

Diseño y evaluación de un generador de resúmenes de texto con modelado de usuario en un entorno de recuperación de información

Manuel J. Maña López
Dpto. de Lenguajes y Sistemas Informáticos
Universidad de Vigo
mjlopez@uvigo.es

Manuel de Buenaga Rodríguez José María Gómez Hidalgo
Departamento de Inteligencia Artificial
Universidad Europea de Madrid – CEES
{buenaga, jmgomez}@dinar.esi.uem.es

Resumen

En este trabajo presentamos un generador de resúmenes que incorpora el modelado de las necesidades de información del usuario con el fin de crear resúmenes adaptados a las mismas. Los resúmenes se generan mediante la extracción de las frases que resultan mejor puntuadas bajo tres criterios: palabras clave, localización y título. El modelado del usuario se consigue a partir de las consultas a un sistema de Recuperación de Información y de la expansión de las mismas utilizando WordNet. Se presenta también un método de evaluación sistemático y objetivo que nos permite comparar la eficacia de los distintos tipos de resúmenes generados. Los resultados demuestran la mayor eficacia de los resúmenes adaptados a las consultas y los de aquellos que emplean WordNet.

1. Introducción

En los últimos años hemos asistido a un crecimiento explosivo de la información a la que cualquier persona puede tener acceso, favorecido por el incremento de los medios de almacenamiento masivo de datos y las redes de computadoras, en especial Internet. Este hecho ha conducido a la popularización de los sistemas de recuperación de información (IR), caracterizados por sugerir una lista de documentos relevantes, de entre una colección de ellos, relativos a una consulta en lenguaje natural realizada por un usuario. De esta manera, la salida típica de estos sistemas es una lista de títulos de documentos, ordenada por el factor de relevancia relativo a la consulta, y las primeras líneas de aquellos. Sin embargo, esta información resulta insuficiente para que en la mayoría de los casos el usuario pueda enjuiciar

la importancia del documento respecto a sus necesidades de información. Esto le obliga a realizar una inspección en la totalidad del documento, con el costo adicional en tiempo que supone esta tarea. La presentación de resúmenes de forma complementaria a los títulos puede permitir disminuir de manera significativa este costo.

El propósito de este trabajo es la construcción de un sistema capaz de generar resúmenes de texto adaptados a las necesidades concretas de información del usuario en un entorno de IR.

Un aspecto fundamental del mismo es el *modelado de las necesidades particulares de información del usuario* al que va dirigido el resumen. Esta idea no es nueva. Así, en [16] ya se menciona la posibilidad de adaptar los resúmenes a determinadas esferas de interés o campos de investigación. Luhn propone puntuar las frases asignando "valores premio" a un conjunto predeterminado de palabras que identifiquen ese ámbito de interés. En [10] también se discute sobre la importancia y necesidad de confeccionar resúmenes adaptados a necesidades particulares de información. Por otra parte, en [21] se constata, al emplear usuarios para evaluar los resúmenes del sistema que construyen, que estos tienden a seleccionar secciones del texto que están muy en relación con sus propios intereses. Nuestro objetivo es modelar estos intereses y construir resúmenes adecuados a los mismos en el ámbito de un sistema de ayuda al acceso a la información.

Tiene, además, un gran interés para nosotros la evaluación de la calidad de los resúmenes generados por nuestro sistema, por lo que pretendemos medir la manera en que ayudan al usuario a decidir sobre la relevancia del documento sin necesidad de inspeccionarlo. Esta evaluación estará orientada a presentar una evidencia importante de que la eficacia de los

resúmenes adaptados a un modelo de usuario es mayor que la de aquellos que se construyen sin tener en cuenta sus necesidades de información.

En el resto del escrito presentamos primero las técnicas utilizadas en la generación de resúmenes y, a continuación, discutimos cómo modelar las necesidades de información del usuario. Después, exponemos los resultados obtenidos en la evaluación de los diferentes tipos de resúmenes y, para finalizar, indicamos nuestras conclusiones y futuros trabajos.

