

Resumen de texto en progreso: una revisión de la literatura

elena lloret·manuel palomar

Publicado en línea: 30 de abril de 2011
© Springer Science+Business Media BV 2011

Resumen Este artículo contiene una gran **revisión de la literatura en el campo de investigación del resumen de texto (TS) basado en tecnologías del lenguaje humano (HLT)**. TS ayuda a los usuarios a **gestionar la gran cantidad de información disponible**, condensando el contenido de los documentos y **extrayendo los hechos o temas más relevantes incluidos en ellos**. El **rápido desarrollo de las tecnologías emergentes** plantea **nuevos desafíos a este campo de investigación**, que aún deben resolverse. Por lo tanto, es **fundamental analizar su progreso a lo largo de los años y proporcionar una visión general** de las direcciones pasadas, presentes y futuras, **destacando los principales avances logrados y destacando las limitaciones restantes**. Con este propósito, se abordan **varios aspectos importantes** dentro del alcance de esta encuesta. Por un lado, el trabajo pretende dar una **visión general del estado del arte, describiendo los principales conceptos**, así como **diferentes enfoques** de síntesis y foros internacionales relevantes. Además, es importante destacar el hecho de que el **nacimiento de nuevos requisitos y escenarios** ha dado lugar a nuevos **tipos de resúmenes con fines específicos** (p. ej., **resúmenes basados en opiniones**) y **dominios novedosos** en los que TS ha demostrado ser también adecuado para (por ejemplo, **blogs**). Además, TS se **combina con éxito con una serie de sistemas inteligentes basados en HLT** (p. ej., **recuperación de información, respuesta a preguntas y clasificación de texto**). Por otra parte, en este trabajo también se realiza un estudio profundo de la **evaluación de resúmenes**, donde se explican las **metodologías y sistemas existentes**, así como las nuevas investigaciones que han surgido en torno a la **evaluación automática de la calidad de los resúmenes**. Por fin, algunas reflexiones sobre TS en general y su futuro animarán al lector a pensar en nuevos enfoques, aplicaciones y líneas para realizar investigaciones en los próximos años. El análisis de estos temas permite al lector tener un **amplio y útil bagaje sobre los principales aspectos importantes de este campo de investigación**.

Palabras clave tecnologías del lenguaje humano·Resumen de texto·Sistemas inteligentes

E. Lloret (✉)·M. Palomar

Departamento de Software y Sistemas Informáticos, Universidad de Alicante, Apdo. de correos, 99, 03080 Alicante, España

correo electrónico: elloret@dlsi.ua.es

M. Palomar

correo electrónico: mpalomar@dlsi.ua.es

1. Introducción

Las tecnologías del lenguaje humano (HLT) cubren una amplia gama de actividades con el objetivo final de permitir que las personas se comuniquen con las máquinas utilizando habilidades de comunicación natural (Cole 1997). Además, el rápido crecimiento de Internet ha resultado en un aumento masivo de la información disponible en diferentes formatos (por ejemplo, texto, video, imágenes) difíciles de manejar. En consecuencia, se pueden desarrollar aplicaciones inteligentes basadas en HLT, como la recuperación de información, la clasificación de textos, la síntesis de textos o el análisis de sentimientos, para tratar de manera eficiente toda esta información. En particular, Text Summarisation (TS), cuyo objetivo es obtener una transformación reductiva del texto fuente a texto resumen a través de la condensación de contenido por selección y/o generalización sobre lo importante en la fuente (Spärck Jones 1999), es fundamental para gestionar de manera eficiente dicha información, permitiendo así a los usuarios ahorrar tiempo y recursos, así como encontrar rápidamente la información específica que buscan dentro de los documentos.

