

# Integración del reconocimiento de la implicación textual en tareas automáticas de resúmenes de textos\*

## *Incorporating Textual Entailment Recognition in Single- and Multi-Document Summarization Systems*

Elena Lloret, Óscar Ferrández, Rafael Muñoz y Manuel Palomar

Dept. Lenguajes y Sistemas Informáticos (Universidad de Alicante)

Carretera San Vicente s/n 03690

San Vicente del Raspeig, Alicante, España

{elloret, ofe, rafael, mpalomar}@dlsi.ua.es

**Resumen:** Este artículo presenta un estudio preliminar sobre la influencia de la implicación textual en tareas de generación automática de resúmenes de textos. Se proponen dos aproximaciones para producir resúmenes a partir de uno o varios documentos de entrada. La primera de ellas se basa únicamente en la frecuencia de las palabras para determinar las frases más relevantes del texto, mientras que la segunda aproximación combina esta técnica con el reconocimiento de la implicación textual para reducir el número de frases del documento de entrada. Los resultados iniciales obtenidos presentan una mejora del 5.85 % para monodocumento y del 6.80 % para multidocumento cuando se incorpora el reconocimiento de la implicación textual.

**Palabras clave:** resumen automático, implicación textual, frecuencia de palabras, multidocumento

**Abstract:** This paper presents a preliminary study of the usefulness of applying textual entailment recognition in text summarization tasks. Two different approaches are proposed. The first one relies on the word's frequency in order to select the most relevant sentences to include in the final summary, whereas the second one combines textual entailment recognition together with the aforementioned approach so that we can reduce the size of the original text or texts first. Results show that this combination has been appropriate, obtaining an increase of 5.85 % for single-document and of 6.80 % for multi-document when textual entailment and the frequency of words are applied together.

**Keywords:** automatic summarization, textual entailment, word-frequency, multi-document

## 1. Introducción

Debido al gran aumento de información digital, especialmente desde el crecimiento exponencial de Internet, la necesidad de herramientas capaces de manejar, controlar y filtrar información es cada vez más importante tanto para la comunidad científica como para los usuarios finales. Consecuentemente, las tareas automáticas de resúmenes de textos, capaces de presentar la información disponible de una manera más con-

cisa para el usuario, están adquiriendo una gran relevancia recientemente. Además, existen otras tareas dentro del campo del Procesamiento del Lenguaje Natural (PLN) que también contribuyen al avance científico en estos aspectos. Tareas como la recuperación de información, extracción de información, búsqueda de respuestas, clasificación de textos e implicación textual están siendo ampliamente investigadas tanto de manera individual como colaborativa. En este artículo exploraremos la posibilidad de usar implicación textual para ayudar a la tarea de resúmenes de textos, estudiando la manera en la que ésta se ve influenciada.

Un **resumen** se define como un texto que se genera a partir de uno o más textos,

\* Esta investigación ha sido financiada por el proyecto TEXT-MESS (TIN2006-15265-C06-01) del Ministerio de Educación y Ciencia y parcialmente financiada bajo el proyecto QALL-ME (FP6-IST-033860), dentro del Sexto Programa Marco de Investigación de la Unión Europea.

que contiene la información más significativa y que no es más extenso que la mitad de los textos originales (Hovy, 2005). Los sistemas de resúmenes de textos pueden estar caracterizados por múltiples factores, pero según (Spärck Jones, 2007) hay principalmente tres que influyen en los resúmenes: la entrada, el propósito y los factores de salida. Por ejemplo, aunque esta tarea tradicionalmente se ha centrado en resúmenes de textos, éstos también pueden obtenerse de información multimedia como imágenes, videos, audios o información on-line o hipertextos. Además, podemos realizar el resumen de un solo documento (resumen monodocumento) o de varios (resumen multidocumento). Respecto a la salida, el resumen puede ser un *extract*<sup>1</sup> o un *abstract*. También es posible distinguir entre resúmenes genéricos, los cuales intentan representar toda la información relevante del texto original, y resúmenes orientados a usuario, en los cuales se representa la información necesaria para un usuario específico. Acerca del estilo de la salida, encontramos una distinción principalmente entre dos tipos: resúmenes indicativos, usados para mostrar los temas de los que trata el texto original y que darían una breve idea sobre lo que se comenta en él, e informativos, destinados a cubrir en mayor medida el contenido del texto a resumir.

