

PRISMA: un modelo interactivo de Síntesis de Información*

Enrique Amigó, Julio Gonzalo, Víctor Peinado, Anselmo Peñas, Felisa Verdejo

Dept. de Lenguajes y Sistemas Informáticos - UNED

c/Juan del Rosal, 16 - 28040 Madrid - Spain

{enrique, julio, victor, anselmo, felisa}@lsi.uned.es

Resumen: En este artículo, describimos un modelo interactivo de resumen orientado a síntesis de información (PRISMA). El usuario interactúa con el sistema por medio de listas de conceptos clave extraídos automáticamente. El modelo hace uso de información sintáctica para identificar conceptos clave, organizar y mostrar la información. Asimismo, proponemos y aplicamos una metodología de evaluación de modelos interactivos, que no requiere interacción entre el usuario y el sistema final.

Palabras clave: Síntesis de Información, sistemas interactivos

Abstract: In this paper, we describe an information synthesis interactive model (PRISMA). The user interacts with the system by means of automatically extracted key concepts lists. The model uses syntactic knowledge to identify key concepts, to organize and display the information pieces. Also, we propose, and therefore we put into practice, a corpora based methodology of interactive models evaluation. The implementation of this methodology doesn't need interaction between user and system.

Keywords: Information Synthesis, interactive systems

1 Introducción

La tarea de resumen multidocumento orientado a consulta consiste en producir, a partir de una necesidad de información expresada en forma de consulta y un conjunto de documentos, un texto que contenga el máximo de información relevante para la consulta en el mínimo de espacio.

Se han realizado diversos trabajos a partir de corpus orientados a resumen multidocumento (Over, 2003; Goldstein et al., 2000; Radev, Hongyan, y Budzikowska, 2000). Sin embargo, en todos ellos, la tarea de resumen se plantea desde conjuntos de no más de 10 o 20 documentos, y sobre resúmenes de no más de 100 palabras. Abordamos en este artículo la elaboración de resúmenes de 50 frases (informes) generados a partir de conjuntos amplios (100 documentos) de textos relacionados. A esta tarea derivada del resumen multidocumento la denominamos *Síntesis de Información*.

La tarea de resumen depende en gran medida de la interpretación que da el usuario, tanto de la necesidad de información como del contenido de los documentos, por lo que existen diversos trabajos que han planteado el problema desde una perspectiva interactiva (Boguraev et al., 1998; Buyukkokten,

García-Molina, y Paepcke, 1999; Neff y Cooper, 1999; Jones, Lundy, y Paynter, 2002; Leuski, Lin, y Stubblebine, 2003). Una característica común en todas estas aproximaciones es que se ofrece al usuario una lista de conceptos clave sobre los que interactuar con el sistema. Aunque las listas de conceptos relevantes son usuales en modelos interactivos de resumen, no se ha evaluado empíricamente la utilidad de este esquema de interacción, ni tampoco se han comparado distintas estrategias de extracción automática de conceptos clave desde el punto de vista del usuario.

A continuación, describimos un modelo interactivo de resumen orientado a síntesis de información, en el que el usuario interactúa con el sistema por medio de listas de términos que representan conceptos clave, extraídos automáticamente, utilizando técnicas de análisis sintáctico superficial. Además, proponemos y aplicamos una metodología de evaluación basada en la descomposición del problema y del modelo interactivo.

El artículo se estructura del modo siguiente: En el apartado 2, discutimos el modelo PRISMA en relación a otros modelos interactivos relacionados. En el apartado 3, describimos el modelo PRISMA y mostramos el prototipo implementado. Tras esto, en el apartado 4, ofrecemos algunos detalles del procesamiento sintáctico empleado en la implementación del prototipo. En el aparta-

* Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología a través del proyecto HERMES (TIC2000-0335-C03-1)

do 5, discutimos la metodología de evaluación propuesta. Presentamos, en el apartado 6, los puntos de la metodología ya implementados. Por último, introducimos algunas conclusiones en el apartado 7.