Aunque empezó a finales de los años cincuenta (Luhn 1958), el ST ha experimentado un gran desarrollo en los últimos años, y se han propuesto un amplio abanico de técnicas y paradigmas para abordar este campo de investigación (Spärck Jones 2007). Sin embargo, producir un resumen automáticamente es muy desafiante. Cuestiones como la redundancia, la dimensión temporal, la correferencia o la ordenación de frases, por citar algunas, deben tenerse en cuenta especialmente a la hora de resumir un conjunto de documentos (multi-document summarisation), lo que dificulta aún más este campo (Goldstein et al. 2000). Además, en los últimos años se ha fomentado la investigación que intenta superar la falta de coherencia que a menudo presentan los resúmenes, lo que ha dado como resultado enfoques combinados que identifican contenido relevante y lo fusionan en nuevos fragmentos de información (Barzilay y McKeown 2005; Zajic et al. 2008). También cabe mencionar que a medida que cambia la sociedad, TS también lo hace, adaptándose a los nuevos requerimientos. Por ejemplo, la Web 2.0 (web social) ha dado lugar a la aparición de nuevos tipos de sitios web, como blogs, foros o redes sociales, donde cualquier persona puede expresar sus sentimientos hacia un tema, entidad, producto o servicio. Esto ha resultado en un nuevo tipo de resúmenes (basados en sentimientos) con el propósito de resumir las opiniones de los usuarios. Otro aspecto importante de TS se refiere a su evaluación. Esto también es muy desafiante porque no está claro, incluso para los humanos, qué tipo de información debe contener un resumen, como se ha demostrado en estudios anteriores (Nenkova 2006). Dependiendo de para qué esté destinado el resumen, la información variará y capturarla automáticamente es muy complicado. Se ha demostrado que los métodos para evaluar automáticamente los resúmenes se correlacionan bien con la evaluación humana, como en el caso de ROUGE (Lin 2004) o AutoSummENG (Giannakopoulos et al. 2008b). Sin embargo, aún persisten algunos problemas relacionados con la calidad de los resúmenes, como la gramaticalidad, la redundancia o la coherencia.

En este trabajo se realiza una revisión del estado del arte en TS, centrándose especialmente en la última década y los nuevos tipos de resúmenes que han aparecido en los últimos años, como los resúmenes basados en sentimientos o los resúmenes de actualización. No se espera que sea una revisión exhaustiva de todos los sistemas y técnicas que se han desarrollado desde el inicio de este campo de investigación, porque ya existen buenas encuestas para tales fines (Spärck Jones 2007; Saggion 2008). Por el contrario, la contribución de este artículo es proporcionar una visión general de TS, enfatizando los tipos de resumen recientes y cómo los resúmenes, aunque no son perfectos, pueden ser de gran ayuda para otros sistemas basados en HLT. Además, la segunda aportación se refiere al proceso de evaluación resumida, que es especialmente interesante, por el margen de mejora que aún permite. Por ello, ofrecemos un análisis de las metodologías y herramientas actuales para evaluar resúmenes. Finalmente, analizamos las direcciones futuras, identificando algunas nuevas tendencias sobre cómo se espera que progrese este campo de investigación en los próximos años.

La estructura del documento es la siguiente. Sección 2 contiene dos partes. En primer lugar, se proporciona una descripción general de los principales enfoques de TS a lo largo de los años (Sect. 2.1), y luego una hoja de ruta de

se describe la **presencia de TS en campañas internacionales de evaluación** (Secc.2.2). **Sección3** aborda la **actualidad de los TS**, donde se explica la descripción de **nuevos tipos de resúmenes**, así como las **técnicas de síntesis empleadas** (Apt.3.1). Además, el **análisis de TS en otros escenarios diferentes de los textos de noticias tradicionales** también se **proporciona en la Secc.3.2**. Más adelante, en la **Secc.4**, estudiamos la **aplicabilidad de los resúmenes de texto en otros sistemas inteligentes basados también en HLT** y, en particular, **cómo pueden beneficiar aplicaciones como la recuperación de información**, la **respuesta a preguntas o la clasificación de textos**. El uso de **resúmenes para este propósito** puede considerarse una **forma extrínseca de evaluarlos**. Sin embargo, la **secta.5** se dedica a la **evaluación**, centrándose principalmente en la **evaluación intrínseca**, distinguiendo entre los **métodos que valoran el contenido** de los resúmenes (Secc.5.1) de aquellos que **evalúan su calidad** (Secc.5.2). Por ello, describimos las **metodologías de evaluación actuales (semiautomáticas o completamente automáticas)**, así como sus **ventajas y sus principales limitaciones**. Finalmente, la **Secc.6** proporciona un **análisis amplio** en forma de **conclusiones y** brinda información sobre las **direcciones futuras de TS**.