Por otro lado, como breve introducción a la implicación textual, decir que ha sido recientemente propuesto como un marco de trabajo genérico para el modelado de la variabilidad del lenguaje en muchas aplicaciones de PLN. La implicación textual se define como una relación unidireccional entre dos textos, de tal manera que del significado de uno de ellos se puede deducir el significado del otro. Siguiendo la metodología propuesta en (Glickman, 2006), el texto inferido sería la *hipótesis* (H) y el texto que permite la inferencia de significados se denomina *texto* (T). Varias aproximaciones han sido propuestas recientemente, la mayoría de ellas y seguramente las más relevantes, han sido presentadas en las tres ediciones de la competición sobre reconocimiento de implicación textual

<sup>1</sup>Adoptamos la terminología inglesa de *extract* para referirnos al resumen formado a partir de la extracción de algunas frases del texto original previamente seleccionadas, mientras que utilizaremos el término *abstract* cuando se genera nuevo lenguaje para componer el resumen final.

(RTE, del inglés *Recognising Textual Entailment*) (Giampiccolo et al., 2007).

La siguiente figura muestra dos ejemplos, uno positivo y otro negativo, de relaciones implicación textual extraídos del corpus de desarrollo de la tercera competición del RTE.

**Par id=109 implicación=SI**

T: ASCAP is a membership association of more than 200,000 US composers, songwriters, lyricists and music publishers of every kind of music.

H: More than 200,000 US composers, songwriters, lyricists and music publishers are members of ASCAP.

**Par id=194 implicación=NO**

T: US President George W. Bush has indicated he will invite Abbas to the United States for talks, something he never did with Abbas's predecessor, the late Yasser Arafat.

H: Yasser Arafat succeeded Abbas.

Figura 1: Ejemplos de implicación textual.

El objetivo de este artículo es valorar cómo el uso de un módulo de implicación textual en un sistema de resúmenes de textos puede mejorar los resultados finales. A lo largo del artículo se detallarán los diferentes experimentos realizados, tanto para monodocumento como para multidocumento, así como los resultados obtenidos.

El artículo está estructurado de la siguiente manera: la sección 2 presenta un breve estado de la cuestión. La sección 3 describe el sistema junto con los diferentes módulos que lo componen. A continuación, la sección 4 muestra la evaluación del sistema y los resultados obtenidos. Por último, la sección 5 presenta las conclusiones y trabajos futuros.

## 2. Estado de la cuestión

La tarea de generación de resúmenes de textos de forma automática se inició hace más de cincuenta años con las investigaciones realizadas por (Luhn, 1958) o (Edmundson, 1969), quienes aplicaron técnicas como la frecuencia de palabras o la posición de una frase dentro de un documento, para producir resúmenes sin depender de la intervención humana. A partir de estas primeras investigaciones, se han desarrollado y perfeccionado multitud de técnicas, algunas de ellas basadas en conocimiento y recursos lingüísticos como (Lin y Hovy, 2002; Gotti et al., 2007) y otras basadas en métodos estadísticos y en aprendizaje automático para indentificar las fras-

es que compondrán el resumen final (Hirao et al., 2002; Svore, Vanderwende, y Burges, 2007). Además, en los últimos años, la generación de resúmenes multidocumento ha cobrado especial interés en este campo del PLN (Goldstein et al., 2000; Mihalcea, 2004; Qiu et al., 2007; Kuo y Chen, 2008) debido a la gran cantidad de información redundante de la que disponemos y la importancia de condensar o fusionar toda esa información de forma breve y resumida.

En la literatura podemos encontrar varias revisiones del estado de la cuestión (Alonso et al., 2004; Spärck Jones, 2007) en las que se describen los aspectos básicos relacionados con la tarea de resúmenes automáticos y las técnicas existentes para dicha tarea (tanto para resúmenes monodocumento como multidocumento). Además, también se exponen y analizan ejemplos de algunos sistemas concretos que han participado en evaluaciones internacionales, tales como las que se han organizado desde el 2001 hasta el 2007 por el *National Institute of Standards and Technology* (NIST) dentro del marco de conferencias DUC (<http://duc.nist.gov/>).