2 Modelos interactivos de resumen

Las listas de conceptos relevantes extraídas automáticamente ya han sido empleadas en varias aproximaciones interactivas para producir resúmenes. En algunas de ellas (Neff y Cooper, 1999; Jones, Lundy, y Paynter, 2002; Leuski, Lin, y Stubblebine, 2003) el resumen es generado automáticamente, tomando como entrada los conceptos clave seleccionados por el usuario a partir de una lista sugerida por el sistema. A este tipo de aproximaciones las denominamos CSBIS (*Concept Selection Based Interactive Summarization*).

En otros casos (Boguraev et al., 1998; Buyukkokten, García-Molina, y Paepcke, 1999; Rao et al., 1995) el usuario accede a piezas de información organizadas por conceptos clave, refinando así progresivamente el resumen. A este tipo de aproximaciones las denominamos CBBIS (*Concept Browsing Based Interactive Summarization*). La diferencia fundamental entre este tipo de esquemas y el anterior, consiste en que en modelos CBBIS es el propio usuario el que, en última instancia, decide qué fragmentos de textos son relevantes, mientras que en modelos CSBIS el usuario actúa como asistente y es el sistema el que realiza el resumen a partir de las directrices del usuario. El modelo PRISMA se ajusta a un esquema de interacción tipo CBBIS.

En el caso de PRISMA, abordamos la tarea de síntesis de información a partir de un conjunto voluminoso de documentos donde es posible aplicar medidas estadísticas no sólo sobre ocurrencias de palabras, sino también sobre el rol sintáctico que desempeñan. Precisamente, la particularidad de PRISMA respecto a otros modelos reside en el uso de conocimiento sintáctico para extraer los conceptos clave y mostrar al usuario la información asociada a cada concepto.

En este sentido, PRISMA mantiene varias semejanzas con el modelo propuesto en (Boguraev et al., 1998):

- El conjunto de candidatos en el proceso de identificación de conceptos clave viene dado por los sintagmas nominales que

aparecen en el documento.

- El rol sintáctico que desempeña el sintagma nominal es un criterio de selección aplicado en el proceso de extracción de conceptos clave.
- Se muestra al usuario el contexto de los conceptos clave en unidades de información más pequeñas que la frase completa.

PRISMA y el modelo de Boguraev difieren en la tarea de alto nivel para la que se destina el sistema. En el caso de la aproximación de Boguraev, el objetivo consiste en identificar información relevante contenida en un único documento, mientras que en el caso de PRISMA tratamos el problema de la síntesis de información a partir de un conjunto de documentos, lo que conlleva las siguientes diferencias:

- En el modelo de Boguraev el rol de sujeto es sólo un criterio más de selección de sintagmas nominales entre otros, como son el número de referencias al concepto o la localización en el documento. En PRISMA, se considera únicamente el número de apariciones de cada sintagma como sujeto sintáctico.
- En el modelo de Boguraev se muestra por orden de aparición todos los contextos en los que aparece el concepto clave. En el caso de PRISMA al partir de un conjunto voluminoso de textos, es necesario seleccionar y organizar las piezas de información asociadas al concepto. En concreto, se muestran aquellas en las que el concepto aparece como sujeto de una oración, ordenadas cronológicamente.
- En PRISMA se ofrece además una vista del contenido completo de un documento resaltando la proposición principal de cada una de las frases del texto.

En el siguiente apartado se describe con más detalle el modelo PRISMA y el prototipo implementado.