2 Resumen de resumen de texto

La **definición de resumen** provista en la Secc.1 **no es estricto**. Por el contrario, **permite una amplia gama de tipos de resumen** en función de **para qué está destinado** el resumen, **cómo es la entrada**, etc. Una de las **taxonomías existentes más conocidas**, fue **propuesta en Spärck Jones (1999)**, donde se toman en consideración **tres clases de factores de contexto** que influyen en los resúmenes: **factores de entrada, propósito y salida**. Los **factores de entrada** se ocupan de **aspectos relacionados con la fuente**, como el **género, el idioma o el registro**. Los segundos, **factores de propósito**, incluyen **audiencia y uso**, por ejemplo, **reseñas literarias o alertas de emergencia**. Finalmente, los **factores de salida** se enfocan en el **estilo y la cobertura**, y normalmente son **impulsados por factores de propósito**. La taxonomía propuesta por Spärck Jones (1999) no es el único propuesto para clasificar diferentes factores de resumen. **hovy y lin (1999)** sugirió también una **taxonomía similar**, donde los **tipos de resúmenes** se clasifican según los **aspectos más relevantes**. Al igual que los factores de entrada, salida y propósito descritos en la taxonomía anterior, esta clasificación **distingue entre características del documento fuente, características del resumen como texto y características del uso del resumen**. La principal **diferencia entre ambas taxonomías** es que **esta última tiene en cuenta factores específicos relacionados con la coherencia y el nivel de subjetividad del resumen**. Además, hay una **taxonomía más sugerida por Mani y Maybury (1999)**, clasificando los sistemas de resumen con respecto al **enfoque adoptado para generar resúmenes**. Se pueden abordar **diferentes enfoques para producir resúmenes**, enfrentando el problema desde **tres niveles**: nivel de **superficie**, de **entidad o de discurso**. Los enfoques de nivel de **superficie** tienen como objetivo **representar la información en términos de características superficiales** que **luego se combinan selectivamente para determinar una función destacada utilizada para extraer la información más importante de un documento**. Tales **características** comprenden **características temáticas** que toman en consideración **términos estadísticamente destacados**, por ejemplo, **basados en conteos de frecuencia; localización** da cuenta de la **posición de una unidad específica (palabra, oración, etc.) en un documento (párrafo, sección); antecedentes** se refiere a la **presencia de términos del título o encabezados en el texto**, las **partes iniciales del documento o la consulta de un usuario**; por fin, **palabras clave y frases** son expresiones como **"En resumen", "nuestra investigación", "en particular", o "En conclusión"**. Los enfoques a nivel de entidad construyen una **representación interna del documento o documentos para modelar las entidades y sus relaciones**. Estos enfoques tienden a **representar patrones de conectividad** en el documento y estas relaciones incluyen, **por ejemplo, semejanza** a través de la **superposición de vocabulario; proximidad**, que se refiere a la **distancia entre unidades de texto; tesauro relaciones** entre palabras como **sinonimia** o

parte de las relaciones; o *relaciones lógicas* tales como *contradicción, vinculación o acuerdo*. Por último, los *enfoques a nivel de discurso* modelan la *estructura global del documento*. Esto incluye características relacionadas con el *formato del documento* (como *esquemas del documento o hipertextos*), los *hilos de temas o subtemas desarrollados en el documento*, o su intento de *capturar la estructura de diferentes tipos de textos*, por ejemplo, *documentos narrativos o argumentativos* estructura.

Las *taxonomías* propuestas en Spärck Jones (1999) y Hovy y Lin (1999) tratan con una *granularidad muy fina*, en el sentido de que los *resúmenes se clasifican con respecto a diferentes criterios de acuerdo con su propia naturaleza*, mientras que el *sugerido por Mani y Maybury (1999)* agrupa *enfoques de resumen*, en lo que se refiere al *tipo de características y técnicas utilizadas para generar resúmenes*. El principal *problema de tener una taxonomía con una granularidad tan fina* surge cuando se *quiere clasificar un sistema de resumen con respecto a esos criterios*, ya que la *mayoría de los sistemas pueden compartir varias características*, aumentando la *dificultad en su clasificación y haciéndola poco clara*. Al mismo tiempo, el *problema con la taxonomía de Mani y Maybury* es la *pureza de la clasificación sugerida*. Actualmente, los *sistemas se basan en enfoques híbridos (Mani y Maybury 1999)*, combinando, por ejemplo, rasgos discursivos y superficiales.