Por otra parte, la tarea de generación de resúmenes automáticos se puede beneficiar de técnicas existentes para otras tareas del PLN, como categorización de textos, recuperación de información o implicación textual. En este artículo nos centraremos en esta última tarea, el reconocimiento de la implicación textual, para estudiar cómo se puede aplicar conjuntamente con la tarea de resúmenes automáticos de texto, obteniendo unos resultados preliminares que son prometedores para futuras líneas de investigación. Se trata de una aproximación novedosa, en la que se utiliza el reconocimiento de la implicación textual para reducir el tamaño del documento original. Hasta ahora, la implicación textual solamente se había utilizado en la evaluación de resúmenes automáticos (Harabagiu, Hickl, y Lacatusu, 2007) sin que influyera directamente en la generación de éstos. En (Tatar et al., 2008) encontramos una aproximación en la que se aplica implicación textual como técnica de segmentación, agrupando las frases que están implicadas entre ellas para extraer posteriormente de esos segmentos, las oraciones que forman el resumen final. Por el contrario, nuestro trabajo muestra una manera diferente de integrar la implicación textual en un sistema de generación de resúmenes,

que consiste en eliminar oraciones que están implicadas entre sí y reducir así el número de frases del documento original. En un proceso posterior, de entre las frases restantes, se seleccionarán aquéllas con mayor puntuación, de acuerdo a los criterios establecidos en la sección 3, para formar parte del resumen final. En la siguiente sección se describen los detalles relacionados con la arquitectura del sistema.

### 3. Arquitectura del sistema

El sistema que hemos desarrollado produce *extracts* de uno o varios documentos de forma automática aplicando dos componentes como fuentes de información: la frecuencia de las palabras y la implicación textual. Nuestra hipótesis de trabajo es que el reconocimiento de la implicación textual puede ser muy útil como paso previo al procesamiento del texto o textos a la hora de generar resúmenes de uno o varios documentos. A partir de la identificación de esas relaciones de implicación podemos reducir el tamaño del texto a procesar. El objetivo de este trabajo consiste en estudiar la influencia positiva que tiene la implicación textual en la generación de resúmenes automáticos. A continuación, se explican las distintas fuentes de conocimiento utilizadas en el sistema (frecuencia de palabras e implicación textual), la relación entre ellas y por último, una primera aproximación para extender el sistema a multidocumento.

#### 3.1. Frecuencia de palabras

La frecuencia de las palabras es la técnica base de nuestro sistema<sup>2</sup>. La idea subyacente es que cuántas más veces aparezca una palabra en un texto (sin tener en cuenta las *stopwords*), más importante será la oración que contenga esa palabra y por tanto, las frases con más puntuación serán las que formen parte del resumen final. Es decir, la puntuación de una oración, tal como se muestra en la Fórmula 1, consistirá en la suma de las frecuencias de las palabras que contenga y se normalizará dividiendo por la longitud de dicha frase:

$$Sc_s = \frac{\sum_{i=1}^n tf_i}{n} \quad (1)$$

<sup>2</sup>Para calcular la frecuencia de las palabras hemos usado la herramienta *WVTool* desarrollada por Michael Wurst (Universidad de Dortmund) y disponible en <http://wvtool.sourceforge.net/>

donde:

$t_{fi}$  = frecuencia de la palabra  $i$ .

$n$  = longitud de la frase (en número de palabras) sin tener en cuenta las stopwords.

Si consideramos, por ejemplo, las dos frases mostradas en la Figura 2, extraídas del corpus del DUC del 2002 (<http://www-nlpir.nist.gov/projects/duc/data.html>), la puntuación para la frase  $a$  es 3.2 y para la  $b$  es 1. Entre paréntesis se indica la frecuencia de cada palabra respecto a todo el documento, y se puede observar como las “stopwords” no se han tenido en cuenta para el cálculo de la frecuencia ni para contar la longitud de la oración.