3 Modelo PRISMA

PRISMA puede descomponerse en los siguientes procesos: visualización de la lista de conceptos clave, acceso a piezas de información asociadas a conceptos clave, visualización de documentos completos y visualización

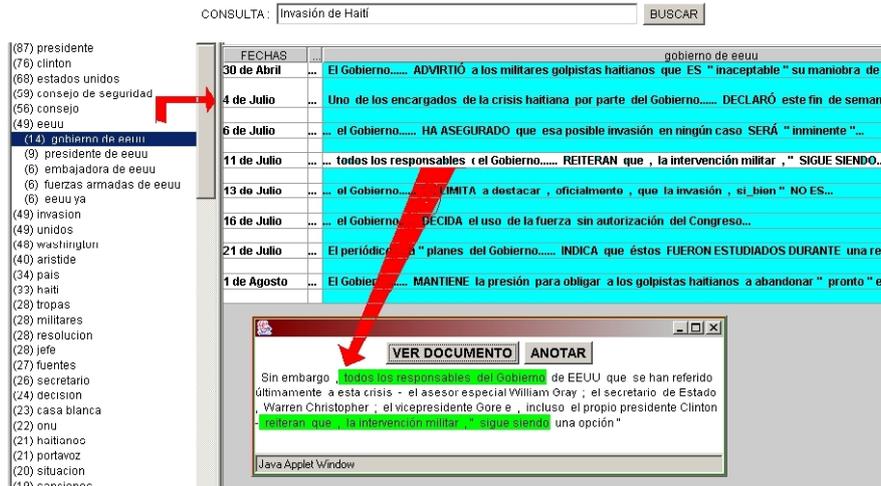


Figura 3: Secuencia de interacción en el prototipo PRISMA

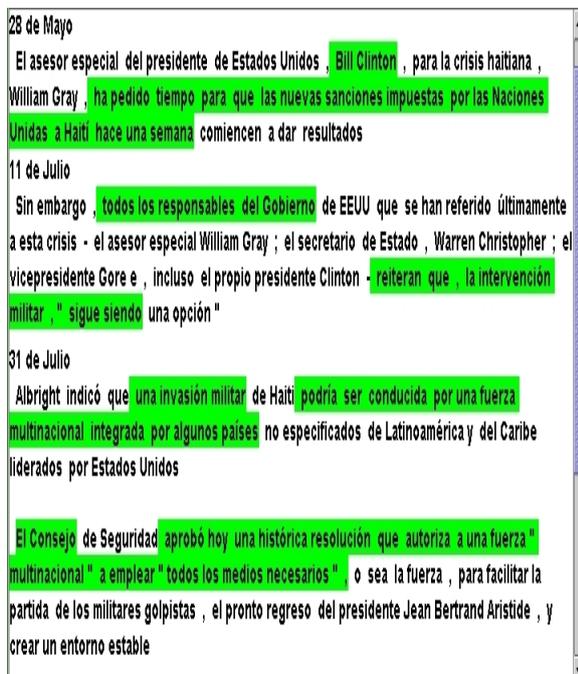


Figura 1: Informe elaborado mediante el prototipo PRISMA

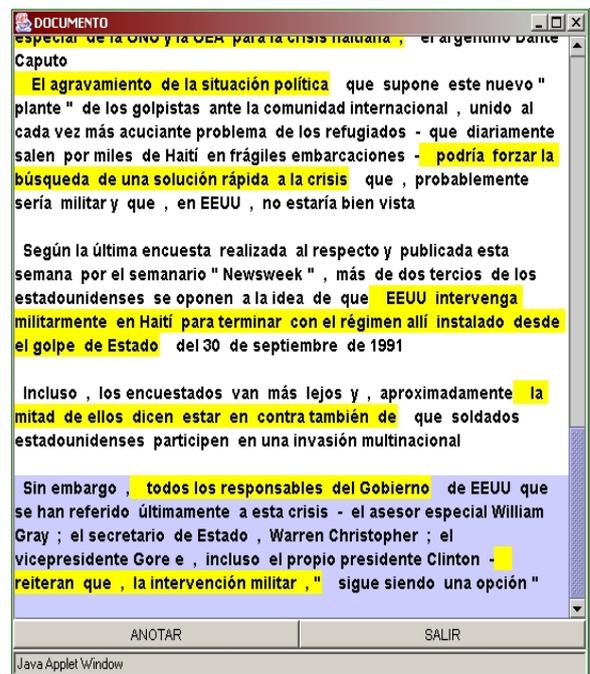


Figura 2: Visualización de un documento en el prototipo PRISMA

del informe final, como veremos a continuación.

