellos términos de la pregunta cuya aparición en un texto es de por sí indicativa de la posibilidad de existencia en sus alrededores de la respuesta buscada. Para la pregunta “¿Qué país limita al norte con España?” el conjunto de palabras clave estaría formado por los términos “limita”, “norte” y “España”.

Por otra parte, el proceso de *generación de patrones de respuesta* es bastante más elaborado. En este caso las consultas estarán formadas por una o varias combinaciones de los términos de la pregunta en forma de expresiones en las que podría encontrarse la respuesta. Posibles consultas derivadas del ejemplo anterior serían las siguientes: “X limita al norte con España”, “X, país que limita al norte con España”, “La frontera norte de España es X”, etc. donde X es una referencia a la respuesta a encontrar. En este caso, el sistema de RI se emplearía en localizar extractos de texto que contengan posibles expresiones de respuesta asociadas a cada tipo de pregunta (Hermjacob *et al.*, 2002; Soubbotin & Soubbotin, 2002). Ambas estrategias no son excluyentes entre sí, existiendo sistemas que combinan ambas aproximaciones (Lin *et al.*, 2002a).

Por lo que respecta al proceso final de extracción de la respuesta, se suelen emplear técnicas de RI o bien, patrones de respuesta en combinación con el uso de *clasificadores de entidades*. Estas herramientas permiten localizar aquellas entidades cuya clase semántica corresponde con aquella que la pregunta espera como respuesta. De esta forma, el sistema extraerá la respuesta de aquellos extractos de texto que contienen alguna entidad del tipo semántico requerido, de forma combinada con la aparición de términos clave en sus cercanías y/o la validación de patrones de respuesta. Finalmente, el sistema ha de elegir de entre las entidades que pueden ser respuesta a la pregunta. Este proceso se lleva a cabo mediante la aplicación de medidas que permitan valorar de alguna forma el grado de “corrección” de cada posible respuesta. Esta valoración se suele realizar aplicando funciones que miden por una parte, el grado de cumplimiento de aquellas características que tiene en cuenta el sistema en el proceso de búsqueda de la respuesta (cercanía de palabras clave en el texto, fiabilidad de los patrones validados, etc.) y por otra, circunstancias generalmente relacionadas con la redundancia de aparición de cada respuesta posible en diferentes documentos.

De entre los sistemas que emplean patrones como base para la tarea de BR podemos destacar el sistema diseñado por InsihtSoft (Soubbotin & Soubbotin, 2002). La base de esta aproximación reside en la identificación y construcción de una serie de *patrones indicativos* (indicative patterns) que dependen del tipo de pregunta a tratar y cuya validación está relacionada con

la posibilidad de encontrar la respuesta correcta. Un patrón indicativo se define como una secuencia o combinación determinada de caracteres, signos de puntuación, espacios, dígitos o palabras. Estos patrones se obtienen de forma totalmente manual mediante el estudio de expresiones que son respuestas a determinados tipos de preguntas. Por ejemplo, la cadena “Mozart (1756-1791)” contiene la respuesta a preguntas relacionadas con los años en que Mozart nació y falleció. A partir de esta observación, se puede construir el siguiente patrón: “[palabra con 1ª letra en mayúsculas; paréntesis; cuatro dígitos; guión; cuatro dígitos; paréntesis]”. Dicho patrón permite detectar respuestas a preguntas acerca del periodo de existencia de una persona. A cada uno de estos patrones se le asigna un valor de forma que el sistema pueda elegir entre varias posibles respuestas a una pregunta en función del grado de fiabilidad de cada patrón con respecto a la pregunta.