$S_a$ : Tropical(2) Storm(6) Gilbert(7) formed(1)  
in(0) the(0) eastern(1) Caribbean(1) and(0)  
strengthened(1) into(0) a(0) hurricane(7)  
Saturday(4) night(2).  
 $S_b$ : There(0) were(0) no(0) reports(1) of(0)  
casualties(1).

Figura 2: Frases de ejemplo del documento AP880911-0016 (corpus DUC 2002).

Si tuviéramos que escoger entre una de ellas para formar el resumen, elegiríamos la primera oración ya que la puntuación que obtenemos (3.2) es mayor que la obtenida para la segunda (1).

### 3.2. Implicación textual

La idea principal que sostiene el uso de implicación textual en tareas automáticas de resúmenes, reside en conseguir un resumen preliminar formado por las oraciones que no tienen relación de implicación con ninguna otra frase del documento. Como ejemplo, supongamos que un documento está formado por el siguiente conjunto de frases:

$$S_1 S_2 S_3 S_4 S_5 S_6$$

y el documento reducido por el cómputo de las relaciones de implicación textual se obtiene de la siguiente manera:

$$\begin{aligned} RESUMEN &= \{S_1\} \\ RESUMEN &\longrightarrow implica \longrightarrow S_2 \Rightarrow NO \\ RESUMEN &= \{S_1, S_2\} \\ RESUMEN &\longrightarrow implica \longrightarrow S_3 \Rightarrow NO \\ RESUMEN &= \{S_1, S_2, S_3\} \\ RESUMEN &\longrightarrow implica \longrightarrow S_4 \Rightarrow SI \\ RESUMEN &= \{S_1, S_2, S_3\} \end{aligned}$$

$$\begin{aligned} RESUMEN &\longrightarrow implica \longrightarrow S_5 \Rightarrow SI \\ RESUMEN &= \{S_1, S_2, S_3\} \\ RESUMEN &\longrightarrow implica \longrightarrow S_6 \Rightarrow NO \\ RESUMEN &= \{S_1, S_2, S_3, S_6\} \end{aligned}$$

Por lo tanto, el documento reducido obtenido por el procesamiento de las inferencias de implicación textual comprende aquellas frases que no son implicadas por el conjunto de oraciones que no han producido relación de implicación previamente (i.e.  $S_1, S_2, S_3, S_6$  en el ejemplo anterior). Para procesar las implicaciones textuales se ha utilizado el sistema de implicación textual presentado en (Ferrández et al., 2007) con algunas mejoras y entrenado con el corpus proporcionado en la Tercera edición del RTE (Giampiccolo et al., 2007). Este sistema se basa en el cómputo de un conjunto de medidas léxicas (e.g. distancia de Leveshtein, Smith–Waterman, similitud del coseno) y semánticas basadas en WordNet3.0, aplicando un clasificador SVM con el objetivo de tomar la decisión final.

### 3.3. Frecuencia de palabras e implicación textual

Con el objetivo de valorar cómo la detección de relaciones de implicación textual mejora la tarea completa de resúmenes de textos, creamos una aproximación que se beneficia del uso de las dos técnicas anteriores (la frecuencia de palabras y la implicación textual).

Inicialmente aplicamos implicación textual para obtener el documento reducido, el cual contendrá sólo aquellas frases que no han producido relación de implicación. En paralelo, aplicamos la técnica de la frecuencia de las palabras sobre las frases del documento obteniendo la frecuencia de cada palabra. Finalmente, creamos el resumen a partir de aquellas frases que están contenidas en el documento reducido y que mayor puntuación obtengan.

### 3.4. Extensión a multidocumento

Partiendo de la hipótesis de que aplicar implicación textual como parte del proceso de generación automática de resúmenes es ventajoso ya que se elimina información redundante del documento de origen, y que la redundancia de información es uno de los problemas más importantes en el ámbito de los resúmenes multidocumento (Radev, Hovy, y McKeown, 2002), se decidió extender el sis-

tema a multidocumento para estudiar el comportamiento de la implicación textual en este nuevo ámbito. Como una primera aproximación muy básica, se optó por considerar todos los documentos de entrada al sistema como un solo documento, concatenando uno a continuación del otro y aplicando, en primer lugar, el reconocimiento de la implicación textual como mecanismo para eliminar frases redundantes. De esta manera, conseguimos reducir considerablemente el tamaño del texto. El cálculo de las frecuencias de palabras se aplicaría después sobre ese nuevo texto, tal y como se describe en la sección 3.1, para determinar las frases que serán extraídas para formar parte del resumen final. En la siguiente sección se describen detalladamente los experimentos y los resultados obtenidos.