Por lo tanto, teniendo en cuenta las taxonomías antes mencionadas, los *enfoques de resumen se pueden caracterizar de acuerdo con muchas características*. Aunque *tradicionalmente se ha centrado en el texto, la entrada al proceso de resumen también puede ser información multimedia*, como *imágenes (Fan et al. 2008); video (Él et al. 1999) o audio (Zechner y Waibel 2000)*, así como *información en línea o hipertextos (sol et al. 2005)*. Además, *podemos hablar de resumir un solo documento (resumen de un solo documento) o múltiples (resumen de varios documentos)*. Con respecto a la *salida*, un resumen puede *ser un extracto* (es decir, cuando se *muestra una selección de frases "significativas"* de un documento), *resumen*, cuando el resumen puede servir como *sustituto del documento original y se añade nuevo vocabulario*, o incluso un *título* (o título). También es posible *distinguir entre genéricos resúmenes y centrado en la consulta* resúmenes (también conocidos como *centrados en el usuario* o *centrados en el tema*). El *primer tipo* de resúmenes puede servir como *sustituto del texto original*, ya que pueden tratar de *representar todos los hechos relevantes de un texto fuente*. En este último, el *contenido* de un resumen está *impulsado por una necesidad del usuario o una consulta*. En cuanto al *estilo del resultado*, normalmente se hace una amplia distinción entre *dos tipos de resúmenes*. *Indicativo* los resúmenes se utilizan para indicar *qué temas se abordan en el texto fuente*. Como resultado, pueden dar una *breve idea de lo que trata el texto original*. El *otro tipo, informativo* resúmenes, están *destinados a cubrir los temas del texto fuente y proporcionar información más detallada*. Aparte de estos dos tipos de resúmenes, también se puede tener en cuenta otro, es decir *resúmenes de evaluación crítica*. Este tipo de resúmenes se enfoca en *expresar los puntos de vista del autor sobre un tema o asunto específico*, e incluyen *reseñas, opiniones, comentarios, recomendaciones, etc.*, con una *fuerte dependencia de la interpretación cultural (Mani 2001a)*. Esa es la razón por la que son tan *difíciles de producir automáticamente* y, por lo tanto, la *mayoría de los sistemas solo intentan generar resúmenes indicativos o informativos*, simplemente *resumiendo lo que aparece en el documento fuente*. En los últimos años han aparecido *nuevos tipos de resúmenes*. Por ejemplo, el *nacimiento de la Web 2.0* ha fomentado *nuevos tipos de géneros textuales*, que contienen un *alto grado de subjetividad*, permitiendo así la *generación de resúmenes basados en sentimientos*. *Actualizar resúmenes*, que *asumen que el usuario ya tiene antecedentes y solo necesita la información más reciente sobre un tema*, son otro ejemplo del nuevo tipo de resumen. Más tipos de resúmenes se explican en la Secc.3.1. Finalmente, en cuanto al *lenguaje del resumen*, se puede distinguir entre *monolingüe, plurilingüe, y multilingüe* resúmenes, *dependiendo del número de idiomas tratados*. Los casos en los que el *idioma de entrada y de salida es*

tabla 1Tipos de integración según varios factores

MEDIOS DE COMUNICACIÓN	Texto
	Imágenes
	Video
	Habla
	hipertexto
APORTE	Documento único
	multidocumento
PRODUCCIÓN	Extracto
	Resumen
OBJETIVO	Titular
	Genérico
	Personalizado
	Centrado en consultas
	Actualizar
	basado en sentimientos
	Indicativo
	Informativo
	Crítico
IDIOMA	Monolingüe
	Plurilingüe
	multilingüe

lo mismo conduce a resúmenes monolingües. Sin embargo, si se trata de diferentes idiomas, el enfoque de resumen se considera multilingüe o multilingüe. Por ejemplo, si un sistema de resumen produce un resumen en español a partir de uno o más documentos en español, ese es el caso de un sistema monolingüe. Por el contrario, si es capaz de tratar varios idiomas, como español, inglés o alemán, y produce resúmenes en el mismo idioma que el documento de entrada, tendríamos un sistema de resumen multilingüe. Más allá de estos enfoques, si el resumen está en español, pero los documentos originales están en inglés, el resumen se ocuparía del cruce de idiomas, ya que los idiomas de entrada y salida son diferentes. Mesa1 resume los factores más comunes con respecto a la síntesis.