En la figura 3, podemos ver la secuencia de pasos seguidos por un usuario en el interfaz del prototipo PRISMA durante el acceso a piezas de información a partir de una lista de conceptos clave. La figura 2 muestra el contenido de un documento completo tal y como se muestra al usuario en el prototipo PRISMA. La figura 1 muestra un hipotético informe generado por un usuario mediante un proceso de extracción de piezas de informa-

ción empleando el prototipo.

3.1 Listas de conceptos clave

Como primer paso en el proceso de síntesis, el usuario introduce una consulta en lenguaje natural. En el ejemplo, la consulta introducida es *Invasión de Haití*. A continuación, el sistema recupera un conjunto de documentos relevantes aplicando técnicas de recuperación de documentos. En el prototipo implementado se realiza una búsqueda booleana.

En la siguiente fase, el sistema sugiere al usuario una lista de conceptos clave, que son

extraídos automáticamente a partir de los sintagmas nominales que con más frecuencia aparecen con rol de sujeto sintáctico en el conjunto de documentos recuperados, como se muestra en la figura 3.

Los conceptos clave se muestran ordenados por frecuencia y organizados jerárquicamente por relaciones de subsunción. Las jerarquías de subsunción permiten al usuario acceder, a partir de un concepto general, a conceptos más precisos. En el ejemplo vemos como *EEUU* subsume a *Gobierno de EEUU*, *presidente de EEUU*, *embajadora de EEUU* y *fuerzas armadas de EEUU*.

3.2 Acceso a piezas de información

En esta fase, el usuario accede a una lista de fragmentos asociados al concepto seleccionado, como se muestra a la derecha en la figura 3. Estos fragmentos se corresponden con proposiciones en las que el término escogido aparece con rol de sujeto. Las oraciones se reducen eliminando las proposiciones subordinadas y complementos del núcleo y del verbo, a partir del analizador sintáctico superficial. Los fragmentos de texto se ordenan cronológicamente según las fechas de los documentos que los contienen.

Considerar la información sintáctica para la selección y visualización de piezas de información asociadas a un concepto permite:

- Seleccionar aquellas piezas en las que el concepto, al formar parte del sujeto, adquiere un papel relevante en la oración.
- Mostrar de forma organizada las piezas de información, facilitando su lectura.

Haciendo clic en uno de los fragmentos se accede a la oración completa. La oración reducida aparece subrayada (margen derecho inferior de la figura 3). Como puede verse en la figura, el sistema resalta la partes más importantes de la frase —el núcleo de sujeto, el verbo y algunos complementos— con el fin de facilitar la lectura.

Desde este punto, el usuario puede acceder al documento completo o bien seleccionar el fragmento para añadirlo al informe final.

3.3 Visualización de documentos completos

La figura 2 refleja la forma en que PRISMA muestra el contenido de un documento

completo. El subrayado de términos, frases o párrafos relevantes ha sido tratado en otros trabajos. El sistema iNeast (Leuski, Lin, y Stubblebine, 2003) subraya aquellos fragmentos considerados relevantes por el sistema. Los criterios empleados se corresponden con algunas técnicas aplicadas en sistemas de Resumen Automático. En otros casos, el subrayado automático está orientado al marcado de términos claves que constituyen puntos de referencia en el documento (Neff y Cooper, 1999). La diferencia de PRISMA con respecto a estas aproximaciones es que no se subraya términos ni frases completas, sino unidades intermedias de texto (proposiciones). Además, en el modelo PRISMA no se presupone que unas oraciones sean más relevantes que otras, sino que se facilita la lectura de cualquiera de las frases distinguiendo sus componentes principales.

Los criterios de PRISMA para la elección de la oración principal dentro de una oración compuesta son: proposiciones principales frente a oraciones subordinadas, estructuras completas frente a proposiciones sin sujeto o sin complementos del verbo, y localización de la proposición. Una vez seleccionada la proposición, se resalta únicamente el núcleo de sujeto, el verbo y los primeros complementos del éste.