#### 4. Evaluación del sistema

Para evaluar nuestro sistema de manera *intrínseca*<sup>3</sup> hemos tomado como datos de prueba los documentos utilizados en el DUC 2002. El corpus de prueba está formado por 59 grupos de noticias que contienen aproximadamente 10 documentos que tratan sobre el mismo tema. En esta edición del DUC, la tarea 1 se orientó a monodocumento mientras que la tarea 2, a multidocumento. La primera de ellas proponía la generación de resúmenes genéricos de 100 palabras de cada uno de los documentos y la segunda, la generación de resúmenes genéricos de diversa longitud (10,50,100,200) de cada uno de los grupos. A partir de la definición de estas tareas, decidimos realizar las siguientes pruebas (véase el Cuadro 1 de resultados):

- Generación de *extracts* de 100 palabras de longitud, para cada uno de los documentos del corpus del DUC 2002 aplicando solamente la frecuencia de las palabras en el proceso de resumen (monodocumento, TS).
- Generación de *extracts* de 100 palabras de longitud, para cada uno de los documentos del corpus del DUC 2002 aplicando el reconocimiento de la implicación textual como paso previo a la frecuencia de las palabras en el proceso de resumen (monodocumento, TE+TS).

- Generación de *extracts* de 100 palabras de longitud, para cada uno de los grupos del corpus del DUC 2002 considerando solamente la frecuencia de las palabras en el proceso de resumen (multidocumento, TS).
- Generación de *extracts* de 100 palabras de longitud, para cada uno de los grupos del corpus del DUC 2002 aplicando el reconocimiento de la implicación textual como paso previo a la frecuencia de las palabras en el proceso de resumen (multidocumento, TE+TS).

Para llevar a cabo estas pruebas, realizamos un preprocesado inicial a los textos proporcionados por el DUC, quedándonos solamente con el cuerpo de la noticia de cada uno de los documentos, y eliminando todas las etiquetas con información adicional. La razón por la que se hizo este tratamiento previo fue que dichas etiquetas introducían mucho ruido en los resúmenes finales y aunque, a priori se pudiera pensar que estamos eliminando información relevante de los documentos, como puede ser el título de la noticia, esta información suele venir resumida de nuevo en la primera frase de la misma, al tratarse de textos periodísticos.

El corpus del DUC 2002 contenía, además, resúmenes generados manualmente por expertos humanos, que nos sirvieron para compararlos con los obtenidos por nuestro sistema. Por otra parte, se desarrollaron para el DUC 2002 dos *baselines*, una para la tarea 1 consistente en extraer las primeras 100 palabras del documento y otra para la segunda tarea, que seleccionaba las primeras  $n$  palabras (en nuestro caso, también 100, ya que los resúmenes multidocumento producidos tienen una longitud de 100 palabras) del documento más reciente. Ambos *baselines* nos sirvieron de referencia y reciben el nombre de *Lead baseline* como se puede ver en el Cuadro 1.

Para la evaluación se utilizó el paquete ROUGE<sup>4</sup> (Lin, 2004), con un intervalo de confianza del 95%. Esta herramienta, desarrollada en 2004, ha sido muy utilizada para la evaluación automática de resúmenes en las últimas competiciones del DUC y proporciona los valores de precisión, cobertura

<sup>3</sup>Este tipo de evaluación difiere de la *extrínseca* en que se evalúa el contenido del resumen, es decir, la calidad del resumen generado atendiendo a diversos criterios.

