CAPÍTULO 7

EXTRACCIÓN DE RESÚMENES CON *BLINDLIGHT*

Las técnicas de resumen automático tienen como misión obtener a partir de un documento o conjunto de documentos más o menos estructurados un único texto mucho más corto que aún contenga los aspectos más relevantes de los originales. Este tipo de técnicas deben trabajar con distintos idiomas y su salida tendría que ser configurable por el usuario no sólo en cuanto a su longitud sino en función de sus necesidades de información. El resumen automático de textos resulta inestimable para aquellos usuarios que tratan con grandes cantidades de documentos y precisan de una herramienta que les permita determinar la información más relevante de un texto o un conjunto de textos a fin de discrimar aquellos a los que dedicar su atención. Durante los años 1950 y 1960 la investigación en este tipo de tecnologías fue intensa para descender considerablemente durante los años siguientes y no recuperarse hasta los años 1990. Desde entonces se trata de un campo muy activo y aunque aún se está lejos de disponer de sistemas capaces de emular a un ser humano (p.ej. produciendo resúmenes indicativos[1]) se ha avanzado enormemente y los sistemas estadísticos y puramente extractivos han demostrado su utilidad. En este capítulo se repasará brevemente la investigación desarrollada hasta el momento en este campo, prestando especial atención al enfoque estadístico. Se describirá la utilización de blindLight como sistema de resumen extractivo y se presentarán los resultados de la evaluación del mismo en un marco estandarizado, resultados que demuestran que la técnica propuesta por el autor resulta más eficaz que muchos de los métodos más avanzados disponibles.

1 Resumen automático

A lo largo de esta disertación se ha mencionado la sobrecarga de información que sufren los usuarios al tratar de localizar información textual y se han señalado distintas tareas que pueden paliar dicho problema como la categorización y clasificación de documentos, la recuperación de información o el resumen automático. Esta última ha atraído mucha

[1] Un **resumen indicativo** tan sólo sugiere el tema de un documento sin revelar sus contenidos.

atención durante los últimos años[1] dado su interés en un entorno con múltiples fuentes que, en mayor o menor medida, se solapan proporcionando mucha información redundante (véase Fig. 107).

Fig. 107 Noticias "agregadas" por *Google* (`http://news.google.com`).

El número de fuentes de las que un usuario puede obtener información es extraordinariamente elevado (existen 1.256 artículos relativos a la primera noticia y 430 para la segunda). Puesto que siempre existirá un "solapamiento" de contenidos sería necesario extraer la información más relevante de las distintas fuentes y proporcionarla de una forma integrada al usuario. Las técnicas de obtención de resúmenes a partir de múltiples documentos tienen un papel muy importante en este escenario.

El **resumen automático** puede definirse como el conjunto de técnicas que producen a partir de un texto de entrada un documento de salida de menor extensión pero que aún contiene los puntos más relevantes del original. Los orígenes de este campo de estudio pueden remontarse a los trabajos de Luhn (1958) y Edmundson (1969) que desarrollaron los primeros sistemas de "extracción de resúmenes".

Dichos sistemas (de **resumen extractivo**) construían los resúmenes a partir de sentencias extraídas del documento original en función de una serie de heurísticos como la utilización de palabras clave o que apareciesen en el título, la posición de las sentencias dentro del documento o la ausencia de palabras "estigma" (p.ej. conjunciones o pronombres al comienzo de la sentencia).

En el extremo opuesto se situarían sistemas de resumen automático que sintetizan un texto totalmente nuevo que recoge las ideas principales del documento original sin incluir necesariamente fragmentos literales del mismo, es decir, estos sistemas trabajarían de manera similar a un ser humano (**resumen abstractivo**). No obstante, el método "abstractivo", mucho más complejo, no es abierto puesto que requiere un importante conocimiento del dominio en que se realizan los resúmenes (Spärck-Jones 1999) y, de hecho, no es previsible la existencia de sistemas prácticos de resumen por abstracción a corto plazo (Hovy 1999, p. 7).

Así pues, la mejora de los métodos de resumen extractivo es un campo de investigación activo debido, por un lado, a las menores exigencias de partida (requieren un

[1] Durante la última década se han celebrado varios "talleres" y se han establecido conferencias dedicadas en exclusiva a las técnicas de resumen automático como *DUC – Document Understanding Conferences* (`http://duc.nist.gov/`).

conocimiento lingüístico nulo o mínimo) y, por otro, al hecho de que la mayor parte de documentos con los que tratan los usuarios en la actualidad no tienen una estructura fija ni pertenecen a un dominio concreto. En semejante escenario la sencillez, flexibilidad y robustez de los métodos de extracción son aspectos valiosos.

Luhn (1958) fue el primero en proponer un método estadístico para extraer las sentencias más significativas de un texto y construir un resumen del mismo. Luhn proponía determinar en primer lugar la significatividad de las distintas palabras[1], suponiendo que las más frecuentes (a excepción de las palabras vacías) serían las más importantes. Posteriormente se asignaría a cada sentencia un peso en función del número de palabras relevantes que incluyese, el peso de las mismas y la distancia entre ellas dentro de la sentencia. Una vez obtenida la puntuación de todas las sentencias del documento sería posible ordenarlas de mayor a menor importancia y seleccionar un subconjunto de las más significativas como resumen del texto original. Resulta interesante notar que Luhn también vio la necesidad de adaptar los resúmenes automáticos a los distintos intereses de los usuarios; proponía para ello asignar una "prima" a las palabras utilizadas por el usuario para describir su necesidad de información de tal modo que las sentencias que contuviesen dichas palabras obtuviesen mayores puntuaciones y pasasen al resumen con mayor facilidad.

Edmundson (1969) emplea cuatro métodos distintos para asignar pesos a las sentencias del documento: (1) un diccionario con palabras que proporcionan pistas sobre la relevancia (*bonus*) o irrelevancia (*estigma*) de las sentencias, (2) la utilización de palabras frecuentes (y no vacías) como indicadores de relevancia, (3) el uso de palabras del título del documento y de los apartados como indicadores positivos y (4) heurísticos basados en la posición de las sentencias en el texto[2]. Cada uno de estos métodos contribuía al peso final de las sentencias de manera independiente y configurable. Cabe señalar que Edmundson fue uno de los primeros en señalar la necesidad de evaluar los sistemas de extracción de resúmenes comparando sus resultados con resúmenes producidos por evaluadores humanos.

Durante los años 1970 y primeros 1980 la investigación en sistemas de resumen automático descendió considerablemente (Spärck-Jones 1999, p. 2) y no resurgiría hasta finales de los 1980 y en especial a partir de los 1990 siendo desde entonces un campo muy activo. Durante este tiempo se abordaron nuevos métodos más allá de la simple extracción de pasajes, por ejemplo, (DeJong 1982), (Tait 1983), (Fum, Guida y Tasso 1985) o (Reimer y Hahn 1988).

No obstante, muchos de estos sistemas tan sólo posibilitaban la generación de resúmenes mediante "plantillas" predefinidas que se rellenaban con datos extraídos de los documentos a resumir y que debían pertenecer a un número de géneros muy limitado (p.ej. Paice y Jones 1993). La técnica del autor, no obstante, sigue un enfoque extractivo puramente estadístico y, en consecuencia, a partir de este punto tan sólo se revisarán algunos de los principales trabajos desarrollados siguiendo métodos análogos; el lector interesado en el enfoque abstractivo puede acudir en primer lugar al capítulo cuarto del libro *"Advances in Automatic Text Summarization"* (Mani y Maybury, eds. 1999).

---

[1] El artículo de Luhn señala la necesidad de llevar a cabo *stemming* a fin de no diferenciar las distintas variantes de un mismo concepto.

[2] Por ejemplo, las sentencias que siguen a un título suelen ser relevantes y las sentencias relevantes tienden a aparecer muy pronto o muy tarde en el documento y en cada párrafo.

Salton y Singhal (1994) aplicaron algunas de las características del sistema de recuperación de información *SMART* a la obtención de resúmenes automáticos. Señalan la necesidad de identificar en primer lugar los distintos temas tratados en un documento así como los párrafos del texto que se refieren al mismo asunto para, posteriormente, emplear una selección de párrafos como resumen del documento. Básicamente se trata de realizar una clasificación automática de párrafos controlada mediante un umbral de similitud que será inversamente proporcional al número de temas que se deseen "descubrir" en el documento. Una vez clasificados los párrafos se extraen en tripletes de similitud máxima hasta alcanzar el tamaño de resumen deseado por el usuario colocándose en el mismo orden en que se encontraban en el texto original.

Salton, Singhal, Buckley y Mitra (1996) ahondan en el mismo tema pero tratando de avanzar desde la identificación (y clasificación) de párrafos hacia la identificación de **pasajes**, esto es, *"fragmentos de texto que exhiben consistencia interna y que pueden distinguirse del resto de texto circundante"* (Salton *et al.* 1996, p. 3). Esta iniciativa es similar a la de la técnica de *TextTiling* (Hearst 1994) con la salvedad de que Hearst no explicitó el interés de la misma para la extracción automática de resúmenes. Salton, Singhal, Mitra y Buckley (1997) desarrollan aún más algunas de las ideas expuestas por Salton y Singhal (1994) y Salton *et al.* (1996) sobre la utilización del grafo de relaciones entre pasajes para la extracción de aquellos más significativos y la construcción de un resumen automático.

Mitra, Singhal y Buckley (1997) evalúan el anterior sistema comparándolo con resúmenes extractivos creados manualmente. En su estudio llegan a una serie de conclusiones que todavía son vigentes: (1) los primeros párrafos de un texto resultan tan efectivos como un resumen obtenido mediante métodos extractivos "inteligentes", (2) esto puede deberse a que los documentos normalmente empleados en los experimentos (artículos periodísticos, técnicos o científicos) están estructurados de tal modo que los primeros párrafos ya son un resumen lo cual explicaría los buenos resultados del *baseline*[1] y (3) aunque *"el resumen por extracción es un método imperfecto parece ser la única técnica que funciona razonablemente con independencia del dominio"*. Brandow, Mitze y Rau (1995) desarrollaron un trabajo similar llegando a conclusiones parecidas, en particular, que los usuarios preferían los resúmenes producidos por el método *baseline*; no obstante, Zechner (1996) señaló que, aunque tal vez menos legibles, los resultados de las técnicas automáticas son mejores en términos de precisión y exhaustividad.

El trabajo de Kupiec, Pedersen y Chen (1995) resulta muy interesante puesto que emplearon un *corpus* de documentos y resúmenes creados manualmente como datos de entrenamiento para un clasificador bayesiano que debía determinar qué sentencias de un documento deberían formar parte de un resumen y cuáles no. El sistema propuesto determinaba para cada sentencia la probabilidad de pertenencia al resumen final y extraía las más probables. Al reducir los documentos a un 25% del tamaño original seleccionaba un 84% de las sentencias elegidas por los expertos humanos y para resúmenes más cortos resultaba sustancialmente superior al *baseline* consistente en presentar el inicio del documento.

Más recientemente, Kraaij, Spitters y van der Heijden (2001) y Kraaij, Spitters y Hulth (2002) también han utilizado clasificadores bayesianos para la extracción de resúmenes. Por su parte, Conroy *et al.* (2001) y Dunlavy *et al.* (2003) han implementado sistemas extractivos mediante modelos de Markov que hacen menos suposiciones que los

[1] Un método *baseline*, literalmente "línea base", es una técnica trivial para resolver un problema en estudio y frente a la cual se comparan los resultados obtenidos mediante las nuevas propuestas.

clasificadores bayesianos sobre la independencia entre elementos. Hirao *et al.* (2002 y 2003) han empleado *SVM's* de manera similar con bastante éxito y Fuentes *et al.* (2003) o Doran *et al.* (2004) árboles de decisión. Alfonseca y Rodriguez (2003), Jaoua y Ben Hamadou (2003, citado por Alfonseca *et al.* 2004) y Alfonseca, Guirao y Moreno Sandoval (2004) han utilizado algoritmos genéticos para la selección de las sentencias.