A partir del marco que muestra el documento completo, el usuario puede seleccionar también oraciones que considere relevantes para ser incluidas en el informe final.

3.4 Generación del informe final

El informe que genera el usuario en el prototipo es de tipo extractivo, y contiene oraciones completas, seleccionadas desde la lista de piezas de información (sección 3.2) o desde la visualización de los documentos completos (sección 3.3).

Los fragmentos se organizan automáticamente en orden cronológico según la fecha del documento del que proceden.

4 Procesamiento lingüístico

Para la implementación del prototipo PRISMA ha sido necesaria la realización de un analizador sintáctico superficial adaptado al modelo. Partimos de un procesamiento lingüístico robusto y eficiente, de coste lineal, basado en autómatas finitos. Aunque el procesamiento no ofrece una precisión del 100%, es suficiente para dar soporte al mode-

lo PRISMA. En primera instancia, se realiza una fragmentación de las oraciones en segmentos que categorizamos del siguiente modo:

- [**N**] Se trata de sintagmas nominales, que se corresponde con nombres o adjetivos precedidos de un determinante, signo de puntuación o inicio de frase.
- [**V**] Etiquetamos de este modo las formas verbales.
- [**Mod**] Etiquetamos de este modo el resto de los sintagmas (adverbiales, preposicionales, etc.).
- [**Sub**] Etiquetamos así las conjunciones que introducen proposiciones subordinadas.
- [**P**] Finalmente, los signos de puntuación se etiquetan como elementos independientes.

Para la identificación de los roles sintácticos, hemos segmentado en primer lugar las oraciones compuestas teniendo en cuenta las conjunciones de subordinación ([Sub]). A continuación, hemos considerado como sujetos las secuencias del tipo [N][Mod]* ocurrentes antes del verbo. Por ejemplo:

El presidente [N] en funciones [Mod] de Haití [Mod] ha afirmado [V] que [Sub]...

Aunque no tenemos en cuenta sujetos de oraciones subordinadas o aquellos que aparecen en posiciones posteriores al verbo, esta aproximación asegura, en la mayoría de los casos, que el sintagma seleccionado como sujeto corresponde con el sujeto real de la oración.

5 Metodología de evaluación

La evaluación de un modelo interactivo de acceso a la información mediante el uso del sistema correspondiente lleva consigo dos grandes retos: los experimentos no son replicables y resultan muy costosos. Por tanto, no es sencillo integrar este tipo de evaluación dentro del ciclo de vida del desarrollo de un sistema, ya que no podemos experimentar con rapidez distintas aproximaciones. En cualquier caso, este tipo de evaluación es adecuado como punto final en el proceso de desarrollo.

Sin embargo, es posible obtener evidencias sobre la idoneidad de diferentes características del modelo desarrollado, sin que necesariamente el usuario interactúe con el sistema completo. Podemos descomponer la pregunta *¿es útil el modelo?* en tres: *¿Es apropiado el tipo de modelo propuesto para la tarea?*, *¿es capaz el sistema, en función de la información que pueda aportar el usuario, de extraer la información que éste necesita? (precisión y cobertura)*, y por último *¿se muestra de manera eficiente la información al usuario?*.

5.1 Fuentes de datos

El análisis de datos en esta metodología parte de las siguientes fuentes:

- Monitorización del proceso de realización de la tarea por parte de sujetos de prueba sin la ayuda del sistema.
- Córpora resultante de la realización de la tarea por los sujetos sin la ayuda del asistente a evaluar. Éstos pueden ser, por ejemplo, resúmenes, selección de documentos, piezas de información, conceptos u otros datos generados de forma manual.
- Experimentos parciales orientados a estudiar la forma en que la información debe de ser mostrada al usuario.