Estas aproximaciones son las que mayor seguimiento están teniendo actualmente y su principal diferencia estriba en la forma de obtención de los patrones de respuesta (también llamados reformulaciones de la pregunta) con respecto a la tipología de preguntas que cada sistema emplea: (1) *manual*, según se ha visto en (Soubbotin & Soubbotin, 2002), o mediante el diseño de gramáticas de transformación (C. Kwok, 2001); (2) *intensiva*, mediante la generación de todas las posibles reformulaciones de la pregunta (Brill *et al.*, 2001); (3) *automática*, empleando algoritmos de aprendizaje (Ravichandran & Hovy, 2002) o bien, (4) de forma *semiautomática* combinando el uso de técnicas de aprendizaje con la revisión manual de resultados (Hermjacob *et al.*, 2002).

Cabe destacar algunas aproximaciones interesantes de entre aquellos sistemas que no emplean estructuras de respuesta. El sistema utilizado por IBM (Chu-Carroll *et al.*, 2002) basa su aproximación en el concepto de *anotación predictiva*. Este sistema utiliza un etiquetador de entidades para anotar en todos los documentos de la colección, la clase semántica de aquellas entidades que detecta. Dicha clase semántica se indexa junto con el resto de términos de los documentos facilitando así, la recuperación preliminar de los extractos de documentos que contienen entidades cuya clase semántica coincide con la esperada como respuesta.

Otras aproximaciones incluidas en este grupo realizan un uso más intensivo de la información sintáctica. Estos sistemas tienen en cuenta la similitud entre las estructuras sintácticas de las preguntas y posibles respuestas como factor importante en el proceso de extracción de la respuesta final (Lee *et al.*, 2001; Oard *et al.*, 1999).

Por otra parte, algunos sistemas como los de IBM (Ittycheriah & Roukos, 2002) y BBN (Xu *et al.*, 2002) se caracterizan principalmente por la aplicación de técnicas de aprendizaje, basadas en modelos de máxima entropía, a los procesos de análisis de la pregunta y de extracción final de la respuesta. En ambos casos, estas técnicas se aplican en un módulo cuya finalidad consiste en validar la corrección de las respuestas suministradas por el sistema mediante la estimación de la probabilidad de que una respuesta sea correcta.

Finalmente cabe destacar la creciente cantidad de sistemas que emplean la densidad de información existente en la Web como fuente de información añadida en el proceso de BR. Estos sistemas suelen emplear Internet para realizar un proceso paralelo de búsqueda que permite recopilar información para expandir la pregunta original (Yang & Chua, 2002; Attardi *et al.*, 2002) o bien, obtener datos de redundancia de posibles respuestas que permitan validar las respuestas obtenidas a partir de la colección de documentos del sistema (Clarke *et al.*, 2002; Magnini *et al.*, 2002; Chu-Carroll *et al.*, 2002).

Esta tendencia va en aumento debido ya que los experimentos realizados hasta la fecha han permitido mejorar de forma notable el rendimiento de aquellos sistemas que los emplean.

5.2.3 Clase 2. Sistemas que utilizan técnicas de análisis profundo.

El uso de técnicas complejas de PLN en tareas de BR es escaso debido fundamentalmente a las dificultades intrínsecas de la representación del conocimiento. De hecho, sólo un grupo reducido de sistemas aplican herramientas que realizan este tipo de análisis.

Estas técnicas se aplican en los procesos de análisis de la pregunta y de extracción final de la respuesta. De forma general, estos sistemas obtienen la representación semántica de la pre-

gunta y de aquellas sentencias que son relevantes a dicha pregunta. La extracción de la respuesta se realiza mediante procesos de comparación y/o unificación entre las representaciones de la pregunta y las frases relevantes.

Los sistemas de la universidad de Pisa (Attardi *et al.*, 2002) y CLR (Litkowski, 2002) utilizan el concepto de *triplezas semánticas* para representar dicha información. Una tripleta semántica está formada por una entidad del discurso, el rol semántico que dicha entidad desempeña y el término con el que dicha entidad mantiene la relación. Con esta notación se representan las preguntas y las frases que contienen respuestas del tipo esperado para proceder a la extracción de la respuesta comparando y puntuando el nivel de relación existente entre las estructuras semánticas obtenidas en preguntas y frases objetivo.