Fukumoto, Suzuki y Fukumoto (1997) asignan a cada palabra del texto un peso que dependerá de su distribución en el propio documento y en un contexto más amplio. Según su criterio una palabra será palabra clave si (1) su dispersión a nivel de párrafo es menor que a nivel de documento y (2) ésta a su vez es menor que la del término en el dominio. Estos criterios se implementan utilizando el método de ponderación χ^2 de Watanabe *et al.* (1996). Así, para cada palabra no vacía se determina su peso χ^2 dentro del párrafo, el documento y el dominio y se seleccionan aquellas que verifican los dos criterios anteriormente expuestos. Posteriormente cada párrafo del documento se representa mediante un vector que sólo incluirá las correspondientes palabras clave y se realiza una clasificación automática de manera análoga a la de Salton *et al.* (1994, 1996 y 1997). El resumen se construirá seleccionando en primer lugar aquellos párrafos que estén incluidos en un mayor número de los grupos resultantes del proceso de clasificación. La principal ventaja de este método radica en la posibilidad de ajustar los resúmenes a distintos contextos pero, al mismo tiempo, es su principal inconveniente al requerir un *corpus* para extraer resúmenes.

Hovy y Lin (1997) describen el sistema *SUMMARIST* que trata de integrar los enfoques extractivos y abstractivos mediante un proceso de tres fases: (1) identificación de tópicos, (2) interpretación y (3) generación. La primera fase se basa en la denominada Política de Posición Óptima (*Optimal Position Policy*) que no es más que una lista que señala las posiciones donde es más probable encontrar los aspectos clave de un texto de un género determinado[1]. Aplicando únicamente esta primera fase de identificación sería posible construir resúmenes extractivos; no obstante, Hovy y Lin plantean las fases de interpretación y generación para evitar algunos de los problemas de este tipo de resúmenes[2]. Sugieren, por ejemplo, emplear *WordNet*[3] en la segunda fase de tal modo que se puedan resolver situaciones como la que se muestra en Fig. 108.

John bought some **vegetables, fruit, bread and milk.**
John bougth some **groceries.**

Fig. 108 Ejemplo de situación que se resolvería en la fase de "interpretación" (Hovy y Lin 1997).

En (Lin y Hovy 2000) se describe otro concepto interesante para las fases de interpretación y generación: las denominadas *topic signatures*. Estas no son más que conjuntos de términos relacionados, susceptibles de ser reemplazados en el resumen final por un único concepto y obtenibles por medios puramente estadísticos a partir de un *corpus*. Por ejemplo, si los términos `mesa`, `menú`, `camarero`, `comida`, `propina`, etc. apareciesen combinados en un documento podrían sustituirse por la frase `visita a restaurante` en el momento de construir el resumen. *NeATS* (Lin y Hovy 2001 y 2002a) es un sistema de resúmenes

[1] Por ejemplo, según Hovy y Lin la política para el *Wall Street Journal* sería `[T1, P1S1, P1S2, …]`, o lo que es lo mismo, los aspectos más relevantes del documento aparecen en el título, la primera sentencia del primer párrafo seguido de la segunda sentencia del primer párrafo, etc. Otros dominios tendrían políticas distintas que habría que descubrir. Lin y Hovy (1997) describe en detalle el modo en que es posible obtener de modo automático una de tales políticas.

[2] Lal y Ruger (2002) han realizado un trabajo similar en lo referente a la simplificación de las sentencias extraídas.

[3] `http://wordnet.princeton.edu`

multidocumento que aplica las ideas anteriores y que obtuvo interesantes resultados en las campañas *DUC* (*Document Understanding Conferences*) de 2001 y 2002.

Barzilay y Elhadad (1997) plantean la utilidad de las cadenas léxicas como elemento facilitador en la extracción de resúmenes. Una **cadena léxica** es una secuencia de palabras semánticamente relacionadas que aparecen en un texto y que pueden ser adyacentes o encontrarse dispersas a lo largo del documento. Para encontrar dichas cadenas léxicas en un texto genérico es necesario utilizar recursos como *WordNet* que proporcionan la información necesaria sobre las posibles relaciones entre distintas palabras. Así pues, Barzilay y Elhadad encuentran en primer lugar cadenas léxicas en el texto, seguidamente asignan a cada cadena léxica una puntuación[1] y, por último, seleccionan aquellas sentencias que mejor satisfacen a las cadenas léxicas de mayor puntuación. Brunn, Chali y Pinchak (2001) desarrollaron un trabajo muy similar concluyendo también que las cadenas léxicas pueden resultar muy interesantes para introducir conocimiento lingüístico en los métodos extractivos. McKeown *et al.* (2001), Fuentes *et al.* (2003) y Doran *et al.* (2004) también han utilizado cadenas léxicas como método de puntuación.

Jing y McKeown (2000) de la Universidad de Columbia estudiaron diversas tareas de post-procesamiento de los resúmenes extractivos para mejorar su calidad. Aun cuando otros autores (Mani, Gates y Bloerdon 1999) ya trataron dicho problema el interés de este trabajo radica en la forma en que se aborda: Jing y McKeown desarrollaron una técnica que permite en primer lugar analizar la relación entre un resumen manual (creado por un humano) y el documento original a fin de determinar, por un lado, las sentencias "extraídas" y, por otro, las fases de reducción, combinación y reordenamiento a que fueron sometidas. De este modo, no sólo obtienen un conocimiento muy interesante sobre la forma en que un ser humano crea resúmenes mediante "corta-y-pega", sino que son capaces de entrenar su sistema a fin de que emule, hasta cierto punto, estas capacidades. Su equipo ha obtenido muy buenos resultados en *DUC* con el método extractivo (McKeown *et al.* 2001 y 2002) aunque en las últimas ediciones tiende más hacia el enfoque generativo al fusionar y reescribir las sentencias extraídas (Nenkova *et al.* 2003). No obstante, para la obtención de resúmenes a partir de texto traducido automáticamente siguen optando por la utilización de técnicas extractivas (Blair-Goldensohn *et al.* 2004).

Hardy *et al.* (2001) describen un sistema para construir resúmenes a partir de varios documentos (decenas o cientos). Para ello, dividen cada documento en párrafos que son clasificados automáticamente empleando una medida de similitud basada en *n*-gramas de palabras. Una vez descubiertos los distintos grupos se selecciona un párrafo de cada uno para construir el documento final.

MEAD (Radev, Blair-Goldensohn y Zhang 2001) también es un sistema extractivo para la obtención de resúmenes a partir de múltiples documentos. Para cada sentencia de cada documento del conjunto a resumir el sistema obtiene tres puntuaciones a partir de (1) la similitud entre la sentencia y el centroide del conjunto, (2) la distancia de la sentencia al inicio de su correspondiente documento y (3) la similitud entre la sentencia y la primera sentencia (o el título) del documento al que pertenece. Estas puntuaciones se normalizan en el intervalo [0, 1] y se combinan linealmente para obtener una única puntuación que permita

[1] Barzilay y Elhadad (1997) determinaron empíricamente que son dos los parámetros de una cadena léxica que resultan buenos predictores acerca de su utilidad para la construcción de un resumen: la "longitud" y el "índice de homogeneidad" entendidas, respectivamente, como el número de ocurrencias en el texto de miembros de la cadena y 1 - el número de distintas ocurrencias dividido entre la longitud. Así, la puntuación de una cadena léxica sería el producto de su longitud por su índice de homogeneidad.

seleccionar las sentencias más relevantes. Por último, se eliminan del resumen aquellas sentencias demasiado similares entre sí. Posteriormente fue adaptado para la extracción de resúmenes guiados por preguntas resultando el mejor participante en dicha tarea de *DUC 2003* (Radev *et al.* 2003). Saggion y Gaizauskas (2004) han llevado a cabo un trabajo similar.

Recientemente, Erkan y Radev (2004a) han desarrollado una nueva medida de "centralidad" para las sentencias, denominada *LexPageRank*, basada en la idea de "prestigio" de las redes sociales y análoga al *PageRank* (Page *et al.* 1998) de *Google*. El valor de *LexPageRank* para una sentencia *S* se define como la suma de los valores *LexPageRank* de aquellas sentencias similares a *S*, donde la similitud se determina mediante la función del coseno. Esta última versión de *MEAD* resultó uno de los mejores participantes en cuatro de las cinco tareas de *DUC 2004* (Erkan y Radev 2004b). Por su parte, Vanderwende, Banko y Menezes (2004) utilizan *PageRank* para determinar qué elementos de un documento son los más relevantes aunque sus resúmenes son generados y no construidos a partir de sentencias extraídas literalmente de los documentos. Ambos trabajos guardan cierta relación con los desarrollados por Salton *et al.* (1996) que también emplearon grafos para analizar los contenidos de un texto.

En resumen, aunque la calidad de los resultados de los métodos puramente extractivos puede ser deficiente en ocasiones, lo cierto es que estas técnicas son mucho más flexibles y generales que las abstractivas (Spärck-Jones 1999) y existen toda una serie de métodos de post-procesamiento sencillos y capaces de mejorar enormemente la legibilidad del texto final.

2 Utilización de *blindLight* para la extracción de resúmenes[1]

Como se recordará, *blindLight* es una técnica bioinspirada (véase pág. 59) construida sobre la idea de un "genoma documental" definido así:

> *El ADN de un documento es un conjunto de genes donde cada gen está formado por un n-grama de caracteres y su correspondiente significatividad dentro del documento de origen.*

En los capítulos anteriores se ha visto cómo se pueden comparar los genomas de distintos documentos (o consultas) para desarrollar sistemas de clasificación (pág. 85), categorización (pág. 117) y recuperación de información (pág. 141). Estas comparaciones requieren la previa "intersección" de dos pares de "secuencias génicas" y se ha mostrado cómo esta misma operación puede resultar útil para desarrollar una pseudo-traducción de textos (pág. 142). Sin embargo, esta idea de un "genoma documental" aún puede llevarse un paso más allá.

En la naturaleza el ADN codifica los distintos aminoácidos que son los elementos constituyentes básicos de las proteínas, las cuales, a su vez, interpretan un papel esencial en la práctica totalidad de procesos biológicos. Así, las células producen las proteínas necesarias en cada momento empleando el ADN a modo de "plano de construcción". No obstante, el ADN no es muy versatil químicamente y, más aún, es demasiado valioso como para trabajar directamente sobre su molécula para construir cada proteína. Por esa razón, las porciones de ADN con el gen o genes que codifican cada proteína son previamente copiadas mediante moléculas de ARN mensajero (ARNm). El ARNm sale del núcleo celular hacia el citoplasma donde los ribosomas lo emplean como plantilla sobre la que se construye, aminoácido a aminoácido, la proteína utilizando ARN transferente (ARNt). El proceso

[1] La técnica aquí descrita es una evolución de la presentada por Gayo Avello *et al.* (2004a).

durante el cual se copia una porción de ADN sobre ARNm se denomina transcripción y la fase en que se construye la cadena de aminoácidos a partir de la molécula de ARNm se conoce como traducción (véase Fig. 109).