<sup>4</sup>La versión utilizada en este trabajo, ROUGE-1.5.5, se encuentra disponible en: <http://haydn.isi.edu/ROUGE>

<b>Monodocumento</b>	ROUGE-1	ROUGE-2	ROUGE-L
Lead baseline	41.132	21.075	37.535
TS	43.210	17.072	39.188
TE+TS	<b>44.759</b> (+3.58 %)	<b>18.840</b> (+10.36 %)	<b>40.606</b> (+3.62 %)
<b>Multidocumento</b>	ROUGE-1	ROUGE-2	ROUGE-W
Lead baseline	28.684	5.283	9.525
TS	29.620	5.200	9.266
TE+TS	<b>31.333</b> (+5.78 %)	<b>5.780</b> (+11.15 %)	<b>9.588</b> (+3.48 %)

Cuadro 1: Resultados sobre corpus DUC 2002 (tareas mono y multidocumento).

<b>Multidocumento</b>	ROUGE-1	ROUGE-2	ROUGE-W
<i>Abstracts-HUM</i>	37.762	8.004	10.436
<i>Extracts-HUM</i>	41.811 (+10.72 %)	13.466 (+68.24 %)	11.503 (+10.22 %)

Cuadro 2: Evaluación de resultados comparando “abstracts” vs. “extracts”.

y F-medida, para diferentes niveles de solapamiento entre distintos textos. En este trabajo sólo tendremos en cuenta el valor para la cobertura (*recall*), debido a que las evaluaciones que se han realizado a posteriori sobre los datos del DUC 2002 (Wan, Yang, y Xiao, 2006; Steinberger et al., 2007) han utilizado la versión anterior de la herramienta ROUGE (versión 1.4.2) que sólo realizaba el cálculo de la cobertura. Las medidas ROUGE que usaremos para nuestra evaluación comprenden ROUGE-N (ROUGE-1 y ROUGE-2), que determinan la cobertura basada en n-gramas entre el resumen candidato y el resumen (o resúmenes) modelo; ROUGE-L, que se basa en obtener la subsecuencia común más larga, consecutiva o no, entre dos textos y ROUGE-W ( $w=1.2$ ), similar a ROUGE-L, con la salvedad de que memoriza los tamaños de los emparejamientos consecutivos para quedarse con el mayor de ellos.

Aprovechando los *extracts* generados por humanos incluidos en el corpus del DUC 2002, se decidió realizar una prueba adicional que consistió en generar resúmenes multidocumento a partir de cada uno de los 59 grupos para analizar las diferencias en los resultados obtenidos con ROUGE, puesto que la mayoría de los sistemas existentes en la actualidad producen *extracts*, pero se suelen comparar con *abstracts* en el momento de su evaluación. Dicha prueba consistió en producir resúmenes de 200 palabras de longitud, utilizando conjuntamente los componentes de

implicación textual y la frecuencia de las palabras. Los resultados de esta prueba se muestran en el Cuadro 2.

En el Cuadro 1 se muestran los resultados para las cuatro pruebas descritas anteriormente. Entre paréntesis se indica el incremento de mejora<sup>5</sup> que se obtiene para cada medida ROUGE, respecto a no incorporar la implicación textual en el sistema. De estos resultados se observa que cuando se incorpora el reconocimiento de la implicación textual en el proceso de generación de resúmenes, se obtiene como media una mejora del 5.85 % para el caso de monodocumento y de 6.80 % para el de multidocumento, frente a aplicar solamente la frecuencia de las palabras en dichos procesos. Cabe destacar el incremento de mejora que obtiene el reconocimiento de implicación textual aplicado a multidocumento. Se ha comprobado experimentalmente que se ha eliminado el 71.57 % de las frases originales del corpus del DUC 2002, por lo que la reducción de los textos es bastante significativa. Por otra parte, las dos aproximaciones propuestas en este trabajo mejoran los resultados obtenidos para las respectivas *baselines* del DUC 2002 para ambas tareas (excepto para la medida ROUGE-2 en el caso monodocumento).