5.2 Fases en el proceso de evaluación

La tabla 1 muestra la lista de cuestiones que debemos plantearnos en el proceso de evaluación de PRISMA. A continuación, mostramos el conjunto de fases que componen la metodología de evaluación propuesta, particularizando para las cuestiones planteadas.

A. Elaboración de corpóra En nuestro caso, un grupo de personas ha elaborado corpóra de informes y listas de conceptos clave. Además, hemos llevado a cabo una monitorización de las acciones realizadas por los sujetos durante el proceso de síntesis sin la ayuda del asistente a evaluar. En el apartado 6.1 se describe el proceso de elaboración de dichos corpóra.

B. Adecuación del modelo a la tarea

En este punto nos cuestionamos si las funcionalidades y el conjunto de fases

Cuestión I	¿Es necesario abordar la tarea de síntesis de información desde una perspectiva interactiva?
Cuestión II	¿Existe relación entre el proceso de síntesis y el tipo de documentos y tema sobre el que se realiza?
Cuestión III	¿Es más apropiado aplicar un esquema del tipo CSBIS o CBBIS?
Cuestión IV	¿Juegan los conceptos clave un papel importante?
Cuestión V	¿Es suficiente un proceso extractivo en la tarea de síntesis?
Cuestión VI	¿Es útil considerar el rol de sintáctico para extraer conceptos clave?
Cuestión VII	¿Es útil considerar el rol de sintáctico para identificar piezas de información asociadas a un término?
Cuestión VIII	¿Facilitamos la exploración de piezas de información al reducir estructuras sintácticas?
Cuestión IX	¿Facilitamos la lectura de un documento resaltando las proposiciones principales?

Tabla 1: Cuestiones planteadas en la evaluación de PRISMA

que componen el proceso de interacción en el modelo propuesto es adecuado para la tarea. En PRISMA, queremos dar respuesta a las cuestiones I, II, III y IV de la tabla 1. Los experimentos realizados se describen en detalle en (Amigo et al., 2004a). En el apartado 6.2 comentamos dichos experimentos y las conclusiones obtenidas.

Queda abierta la cuestión de si es suficiente un proceso extractivo para la tarea de síntesis o si es necesario facilitar al usuario la organización y edición el informe final (cuestión V).

C. Evaluación de procesos automáticos de identificación y extracción de información Se evalúa en este punto la capacidad del modelo de dar acceso a la información que el usuario requiere en las distintas fases de interacción. Para ello, se parte de los corpórea de productos generados manualmente por usuarios de prueba. En el caso de PRISMA, las funcionalidades implicadas son:

- 1. Identificación de conceptos clave** Caracterizamos los conceptos clave identificados manualmente durante el proceso de síntesis de información. Es decir, nos planteamos la cuestión VI de la tabla 1. En el apartado 6.3 describimos algunos de los experimentos y conclusiones, presentados en (Amigo et al.,

2004b).

2. Acceso a piezas de información

Esta fase daría respuesta a la cuestión VII. Actualmente estamos elaborando experimentos en esta dirección.

D. Evaluación de procesos de visualización En esta fase se evalúa la capacidad del sistema de mostrar debidamente la información a la que se da acceso. Esta evaluación se realiza a partir de experimentos parciales comparando distintas estrategias de visualización. En nuestro caso, las funcionalidades evaluadas son:

- 1. Visualización de piezas de información** Mostrando a sujetos de prueba fragmentos organizados y reducidos siguiendo criterios sintácticos, podemos comparar el modelo con la aproximación de mostrar fragmentos completos de información (cuestión VIII). Estos experimentos se encuentran aún en periodo de implementación.
- 2. Visualización de un documento completo** Esta evaluación requiere experimentos parciales en donde se compare la exploración de documentos con y sin subrayado (cuestión IX).

6 Implementación de la metodología de evaluación en PRISMA

En los siguientes apartados describimos con más detalle la implementación de los puntos A, B y C.1 de la metodología descrita.