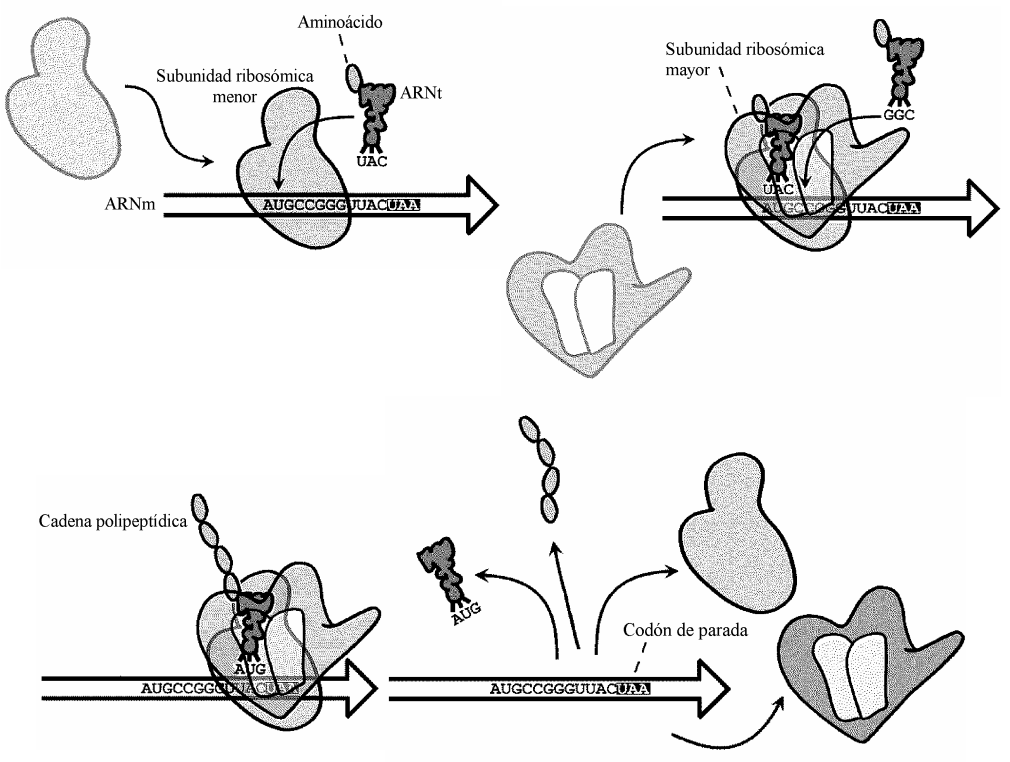

Fig. 109 (De izquierda a derecha y de arriba abajo) **Inicio de la traducción, comienzo de la elongación, fin de la elongación, fin de la traducción.**

La síntesis de una proteína comienza con la unión de la subunidad ribosómica menor a la cadena de ARNm. Entonces la molécula de ARNt iniciador se une al codón de inicio AUG. La subunidad ribosómica mayor es atraída por el ARNt iniciador completando el ribosoma que comenzará a desplazarse a lo largo del ARNm codón a codón. Cada uno de los codones en el ARNm tiene un anticodón complementario en las correspondientes moléculas de ARNt. Cada una de estas moléculas transporta un aminoácido que es añadido a la creciente cadena polipeptídica. Al llegar al codón de parada finaliza el proceso de traducción, el ribosoma se disgrega y la cadena polipeptídica queda libre, plegándose y formando la proteína final.

La utilización de *blindLight* como técnica de extracción de resúmenes se inspira en este proceso de traducción y síntesis de las proteínas para lo cual emplea tanto el vector de *n*-gramas obtenido a partir del documento como el texto plano original del mismo. La ideas subyacentes son muy sencillas:

1. El "ADN documental" está codificado mediante un vector de *n*-gramas de caracteres, cada uno de los cuales tiene asociado un peso, su significatividad. Cada uno de estos pares `(n-grama, significatividad)` puede emplearse a modo de ARNt (véase Fig. 110).
2. El texto plano no proporciona ninguna información al ordenador sobre la relevancia de los distintos pasajes. Sin embargo, puede procesarse de manera secuencial y junto con el "ARNt documental" se puede transferir significatividad a dicho texto (véase Fig. 111).
3. El proceso de transferencia de significatividad del "ARNt documental" al texto no se realiza en una única fase sobre el texto completo sino en varias pasadas

garantizando que la significatividad media por carácter sea creciente. De este modo, el texto de partida es "troceado" en fragmentos (*chunks*) de máxima significatividad. Dichos fragmentos pueden ser utilizados posteriormente para obtener palabras clave o para facilitar la extracción de las sentencias más relevantes (véase Fig. 112 y Fig. 113).

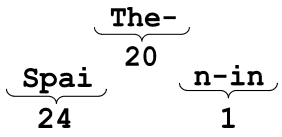

Fig. 110 Cada componente del vector de *n*-gramas puede utilizarse a modo de ARNt.

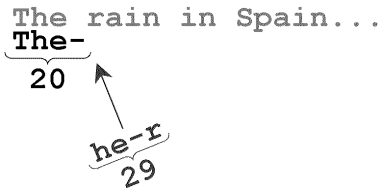

Fig. 111 El "ARNt documental" permite asociar al texto información sobre su significatividad.

El proceso en sí mismo también es muy simple. En primer lugar debe obtenerse un vector de *n*-gramas con sus correspondientes significatividades tal y como se hace para cualquier otra aplicación de la técnica. Una vez hecho esto se pasa a trabajar sobre el texto original que se habrá segmentado en frases y sentencias[1]. Tal segmentación es necesaria por dos razones: (1) al dividir el texto en fragmentos de máxima significatividad no se debe saltar entre dos frases y (2) es necesario conocer los límites entre sentencias para garantizar una mínima coherencia en el resumen extraído.

La Comisión ha adoptado hoy propuestas relativas a un paquete de medidas destinadas a reforzar la capacidad de respuesta de la Unión Europea en caso de catástrofes. Estas medidas se destinan a financiar nuevos equipos especializados en materia de planificación para agilizar el suministro eficaz de ayuda a largo plazo; a reforzar la capacidad de la Unión de facilitar equipos de expertos civiles y de equipo y a suministrar ayuda humanitaria. La Comunicación adoptada hoy también presenta un informe detallado sobre la utilización de los 450 millones de euros anunciados por la UE tras la catástrofe del tsunami. Las propuestas adoptadas hoy constituyen la contribución de la Comisión al plan de acción tras el tsunami propuesto por la Presidencia luxemburguesa el 31 de enero.

«Vistas las situaciones anteriores y nuestra capacidad de responder inmediatamente ante la catástrofe del tsunami, la Comisión propone ahora medidas que nos ayudarán, en el futuro, a contribuir de forma rápida y eficaz a las tareas de reconstrucción tras una catástrofe» ha declarado la Comisaria de Relaciones Exteriores y Política de Vecindad, Benita Ferrero-Waldner, que propone dichas medidas conjuntamente con los Comisarios Michel y Dimas. Stavros Dimas, Comisario Europeo responsable de Protección Civil ha dicho: «Nuestra reacción ante el Tsunami ha demostrado el claro valor añadido que la dimensión europea aporta a la asistencia en materia de protección civil. Las propuestas de hoy hacen avanzar un paso más al Mecanismo actual... Tomadas en su conjunto, permitirán disponer de un instrumento que garantiza una reacción europea eficaz ante futuras catástrofes».

Fig. 112 División de un texto en fragmentos de máxima significatividad.

Con fondo gris se muestran los 50 fragmentos más significativos, con borde negro los 10 más significativos.

Una de las primeras tareas a realizar sobre el texto del documento es la obtención de la significatividad media por carácter de cada sentencia que no es más que el cociente de la suma de las significatividades de todos los *n*-gramas que aparecen en la sentencia entre la longitud de la misma. Este valor será una de las varias "puntuaciones" que se utilizarán para determinar la relevancia de las distintas sentencias.

La siguiente fase es la ya citada división del texto en fragmentos de máxima significatividad, fragmentos que, como también se ha dicho, deben estar enteramente contenidos en una única frase. El algoritmo para llevar a cabo esta fragmentación se muestra

[1] El modo en que se lleve a cabo la segmentación es irrelevante. Para utilizar *blindLight* tan sólo se precisa de sentencias y frases entendiendo las primeras como oraciones gramaticales y las segundas como secuencias de palabras dotadas de sentido pero que no forman oración.

en Fig. 115; no obstante se describirá brevemente a continuación y se proporciona un ejemplo ilustrativo en Fig. 116.

os 450 m	la Comisión p	catástrofes
luxemburgues	e la Comisión	catástrofe
trofe» h	agilizar	a Comisión
Políti	el 31 de	la Comisión
también p	La Comisión h	el tsunam
catástrofes»	ión Europe	eficaz
ro-Wald	el Tsunam	medidas
catástrofes	tilización	equipos
hoy propuest	conjunt	la capacida
catástrofe	nstrucción	cción

Fig. 113 (A la izquierda) **20 fragmentos más significativos de un documento y** (a la derecha) **10 primeras "palabras clave" obtenidas al desplazar una ventana sobre los primeros.**

Este algoritmo trabaja sobre dos estructuras diferentes: por una parte, una lista que inicialmente contiene las frases extraídas del texto original y, por otra, una pila con los *n*-gramas del documento ordenados por significatividad decreciente. En cada iteración se extrae un *n*-grama de la pila, que será el *n*-grama más significativo del texto disponible en la lista de frases, y se buscan aquellas frases que lo contengan. Posteriormente, para cada una de las frases se localizan los *n*-gramas anterior y posterior al fragmento localizado y se aumenta dicho fragmento añadiéndole el *n*-grama más significativo del par. Este proceso se repite para cada frase extraída mientras la significatividad por carácter no decrezca. En el momento en que el fragmento de texto no pueda crecer sin disminuir su significatividad se detiene la fase de crecimiento, se extrae el fragmento y se elimina la frase de la lista, sustituyéndola por las secciones anterior y posterior al fragmento extraído. Una vez se ha terminado de procesar todas las frases correspondientes a un *n*-grama se repite todo el proceso para el siguiente *n*-grama finalizando cuando la pila quede vacía. En ese momento se habrá segmentado todo el texto del documento en fragmentos de máxima significatividad.

Una vez hecho esto se puede obtener una nueva puntuación para cada sentencia; dicha puntuación no es más que la suma de la significatividad media por carácter de cada fragmento presente en la sentencia. A partir de la lista de fragmentos también es posible obtener las "palabras clave" del documento. Para ello basta con recorrerla con una ventana de tamaño *K* extrayendo como claves las subcadenas más largas que resulten de la intersección de cualquier par de fragmentos contenidos en la ventana (véase Fig. 114).

la Comisión	catástrofes	catástrofes	catástrofes
el tsunam	catástrofe	catástrofe	catástrofe
medidas	a Comisión	a Comisión	a Comisión
cción	la Comisión	la Comisión	la Comisión
uest	el tsunam	propuest	la Comis
teri	eficaz	el tsunam	a Comis
ión	medidas	eficaz	eficaz a
uro	equipos	medidas	propuest
ida	cción	equipos	el tsunam
aci	suministr	acción	La Com
	trofe	la capacida	civil
	ción	cción	eficaz
	Com	iliza	Comisa
	la U	suministr	Comis
	uest	ación	medidas

Fig. 114 (De izquierda a derecha) **15 primeras "palabras clave" obtenidas con ventanas de tamaño 2, 4, 8 y la longitud total de la lista de fragmentos. El peso de cada palabra clave es su significatividad por carácter.**

Naturalmente, las claves obtenidas pueden ser fragmentos de palabras o frases (véase Fig. 113), no obstante, esta solución es razonablemente flexible puesto que en el caso de textos pertenecientes a idiomas occidentales siempre se pueden refinar las claves encontradas (esto es, asegurarse de que se trata de palabras o frases completas) y para los idiomas orientales que no separan las palabras no requeriría una fase de segmentación previa. Sin embargo, la utilización de *blindLight* como sistema de extracción de palabras clave aún requiere una experimentación más rigurosa puesto que sus resultados aún no son todo lo satisfactorios que se desearía como se verá en próximos apartados.