En el resto de esta sección, se describirán varios enfoques bien conocidos para TS (Sect.2.1). Los enfoques se agruparán de acuerdo con las técnicas predominantes empleadas. Además, en la Secc.2.2, junto con un análisis de las tendencias de TS que ha ido cambiando a lo largo de los años y que se reflejan en las pistas propuestas para la conferencia.

2.1 Enfoques comunes para generar resúmenes

El proceso de resumen se puede descomponer en tres subtarear principales: identificación del tema, interpretación del tema, y generación de resumen(Radev et al. 2002). La primera etapa determina la estructura temática de un texto, es decir, una representación que indica qué temas se incluyen en un texto.

y cómo esos temas cambian dentro del texto. Una vez que se han identificado los temas principales de un documento, se necesita una segunda etapa para comprender su significado y, por lo tanto, distinguir entre información relevante e irrelevante. La etapa final aborda la generación de un resumen mediante la fusión y fusión de la información previamente identificada. Sin embargo, dado que la etapa de generación de resúmenes no es fácil de abordar, la mayoría de los enfoques solo se enfocan en las dos primeras etapas, simplemente extrayendo las oraciones tal como aparecen en los documentos, produciendo extractos como consecuencia. A continuación, se describen varios enfoques extractivos. Específicamente, distinguimos entre cinco tipos de enfoques adecuados para TS, dependiendo de la naturaleza de las técnicas empleadas, que son: *basado en estadísticas* (Secta.2.1.1); *basado en temas* (Secta. 2.1.2); *basado en gráficos* (Secta.2.1.3); *basado en el discurso* (Secta.2.1.4); y *basado en aprendizaje automático* (Secta.2.1.5).

2.1.1 Enfoques basados en estadísticas

Luhn(1958) utilizaron recuentos de frecuencia de términos para producir resúmenes de documentos científicos con el objetivo de determinar la relevancia de una oración en un documento. La suposición subyacente es que las palabras más frecuentes son indicativas del tema principal de un documento. Sin embargo, no todas las palabras se tienen en cuenta. Por el contrario, las palabras vacías, es decir, las palabras que no llevan ninguna información semántica, como "un" o "el", no se utilizan para calcular la frecuencia del término. Bajo la misma suposición, se han empleado en TS varias técnicas basadas en conteos de frecuencia de términos. Por ejemplo, (Lloret y Palomar 2009) usan la frecuencia de las palabras en combinación con la longitud de las frases nominales para calcular la relevancia de una oración, superando el resultado del estado del arte en el resumen de un solo documento para el dominio de noticias. En McCargar(2005), varios enfoques estadísticos, como la frecuencia de términos o la frecuencia inversa de documentos ($tf*idf$), se analizan brevemente, así como los posibles problemas que pueden tener este tipo de características. La idea detrás de $tf*idf$ es que los términos frecuentes en un documento son importantes solo si no son muy frecuentes en toda la colección. Esta técnica también se ha empleado para puntuar oraciones, por ejemplo, en Gotti et al.(2007). Como se afirma en Filátova y Hatzivassiloglou(2004) este tipo de métodos pueden no ser suficientes para construir resúmenes de alta calidad, y otros tipos de conocimiento, por ejemplo, eventos, información semántica, basada en temas o discursiva, pueden ser los más apropiados para abordar el ST. Sin embargo, una revisión más profunda de las técnicas estadísticas para TS se lleva a cabo en Orásan et al.(2004) y Orásan(2009), donde se demuestra que estas técnicas, a pesar de ser sencillas y no requerir un nivel profundo de análisis de conocimientos, son apropiadas para construir buenos resúmenes. Además de las técnicas mencionadas anteriormente, también se analizan la información mutua, la ganancia de información y la frecuencia del documento inverso residual. La información mutua se puede utilizar para medir la dependencia o la información común entre dos palabras, mientras que la ganancia de información es una buena métrica para decidir la relevancia de un atributo y, en este caso, podría aplicarse perfectamente a los términos o frases de un documento. La frecuencia de documento inversa residual es una variante de la frecuencia de documento inversa, que calcula el término frecuencia de documento de acuerdo con el *veneno* distribución. Cada técnica establece una forma de asignar pesos a las palabras incluidas en el documento, y luego, las oraciones se puntúan en función de estos pesos, para determinar su relevancia. El enfoque sugerido en Mori(2002) también emplea la ganancia de información para determinar el peso de los términos del documento y luego la utiliza para resumir correctamente los documentos. La idea es construir primero grupos de documentos de acuerdo con la similitud entre ellos y luego calcular el peso de cada palabra en los grupos. El resumen final se elaborará seleccionando aquellas frases mejor puntuadas en función del peso de palabras que contenga, calculado previamente mediante la ganancia de información.