6.1 Elaboración de corpórea y monitorización de la tarea de usuario

Para la evaluación de PRISMA, 9 sujetos de prueba han elaborado, mediante un interfaz, un corpus de 72 informes y listas de conceptos clave, a partir de 8 conjuntos de documentos. Cada conjunto está compuesto de 100 documentos relevantes asociados a consultas del corpus empleado en CLEF 2001-2002 (Peters et al., 2002). Hemos seleccionado las 8 consultas con más documentos relevantes.

El interfaz empleado permite el acceso a documentos a partir de listas de títulos y recopilación de frases relevantes a partir del

contenido de los documentos. Hemos monitorizado el comportamiento de los sujetos de prueba en el proceso de síntesis.

Para la identificación de conceptos clave, los sujetos de prueba han respondido a las siguientes preguntas tras realizar el informe:

- ¿Quiénes son las principales personas que participan en el asunto?
- ¿Cuáles son las principales organizaciones o instituciones involucradas?
- ¿Cuáles son los principales factores que actúan en el asunto?

6.2 Adecuación del modelo a la tarea

Hemos estudiado el contenido de los informes y la distribución de conceptos clave para comprobar en qué medida y de qué forma, los informes del corpus se asemejan entre sí (Amigo et al., 2004a). Los resultados mostraron que los informes discrepan en los contenidos seleccionados, mientras que los conceptos clave parecen ser un rasgo característico común.

Estos datos apuntan a que realmente es necesario plantear la tarea de síntesis de información desde una perspectiva interactiva, dado que no parece existir un criterio unificado entre los usuarios acerca de cuál es la información relevante y cuál no (cuestión I de la tabla 1).

Por otro lado, el tratamiento de los conceptos clave parece ser un elemento característico de los informes elaborados manualmente, y por tanto, puede ser un buen criterio para organizar la información en un modelo interactivo (cuestión IV).

Además, los resultados obtenidos en nuestros experimentos apuntan a aproximaciones de tipo CBBIS, como el modelo PRISMA, dado que, aunque la distribución de conceptos clave es un rasgo común de los informes manuales, no lo es la información que contienen. Por tanto, debe ser el propio usuario el que seleccione, en última instancia, las piezas de información de las que consta su informe (cuestión III).

Por último, analizando la similitud entre informes y a partir de la monitorización de los procesos de síntesis, hemos podido comprobar que, dentro del conjunto de temas tratados se distinguen claramente dos subconjuntos: temas en los que un único asunto evoluciona a los largo del tiempo, por ejemplo, *Invasión de Haití*, y temas en los que se tratan

varias instancias de un mismo tipo de evento, por ejemplo *huelgas de hambre*, siendo el primer tipo de temas los que con diferencia requieren más esfuerzo de análisis por parte del usuario, y donde los conceptos clave juegan un papel más relevante (cuestión II).

6.3 Rol sintáctico y extracción de conceptos clave

Para responder a la cuestión VI de las descritas en la tabla 1, hemos realizado, en primer lugar, un experimento con el fin de medir la frecuencia de aparición de los conceptos relevantes en distintos puntos de las estructuras sintácticas (Amigo et al., 2004b). Hemos podido comprobar que la frecuencia de palabras pertenecientes a la lista de conceptos clave aumenta hasta un 30% en sintagmas con función de sujeto frente a otros sintagmas nominales.

En segundo lugar, hemos comparado dos tipos de aproximaciones al problema de la extracción automática de conceptos clave:

- Listas de términos extraídos mediante estrategias puramente estadísticas (OKAPI, *likelihood ratio*, frecuencia).
- Listas de términos extraídos considerando la frecuencia con la que ocurren justo antes de un verbo.

Esta segunda aproximación ofrece una cobertura sobre los conceptos clave superior a la primera en torno a un 15%, considerando listas de entre 20 y 70 términos.

Estos resultados indican que la identificación de sujetos sintácticos resulta útil en el proceso de extracción automática de conceptos clave (cuestión VI).