Por último, del mismo modo que se puede asignar a cada sentencia una puntuación en función de los fragmentos que contiene también es posible elaborar otra puntuación partiendo de las palabras clave: este valor sería la suma de la significatividad media por carácter de cada palabra clave contenida en la sentencia.

Algoritmo ribosomalTranslation *(sentences, ngramStack, sizeNgram)*
Input: *sentences*, el texto del documento segmentado en sentencias, *ngramStack* una pila con los *n*-gramas del documento donde el tope contendrá el n-grama más significativo aún sin tratar y *sizeNgram*, el tamaño de *n*-grama empleado.

```
1.  while ngramStack ≠ λ do
2.      ngram ← pop (ngramStack)
3.      w ← ngramWeight (ngram)
4.      sentenceSubset ← searchSentencesContaining (sentences, ngram)
5.      for each sentence in sentenceSubset do
6.          chunk ← ngram
7.          totalSig ← w
8.          currentAvgSig ← 0
9.          do
10.             previousAvgSig ← currentAvgSig
11.             lead ← leadNgram (chunk, sentence, sizeNgram)
12.             tail ← tailNgram (chunk, sentence, sizeNgram)
13.             w_L ← ngramWeight (lead)
14.             w_T ← ngramWeight (tail)
15.             if (w_L > w_T)
16.                 chunk ← charAt (lead, 0) + chunk
17.                 totalSig ← totalSig + w_L
18.             else
19.                 chunk ← chunk + charAt (tail, sizeNgram - 1)
20.                 totalSig ← totalSig + w_T
21.             end if
22.             currentAvgSig ← totalSig / strlen (chunk)
23.         while (currentAvgSig > previousAvgSig)
24.         chunks (chunk) ← currentAvgSig
25.         sentenceFragments ← explode (sentence, chunk)
26.         for each newSentence in sentenceFragments do
27.             insertInto (sentences, newSentence)
28.         loop
29.     loop
30. loop
31. return chunks
```

Fig. 115 Algoritmo que divide el texto del documento en fragmentos de máxima significatividad.

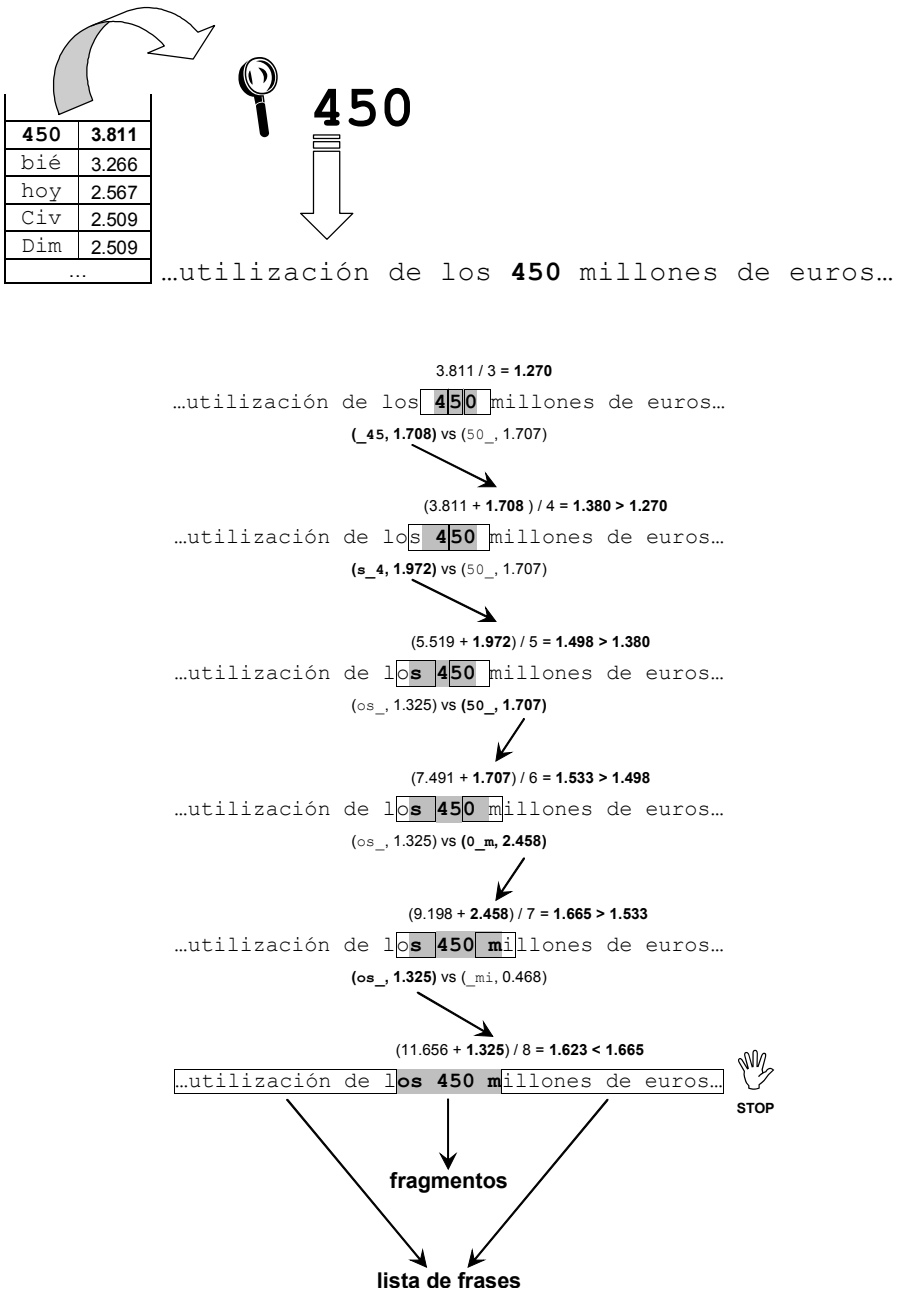

Fig. 116 Proceso mediante el cual se divide el texto en fragmentos de máxima significatividad.

Así pues, a partir del vector de *n*-gramas y del texto del documento segmentado en frases y sentencias es posible (1) obtener la significatividad media por carácter de cada sentencia, (2) segmentar el texto original en fragmentos de máxima significatividad, (3) utilizar dichos fragmentos para obtener una segunda puntuación para cada sentencia, (4) extraer a partir de los fragmentos las palabras clave del documento y (5) emplear las palabras clave para calcular una tercera puntuación para cada sentencia del texto. De acuerdo con el enfoque extractivo "clásico" (Edmundson 1969) para construir el resumen del documento tan sólo hay que extraer las sentencias con mayor puntuación mientras no se supere el tamaño máximo del resumen y colocarlas en el mismo orden en que aparecen en el texto original (véase Fig. 117). Por el momento la utilización de *blindLight* como sistema de extracción de resúmenes queda limitado a su utilización con un único documento; aún no se

ha implementado ningún sistema de resumen guiado mediante consultas y tan sólo se ha desarrollado una versión preliminar de extracción de resúmenes a partir de varios documentos que se describirá en un apartado posterior.

Esta técnica tiene un parecido superficial con la propuesta por Cohen (1995), *Highlights* (véase pág. 55). No obstante, la técnica presentada por el autor es, por un lado, más sencilla al no requerir el uso de un contexto externo y, por otro, más potente puesto que *Highlights* tan sólo permitía determinar (de una manera un tanto complicada) qué secuencias de caracteres eran relevantes para, posteriormente, extraer unos pocos términos clave mientras que la nueva técnica aquí descrita no sólo permite extraer dichos términos sino generar resúmenes extractivos de manera sencilla.

Significatividad media por carácter

La Comunicación adoptada hoy también presenta un informe detallado sobre la utilización de los 450 millones de euros anunciados por la UE tras la catástrofe del tsunami. Las propuestas de hoy hacen avanzar un paso más al Mecanismo actual... Tomadas en su conjunto, permitirán disponer de un instrumento que garantiza una reacción europea eficaz ante futuras catástrofes.

Puntuación por *chunks*

«Vistas las situaciones anteriores y nuestra capacidad de responder inmediatamente ante la catástrofe del tsunami, la Comisión propone ahora medidas que nos ayudarán, en el futuro, a contribuir de forma rápida y eficaz a las tareas de reconstrucción tras una catástrofe» ha declarado la Comisaria de Relaciones Exteriores y Política de Vecindad, Benita Ferrero-Waldner, que propone dichas medidas conjuntamente con los Comisarios Michel y Dimas.

Puntuación por palabras clave (ventana de tamaño 6)

La Comisión ha adoptado hoy propuestas relativas a un paquete de medidas destinadas a reforzar la capacidad de respuesta de la Unión Europea en caso de catástrofes. Las propuestas adoptadas hoy constituyen la contribución de la Comisión al plan de acción tras el tsunami propuesto por la Presidencia luxemburguesa el 31 de enero.

Significatividad media por carácter y puntuación por *chunks*

La Comisión ha adoptado hoy propuestas relativas a un paquete de medidas destinadas a reforzar la capacidad de respuesta de la Unión Europea en caso de catástrofes. La Comunicación adoptada hoy también presenta un informe detallado sobre la utilización de los 450 millones de euros anunciados por la UE tras la catástrofe del tsunami.

Fig. 117 Resúmenes del 20% (aprox.) empleando las distintas formas de puntuación descritas.

3 Evaluación de los sistemas de resumen automático

Medir de un modo objetivo la idoneidad de un resumen (automático o no) no es una tarea trivial. Según Hovy (1999) habría que determinar, al menos, dos medidas: la ratio de compresión (la relación entre la longitud del resumen y del texto original) y la ratio de retención[1] (cuánta información del original ha sido retenida por el resumen). Está claro que un buen resumen tendría ratios de compresión y retención próximas a cero y a uno respectivamente (Hovy y Lin 1998).

No obstante, mientras que calcular la primera es sencillo hacer lo propio para la segunda resulta mucho más complicado. Hovy y Lin (1998) describen tres tipos de experimentos que permitirían obtener sendas medidas de la ratio de retención. Se trata de los denominados juegos de Shannon, de la Pregunta y de Clasificación (*Shannon Game*, *Question Game* y *Classification Game*). En todos ellos era necesario recurrir a sujetos humanos que debían llevar a cabo una tarea que requería el conocimiento previo del texto original. Por ejemplo, en el caso del juego de Shannon los sujetos debían reconstruir el documento original de manera literal; algunos de los participantes habían tenido acceso al mismo mientras que otros sólo habían leído el resumen. En todos los casos se informaba a los sujetos cuando se equivocaban en una letra y se les permitía un nuevo intento; la relación

[1] En realidad, Hovy (1999) la denomina, paradójicamente, ratio de omisión pero con idéntico sentido.

entre el número de intentos requeridos en ambos grupos permitía calcular la ratio de retención.