Aunque son escasos, otros sistemas profundizan aún más en el análisis del LN mediante la aplicación de técnicas de análisis contextual. Estos sistemas incorporan conocimiento general del mundo asociado a mecanismos inferenciales que facilitan el proceso de extracción de respuestas.

El sistema de la Universidad de Sheffield (Greenwood *et al.*, 2002) utiliza fórmulas lógicas (FLs) para representar las preguntas y los pasajes candidatos a contener la respuesta y los incorpora en un modelo de discurso. El modelo de discurso es una especialización de una red semántica que codifica el conocimiento general del mundo y que se enriquece con el conocimiento específico codificado en las FLs de la pregunta y los pasajes candidatos. Una vez que se ha generado el modelo de discurso para una pregunta, se aplican sistemas de resolución de correferencias para integrar en una, todas las referencias que aparecen en el modelo a una misma entidad. A pesar de la complejidad de esta aproximación, el sistema no utiliza ningún tipo de inferencia y por tanto, la selección de la respuesta final se realiza mediante la aplicación de sistemas de puntuación que valoran principalmente la redundancia observada en cada una de las respuestas posibles.

Por otra parte, cabe destacar que el sistema de LCC (Moldovan *et al.*, 2002) es el que mejor rendimiento obtiene de la aplicación de técnicas de análisis semántico y contextual en el proceso de extracción final de la respuesta. Para ello,

tanto las preguntas como las frases susceptibles de contener la respuesta son representadas mediante fórmulas lógicas a las que se aplica un proceso de unificación que permite detectar aquellas las respuestas posibles a la pregunta. Posteriormente, a estas respuestas se añaden un conjunto de axiomas que representan el conocimiento general del mundo (obtenidos de WordNet) junto con otros derivados de la aplicación de técnicas de resolución de correferencias. Toda esta información se utiliza para determinar la corrección de dichas respuestas a través de un sistema de inferencia abductiva.

6 La evaluación

La investigación en sistemas de BR ha propiciado, de forma colateral, un creciente interés en el desarrollo de técnicas que permitan evaluar de forma automática el rendimiento de estos sistemas.

Esta tarea se está afrontando desde diversas perspectivas: la utilización de colecciones de test (Voorhees & Tice, 2000), el uso de tests de lectura y comprensión de textos (ver sección 3.2) y la aplicación de sistemas automáticos que evalúan la corrección de las respuestas suministradas por los sistemas mediante su comparación con las respuestas generadas por humanos a las mismas preguntas (Breck *et al.*, 2000).

La propuesta que mayor éxito ha tenido hasta el momento consiste en la utilización de *colecciones de test*. Una colección de test comprende un conjunto de documentos, un conjunto de preguntas junto a sus correspondientes respuestas, una medida de rendimiento del sistema y un programa que permite de forma automática, comprobar la corrección de las respuestas suministradas por el sistema de BR y evaluar su rendimiento global.

Las colecciones de test más importantes disponibles en la actualidad se han generado a partir de los datos y resultados de las evaluaciones desarrolladas en el ámbito de las conferencias TREC. De hecho, su uso ha sido aceptado de forma general por los investigadores en la materia, convirtiéndose así en la principal base de comparación entre este tipo de sistemas.

6.1 Las conferencias TREC y los sistemas de BR

En 1999, en el seno de la conferencia TREC-8, se presentó la primera tarea específica para la evaluación de sistemas de BR: “*The first Question Answering track*”.

En esta primera convocatoria, se evaluó el rendimiento de los sistemas participantes sobre 200 preguntas de test elaboradas por la organización con la seguridad de que la respuesta se encontraba en algún documento de la colección. Para cada pregunta, los sistemas debían devolver una lista ordenada con un máximo de 5 respuestas posibles. Cada respuesta consistía en un fragmento de texto extraído de la base documental en el que debería aparecer la respuesta a la pregunta. Se diseñaron dos categorías en función del tamaño máximo permitido del fragmento de texto respuesta (250 y 50 caracteres).