2. Técnicas para la generación de resúmenes

La técnica más empleada en la generación de resúmenes es la de puntuar las frases del texto original. Las frases mejor valoradas se seleccionan e incluyen en el extracto, en el orden en que aparecen en el documento [16, 9, 19, 15, 27, 28]. Otras técnicas intentan realizar una representación semántica del significado del texto y una generación posterior del resumen, pero crean sistemas para dominios muy concretos [8, 3]. Por el contrario, la extracción de sentencias crea sistemas generadores de resúmenes independientes del dominio. Sin embargo, tiene algunos inconvenientes. Los resúmenes resultantes pueden ser *inconsistentes* [9, 20], debido a la presencia de referencias anafóricas a elementos del discurso que no se han incluido en el resumen o, simplemente, a que las frases extraídas no aparecían de manera consecutiva en el texto original. También pueden resultar *desequilibrados*, por no reflejar todos los conceptos importantes que se tratan en el documento o su organización estructural [20]. Algunos trabajos proponen soluciones para estos problemas, identificando y resolviendo las referencias anafóricas [20, 14] o empleando reglas semánticas, dependientes del dominio, para mejorar el equilibrio [21].

2.1 Criterios para la selección de frases

Para medir la significación de las frases pueden emplearse varios criterios. El método de las *palabras clave* [16] valora positivamente las frases que contienen palabras con una frecuencia de aparición alta, entendiendo que se trata de buenas indicadoras del contenido del texto. En [9] se emplea un criterio de *localización* por el cual se valoran positivamente las frases de los

párrafos iniciales o finales del documento, y especialmente aquellas que aparecen al principio o final de cada párrafo, o las situadas bajo encabezamientos como "Introducción", "Propósito" o "Conclusiones". Otra heurística consiste en emplear *expresiones indicadoras* [19] para localizar la presencia de los tópicos principales del documento, como por ejemplo "El propósito de este artículo es...". Muy relacionada con ésta es la de *palabras indicadoras* [9], en la que se dispone de dos conjuntos de palabras, denominadas *bonus* y *stigma*, de modo que cuando una de ellas aparece en una frase incrementa o decrementa su valor, dependiendo del conjunto al que pertenezca. Un último método es el que considera que las palabras que aparecen en los *títulos* son buenas candidatas para formar parte de la lista de términos índice [9] que se empleará para identificar frases significativas.

En el resto del apartado realizaremos una discusión más detallada sobre los métodos de palabras clave, localización y títulos, seleccionados para ser implementados en nuestro sistema generador de resúmenes. Los criterios de expresiones y palabras indicadoras no han sido elegidos, debido a que estos métodos de selección resultan necesariamente dependientes del dominio. Además, la técnica de expresiones indicadoras es de difícil aplicación en una colección de documentos periodísticos como el nuestro, donde es complicado encontrar construcciones léxicas que puedan indicar la importancia de la frase en la que se incluyen. De manera análoga la elección de palabras que otorguen un valor positivo o negativo a la importancia de una frase en el dominio mencionado es una tarea muy compleja.

2.2 Palabras clave

Para puntuar las frases, Luhn [16] define grupos (o *clusters*) de términos, como conjuntos de *palabras clave* separadas por un máximo de otras cuatro palabras. En este sentido en [1] se constata que el 98% de las relaciones léxicas en inglés se producen entre palabras comprendidas en un intervalo de cinco. Sin embargo, en otros trabajos [15, 27] sólo tienen en cuenta las apariciones por separado de palabras clave.

Para nuestros experimentos seleccionamos, como palabras clave, los diez términos con mayor $tf*idf$ (frecuencia del término * inversa de la frecuencia de documentos) de cada documento

[27] y el grupo con más palabras clave de cada frase. La puntuación de cada una de las frases se obtiene multiplicando la suma del peso de las palabras clave del grupo por el cociente entre el cuadrado del número de palabras clave y el número total de palabras del grupo. De esta forma perseguimos conseguir una doble ventaja. Con la utilización del factor $t_f \cdot idf$ pretendemos descartar palabras de uso común en el dominio de la colección. Con la identificación de grupos, entendidos como conjuntos de palabras clave separadas por un máximo de otras cinco palabras, aspiramos a seleccionar las frases más importantes.