2.1.2 Enfoques basados en temas

En Edmundson (1969) los resúmenes se producen por medio de la identificación de palabras claves. Esta técnica consiste en determinar la relevancia de una oración por medio de las frases o palabras que contiene. Oraciones que contienen frases como "En conclusión" o "el objetivo de este trabajo" pueden ser buenos indicadores de información relevante. Además, otros enfoques, como Boguraev y Neff (2000), Neto et al. (2000), Angheluta et al. (2002), o Harabagiu y Lacatusu (2005) aprovechan las ventajas de combinar la identificación y segmentación de temas. Particularmente en Harabagiu y Lacatusu (2005), la estructura del tema se caracteriza en términos de temas, que son representaciones de eventos que se reiteran a lo largo de la colección de documentos y, por lo tanto, representan información repetitiva. Se analizan cinco formas diferentes de representar los temas: (1) a través de firmas de temas. Esta idea viene de Lin y Hovy (2000), donde se supone que el tema de un documento se puede representar mediante un conjunto de términos; (2) a través de firmas de temas mejoradas. Este se diferencia del anterior en que ahora se pretende descubrir relaciones relevantes entre dos conceptos tópicos; (3) a través de firmas temáticas, que se realiza mediante la segmentación de documentos mediante el algoritmo TextTiling (Hearst 1997) primero, y luego asignando etiquetas a los temas para poder clasificarlos más tarde; (4) mediante el modelado de la estructura de contenido de los documentos. La suposición aquí es que todos los textos que describen un tema dado son generados por un solo modelo de contenido (en este caso, un Modelo Oculto de Markov). Finalmente, el último método para representar temas dentro de un texto es usar (5) plantillas, siguiendo la idea del campo de extracción de información, identificando entidades o hechos específicos. Además, en Teng et al. (2008), se sugiere un enfoque de resumen de un solo documento que combina la identificación del tema local con la frecuencia de los términos. La metodología propuesta calcula primero la similitud de las oraciones y luego realiza la identificación del tema mediante la agrupación de oraciones. En un segundo paso, se seleccionan frases de temas locales según el valor de frecuencia del término. Además, no solo se utilizan palabras temáticas para detectar información relevante dentro de un documento. En otros enfoques, (Kuo y Chen 2008), por ejemplo, también se tienen en cuenta las palabras informativas y de eventos para producir resúmenes de varios documentos. La idea subyacente es que este tipo de palabras indican los conceptos y relaciones importantes, y pueden usarse para detectar oraciones relevantes dentro de un conjunto de documentos. Además, se utiliza un algoritmo de resolución temporal, de modo que las fechas y otras expresiones temporales se pueden traducir a formas calendáricas. La identificación de múltiples temas dentro de una colección heterogénea de documentos se aborda en Ando et al. (2005) por medio de representaciones de espacio vectorial; en particular, Cambio de escala residual iterativo utiliza que se ha demostrado que es adecuado para la construcción de modelos espaciales para objetos lingüísticos.

2.1.3 Enfoques basados en gráficos

También se ha demostrado que el uso de algoritmos de clasificación basados en gráficos es efectivo en TS. Básicamente, los nodos del gráfico representan elementos de texto (es decir, normalmente palabras u oraciones), mientras que los bordes son enlaces entre esos elementos de texto, previamente definidos (por ejemplo, relaciones semánticas, como la sinonimia). Sobre la base de la representación del texto como un gráfico, la idea es que la topología del gráfico revele cosas interesantes sobre los elementos más destacados del texto, por ejemplo, sobre la conectividad de los diferentes elementos. LexRank (Erkan y Radev 2004) es un sistema de resumen de varios documentos, en el que todas las oraciones candidatas que pueden incluirse potencialmente en el resumen se representan en un gráfico. En esta representación gráfica, dos oraciones están conectadas si la similitud entre ellas está por encima de un umbral predefinido. Luego, una vez que se construye la red, el sistema encuentra las oraciones más centrales realizando un recorrido aleatorio en el gráfico. En Mihalcea (2004), un análisis de