En las siguientes convocatorias se introdujeron progresivamente nuevos requerimientos relacionados principalmente, con el incremento del tamaño de la base documental, con la cantidad y complejidad de las preguntas de test realizadas y con el endurecimiento de los requisitos que deben cumplir las respuestas para ser consideradas correctas.

Con la finalidad de obtener una visión de la evolución de estas evaluaciones, extractaremos las principales novedades introducidas en las sucesivas convocatorias. La conferencia TREC-9 (2000) supuso un considerable aumento del tamaño de la colección de test tanto en el número de preguntas a evaluar como en la cantidad de documentos. En la conferencia TREC-10 (2001) se limitó a 50 caracteres el tamaño máximo de texto permitido como respuesta. Además, no se garantizó la existencia de respuesta a las preguntas en la base de datos documental. De esta forma, la única contestación correcta a preguntas cuya respuesta no existía en estos documentos era “No hay respuesta”. Esta circunstancia favoreció la investigación en herramientas que afrontasen el problema de la validación de respuestas. Por otra parte, se incrementó la complejidad de las preguntas de test. Se incluyeron preguntas en las que se especificaba un número de instancias a recuperar como respuesta (“Dime 3 películas interpretadas por Antonio Banderas”) y también, se pro-

pusieron series de preguntas formuladas sobre un mismo contexto. Estas series estaban formadas por preguntas relacionadas entre sí de forma que la interpretación de cada pregunta dependía tanto del significado de las preguntas realizadas previamente como de sus respectivas contestaciones. La última evaluación realizada (TREC-11, 2002) también aportó nuevos retos. En este caso, el sistema sólo podía responder con una única respuesta a cada pregunta y además, la cadena respuesta debía estar formada exacta y exclusivamente por la respuesta concreta. De esta forma, se consideraron erróneas tanto las cadenas respuesta que sólo contemplaran parte de la respuesta, como aquellas que incluyeran, además de la respuesta, cualquier otro texto ajeno a ella.

Por otra parte, la evolución de los sistemas se ha desarrollado (quizás más allá de lo inicialmente previsto) de forma muy satisfactoria y en total concordancia con el incremento de complejidad propuesto en las sucesivas evaluaciones. Una medida de este progreso se fundamenta en el hecho de que, a pesar del incremento de las dificultades, los mejores sistemas son capaces de contestar correctamente más de dos tercios de las preguntas formuladas. La tabla 1 muestra de forma resumida la evolución de las características más interesantes de estas evaluaciones junto al resultado obtenido por el mejor sistema en cada convocatoria (medido en % de respuestas correctas contestadas).

7 Perspectivas de futuro

Llegados a este punto, y en base a las perspectivas abiertas en torno a la investigación en este campo, cabría plantear las siguientes preguntas: ¿Cómo debe avanzar la investigación desde la situación actual?, ¿En qué aspectos se debe profundizar?, ¿Se puede organizar la investigación en estos aspectos en tareas de creciente complejidad?, ¿Puede programarse este proceso en el tiempo?

Estos interrogantes han sido objeto de estudio por un comité creado a tal efecto (*the Roadmap Committee*) cuyo trabajo ha permitido establecer los objetivos principales en este campo de investigación de entre los que extractamos los siguientes (Burger *et al.*, 2000):

Características	TREC-8	TREC-9	TREC-10	TREC-11
Número de documentos	528.000	978.952	978.952	1.033.000
Documentos en megabytes	1.904	3.033	3.033	3.000
Preguntas propuestas	200	693	500	500
Preguntas evaluadas	198	682	496	500
Respuestas permitidas por pregunta	5	5	5	1
¿Permite preguntas sin respuesta?	No	No	Si	Si
Número de preguntas sin respuesta	0	0	49	46
Tamaño máximo respuestas (caracteres)	250 - 50	250 - 50	50	Exacta
Preguntas contestadas (% máximo)	77 % - 72 %	86 % - 66 %	69 %	83 %