2.3 Títulos

Para puntuar la frase debemos tener en cuenta que lo normal es que la frecuencia de aparición de cada término en el título sea uno y que en este caso no nos interesa su frecuencia en títulos de otros documentos. Por tanto, el valor que consideramos es el cociente entre el cuadrado del número de palabras del título que aparecen en la frase y el número de palabras significativas (i.e. no pertenecen a la lista de parada) que forman el título.

2.4 Localización

Para la aplicación de este criterio hemos tenido en cuenta las características del corpus de documentos que vamos a emplear para nuestras pruebas. Se trata de un conjunto de documentos periodísticos, de modo que lo usual es que la información más importante aparezca al comienzo. De esta forma, nuestro sistema valora positivamente y de manera decreciente las diez primeras frases de cada documento.

3. Modelado de las necesidades de información del usuario

La incorporación del modelado de usuario a la generación de resúmenes puede presentar numerosas ventajas respecto a los extractos genéricos. Estos últimos se limitan a seleccionar frases importantes del documento pero sin tener en cuenta qué información puede ser importante para el usuario o cuál es su dominio de conocimiento. De esta forma, si los resúmenes se utilizan en un entorno de IR, puede suceder que el resumen, al igual que las primeras líneas del documento no ayuden al usuario a decidir sobre la relevancia del mismo. En cambio, si el sistema consigue representar adecuadamente las

necesidades de información del usuario podrá seleccionar las frases con mayor interés y construir un resumen conforme a su perfil que le será mucho más útil.

3.1 Técnicas para modelar el perfil de usuario

Para modelar el perfil del usuario proponemos el siguiente conjunto de técnicas, empleadas con éxito en diversas áreas de la clasificación automática de textos.

1. *Procesamiento de la consulta.* En un entorno de recuperación de información la consulta de usuario es un elemento fundamental para averiguar sus necesidades de información. Las palabras empleadas por el usuario pueden ser utilizadas por el generador automático de resúmenes con el propósito de seleccionar frases con alto contenido semántico en relación con su consulta y, por ende, respecto a sus necesidades de información. En [28] se sigue este enfoque, asignando a cada frase el resultado de dividir el cuadrado del número de términos de la consulta que aparecen en la frase entre el número total de términos de la consulta.
2. *Utilización de recursos léxicos: como thesaurus, diccionarios y bases de datos léxicas.* Se han empleado con éxito en otras tareas de clasificación de texto, como categorización [5] o recuperación de información [26]. La idea es utilizar, por ejemplo, las relaciones léxicas de sinonimia que proporcionan bases de datos léxicas como WordNet [18] para construir descriptores de los términos empleados por el usuario en la consulta. Con esto conseguiremos, sin duda, aumentar la capacidad de seleccionar frases en el documento con alto interés para el usuario.
3. *Realimentación por relevancia.* En multitud de trabajos tanto en recuperación de información [6, 7] como en filtrado [2] se ha utilizado la realimentación por relevancia con notable éxito. Esta técnica consiste en utilizar el juicio del usuario sobre la relevancia o no de los documentos recuperados con el objetivo de mejorar la efectividad de la consulta inicial. Rocchio [22] propone un algoritmo, dentro del modelo de espacio vectorial, para derivar el vector de consultas óptimo a partir de los vectores de los documentos que se saben relevantes o

irrelevantes. Este vector de consultas óptimo puede ser utilizado por el generador de resúmenes como un modelo adecuado de las necesidades de información del usuario.

4. *Utilización de "estereotipos"*. Otra forma posible de adquirir conocimiento sobre las necesidades de información del usuario es asociándolo con uno o varios grupos de usuario entre un conjunto de ellos. El objetivo es que cada uno de estos grupos identifique un posible estereotipo, entendido como conjunto de suposiciones sobre las características de unos determinados intereses. En algunos sistemas [11] emplean una entrevista inicial con el usuario sobre la que basan la decisión de la asignación de grupos. Relacionado con este enfoque está el trabajo [6], en el que cada usuario o grupo de usuarios tiene asociado una red de relevancia que puede compartirse con otros usuarios de perfil similar.