No parece necesario decir que este tipo de experimentos son enormemente costosos en tiempo y recursos y, por otro lado, evalúan la "calidad" de los resúmenes de manera indirecta a través de su influencia en la ejecución de una o más tareas. Esto es lo que se conoce como evaluación extrínseca. No obstante, un sistema de resumen automático, y en general cualquier sistema PLN, puede ser evaluado también de manera intrínseca (Galliers y Spärck-Jones 1993). En este caso se trata de evaluar directamente la calidad del resúmen comparándolo con un resumen "modelo" pre-existente o creado a tal efecto por seres humanos. Este método ya fue reivindicado por Edmundson (1969) y es el más utilizado (Hovy 1999), por ejemplo, (Kupiec, Pedersen y Chen 1995), (Brandow, Mitze y Rau 1995), (Mitra, Singhal y Buckley 1997) o (Jing *et al.* 1998). A su vez, la evaluación intrínseca puede ser manual cuando la comparación entre resultados y modelo la realiza un ser humano o automática cuando se reemplaza al evaluador por algún tipo de algoritmo.

No obstante, no es suficiente con evaluar un sistema de resumen automático, es necesario que la evaluación sea aplicable a distintos sistemas y, en consecuencia, permita la comparación de diferentes técnicas. La serie de conferencias *DUC* (*Document Understanding Conferences*) surgió como una iniciativa[1] para el desarrollo de un marco común para la evaluación (y consecuente mejora) de los sistemas de resumen automático. Desde hace ya algún tiempo se trata del principal foro de evaluación de este tipo de sistemas, los documentos y modelos de ediciones anteriores están disponibles y desde 2004 la evaluación se hace de modo automático[2].

En todas las ediciones *DUC* se presentan diversas tareas que incluyen, entre otras, la obtención de resúmenes genéricos a partir de un único documento o de conjuntos de textos relativos a un tema común. La organización prepara los conjuntos de documentos y elabora los correspondientes resúmenes modelo para la posterior evaluación de los resultados.

Durante las ediciones 2001-2003 la evaluación se realizó de manera manual; es decir, una serie de jueces humanos debían "comparar" los resultados de los distintos participantes con los modelos disponibles. Por supuesto la comparación no era totalmente subjetiva sino que se realizaba con la asistencia de una herramienta (*SEE*)[3]. Inicialmente los textos a comparar se dividen en "unidades de discurso" (p.ej. frases) de tal manera que el revisor puede seleccionar distintas unidades del resumen a evaluar, asociarlas a unidades del modelo e indicar si los contenidos de la unidad en el resumen conciden total o parcialmente con aquellos de la unidad en el modelo. El revisor puede también indicar la calidad gramatical de cada unidad y, por último, evaluar de manera global la coherencia, gramática y organización del resumen automático. Finalmente, la herramienta calcula la coincidencia entre el resumen y el modelo como valores de precisión y exhaustividad (Lin y Hovy 2002b).

---

[1] Anteriormente ya se había realizado una experiencia de evaluación a gran escala, *SUMMAC* (Mani *et al.* 1998). Sin embargo, el enfoque seguido fue fundamentalmente extrínseco ya que se consideró que *"los resúmenes ideales son difíciles de conseguir y raramente únicos*". Por ello, el mérito de esta iniciativa no fue tanto la posibilidad de reutilizar sus productos para posteriores evaluaciones sino la demostración, por un lado, del interés de un marco de evaluación común e independiente de los desarrolladores y, por otro, de la enorme utilidad de las técnicas de resumen automático para otras tareas de tratamiento de información.

[2] Estos motivos han llevado al autor a evaluar su técnica mediante los datos de la última edición. Los resultados de dicha evaluación se presentan en un apartado posterior.

[3] *SEE* (*Summary Evaluation Environment*) disponible en: `http://www.isi.edu/~cyl/SEE/`

Harman y Over (2004) analizan los efectos de los distintos "factores humanos" que implica este tipo de evaluación[1] y concluyen que, a pesar de existir grandes diferencias entre distintos evaluadores y entre diferentes modelos, la clasificación de los sistemas participantes apenas cambia cuando se promedian los resultados obtenidos al trabajar con varios conjuntos de documentos, modelos y jueces. Señalan que, naturalmente, habrá diferencias en los resúmenes de documentos individuales pero que la única forma de mejorar la tecnología de resumen automático es mejorando los resultados promedio en este tipo de evaluaciones.

Lin y Hovy (2002b) también analizaron la influencia del factor humano en *DUC 2001* y estudiaron la posibilidad de sustituir este tipo de evaluaciones por métodos automáticos. Llegaron a las siguientes conclusiones[2]: (1) las evaluaciones humanas son "inestables", es decir, dos revisores pueden asignar puntuaciones distintas al comparar la misma sentencia con un modelo; (2) los distintos sistemas evaluados quedan separados en distintos "grupos de rendimiento" por lo que, a pesar de todo, los revisores humanos, tomados en conjunto, demuestran un criterio que permite establecer comparaciones entre sistemas; (3) es posible desarrollar un método automático que otorgue puntuaciones basándose en la coincidencia de *n*-gramas de palabras entre resumen y modelo; (4) la clasificación de los participantes obtenida mediante dicho sistema automático muestra una elevada correlación con la clasificación producida por los revisores humanos y (5) los autores de los modelos suelen construir sentencias nuevas por lo que la única forma en que un sistema de evaluación automática puede enfrentarse a estos problemas es empleando varios (probablemente muchos) modelos.

Posteriormente (Lin y Hovy 2003) estudiaron la posibilidad de utilizar *BLEU* (Papineni *et al.* 2002), una herramienta para la evaluación de sistemas de traducción automática, para la evaluación de resúmenes automáticos. *BLEU* emplea una media ponderada del número de *n*-gramas de palabras de longitud variable que coinciden entre una traducción automática y un modelo de traducción. Comprobaron que esta medida no siempre exhibía una correlación con las clasificaciones producidas por evaluadores humanos mientras que una medida basada en la coincidencia de unigramas (esto es, palabras aisladas) mostraba un mejor comportamiento y, en consecuencia, abría las puertas al desarrollo de métricas que pudiesen ser obtenidas de modo automático y, al mismo tiempo, garantizar una evaluación similar a la que podría realizar un revisor humano.

A raíz de estos trabajos se implementó un sistema de evaluación automático denominado *ROUGE* (Lin 2004a) que comenzó a utilizarse en *DUC 2004*. Esta herramienta permite calcular diversas medidas, principalmente *ROUGE-N*, *ROUGE-L* y *ROUGE-W*. La primera se basa en el número de *n*-gramas de palabras que coinciden entre un resumen candidato y uno o más modelos por lo que existen las medidas *ROUGE-1*, *-2*, *-3*, etc. no siendo habitual emplear más alla de los 4-gramas. *ROUGE-L* emplea la longitud de las subcadenas más largas que son comunes en el candidato y en el modelo mientras que *ROUGE-W* es una versión ponderada de *ROUGE-L* que además de la longitud de la subcadena valora la ausencia de "huecos" en la misma (véase Fig. 118).

---

[1] La variabilidad tanto entre revisores como entre los modelos construidos para realizar la evaluación.

[2] Estas conclusiones encuentran apoyo en otros autores, así van Halteren (2002) en relación a *DUC 2002* afirma *"no está claro si uno o dos extractos creados manualmente constituyen una referencia suficiente"*. Por su parte, Santos, Mohamed y Zhao (2004) también propusieron un sistema de evaluación automática aunque en su caso los resúmenes no eran comparados con ningún modelo sino con los propios documentos de partida.

1. Opposition leaders Prince Norodom **Ranariddh and Sam Rainsy**
2. **Ranariddh and Sam Rainsy** have charged that Hun Sen's victory in the elections was…
3. **Ranariddh and** his opposition ally, **Sam Rainsy**, refused to accept the election results
4. **Ranariddh and** former finance minister **Sam** Ram **Rainsy** have refused to enter into a coalition

Fig. 118 Diferencias entre las puntuaciones *ROUGE-L* y *ROUGE-W*.

Tomando la primera sentencia como referencia las tres siguientes tienen la misma puntuación *ROUGE-L* puesto que en los tres casos la longitud de la subcadena común más larga (en negrita) es la misma. No obstante, la "similitud" con la sentencia modelo es mayor en el segundo caso que en el tercero y en éste que en el cuarto debido a que las correspondientes subcadenas no son contiguas. La medida *ROUGE-W* permite capturar estas particularidades.

Lin (2004b) volvió a analizar el método de evaluación empleado en *DUC* a raíz de las críticas al mismo hechas por Nenkova y Passonneau (2004): *"las puntuaciones de DUC no pueden usarse para distinguir un buen resumen humano de uno malo; además, el método de DUC no es suficiente para diferenciar entre sistemas automáticos"*. Lin demuestra que dichas conclusiones no son correctas y que la metodología empleada en *DUC* es válida, en particular en lo que se refiere al número de modelos y muestras enviadas por cada participante. Además de esto, utilizando los datos de las ediciones de 2001, 2002 y 2003 concluye que las medidas obtenidas empleando *ROUGE* muestran una elevada correlación con las evaluaciones humanas hechas en *DUC* y puesto que éstas ofrecen resultados significativos concluye que no sólo la metodología de evaluación es válida sino que es posible llevarla a cabo de modo completamente automático.

En resumen, en la actualidad es posible evaluar de manera sistemática métodos de obtención de resúmenes (fundamentalmente extracción) a partir de textos de estilo periodístico. Sin embargo, aún quedan muchos aspectos a resolver en el campo del resumen automático como, por ejemplo, la necesidad de afrontar otros estilos más allá del periodístico o de adaptar los resúmenes a los propósitos específicos de los usuarios. Sin abandonar la evaluación intrínseca automática todo esto requeriría, sin duda, metodologías de evaluación extrínsecas basadas en tareas realistas (Spärck-Jones *et al.* 2004).

4 Resultados obtenidos por *blindLight*

A fin de analizar la viabilidad de la nueva técnica para la extracción automática de resúmenes se decidió utilizar los productos correspondientes a *DUC 2004*. Las razones fueron dos: por un lado se trataba de la primera campaña en que se había llevado a cabo una evaluación automática facilitando su reutilización y por otro, al tratarse de la edición más reciente, permitiría comparar la propuesta del autor con las técnicas más avanzadas disponibles.

En dicha edición se propusieron cinco tareas:

1. Resúmenes muy cortos, máximo 75 caracteres, a partir de un único documento.
2. Resúmenes cortos, máximo 665 caracteres, a partir de un conjunto de documentos.
3. Resúmenes muy cortos a partir de traducciones automáticas y manuales de árabe a inglés.
4. Resúmenes cortos a partir de un conjunto de traducciones automáticas y manuales de árabe a inglés.
5. Resúmens cortos creados a partir de un conjunto de documentos y "guiados" por consultas del tipo `"Who is X?"` (`"¿Quién es X?"`).

Las tareas 1 a 4 fueron evaluadas empleando únicamente *ROUGE* y la última mediante *SEE* (véase nota al pie en pág. 170), es decir, las cuatro primeras tareas se

evaluaron automáticamente y la quinta de manera manual. Para las tareas 1 y 2 se emplearon 50 conjuntos, alrededor de 500 documentos, pertenecientes a la colección *TDT*[1] en inglés. Para las tareas 3 y 4 se usaron 25 conjuntos, del orden de 250 documentos, pertenecientes a la misma colección en árabe y para la última tarea se utilizaron 50 conjuntos, aproximadamente 500 documentos, de la colección *TREC*[2] en inglés. En todos los casos los documentos eran noticias procedentes de agencias de prensa (p.ej. *Associated Press* y *New York Times* en el caso de la *TDT*). Para cada tarea se definieron unos métodos *baseline* que también participaron en la evaluación. En el caso de la primera tarea el resumen *baseline* consistía en los 75 primeros caracteres del documento y en la segunda se construía con los primeros 665 caracteres del documento más reciente.