7 Conclusiones

En este artículo, hemos propuesto un modelo interactivo basado en conocimiento sintáctico y orientado a síntesis de información, para el que hemos implementado un primer prototipo. El modelo se basa en listas de conceptos relevantes a partir de los que el usuario interactúa con el sistema. Además, el prototipo facilita la lectura de documentos completos mediante el resaltado automático de proposiciones principales dentro de cada frase.

Hemos propuesto una metodología de evaluación de modelos interactivos de acceso a la información para la que no es necesaria la interacción del usuario con el sistema completo.

Siguiendo esta metodología, hemos llegado a las siguientes conclusiones:

- Los contenidos seleccionados en la elaboración de informes dependen en gran medida de la interpretación del usuario, lo que implica la necesidad de aplicar modelos interactivos en la tarea de síntesis.
- Los conceptos clave del asunto, que la mayoría de los usuarios comparten, juegan un papel fundamental en el proceso de síntesis.
- La identificación del rol sintáctico de los sintagmas nominales es útil para el proceso de extracción automática de dichos conceptos clave.

A la vista de los resultados obtenidos hasta el momento, la metodología propuesta nos permite integrar el proceso de evaluación dentro del ciclo de vida de desarrollo del sistema, estableciendo un marco sobre el que validar y refinar los distintos componentes del modelo PRISMA.

Bibliografía

- Amigo, E., V. Peinado, J. Gonzalo, A. Peñas, y F. Verdejo. 2004a. Information synthesis: an empirical study. En *Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics (ACL)*, Barcelona, July.
- Amigo, E., V. Peinado, J. Gonzalo, A. Peñas, y F. Verdejo. 2004b. Using syntactic information to extract relevant terms for multi-document summarization. En *Proceedings of the 36th Annual Conference on Computational Linguistics for Computational Linguistics (Coling'04)*, Geneva, August.
- Boguraev, B., C. Kennedy, R. Bellamy, S. Brawer, Y. Wong, y J. Swartz. 1998. Dynamic presentation of document content for rapid on-line skimming. En *Proceedings of the AAAI Spring 1998 Symposium on Intelligent Text Summarization*.
- Buyukkokten, O., H. García-Molina, y A. Paepcke. 1999. Seeing the whole in parts: Text summarization for web browsing on handheld devices. En *Proceedings of 10th International WWW Conference*.
- Goldstein, J., V. O. Mittal, J. G. Carbonell, y J. P. Callan. 2000. Creating and evaluating multi-document sentence extract summaries. En *CIKM*, páginas 165–172.
- Jones, S., S. Lundy, y G. W. Paynter. 2002. Interactive document summarization using automatically extracted keyphrases. En *Proceedings of the 35th Hawaii International Conference on System Sciences*.
- Leuski, A., C. Y. Lin, y S. Stubblebine. 2003. Ineats: Interactive multidocument summarization. En *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*.
- Neff, M. S. y J. W. Cooper. 1999. Ashram: active summarization and markup. En *Proceedings of the Hawaii International Conference on System Sciences (HICSS-32): Understanding Digital Documents*.
- Over, P. 2003. Introduction to duc-2003: An intrinsic evaluation of generic news text summarization systems. En *Proceedings of Workshop on Automatic Summarization (DUC 2003)*.
- Peters, C., M. Braschler, J. Gonzalo, y M. Kluck, editores. 2002. *Evaluation of Cross-Language Information Retrieval Systems*, volumen 2406 de *Lecture Notes in Computer Science*. Springer-Verlag, Berlin-Heidelberg-New York.
- Radev, D. R., J. Hongyan, y M. Budzikowska. 2000. Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies. En Udo Hahn Chin-Yew Lin Inderjeet Mani, y Dragomir R. Radev, editores, *Proceedings of the Workshop on Automatic Summarization at the 6th Applied Natural Language Processing Conference and the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, April.
- Rao, R., J. Pedersen, M. A. Hearst, J. D. Mackinlay, S. K. Card, L. Masinter, P. Halvorsen, y G. G. Robertson. 1995. Rich interaction in the digital library. *Communications of the ACM*, 38(4):29–39.