En este primer prototipo que presentamos el modelado de usuario se limita a la utilización de las dos primeras técnicas. En los puntos siguientes se discute más detalladamente cómo se adoptan.

3.2 Procesamiento de la consulta

Para los experimentos de evaluación de nuestro sistema se eligieron las consultas TREC (Text REtrieval Conferences) [25], formada por tres partes: título, descripción y narrativa. El título correspondería a una consulta formulada por un usuario de un sistema IR. La descripción y narrativa son una explicación en detalle sobre las propiedades que deben tener los documentos relevantes. Al no considerarlas representativas de una consulta de usuario, las descartamos para su posible tratamiento.

Nuestro sistema entonces genera para cada documento un resumen adaptado a la consulta de usuario, esto es, al título de la consulta TREC. Por resúmenes adaptados entendemos que incluyen frases con grupos de palabras que aparecen en la consulta, en el sentido de los *clusters* de Luhn [16]. El objetivo que se persigue es incluir sólo frases verdaderamente significativas a la consulta y no aquellas que tengan sólo alguna palabra.

El modo de puntuación de las frases mediante este método es análogo al de palabras clave, pero utilizando las palabras de la consulta. Así, primero se localiza el grupo que incluye más

palabras de la consulta. La puntuación de la frase se obtiene dividiendo el cuadrado del número de palabras de este grupo que también lo son de la consulta entre el número total de palabras del mismo.

3.3 Utilización de WordNet

En un último conjunto de experimentos hemos expandido las consultas TREC empleando las relaciones de sinonimia contempladas en WordNet. Hemos realizado el proceso de expansión de cuatro formas distintas. A continuación indicamos las características de cada una de ellas:

1. Para cada palabra de la consulta se toman todos los términos correspondientes a todas las categorías gramaticales con las que aparece y todos los significados posibles.
2. También se realiza una expansión máxima pero no a partir de términos aislados, sino de las posibles subexpresiones que forman la consulta.
3. Partiendo de la colección de sinónimos anterior se realiza manualmente una desambiguación. Para ello se elige una sola categoría gramatical y un solo significado para cada término o expresión. Esto, naturalmente, hace que se restrinjan el número de sinónimos.
4. Se realiza una desambiguación similar a la que se explica en el proceso anterior pero proporcionándole a la persona que la hace, no experta en la colección de textos, la parte de descripción y narrativa que acompaña al título de la consulta TREC.

Los términos que maneja WordNet pueden ser palabras aisladas o *collocations* (frase hecha, verbo preposicional, ...). Nuestro generador, sin embargo, sólo tiene en cuenta las palabras por separado y después de eliminar aquellas que pertenecen a la lista de parada que se emplea (debemos pensar que muchas *collocations* están formadas por preposiciones, artículos e incluso algunos verbos de uso muy común). De cualquier forma, esta pérdida de información la compensamos puntuando sólo la aparición de grupos de palabras y no palabras aisladas en las frases.

4. Evaluación del sistema

La medida de la calidad de los resúmenes es un problema de difícil solución, debido a que es muy complicado establecer cuáles deben ser las