Para la evaluación de *blindLight* como técnica de extracción de resúmenes se optó por realizar únicamente los experimentos relativos a las dos primeras tareas: resúmenes muy cortos o cortos de textos escritos (no traducidos) en inglés. No obstante, la resolución de ambas tareas requirió algunas modificaciones en la aplicación de la técnica. Por un lado, los resúmenes muy cortos podían ser simples listas de palabras clave (*NIST* 2004) por lo que parecía razonable añadir, además de la extracción de una única sentencia, la selección de palabras clave a los métodos de producción de resúmenes. Por otro, *blindLight* no contempla, a día de hoy, la extracción de resúmenes multidocumento por lo que para afrontar la segunda tarea fue necesario fusionar[3] las noticias de cada conjunto en un único documento para su resumen; naturalmente, ese no es el enfoque más adecuado (véase Fig. 119).

> **A fire turned a dance hall jammed with teen-age Halloween revelers into a deathtrap, killing at least 60 people and injuring about 180 in Sweden's second-largest city.** The building had just two exits, one of which was blocked by fire, city police technician Stephen Holmberg was quoted as saying by the Swedish news agency TT. Jamal Fawz, graf 19 pvs. **A fire turned a dance hall jammed with teen-age Halloween revelers into a deathtrap, killing 65 people and injuring 157 others in Sweden's second-largest city.** The fast-spreading fire that broke out just a few minutes before midnight Thursday gutted the building and left rescuers facing a hideous scene that local rescue service leader Lennart Olin likened to a "gas chamber".

Fig. 119 Ejemplo de resumen multidocumento obtenido con *blindLight*.

En este ejemplo puede apreciarse uno de los inconvenientes del sistema de resúmenes multi-documento descrito. Como se puede ver hay dos sentencias (la primera y la cuarta) prácticamente idénticas. Esto debería solucionarse en el futuro, sin embargo, no debería plantear excesivos inconvenientes puesto que podrían utilizarse las consabidas medidas Π y P para determinar qué sentencias se "solapan" y no deben, por tanto, formar parte del resumen final.

Así, para la construcción de resúmenes muy cortos con *blindLight* se plantearon los siguientes métodos:

- Extraer la sentencia con mayor significatividad media por caracter.
- Extraer la sentencia con mayor puntuación por fragmentos.
- Extraer la sentencia más relevante mediante la combinación de las puntuaciones por fragmentos y por significatividad media por carácter.

---

[1] *TDT – Topic Detection and Tracking* (Detección y Seguimiento de Temas) hace referencia a una iniciativa *DARPA* para el desarrollo de métodos para la detección y agrupamiento de material relativo a un tema específico extraído a partir de artículos o locuciones periodísticas tanto en inglés como en mandarín <http://www.nist.gov/speech/tests/tdt>.

[2] La colección de documentos desarrolladas para las *Text REtrieval Conferences* <http://trec.nist.gov>.

[3] El orden en que se unieron los documentos es aquel en el que aparecen dentro de su directorio; puesto que la fecha forma parte del nombre del archivo, en la mayor parte de los casos el orden fue cronológico a excepción de aquellos conjuntos que combinan noticias de distintas agencias (p.ej. APW19981028.0444 y NTY19981026.0292).

- Extraer las frases[1] (no sentencias) con mayor puntuación por palabras clave ordenándolas según su orden de aparición en el texto.
- Extraer una lista de palabras clave ordenadas por primera aparición en el texto. Dado que las "palabras clave" obtenidas por *blindLight* pueden ser frases el resumen podría contener palabras repetidas.
- Extraer una lista de palabras clave no repetidas.

Además, se experimentó con distintos tamaños de *n*-grama, de ventana (en el caso de los métodos que implican la extracción de palabras clave) y de estadísticos para el cálculo del peso de los *n*-gramas (información mutua, *SCP*, ganancia de información, etc.). Por otro lado, se probó un sistema de **compresión de sentencias**[2] trivial consistente en eliminar los *n*-gramas menos significativos hasta reducir la longitud de la sentencia extraída a 75 caracteres; en aquellos casos en que no se aplicó esta "compresión" simplemente se truncaba el resumen.

Los resultados detallados de la evaluación se presentan en un anexo mientras que las conclusiones a las que se ha llegado son las siguientes:

1. Los estadísticos más adecuados para la ponderación de los *n*-gramas de cara a la extracción de resúmenes son información mutua, *SCP* y Dice.
2. La utilización de *blindLight* para la construcción de resúmenes muy cortos (máximo 75 caracteres) no parece adecuada puesto que en todas las medidas a excepción de *ROUGE-1* el rendimiento es sustancialmente inferior a la media de participantes. En cambio, las palabras clave extraídas sí parecen ser relativamente útiles puesto que empleando *ROUGE-1* (unigramas) su rendimiento es ligeramente superior a la media.
3. El método de compresión de sentencias propuesto resulta contraproducente para los resúmenes muy cortos al ofrecer un rendimiento mucho peor que el simple truncamiento. Estos resultados coinciden con los de Lin (2003), por lo que antes de plantearse la eliminación del método de compresión habría que estudiar su aplicabilidad al texto de manera global y no local.
4. Debido a estos resultados es difícil afirmar qué método de los anteriormente mencionados resulta más útil para la construcción de resúmenes extractivos tan cortos. No obstante, a la luz de los resultados generales de *DUC 2004* en que el *baseline* (primeros 75 caracteres del documento) ofreció sistemáticamente resultados inapreciablemente peores que el mejor participante, inapreciablemente superiores a los 5 mejores participantes, sustancialmente mejores que la media de participantes y sólo fue superado de manera rotunda por los seres humanos parece necesario

[1] Recuérdese que por sentencia entendemos una oración gramatical mientras que una frase tan sólo es una secuencia de palabras dotadas de sentido pero que no forman oración. Por ejemplo, la sentencia "`Welcome to Wikipedia, the free-content encyclopedia that anyone can edit`" constaría de dos frases separadas por una coma.

[2] Mani, Gates y Bloerdom (1999) describen un sistema para la eliminación de frases innecesarias. Jing (2000) describe un sistema similar aunque más flexible (no se aplicaría a todas las sentencias extraídas) y general (se basa en distintas fuentes de conocimiento y no únicamente en heurísticos). Knight y Marcu (2000) desarrollaron un sistema para la compresión de sentencias empleando pares `<abstract, texto>` como datos de entrenamiento. Resulta interesante el estudio de Lin (2003) según el cual un resumen extractivo construido a partir de sentencias comprimidas no resulta necesariamente mejor, aún así *"existe potencial en la compresión de sentencias pero es necesario encontrar un mejor sistema de compresión que tenga en cuenta para la optimización aspectos globales entre distintas sentencias"*.

plantearse la utilidad de métodos tan complejos para extraer resúmenes de semejante longitud a partir de textos de estilo periodístico.

5. En en el caso de los resúmenes cortos (máximo 665 caracteres) sí hay un método preferible sobre los demás: la combinación de las puntuaciones obtenidas a partir de los fragmentos y de la significatividad media por carácter. Empleando dicho método, *blindLight* ha ofrecido rendimientos entre apreciable y sustancialmente superiores a la media de participantes en *DUC 2004* y, dependiendo de la medida empleada, entre sustancial y apreciablemente inferiores a los alcanzados por los 5 mejores participantes.
6. Por lo que respecta al tamaño de *n*-grama empleado parece ser preferible la utilización de 4-gramas sobre 3-gramas. No obstante, las diferencias no son apreciables para ninguna medida *ROUGE* a excepción de *ROUGE-3* y *ROUGE-4* en que resultan sustanciales.

En resumen, la utilización de *blindLight* como método de extracción de palabras clave sin tratar de construir un "titular" legible ofrece resultados muy similares a los de la mayor parte de tecnologías disponibles y, al igual que éstas, se encuentra muy lejos de alcanzar los resultados de un sistema tan sencillo como extraer los primeros caracteres de un artículo. Estaría por determinar si este fenómeno se limita a textos periodísticos o es generalizable a otros estilos.

En cuanto a resúmenes de mayor longitud pero igualmente cortos (máximo 665 caracteres) se puede afirmar que *blindLight* es, cuando menos, apreciablemente mejor que muchas de las tecnologías disponibles. De las diversas configuraciones admitidas por la técnica propuesta por el autor la que ofrece de manera sistemática los mejores resultados es aquella que emplea 4-gramas, información mutua como método de ponderación de los mismos y determina la relevancia de las sentencias a extraer combinando la puntuación obtenida a partir de los fragmentos y de la significatividad media por carácter.

Ejemplo de resumen humano

Lebanon's Parliament voted the country's top military man, Gen. Emile Lahoud, president. Lahoud, who promises to clean up a graft-riddled government, is popular and is backed by powerful Syria. It is unclear, though, whether Prime Minister Hariri, in office since 1992 and credited with the country's economic recovery, will continue to head the cabinet. 31 of 128 legislators chose not to support him, leaving it to the president to name the next prime minister. Consequently, Hariri withdrew his candidacy, claiming the president acted unconstitutionally when he accepted the mandate to name a prime minister. Hariri's administration was plagued by nepotism.

Resultado del mejor participante en *DUC 2004* *(ROUGE-L)*

The commander of Lebanon's army will become the country's next president after winning the crucial backing of Syria, the powerbroker in Lebanon. According to the constitution, a president must begin his term by appointing the prime minister and a Cabinet through a decree issued after consulting with Parliament members. Sources close to the prime minister, whose Cabinet has been in care-taker capacity since last Tuesday's swearing in of Emile Lahoud as president, said Hariri turned down the invitation from Lahoud to select a new Cabinet. Hariri is credited with restoring economic confidence and stabilizing the national currency.

Resumen *blindLight*

The 128-member legislature is expected to meet Thursday to elect Lahoud after outgoing President Elias Hrawi signs the constitutional amendment. Under a formula aimed at preventing the recurrence of the 1975-90 civil war, power in Lebanon is shared equally by a Maronite Christian president, a Sunni Muslim prime minister and a Shiite Parliament speaker. Prime Minister Rafik Hariri has declined an informal invitation from Lebanon's new president to form the next government, sparking a political crisis in this country as it rebuilds from its devastating civil war. Lahoud had been expected to issue a presidential decree last week asking Hariri to form the next government after the president polled members of the 128-seat Parliament on their choice for prime minister.

Fig. 120 Comparación de uno de los peores (*ROUGE-L*) resúmenes *blindLight* con los correspondientes a un ser humano y al mejor participante en *DUC 2004*.

Ejemplo de resumen humano

In a move widely viewed as an effort to placate the far right as he moves to withdraw from more West Bank land, Israeli Prime Minister Netanyahu named hardliner Ariel Sharon foreign minister and chief peace negotiator. Sharon, a military leader with legendary victories in the 1967 and 1973 Mideast wars, is infamous in the Arab world as the defense minister in the 1982 invasion of Lebanon during which Lebanese Christian militiamen, Israeli allies, slaughtered hundreds of unarmed Palestinians. His appointment as lead negotiator was denounced as a "disaster" in the Lebanese press and "a bullet of mercy to the peace process" in a Syrian paper.

Resultado del mejor participante en *DUC 2004* *(ROUGE-L)*

Sharon is the general who led Israel's 1982 invasion of Lebanon, and a former housing minister who strengthened Jewish settlement in territories Israel captured from Syria, Jordan and Egypt in the 1967 Mideast war. Ariel Sharon has a law degree, and fancies himself a farmer. 1996: Named infrastructure minister in Netanyahu government. Prime Minister Benjamin Netanyahu named Sharon foreign minister on Friday, effectively putting the hard-liner in charge of negotiating Israel's final borders with the Palestinians. Palestinians also were expressing growing unease over the naming of hawkish former Israeli general Ariel Sharon as Netanyahu's foreign minister.