Tabla 1: Características de las evaluaciones TREC

Clases de preguntas. Obtención de una buena taxonomía. Una parte importante en el proceso de interpretación de las preguntas reside en poder relacionar el tipo de pregunta que se está realizando con las características de la respuesta que espera. Aunque se han propuesto muchas clasificaciones ninguna de ellas se ha realizado teniendo en cuenta el concepto y características del sistema de BR “futuro” introducido anteriormente. Por ello, se requiere la definición de una tipología de preguntas basada en principios bien definidos y que asuma los requerimientos anteriormente especificados.

Análisis de la pregunta. Comprensión y resolución de ambigüedades. Dado que una misma pregunta puede realizarse de muy diversas formas (interrogativa, afirmativa, con diferentes palabras y estructuras, . . .), se necesita un modelo semántico que permita reconocer preguntas equivalentes y facilite su traducción al lenguaje utilizado por el sistema para su correcto procesamiento.

El contexto en los sistemas de BR. El análisis del contexto en el que se hace una pregunta debe poder utilizarse para resolver ambigüedades y facilitar la investigación en un tema a través de series de preguntas relacionadas.

Integración de diferentes fuentes de información. Existen grandes cantidades de información distribuida en ficheros y bases de datos con diferentes formatos y estructuras. El modelo a realizar debería ser capaz de integrar y utilizar dicha información en el proceso de BR de igual forma que actualmente se trata la información textual. Algunos trabajos recientes ya están profundizando en estos aspectos. En particular, en la combinación de respuestas proce-

dentales de bases de datos estructurada con otras extraídas de bases textuales (Chu-Carroll *et al.*, 2002; Clarke *et al.*, 2002; Lin *et al.*, 2002b).

Extracción de respuestas a través de información distribuida. Justificación y evaluación de la corrección. Un aspecto a potenciar consiste en el diseño de modelos que permitan detectar evidencias puntuales en diferentes fuentes y cuya integración y combinación permita la obtención de la respuesta. Sin duda, las técnicas que faciliten esta integración estarán muy relacionadas con modelos de justificación y evaluación de la corrección de las respuestas.

Generación y presentación de respuestas. Consiste en el estudio de modelos de generación de lenguaje natural que permitan presentar las respuestas al usuario de una forma natural y coherente.

Búsqueda de respuestas en tiempo real. Además de la efectividad, se pretende que un sistema de BR sea capaz de obtener resultados en un tiempo limitado independientemente de las características de la pregunta y la cantidad de recursos que utilice. Las investigaciones en este ámbito se dirigen a la detección de cuellos de botella en los procesos de BR y al estudio de modelos rápidos de recuperación y extracción.

Integración de información multilingüe. Se considera muy importante el desarrollo de sistemas de BR para otros lenguajes diferentes del inglés. Por extensión, se pretende investigar en sistemas que soporten la BR en fuentes de información disponibles en varios lenguajes. Este es uno de los objetivos en los que mayores esfuerzos se están invirtiendo en la actualidad. Prueba de ello reside en la organización, en el

ámbito de la conferencia *Cross-Language Evaluation Forum*⁹ (CLEF), de una nueva tarea denominada *Multiple Language Question Answering* orientada a la evaluación de sistemas de BR en varios idiomas. Esta evaluación propone varias subtareas: BR monolingüe en español, italiano y holandés, y BR bilingüe. La tarea bilingüe, consiste en el tratamiento de preguntas en español e italiano que deben ser contestadas a partir de documentos en lengua inglesa.

Interactividad en la BR. Se pretende conseguir sistemas interactivos que permitan un diálogo sistema-usuario. Esta interacción ha de facilitar la adaptación del proceso de búsqueda según las sugerencias, comentarios e indicaciones progresivas del usuario.