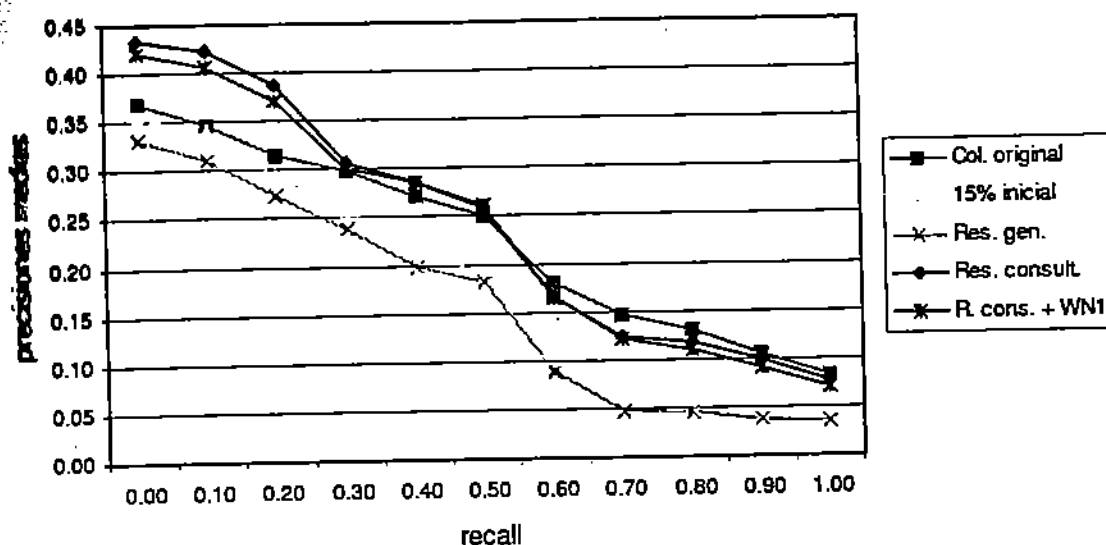


Figura 1. Precisiones medias en diferentes niveles de *recall* para la colección original, el segmento del 15% inicial de los documentos y los tres tipos de resúmenes generados.

propiedades de un buen resumen. En algunos trabajos la evaluación se ha realizado comparando el resumen generado por el sistema con el realizado por personas o con frases del texto consideradas importantes por jueces humanos [9, 15, 27]. Sin embargo, se ha constatado que el nivel de coincidencia entre diferentes resúmenes realizados por personas distintas no es muy alto, por lo que no parece que éste sea un sistema de evaluación adecuado. Un enfoque muy distinto es el que se plantea en [13], donde se sugiere evaluar la calidad del resumen respecto a una tarea determinada, con el objetivo de medir la utilidad para el usuario. Esta orientación es la seleccionada en la Fase III de TIPSTER, en la que proponen una evaluación basada en categorización y otra en IR, donde un conjunto de jueces examinará la relevancia del resumen respecto a los tópicos utilizados en una consulta. [17, 28] son trabajos en los que se ha empleado este tipo de evaluación en entornos IR.

En nuestro trabajo proponemos una evaluación indirecta, pero sistemática, de los resúmenes en un entorno de ayuda al acceso a la información. No hemos empleado usuarios como en otros trabajos [28], sino un sistema de recuperación que actúa sobre una colección de documentos y su conjunto de resúmenes generados de manera automática. De esta forma, al poder evaluar cuantitativamente la efectividad en la recuperación de documentos, es posible comparar los resultados. También podemos extraer conclusiones sobre la capacidad de retención de información importante de los

resúmenes y la medida en que reflejan las necesidades de información expresadas por el usuario en su consulta. En estas pruebas hemos utilizado el sistema de recuperación de información Smart [4], basado en el modelo de espacio vectorial y dotado de un módulo para la evaluación de la eficacia sobre colecciones de prueba.

Para poder llevar a cabo las evaluaciones mencionadas necesitamos no sólo una colección de documentos, sino también un conjunto de consultas y los correspondientes juicios sobre la relevancia de cada documento para cada consulta. El corpus elegido pertenece a la colección TIPSTER, una de las más empleadas en IR. En concreto se eligieron al azar 5000 documentos pertenecientes a la colección Wall Street Journal. También al azar se escogieron 50 consultas TREC de entre aquellas que tenían algún documento relevante en los 5000 seleccionados. Nuestro método de evaluación nos ha permitido utilizar una colección de documentos de dimensiones mayores que las empleadas en otros trabajos [15, 27, 28].