Resumen *blindLight*

Just days before heading to the United States for critical negotiations with Palestinian leaders, Prime Minister Benjamin Netanyahu jolted the Middle East peace effort with the appointment of Ariel Sharon as Israeli foreign minister. Ariel Sharon's appointment as the Israeli foreign minister serves as "the bullet of mercy" for the Middle East peace process, an official Syrian newspaper said Saturday. Prime Minister Benjamin Netanyahu named Sharon foreign minister on Friday, effectively putting the hard-liner in charge of negotiating Israel's final borders with the Palestinians. An Israeli tribunal looking into the invasion found him indirectly responsible for the massacre of hundreds of Palestinian refugees by Christian Lebanese militiamen at two Beirut camps.

Fig. 121 Comparación de uno de los mejores (*ROUGE-L*) resúmenes *blindLight* con los correspondientes a un ser humano y al mejor participante en *DUC 2004*.

4.1 Variabilidad de los resultados entre distintos idiomas

Parece claro que el resumen automático no debe limitarse a unos pocos idiomas:

> *Si un sistema de resumen automático emplea métodos tomados de la recuperación de información y, por tanto, independientes del idioma o si los métodos para un idioma específico pueden simplificarse y adaptarse a otros se podrá adaptar rápidamente un sistema de extracción de resúmenes en un idioma para que funcione con otros.* (Hovy 1999)

Dentro de este entorno multilingüe hay dos escenarios que reciben mucha atención[1]. Por un lado, estarían aquellos sistemas que deben resumir en un único idioma "objetivo" documentos escritos en una variedad de idiomas "fuente". Por ejemplo, Evans y Klavans (2003) o Evans, Klavans y McKeown (2004) generan resúmenes en inglés a partir de noticias escritas en inglés o traducidas automáticamente[2] al inglés desde otros lenguajes. En segundo lugar se encontrarían aquellos en los que deben producirse resúmenes en un idioma arbitrario a partir de documentos escritos en dicho idioma u otro diferente, por ejemplo, el proyecto *MUSI* (Busemann 2001) o (Lenci *et al.* 2002) que a partir de artículos de medicina escritos en inglés o italiano obtiene resúmenes en alemán y francés, o el trabajo desarrollado por Gawronska (2002) que permite el resumen de noticias escritas en inglés al danés, sueco y polaco. En este segundo escenario el enfoque extractivo no parece ser el preferido por los investigadores y se tiende a la utilización de representaciones abstractas de los documentos que permiten la generación del resumen final en el idioma seleccionado (Busemann 2001), (Gawronska 2002).

No obstante, y como paso previo al desarrollo de tales sistemas, resulta interesante el estudio de técnicas estadísticas de resumen automático que sean fácilmente adaptables a distintos idiomas y ofrezcan resultados consistentes con independencia del lenguaje a

[1] En particular en la Unión Europea.

[2] Las tareas tercera y cuarta de *DUC 2004* se correspondían con este escenario, los sistemas participantes debían resumir tanto documentos escritos en inglés como traducidos al inglés desde el árabe (*NIST* 2004).

resumir. En este sentido cabría señalar el trabajo realizado con dos sistemas ya mencionados, ambos extractivos y basados en técnicas independientes del idioma: *SUMMARIST* (Hovy y Lin 1997) y *MEAD* (Radev, Blair-Gondensohn y Zhang 2001). El primero de ellos fue adaptado con facilidad para su aplicación al bahasa Indonesia[1] (Lin 1999) mientras que el segundo se ha extendido para su funcionamiento en chino e inglés y en teoría con cualquier otro lenguaje natural (Radev *et al.* 2004). Otro sistema de resumen multilingüe, según esta interpretación, fue *MINDS* (Cowie *et al.* 1998) que permitía la obtención de resúmenes a partir de documentos escritos en coreano, español, japonés o ruso.

Por último, hay que destacar el trabajo llevado a cabo por Radev *et al.* (2002) para la evaluación de sistemas de resumen automático en entornos multilingües. Estos investigadores han publicado[2] un *corpus* paralelo de textos en chino e inglés junto con resúmenes creados a partir de cada documento individual y de distintos subconjuntos del *corpus*.

Por su propia naturaleza *blindLight* debería ser aplicable a distintos idiomas sin mayores problemas. No obstante, la cuestión no es su aplicabilidad sino su variabilidad, esto es, la influencia que tiene el idioma en que está escrito un documento en el momento de extraer su resumen. Dicho de otro modo, dadas dos traducciones literales de un texto ¿son también traducciones literales los resúmenes extraídos?

Sin lugar a dudas, hubiera resultado muy interesante la utilización del *corpus* paralelo antes mencionado para la evaluación de la nueva técnica y en el futuro se llevará a cabo tal evaluación. No obstante, el autor ha realizado por el momento una pequeñísima "prueba de concepto" para verificar hasta qué punto los resultados de la técnica que propone son invariables respecto a los distintos idiomas.

Para ello se seleccionaron cinco[3] notas de prensa de la Comisión Europea[4] que disponían de versiones en alemán, danés, francés, húngaro e inglés. Todos estos idiomas, a excepción del húngaro que es urálico, son indoeuropeos de los cuales todos, a excepción del francés que es romance, son germánicos siendo, a su vez, el danés un idioma germánico nórdico y alemán e inglés germánicos occidentales. La longitud media de los documentos empleados es de 2800 caracteres pero hay diferencias sustanciales entre los distintos idiomas. Por ejemplo, el francés es el más verboso con un promedio de 3.073 caracteres por documento frente al inglés con 2.506 caracteres. Por último, aun cuando las traducciones son literales hasta donde ha podido comprobar el autor, debido al uso de la puntuación el número de sentencias puede variar ligeramente de un idioma a otro. Por ese motivo los documentos originales fueron "corregidos" para garantizar que el número de sentencias era idéntico en todas las traducciones y permitir así la utilización del coeficiente de Spearman[5] para comprobar la correlación entre los distintos resúmenes.

---

[1] La variedad de malayo hablada en ese país.

[2] `http://www.clsp.jhu.edu/ws2001/groups/asmd`

[3] "*French journalists win first EU 'For Diversity. Against Discrimination' Award*", *"The European Union on your doorstep: new generation of information relays launched"*, *"Europeans want policy makers to consider the environment as important as economic and social policies"*, *"European Commission launches investigations into sharp surge in Chinese textiles imports"* y *"Post tsunami: the Commission reinforces its disaster response capacity"*.

[4] `http://europa.eu.int/comm/press_room/index_en.htm`

[5] Véase página 75 y posteriores.

El método empleado en este pequeño experimento es similar al empleado por Radev *et al.* (2002) para obtener resúmenes extractivos "manuales" para cada documento del *corpus* paralelo chino-inglés que, recordemos, estaba alineado a nivel de sentencia. En su caso recurrieron a revisores que para cada sentencia asignaron una puntuación de 0 a 10 en función de su utilidad para un resumen extractivo. Una vez hecho esto, para obtener un resumen extractivo "manual" de un documento tan sólo era necesario indicar el porcentaje de compresión y seleccionar aquellas sentencias más útiles hasta completar la longitud deseada; el resumen equivalente en el otro idioma se construía mediante las sentencias homólogas. De este modo, fue posible disponer de cientos de resúmenes elaborados según un criterio humano y garantizando que los resúmenes en ambos idiomas eran traducciones literales.

En este caso también se dispone de un mínimo conjunto de documentos alineado a nivel de sentencia (después de corregir la puntuación en algún caso concreto) y de un evaluador automático: *blindLight*. Lo que se desea saber es hasta qué punto la técnica es invariable frente al idioma, es decir, en qué medida se acerca al ideal según el cual siempre determinaría la sentencia más relevante con independencia del idioma. Para ello, se procedió a procesar cada documento mediante la nueva técnica obteniendo para cada una de sus sentencias la significatividad media por carácter.

A partir de aquí resulta inmediato elaborar para cada documento una lista de sentencias ordenadas por relevancia decreciente, listas que pueden ser comparadas mediante el coeficiente de Spearman para comprobar su correlación. Recordemos que dicho coeficiente varía en el intervalo [-1, 1] donde -1 significa que hay una correlación negativa perfecta, 0 la ausencia de correlación y 1 la existencia de una correlación positiva perfecta.

Por ejemplo, si al comparar la lista de sentencias ordenadas para un documento d_i en los lenguajes L_1 y L_2 se obtuviese -1 significaría que las sentencias más importantes en un idioma serían sistemáticamente elegidas como las menos relevantes en el otro y viceversa; si el coeficiente fuese 0 no habría ninguna correlación y en caso de obtener la unidad la técnica funcionaría de un modo "ideal" puesto que habría otorgado a cada sentencia la misma relevancia con independencia del idioma.

En la Tabla 24 se muestran los resultados obtenidos al analizar la correlación existente entre las ordenaciones producidas por *blindLight* para cada idioma. Como se puede ver, en todos los casos la correlación es superior a 0,5 y ciertamente es elevada aunque lejos de ser ideal. Por otro lado, el grado de correlación parece venir marcado por el "parentesco" de los distintos idiomas[1], lo cual era de esperar, y en cierta medida por la longitud del documento pues en los más largos la correlación es menor[2]. La misma prueba fue llevada a cabo para el sistema *MEAD* (Radev, Blair-Goldensohn y Zhang 2001) y como se puede comprobar (véase Tabla 25) el comportamiento de ambos sistemas es muy similar.

En resumen, la técnica *blindLight* es una herramienta útil para la extracción de resúmenes automáticos a partir de texto libre escrito en cualquier lenguaje natural. Su aplicación para la obtención de resúmenes muy cortos (máximo 75 caracteres) no parece adecuada pero, a la luz de las evaluaciones realizadas hasta la fecha, muy pocas técnicas consiguen superar la eficacia de un método tan sencillo como extraer los primeros

[1] Por ejemplo, alemán, danés e inglés presentan los valores más elevados mientras que el húngaro es el que obtiene los menores valores en todos los casos.

[2] Aunque el autor no ha llevado a cabo ningún experimento en relación con la influencia del tamaño de los documentos es muy probable que la previa segmentación del documento en pasajes, mediante técnicas análogas a *TextTiling* (Hearst 1994), influya positivamente en la calidad del resumen final.

caracteres del documento. Por lo que respecta a la extracción de resúmenes cortos (máximo 665 caracteres) su rendimiento es superior a muchas de las técnicas más avanzadas disponibles y aunque aún no se ha implementado un verdadero sistema de resumen multidocumento su desarrollo resulta natural debido a las características de la técnica para la detección de similitudes "semánticas". Por lo que respecta a la invariabilidad de los resultados respecto al idioma a resumir las pruebas preliminares indican que es bastante elevada aunque la naturaleza del idioma influye de manera importante. No obstante, con base en la experiencia previa, el autor tiene confianza en que empleando marcos de evaluación multilingüe o monolingüe en idiomas distintos al inglés será posible alcanzar buenos resultados en cualquier lenguaje natural.

EN/FR	EN/DE	EN/DA	EN/HU	FR/DE	FR/DA	FR/HU	DE/DA	DE/HU	DA/HU
0,50	0,68	0,77	0,81	0,63	0,71	0,35	0,90	0,58	0,55
0,72	0,74	0,73	0,61	0,67	0,78	0,67	0,81	0,55	0,52
0,36	0,43	0,52	0,15	0,42	0,59	0,63	0,71	0,12	0,20
0,60	0,83	0,80	0,72	0,64	0,68	0,61	0,82	0,67	0,77
0,43	0,89	0,93	0,93	0,71	0,46	0,54	0,82	0,93	0,79
0,52	**0,71**	**0,75**	**0,64**	**0,61**	**0,64**	**0,56**	**0,81**	**0,57**	**0,57**

Tabla 24. Coeficientes de correlación de Spearman entre los resúmenes obtenidos empleando información mutua y 3-gramas. La última fila es el valor medio.