En cuanto a los resultados de la salida generada por nuestro sistema aplicando con-

<sup>5</sup>Se ha comprobado, aplicando el “sign-test” con un nivel de significación del 5 %, que la mejora obtenida con el reconocimiento de la implicación textual es estadísticamente significativa.

juntamente implicación textual y frecuencia de palabras sobre cada grupo de noticias, y evaluada respecto a los *abstracts* y los *extracts* manuales proporcionados en el corpus del DUC 2002 (Cuadro 2), se observa que si comparamos las salidas obtenidas por nuestro sistema con *extracts* producidos por expertos humanos obtenemos una mejora del 29.73 %, frente a los resultados que se derivan de la comparación con *abstracts*. Esto simplemente indica que, puesto que la mayoría de sistemas en la actualidad se basan en la producción de *extracts* (Radev, Hovy, y McKeown, 2002) y no de *abstracts*, si dispusiéramos de *extracts* como resúmenes modelo para poder evaluar nuestros sistemas, los resultados serían más representativos. No obstante, a pesar de que han habido intentos para la generación de *abstracts* de forma automática (Kan, McKeown, y Klavans, 2001; Daumé III et al., 2002) esta tarea es todavía un reto.

### 5. Conclusión y trabajo futuro

En este artículo se presentan dos aproximaciones para la tarea de resúmenes automáticos, tanto para resúmenes monodocumento como para multidocumento. La primera aproximación consiste en considerar solamente la frecuencia de las palabras para que, a partir de éstas, asignemos una puntuación a las frases del texto y seleccionemos aquéllas con mayor puntuación para formar parte del resumen final. La segunda aproximación consiste en incorporar, como tarea de preprocesamiento, el reconocimiento de la implicación textual entre frases, para reducir el tamaño del documento o documentos y analizar, a partir de la combinación de implicación textual con la frecuencia de las palabras, si se obtienen resultados que corroboren que el uso conjunto de estas dos fuentes de información es útil en el proceso de generación de resúmenes. Los resultados presentados en la sección 4 verifican nuestra hipótesis de trabajo, obteniendo unos resultados preliminares que son el punto de partida para futuras investigaciones.

Como trabajos futuros, se plantea la incorporación de otras técnicas en el sistema para dotarlo de más conocimiento, como por ejemplo, la resolución de la anáfora y de expresiones temporales, así como también, conocimiento basado en semántica, haciendo uso de WordNet, que proporciona relaciones tales como las de sinonimia o hiper-

onimia. Además, unos de los trabajos prioritarios para un futuro sería estudiar más en profundidad los mecanismos existentes para detectar y eliminar la redundancia en uno o varios documentos, y compararlos con la técnica de implicación textual propuesta en este artículo. Otro aspecto a considerar consistiría en evaluar los resúmenes generados con otras medidas complementarias, utilizando herramientas de evaluación automática que no se basen únicamente en la cobertura.

### Bibliografía

- Alonso, Laura, Irene Castellón, Salvador Climent, María Fuentes, Lluís Padró, y Horacio Rodríguez. 2004. *Approaches to Text Summarization: Questions and Answers*. *Revista Iberoamericana de Inteligencia Artificial*, ISSN 1137-3601, (22):79–102.
- Daumé III, Hal, Abdessamad Echihabi, Daniel Marcu, Dragos Stefan Munteanu, y Rado Soricut. 2002. *GLEANS: A Generator of Logical Extracts and Abstracts for Nice Summaries*. En *Workshop on Text Summarization (In conjunction with the ACL 2002 and including the DARPA/NIST sponsored DUC 2002 Meeting on Text Summarization)*, Philadelphia.
- Edmundson, H. P. 1969. *New Methods in Automatic Extracting*. En *Inderjeet Mani and Mark Maybury, editors, Advances in Automatic Text Summarization*, páginas 23–42. MIT Press.
- Ferrández, Óscar, Daniel Micol, Rafael Muñoz, y Manuel Palomar. 2007. A perspective-based approach for solving textual entailment recognition. En *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, páginas 66–71, Prague, June. Association for Computational Linguistics.
- Giampiccolo, Danilo, Bernardo Magnini, Ido Dagan, y Bill Dolan. 2007. The third pascal recognizing textual entailment challenge. En *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, páginas 1–9, Prague, June. Association for Computational Linguistics.
- Glickman, Oren. 2006. *Applied Textual Entailment*. Ph.D. tesis, Bar Ilan University.