Integración de sistemas de razonamiento. Estos sistemas responderían a las expectativas de usuarios profesionales. Se debe profundizar por tanto, en aspectos relacionados con la integración de componentes que permitan un elevado nivel de razonamiento sobre diferentes bases de conocimiento incluyendo, desde el conocimiento general del mundo hasta el conocimiento específico de determinados dominios.

Integración y gestión de perfiles de usuarios. El sistema debe de poder capturar información del usuario relativa por ejemplo, a dominios de interés, esquemas de razonamiento frecuentemente utilizados, nivel de profundidad de búsqueda, etc. Esta integración permitiría la adaptación del sistema a la forma de trabajar del usuario y en consecuencia, facilitaría aún más su tarea.

A modo de conclusión podemos afirmar que los sistemas de BR son los potenciales sucesores de los buscadores de información tradicionales si bien, una mera revisión de los objetivos descritos previamente nos da una idea de las grandes posibilidades de investigación que este campo todavía presenta y de la necesidad de dirigir las investigaciones desarrolladas en otros campos (como el procesamiento del lenguaje natural, la representación del conocimiento, el procesamiento multimedia, la interacción usuario-computador, etc.) hacia su adaptación a entornos de BR.

⁹<http://clef-qa.itc.it/>

Referencias

- Allan, J., Connel, M., Croft, W., Feng, F., Fisher, D., & Li, X. 2000. INQUERY and TREC-9. *Pages 551–562 of: Ninth Text REtrieval Conference*. NIST Special Publication, vol. 500-249. Gaithersburg, USA: National Institute of Standards and Technology.
- Attardi, G., Cisternino, A., Formica, F., Simi, M., & Tommasi, A. 2002. PIQASs 2002. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Berger, Adam, Caruana, Rich, Cohn, David, Freitag, Dayne, & Mittal, Vibhu. 2000. Bridging the Lexical Chasm: Statistical Approaches to Answer-Finding. *Pages 192–199 of: Proceedings of the 23th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Question Answering.
- Breck, Eric, Burger, John, Ferro, Lisa, Hirschman, Lynette, House, David, Light, Marc, & Mani, Inderjeet. 2000. How to Evaluate Your Question Answering System Every Day ... and Still Get Real Work Done. *In: Proceedings of Second International Conference on Language Resources and Evaluation. LREC-2000*.
- Brill, E., Lin, J., Banko, M., & Dumais, S. 2001. Data-Intensive Question Answering. *In: Tenth Text REtrieval Conference*. NIST Special Publication, vol. 500-250. Gaithersburg, USA: National Institute of Standards and Technology.
- Burger, John, Cardie, Claire, Chaudhri, Vinay, Gaizauskas, Robert, Harabagiu, Sanda, Israel, David, Jacquemin, Christian, Lin, Chin-Yew, Maiorano, Steve, Miller, George, Moldovan, Dan, Ogden, Bill, Prager, John, Riloff, Ellen, Singhal, Amit, Shrihari, Rohini, Strzalkowski, Tomek, Voorhees, Ellen, & Weishedel, Ralph. 2000. *Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A)*. http://www-nlpir.nist.gov/projects/duc/papers/qa-Roadmap-paper_v2.doc.