4.1 Experimentos realizados

Como paso previo se realizó una evaluación sobre la colección original de documentos. Los resultados que muestra Smart se basan en tres medidas de efectividad: *recall*, precisión y precisión truncada [23, 24]. La evolución pareja de precisión y *recall* se observa en la figura 1, que comentamos más adelante. Las tablas 1 y 2 contienen datos concretos del *recall* y la precisión truncada. *Recall* es la proporción de

	Colección original	15% inicial col. orig.	Resúmenes genéricos	Res. con consultas	Res. cons. + WN1
Recall					
Exacto	0.7208	0.5364	0.4964	0.5979	0.6250
con 5 docs.	0.1847	0.1528	0.1455	0.1970	0.2197
con 10 docs.	0.2848	0.1932	0.1897	0.2856	0.2873
con 15 docs.	0.3213	0.2215	0.2189	0.3147	0.3234
con 30 docs.	0.4243	0.3010	0.2824	0.3988	0.4170
Precisión truncada					
Exacto	0.0999	0.0721	0.0705	0.1032	0.0949
con 5 docs.	0.2170	0.1700	0.1660	0.2267	0.2200
con 10 docs.	0.1975	0.1460	0.1380	0.1987	0.1800
con 15 docs.	0.1755	0.1300	0.1233	0.1727	0.1571
con 30 docs.	0.1479	0.1040	0.0975	0.1430	0.1354

Tabla 1. Precisión truncada y *recall* para la colección original, el segmento del 15% inicial de los documentos y los tres tipos de resúmenes generados.

documentos relevantes de la colección que se han recuperado. *Precisión* es la proporción de documentos recuperados que son relevantes. Ambas medidas entienden por "recuperados" aquellos documentos con un rango inferior al número máximo de documentos esperados, 200 para nuestros experimentos. Sin embargo, la *precisión truncada* entiende por número de documentos recuperados el mínimo entre el rango del peor documento relevante recuperado y el número de documentos esperados. El propósito es utilizar un valor de precisión que no disminuya después de que se hayan recuperado todos los documentos relevantes para una consulta. Los resultados finales que muestra Smart los obtiene haciendo la media para todas las consultas.

Para demostrar la validez y eficacia de los resúmenes generados hemos construido una colección formada por el segmento del 15% inicial de los documentos originales (medido en número de frases). Esa es la longitud de los resúmenes con la que hemos experimentado. De esta forma podremos comparar la eficacia de los resúmenes respecto a la salida de un sistema de IR convencional, esto es, las primeras líneas del documento.

En la obtención de los resúmenes genéricos se utilizan, como ya se ha dicho, los métodos de palabras clave, localización y título. La puntuación final de una frase se obtiene normalizando el peso de los métodos en el intervalo 0-1, para lo que se divide el valor obtenido con cada método por el máximo respectivo en cada documento, y multiplicándolo por un determinado peso [9]. Los valores que muestran la figura 1 y la tabla 1 se obtuvieron con pesos de 1.5, 0.5 y 1 respectivamente.

Para evaluar los resúmenes adaptados a las consultas se crearon 50 colecciones de 5000 documentos cada una. Cada colección contenía un resumen de cada documento del corpus original adaptado a la misma consulta. Se utilizaron los pesos anteriores añadiendo 3.0 para la puntuación de las consultas. De esta manera se evaluó por separado cada colección con su consulta asociada. Las medias de los valores obtenidos se presentan en la figura 1 y en la tabla 1.

En cuanto a la experimentación con las consultas y WordNet se hicieron cuatro experimentos correspondientes a los cuatro procesos de expansión que hemos mencionado. La tabla 2 recoge un resumen de los resultados obtenidos. En la figura 1 y en la tabla 1, para poder compararlos con los otros resúmenes, se han incluido los datos correspondientes al primer proceso de expansión por ser el de más fácil automatización.

4.2 Interpretación de los resultados

La figura 1 nos da una visión general muy clara del comportamiento de todos los sistemas. Las mejores precisiones para los diferentes niveles de *recall* las obtienen los resúmenes que emplean consultas y sinónimos de WordNet. Por debajo aparecen las precisiones para el segmento inicial y los resúmenes genéricos. La precisión en la colección original está por debajo de los resúmenes adaptados al usuario en los niveles de *recall* inferiores y es similar a éstos en el resto de niveles.