- Goldstein, Jade, Vibhu Mittal, Jaime Carbonell, y Mark Kantrowitz. 2000. *Multi-document summarization by sentence extraction*. En *NAACL-ANLP 2000 Workshop on Automatic summarization*, páginas 40–48, Morristown, NJ, USA.
- Gotti, Fabrizio, Guy Lapalme, Luka Nerima, y Eric Wehrli. 2007. *GOFASUM: A Symbolic Summarizer for DUC*. En *the Document Understanding Workshop (presented at the HLT/NAACL)*, Rochester, New York USA.
- Harabagiu, Sanda, Andrew Hickl, y Finley Lacatusu. 2007. *Satisfying information needs with multi-document summaries*. *Information Processing & Management*, 43(6):1619–1642.
- Hirao, Tsutomu, Yutaka Sasaki, Hideki Isozaki, y Eisaku Maeda. 2002. *NTT's Text Summarization system for DUC-2002*. En *Workshop on Text Summarization (In conjunction with the ACL 2002 and including the DARPA/NIST sponsored DUC 2002 Meeting on Text Summarization)*, Philadelphia.
- Hovy, Eduard, 2005. *The Oxford Handbook of Computational Linguistics*, capítulo Automated Text Summarization, páginas 583–598. Oxford University Press.
- Kan, Min-Yen, Kathleen R McKeown, y Judith L. Klavans. 2001. *Applying Natural Language Generation to Indicative Summarization*. En *8th European Workshop on Natural Language Generation*, Toulouse, France.
- Kuo, June-Jei y Hsin-Hsi Chen. 2008. *Multidocument Summary Generation: Using Informative and Event Words*. *ACM Transactions on Asian Language Information Processing (TALIP)*, 7(1):1–23.
- Lin, Chin-Yew. 2004. *ROUGE: a package for automatic evaluation of summaries*. En *Proceedings of ACL Text Summarization Workshop*, páginas 74–81.
- Lin, Chin-Yew y Eduard Hovy. 2002. *Automated Multi-document Summarization in NeATS*. En *Proceedings of the Human Language Technology (HLT) Conference*. San Diego, CA, páginas 50–53.
- Luhn, H. P. 1958. *The Automatic Creation of Literature Abstracts*. En *Inderjeet Mani and Mark Maybury, editors, Advances in Automatic Text Summarization*, páginas 15–22. MIT Press.
- Mihalcea, Rada. 2004. *Graph-based ranking algorithms for sentence extraction, applied to text summarization*. En *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, página 20.
- Qiu, Li-Qing, Bin Pang, Sai-Qun Lin, y Peng Chen. 2007. *A Novel Approach to Multi-document Summarization*. En *Proceedings of the 18th International Workshop on Database and Expert Systems Applications (DEXA 2007), 3-7 September 2007, Regensburg, Germany*, páginas 187–191.
- Radev, Dragomir R., Eduard Hovy, y Kathleen McKeown. 2002. *Introduction to the special issue on summarization*. *Computational Linguistics*, 28(4):399–408.
- Spärck Jones, Karen. 2007. *Automatic summarising: The state of the art*. *Information Processing & Management*, 43(6):1449–1481.
- Steinberger, Josef, Massimo Poesio, Mijail A. Kabadjov, y Kerel Ježek. 2007. *Two uses of anaphora resolution in summarization*. *Information Processing & Management*, 43(6):1663–1680.
- Svore, Krysta M., Lucy Vanderwende, y Christopher J.C. Burges. 2007. *Enhancing Single-Document Summarization by Combining RankNet and Third-Party Sources*. En *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, páginas 448–457.
- Tatar, Doina, Emma Tamaianu-Morita, Andreea Mihiş, y Dana Lupşa. 2008. *Summarization by Logic Segmentation and Text Entailment*. En *Proceedings of Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2008)*, páginas 15–26.
- Wan, Xiaojun, Jianwu Yang, y Jianguo Xiao. 2006. *The Great Importance of Cross-Document Relationships for Multi-document Summarization*. En *Proceedings of the 21st International Conference Computer Processing of Oriental Languages (ICCPOL 2006)*, Singapore, páginas 131–138.