- Burke, R., Hammond, K., Kulyukin, V., Lytinen, S., Tomuro, N., & Schoenberg, S. 1997. Question Answering from Frequently Asked Question Files. *AI Magazine*, **18**(2), 57–66.
- C. Kwok, O. Etzioni, D. Weld. 2001. Scaling Question Answering to the Web. *In: Proceedings of the Tenth World Wide Web Conference*.
- Callan, James P. 1994. Passage-Level Evidence in Document Retrieval. *Pages 302–310 of: Proceedings of the 17th Annual International Conference on Research and Development in Information Retrieval*. London, UK: Springer Verlag.
- Charniak, E., Altun, Y., Braz, R., Garrett, B., Kosmala, M., Moscovich, T., Pang, L., Pyo, C., Sun, Y., Wy, W., Yang, Z., Zeller, S., & Zorn, L. 2000 (may). Reading Comprehension Programs in a Statistical-Language-Processing Class. *Pages 1–5 of: ANLP/NAACL Workshop on Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems*.
- Chu-Carroll, Jennifer, Prager, John, Welty, Chirtopher, Czuba, Krzysztof, & Ferrucci, David. 2002. A Multi-Strategy and Multi-Source Approach to Question Answering. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Cinchor, Nancy A. 1998. Overview of MUC/MET-2. *In: Seventh Message Understanding Conference*.
- Clarke, Charles L., Cormack, G. V., Kemkes, G., Laszlo, M., Lynam, T. R., Terra, E. L., & Tilker, P. L. 2002. Statistical Selection of Exact Answers (MultiText Experiments for TREC 2002). *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Cormack, Gordon V., Clarke, Charles L. A., Palmer, Christopher R., & Kisman, Derek I. E. 1999. Fast Automatic Passage Ranking (MultiText Experiments for TREC-8). *Pages 735–742 of: Eighth Text REtrieval Conference*. NIST Special Publication, vol. 500-246. Gaithersburg, USA: National Institute of Standards and Technology.
- Fuller, M., Kaszkiel, M., Kimberley, S., Zobel, J., Wilkinson, R., & Wu, M. 1999. The RMIT/CSIRO Ad Hoc, Q&A, Web, Interactive, and Speech Experiments at TREC-8. *Pages 549–564 of: Eighth Text REtrieval Conference*. NIST Special Publication, vol. 500-246. Gaithersburg, USA: National Institute of Standards and Technology.
- Greenwood, M.A., Roberts, I., & Gaizauskas, R. 2002. The University of Sheffield TREC 2002 Q&A System. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Hermjacob, U., Echihabi, A., & Marcu, D. 2002. Natural Language Based Reformulation Resource and Wide Exploitation for Question Answering. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Hirschman, L., Light, M., Breck, E., & Burger, J. 1999. Deep read: a reading comprehension system. *Pages 325–332 of: 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*.
- Hovy, E., Gerber, L, Hermjacob, U., Junk, M., & Lin, C. 2000. Question Answering in Webclopedia. *Pages 655–664 of: Ninth Text REtrieval Conference*. NIST Special Publication, vol. 500-249. Gaithersburg, USA: National Institute of Standards and Technology.
- Ittycheriah, Abraham, & Roukos, Salim. 2002. IBM's Statistical Question Answering System-TREC 11. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Kaszkiel, Marcin, Zobel, Justin, & Sacks-Davis, Ron. 1999. Efficient passage ranking for document databases. *ACM Transactions on Information Systems*, **17**(4), 406–439.