Como estábamos apuntando, no existe una gran diferencia entre los resultados obtenidos por los resúmenes genéricos en comparación con el 15% inicial: esto sin duda se debe a las

	Res. cons. + WN1	Res. cons. + WN2	Res. cons. + WN3	Res. cons. + WN4
Recall				
Exacto	0.6250	0.6240	0.5977	0.5977
con 5 docs.	0.2197	0.2187	0.2020	0.2020
con 10 docs.	0.2873	0.2965	0.2769	0.2769
con 15 docs.	0.3234	0.3326	0.3116	0.3116
con 30 docs.	0.4170	0.4210	0.3962	0.3962
Precisión truncada				
Exacto	0.0949	0.0946	0.1022	0.1023
con 5 docs.	0.2200	0.2200	0.2307	0.2307
con 10 docs.	0.1800	0.1860	0.1947	0.1947
con 15 docs.	0.1571	0.1611	0.1691	0.1691
con 30 docs.	0.1354	0.1361	0.1414	0.1414

Tabla 2. Precisión truncada y *recall* correspondientes a los distintos procesos de expansión de las consultas con WordNet.

características del corpus de documentos con el que se han hecho las pruebas. Los documentos periodísticos tienden a condensar la información importante al comienzo de los mismos. Con otro tipo de corpus probablemente se habrían obtenido peores resultados para las pruebas del segmento inicial.

Otra conclusión importante son los buenos resultados que obtienen los resúmenes adaptados a las consultas respecto al segmento inicial y a los resúmenes genéricos, como reflejan los datos de la tabla 1. Resultan especialmente significativos los valores exactos. En cuanto a *recall*, mientras en los resúmenes genéricos baja en un 7.46% respecto al segmento inicial, en los resúmenes que emplean las consultas aumenta en un 11.46% respecto al mismo valor. Para la precisión truncada los datos son aún más contundentes: disminuye en un 2.22% en los resúmenes genéricos y crece en un 43.13% en los que emplean consultas.

Al incluir los sinónimos de WordNet se mejora el *recall* en un 16.52%, mientras que la precisión truncada, un poco por debajo de la conseguida en los resúmenes por consulta, aumenta en un 31.62%, siempre respecto a la colección que contiene el segmento inicial.

Por último, como puede observarse en la tabla 2, las diferencias obtenidas dependiendo del proceso de expansión utilizado son mínimas. Como era previsible, aumenta la precisión con los procesos 3 y 4, que utilizan desambiguación, a costa de perder algo en *recall*.

5. Conclusiones y futuros trabajos

En este trabajo hemos presentado un generador de resúmenes cuya característica

principal es la incorporación de un modelo de usuario con el objetivo de adaptar los resúmenes a las necesidades de información del mismo. El modelado del usuario se realiza a partir de las consultas a un sistema de IR y de la expansión de las mismas utilizando WordNet. El objetivo final que se persigue es proporcionar al usuario de un sistema de IR un resumen adecuado a sus necesidades que le ayude a decidir sobre la relevancia del mismo sin necesidad de inspeccionarlo, tan solo estudiando un porcentaje mínimo de las frases del mismo que le ofrece nuestro sistema.

Se presenta, además, un método de evaluación sistemático y objetivo que nos permite comparar la eficacia de los resúmenes genéricos, adaptados a las consultas y que emplean los sinónimos de WordNet. Los resultados obtenidos demuestran la superioridad de los dos últimos.

En futuros trabajos seguiremos mejorando el modelado de usuario incluyendo la realimentación por relevancia y los estereotipos. En cuanto a la realimentación estamos muy interesados en comprobar como se comportan los resúmenes, ya que en [12] se comprueba que se obtienen mejores resultados con colecciones cuyos documentos son títulos y resúmenes (CACM) que con otras que contienen textos completos (WEST).

Referencias

- [1] Abracos, J. y G.P. Lopes. 1997. Statistical methods for retrieving most significant paragraphs in newspaper articles. En *Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization*, 51-57.