EN/FR	EN/DE	EN/DA	EN/HU	FR/DE	FR/DA	FR/HU	DE/DA	DE/HU	DA/HU
0,49	0,71	0,86	0,54	0,30	0,55	0,55	0,53	0,20	0,56
0,48	0,37	0,46	0,84	0,60	0,81	0,46	0,48	0,44	0,47
0,53	0,67	0,49	0,82	0,41	0,41	0,55	0,24	0,75	0,48
0,56	0,68	0,81	0,45	0,63	0,63	0,41	0,79	0,53	0,41
0,89	0,94	-0,26	0,94	0,94	0,09	0,83	-0,20	0,89	-0,09
0,59	**0,67**	**0,47**	**0,72**	**0,58**	**0,50**	**0,56**	**0,37**	**0,56**	**0,37**

Tabla 25. Resultados de la misma prueba para el sistema *MEAD*.

La Comisión ha adoptado hoy propuestas relativas a un paquete de medidas destinadas a reforzar la capacidad de respuesta de la Unión Europea en caso de catástrofes. Estas medidas se destinan a financiar nuevos equipos especializados en materia de planificación para agilizar el suministro eficaz de ayuda a largo plazo; a reforzar la capacidad de la Unión de facilitar equipos de expertos civiles y de equipo y a suministrar ayuda humanitaria. La Comunicación adoptada hoy también presenta un informe detallado sobre la utilización de los 450 millones de euros anunciados por la UE tras la catástrofe del tsunami. Las propuestas adoptadas hoy constituyen la contribución de la Comisión al plan de acción tras el tsunami propuesto por la Presidencia luxemburguesa el 31 de enero.

«Vistas las situaciones anteriores y nuestra capacidad de responder inmediatamente ante la catástrofe del tsunami, la Comisión propone ahora medidas que nos ayudarán, en el futuro, a contribuir de forma rápida y eficaz a las tareas de reconstrucción tras una catástrofe» ha declarado la Comisaria de Relaciones Exteriores y Política de Vecindad, Benita Ferrero-Waldner, que propone dichas medidas conjuntamente con los Comisarios Michel y Dimas.

Stavros Dimas, Comisario Europeo responsable de Protección Civil ha dicho: «Nuestra reacción ante el Tsunami ha demostrado el claro valor añadido que la dimensión europea aporta a la asistencia en materia de protección civil. Las propuestas de hoy hacen avanzar un paso más al Mecanismo actual... Tomadas en su conjunto, permitirán disponer de un instrumento que garantiza una reacción europea eficaz ante futuras catástrofes».

The Commission has today adopted proposals for a package of measures to reinforce the European Union's disaster response capacity. The package will: fund new specialist planning teams to speed up the effective delivery of long term aid; reinforce the Union's capacity to provide specialised civil expertise units and equipment; and strengthen the Union's capacity to deliver humanitarian aid. The Communication adopted today also provides a detailed progress report of how the 450 million Euro pledged by the EU after the tsunami disaster is being spent. The proposals agreed today are the Commission's contribution to the post-tsunami Action Plan proposed by the Luxembourg Presidency on 31st January.

"Against our background and success to respond immediately to the Tsunami disaster the Commission proposes now measures that will help us to respond swiftly and effectively to post crisis reconstruction in the future" said Commissioner for External Assistance and European Neighbourhood Policy, Benita Ferrero-Waldner, who proposed the steps with Commissioners Michel, and Dimas.

Stavros Dimas, the European Commissioner responsible for Civil Protection, said: "The response to the Tsunami demonstrated the clear added value that the European dimension brings to civil protection assistance. The proposals made today take the existing Mechanism one step further. Taken together they will result in an instrument that guarantees an effective European reaction to future disasters."

La Commission a approuvé ce jour plusieurs propositions relatives à un train de mesures destinées à renforcer la capacité de réaction de l'Union européenne en cas de catastrophes. Il s'agit des mesures suivantes: financement accordé pour la mise en place d'équipes de spécialistes en matière de planification pour accélérer la fourniture efficace d'une aide à long terme et renforcement de la capacité de l'Union à mettre à disposition des équipes d'experts civils et du matériel et à effectuer des opérations d'aide humanitaire. La Communication adoptée ce jour fournit également un rapport détaillé sur l'utilisation de la contribution de 450 millions d'euros annoncée par l'Union européenne au lendemain du tsunami. Les propositions approuvées aujourd'hui constituent la contribution de la Commission au plan d'action après-tsunami proposé le 31 janvier dernier par la Présidence luxembourgeoise.

"Au regard de nos expériences antérieures et de notre capacité à réagir sans délai après la survenue du tsunami, la Commission propose maintenant des mesures qui nous permettront, dans l'avenir, de contribuer rapidement et en toute efficacité aux travaux de reconstruction faisant suite à des crises", a déclaré Benita Ferrero-Waldner, commissaire chargée des relations extérieures et de la politique européenne de voisinage, qui a présenté les mesures en question conjointement avec les commissaires Michel et Dimas.

"Notre réaction lors du tsunami témoigne clairement de la valeur ajoutée que la dimension européenne confère à l'assistance en matière de protection civile. Les mesures proposées aujourd'hui constituent une nouvelle avancée du mécanisme existant et elles permettront de disposer d'un instrument garantissant une réaction européenne efficace lors de prochaines catastrophes" a ajouté Stavros Dimas, commissaire européen chargé de la protection civile.

Die Kommission hat heute Vorschläge für ein Maßnahmenpaket angenommen, das die Katastrophenabwehrkapazitäten der Europäischen Union stärken soll, indem neue spezialisierte Planungsteams zur Beschleunigung der wirksamen Erbringung langfristiger Hilfe finanziert werden, die Kapazitäten der EU für die Bereitstellung spezialisierter zivilen Expertenteams und Ausrüstung verstärkt werden und die Kapazitäten der EU für die Erbringung humanitärer Hilfe ausgebaut werden. Die heute angenommene Mitteilung enthält außerdem einen ausführlichen Fortschrittsbericht darüber, wie die von der EU nach der Tsunami-Katastrophe bereitgestellten 450 Mio. EUR eingesetzt werden. Die heute gebilligten Vorschläge stellen den Beitrag der Kommission zum dem Aktionsplan dar, den der luxemburgische Vorsitz am 31. Januar infolge des Tsunami vorgelegt hatte.

"Vor dem Hintergrund des Erfolgs unserer unmittelbaren Reaktion auf die Tsunami-Katastrophe schlägt die Kommission nun Maßnahmen vor, die uns in die Lage versetzen werden, künftig rasch und wirksam zu reagieren, wenn es um Wiederaufbaumaßnahmen nach Krisen geht", kommentierte die für die Außenhilfe und die Europäische Nachbarschaftspolitik zuständige Kommissarin Benita Ferrero-Waldner, die die Maßnahmen gemeinsam mit den Kommissaren Michel und Dimas vorgestellt hat.

Der für den Katastrophenschutz zuständige Kommissar Stavros Dimas äußerte sich wie folgt: "Die Reaktion auf den Tsunami hat den deutlichen Mehrwert gezeigt, den die europäische Dimension für die Katastrophenhilfe erbringt. Die heute unterbreiteten Vorschläge gehen einen Schritt weiter als das bestehende Verfahren. Gemeinsam werden sie ein Instrument bilden, das eine effiziente europäische Reaktion auf künftige Katastrophen ermöglicht."

I dag vedtog Kommissionen en række forslag som led i en pakke af foranstaltninger for at styrke EU's katastrofeberedskab. Med pakken finansieres nye eksperthold, hvis planlægning skal sikre, at den langsigtede bistand bliver effektiv, EU's muligheder for at tilbyde specialiserede civile ekspertenheder og udstyr forbedres, og EU's muligheder for at yde humanitær bistand øges. Meddelelsen, som blev vedtaget i dag, rummer også en detaljeret statusrapport med oplysninger om, hvad de 450 mio. EUR, som EU har stillet til rådighed efter flodbølge-katastrofen, bruges til. Forslagene, der blev vedtaget i dag, er Kommissionens bidrag til den handlingsplan, det Luxembourgske formandskab foreslog den 31. januar 2005 som opfølgning på flodbølgekatastrofen.

"I lyset af den hurtige og vellykkede indsats i forbindelse med flodbølgekatastrofen foreslår Kommissionen nu foranstaltninger, som sætter os i stand til fremover hurtigt og effektivt at yde en genopbygningsindsats efter en nødsituation", sagde kommissæren for eksterne forbindelser og den europæiske naboskabspolitik, Benita Ferrero-Waldner, der sammen med kommissær Louis Michel og Stavros Dimas foreslog dette tiltag.

Stavros Dimas, der er kommissær med ansvar for civilbeskyttelse, udtalte: "Indsatsen i forbindelse med flodbølgen viste, at den europæiske dimension skaber en klar merværdi i forbindelse med civilbeskyttelsesbistand. Den eksisterende ordning udvikles yderligere med de forslag, som er fremsat i dag. De vil samlet set munde ud i et beredskab, som garanterer en effektiv europæisk indsats i forbindelse med katastrofer i fremtiden."

A Bizottság ma javaslatokat fogadott el az Európai Unió katasztrófaelhárító képességének fokozására hivatott intézkedéscsomagot illetően. A csomag a hosszú távú segítségnyújtás hatékony felgyorsítása érdekében biztosítja az új szakértői tervezőcsoportok működéséhez szükséges anyagi fedezetet; fokozza az Unió azon képességét, hogy speciális civil szakértői csoportokat és felszerelést biztosítson; elősegíti, hogy az Unió nagyobb részt vállalhasson a humanitárius segítségnyújtásban. A ma elfogadott közlemény ezenkívül részletesen beszámol az EU által a cunami-katasztrófát követően felajánlott 450 millió euró felhasználásáról is. A javaslatok elfogadásával a Bizottság ahhoz a cselekvési tervhez kíván hozzájárulni, melyet a luxemburgi elnökség január 31-én, a cunamit követően terjesztett elő.

„Meglévő lehetőségeinkre és a cunami-katasztrófára adott gyors válaszintézkedéseinkre alapozva a Bizottság olyan intézkedéseket javasol, amelyeknek köszönhetően a jövőben gyorsan és hatékonyan reagálhatunk a válságokat követő újjáépítési szükségletekre" – nyilatkozta Benita Ferrero-Waldner, az EU külkapcsolatokért és európai szomszédsági politikáért felelős biztosa, aki Louis Michel és Stavros Dimas európai biztossal közösen javasolja ezeket az intézkedéseket.

Stavros Dimas, a polgári védelemért felelős biztos kijelentette: „A cunamira adott válaszlépéseink bebizonyították, mekkora többletet ad az európai dimenzió a polgári védelem által nyújtott segítséghez. A ma beterjesztett javaslatoknak köszönhetően újabb lépéssel viszik előre a meglévő mechanizmust és együttesen olyan eszköz létrejöttét eredményezik, amely garantálja, hogy Európa hatékonyan reagál a jövőbeni katasztrófákra."

Fig. 122 Resúmenes de aproximadamente el 25% de la última nota de prensa en español, inglés, francés, alemán, danés y húngaro (se ha conservado la puntuación original y el número de sentencias varía).