- Kupiec, J.M. 1993 (jun). MURAX: A robust linguistic approach for question-answering using an on-line encyclopedia. *Pages 181–190 of: Sixteenth International ACM SIGIR Conference on Research and Development in Information Retrieval.*
- Lee, G., Seo, J., Lee, S., Jung, H., Cho, B., Lee, C., Kwak, B., Cha, J., Kim, D., An, J., Kim, H., & Kim, K. 2001. SiteQ: Engineering High Performance QA system Using Lexico-Semantic Pattern Matching and Shallow NLP. *In: Tenth Text REtrieval Conference.* NIST Special Publication, vol. 500-250. Gaithersburg, USA: National Institute of Standards and Technology.
- Lehnert, Wendy G. 1977. Human and computational question answering. *Cognitive Science*, 47–63.
- Lehnert, Wendy G. 1980. Question answering in natural language procesing. *Pages 9–71 of: Verlag, Carl Hansen (ed), Natural Language Question Answering Systems.*
- Levine, J. M., & Fedder, L. 1989 (October). *The theory and implementation of a bidirectional question answerin system.* Tech. Rep. 182. University of Cambridge.
- Light, Marc, Mann, Gideon S., Riloff, Ellen, & Breck, Eric. 2001. Analyses for Elucidating Current Question Answering Technology. *Journal of Natural Language Engineering*, 7(4), 325–342.
- Lin, J., Fernandes, A., Katz, B., Marton, G., & Tellex, S. 2002a. Extracting Answers from the Web Using Data Annotation and Knowledge Mining Techniques. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Lin, J., Fernandes, A., Katz, B., Marton, G., & Tellex, S. 2002b. Extracting Answers from the Web Using Data Annotation and Knowledge Mining Techniques. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Litkowski, K.C. 2002. Question Answering Using XML-Tagged Documents. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Magnini, B., Negri, M., Prevete, R., & Tanev, H. 2002. Mining Knowledge from Repeated Co-Occurrences: DIOGENE at TREC 2002. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Moldovan, Dan, Harabagiu, Sanda, Pasca, Marius, Mihalcea, Rada, Goodrum, Richard, Gîrju, Roxana, & Rus, Vasile. 1999. LASO: A Tool for Surfing the Answer Net. *Pages 175–184 of: Eighth Text REtrieval Conference.* NIST Special Publication, vol. 500-246. Gaithersburg, USA: National Institute of Standards and Technology.
- Moldovan, Dan, Harabagiu, Sanda, Gîrju, Roxana, Morarescu, Paul, Lacatusu, Finley, Novischi, A., Badulescu, A., & Bolohan, O. 2002. LCC Tools for Question Answering. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.
- Oard, Douglas W., Wang, Jianqiang, Lin, Dekang, & Soboroff, Ian. 1999. TREC-8 Experiments at Maryland: CLIR, QA and Routing. *Pages 623–636 of: Eighth Text REtrieval Conference.* NIST Special Publication, vol. 500-246. Gaithersburg, USA: National Institute of Standards and Technology.
- Odgen, Bill, Cowie, Jim, Ludovik, Eugene, Molina-Salgado, Hugo, Nirenburg, Sergei, Sharples, Nigel, & Sheremtyeva, Svetlana. 1999. CRL’s TREC-8 Systems Cross-Lingual IR and Q&A. *Pages 513–522 of: Eighth Text REtrieval Conference.* NIST Special Publication, vol. 500-246. Gaithersburg, USA: National Institute of Standards and Technology.
- Ravichandran, D., & Hovy, E. 2002. Learning Surface Text Patterns for a Question Answering System. *In: 40th Annual Meeting of the Association for Computational Linguistics (ACL-02).*
- Soubbotin, M., & Soubbotin, S. 2002. Use of Patterns for Detection of Likely Answer

Strings: A Systematic Approach. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.

Spark-Jones, Karen. 1999. What is the Role of NLP in Text Retrieval? *Chap. 1, pages 1–24 of: Natural Language Information Retrieval*. New York, USA: Kluwer Academic.

Strzalkowski, T., Stein, G., Wise, G. Bowden, Perez-Carballo, J., Tapananinen, P., Jarvinen, T., Voutilainen, A., & Karlgren, J. 1998. Natural language information retrieval: TREC-7 report. *Pages 217–226 of: Seventh Text REtrieval Conference*. NIST Special Publication, vol. 500-242. Gaithersburg, USA: National Institute of Standards and Technology.

Voorhees, Ellen M., & Tice, Dawn M. 2000. Building a question answering test collection. *Pages 200–207 of: Proceedings of the 23th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Question Answering.

Xu, J., Licuanan, A., May, J., Miller, S., & Weischedel, R. 2002. TREC 2002 QA at BBN: Answer Selection and Confidence Estimation. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.

Yang, Hui, & Chua, Tat-Seng. 2002. The Integration of Lexical Knowledge and External Resources for Question Answering. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. Gaithersburg, USA: National Institute of Standards and Technology.