- [2] Allan, J. 1996. Incremental Relevance Feedback for Information Filtering. En *Proceedings of the 19th ACM/SIGIR*, 270-278.
- [3] Aretoulaki, M. 1997. COSY-MATS: An Intelligent and Scalable Summarisation Shell. En *Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization*.
- [4] Buckley, C. 1985. *Implementation of the Smart Information Retrieval System*. Technical Report 85-686, Cornell University.
- [5] Buenaga, M., J.M. Gómez y B. Díaz. 1997. Using WORDNET to Complement Training Information in Text Categorization. En *Proceedings of the 2nd RANLP*.
- [6] Chen, J.R. y Nathalie Mathé. 1995. Learning Subjective Relevance to Facilitate Information Access. En *Proceedings of the Fourth CIKM*.
- [7] Croft, W.B. 1995. Effective Text Retrieval Based on Combining Evidence from the Corpus and Users. *IEEE Expert*, 10(6):59-63.
- [8] DeJong, G. 1982. An overview of the FRUMP system. En Lehnert, W.G. y M.H. Ringle (eds.), *Strategies for Natural Language Processing*, 149-172. Lawrence Erlbaum, Londres.
- [9] Edmundson, H.P. 1969. New Methods in Automatic Abstracting. *Journal of the ACM*, 16(2):264-285.
- [10] Endres-Niggemeyer, B., J. Hobbs y K.S. Jones. 1993. *Summarizing text for intelligent communication*. Technical Report Dagstuhl Seminar Report 79, 13.12-19. 12.93 (9350), IBFI, Dagstuhl.
- [11] Fink, J., A. Kobsa y A. Nill. 1997. Adaptable and Adaptive Information Access for All Users, Including the Disabled and the Elderly. En *The Sixth Conference on User Modeling*, Springer Wien.
- [12] Haines, D. y W.B. Croft. 1993. Relevance Feedback and Inference Networks. En *Proceedings of the 16th ACM/SIGIR*, 2-11.
- [13] Hand, T.F. 1997. A Proposal for a Task-based Evaluation of Text Summarization Systems. En *Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization*.
- [14] Johnson, F.C., C.D. Paice, W.J. Black y A.P. Neal. 1993. The application of linguistic processing to automatic abstract generation. *Journal of Document and Text Management*, 1(3):215-241.
- [15] Kupiec, J., J.O. Pedersen y F. Chen. 1995. A Trainable Document Summarizer. En *Proceedings of the 18th ACM/SIGIR*, 68-73.
- [16] Luhn, H.P. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159-165.
- [17] Mani, I. y E. Bloedorn. 1997. Multidocument Summarization by Graph Search and Matching. En *Proceedings of AAAI*.
- [18] Miller, G. 1995. WordNet: a lexical database for english. *Communications of the ACM*, 38(11).
- [19] Paice, C.D. 1981. The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self-Indicating Phrases. *Information Retrieval Research*, 172-191.
- [20] Paice, C.D. 1990. Constructing Literature Abstracts by Computer: Techniques and Prospects. *Information Processing and Management*, 26(1):171-186, UK.
- [21] Paice, C.D. y P.A. Jones. 1993. The Identification of Important Concepts in Highly Structured Technical Papers. En *Proceedings of the 16th ACM/SIGIR*, 69-78.
- [22] Rocchio, J.J. 1971. Relevance Feedback in Information Retrieval. En G. Salton (ed.), *The SMART Retrieval System: Experiments in Automatic Document Processing*, 313-323. Prentice-Hall, Inc.
- [23] Salton, G. 1989. *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Addison Wesley.
- [24] Salton, G. y M.J. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.
- [25] Sparck Jones, K. 1995. Reflections on TREC. *Information Processing and Management*, 31(5):291-314.
- [26] Spink, A. 1994. Term Relevance Feedback and Query Expansion: Relation to Design. En *Proceedings of the 17th ACM/SIGIR*, 81-90.
- [27] Teufel, S. y M. Moens. 1997. Sentence extraction as a classification task. En *Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization*.
- [28] Tombros, A. 1997. *Reflecting User Information Needs through Query Biased Summaries*. MSc Thesis, Technical Report (TR-1997-35), Department of Computing Science, University of Glasgow, UK.