Se lleva a cabo varios algoritmos basados en grafos, evaluando también su aplicación a la extracción automática de oraciones en el contexto del TS. Además, en Wan et al.(2007), se sugiere un enfoque basado en gráficos de afinidad, tanto para resúmenes de documentos múltiples genéricos como centrados en consultas. La idea aquí es extraer oraciones con alta riqueza de información y novedad. Esto se logra tomando en consideración la similitud entre cada par de oraciones, incorporando información temática, diferenciando los vínculos intradocumento e interdocumento entre oraciones y, finalmente, penalizando la información redundante. En Giannakopoulos et al.(2008a) los gráficos de n-gramas de caracteres y palabras se utilizan para extraer información relevante de un conjunto de documentos, mientras que en Plaza et al.(2008) los gráficos se construyen utilizando conceptos identificados con Wordnet (Felbaum 1998) y sus relaciones, que luego se utilizan para construir una representación gráfica para cada oración en un documento. Este enfoque se ha probado con éxito en diferentes dominios, como noticias, documentos biomédicos o leyendas de imágenes.

2.1.4 Enfoques basados en el discurso

Además de todas las técnicas mencionadas anteriormente, también es posible abordar el problema de la síntesis desde un punto de vista lingüístico, por ejemplo, explotando las relaciones discursivas. Teoría de la Estructura Retórica (RST) propuesta en Mann y Thompson(1988) sirvió como base para el enfoque de resumen desarrollado en Marcu(1999), ampliando las relaciones retóricas, y utilizando este tipo de representación discursiva (relaciones de núcleo y satélite, según la relevancia de la información) para determinar las unidades textuales más importantes de un documento. Además, en Khan et al.(2005) el RST se combina dentro de un resumen genérico para agregar conocimiento lingüístico al proceso de resumen. Aunque los resultados obtenidos para este enfoque mixto no mejoraron los obtenidos por el resumen genérico, se afirmó que el inconveniente de este enfoque se basaba en el analizador que no podía detectar todas las relaciones RST; de lo contrario, el conocimiento lingüístico podría haber mejorado el resumen general. actuación. Además, en Cristea et al.(2005) se describe un enfoque similar a RST, que difiere de los anteriores en la ausencia de nombres de relación y el uso de árboles binarios. Este enfoque de resumen pretende explotar la coherencia y la cohesión de un documento.

La cohesión y la coherencia son dos de los principales desafíos para TS. Algunos enfoques se basan en la identificación de tales relaciones para mejorar la calidad de los resúmenes generados. Cunha et al.(2007) combina técnicas estadísticas y lingüísticas para demostrar que los resultados mejoran con respecto al uso de un solo tipo de técnicas. En Gonçalves et al.(2008), las cadenas de correferencia se utilizan para tratar los problemas de cohesión referencial que son frecuentes en el enfoque de resumen extractivo. Se desarrolla un sistema de postprocesamiento para reescribir expresiones referenciales de la forma más coherente posible, y se aplica después de generado el resumen, obteniendo mejoras considerables respecto a los resúmenes originales. Para garantizar la coherencia de un resumen, un enfoque generalizado es el uso de cadenas léxicas o de correferencia. Sin embargo, el uso de cadenas de correferencia no es novedoso en TS. Las primeras aproximaciones se encuentran en baldwin y morton(1998), y Azzam et al.(1999). La suposición principal es que la cadena de correferencia más larga indica el tema principal del documento y las cadenas más cortas representan subtemas. Por lo tanto, una estrategia posible para construir resúmenes es seleccionar solo aquellas oraciones relacionadas en la cadena más larga. Esta estrategia ayuda a mantener la coherencia del texto. Una idea similar es usar cadenas léxicas, que consiste en determinar secuencias de palabras relacionadas semánticamente (por ejemplo, por repetición de conceptos o relaciones de sinonimia). Mediante el uso de cadenas léxicas, también se pueden detectar los temas principales de un documento. Esta técnica también ha sido ampliamente utilizada en resúmenes, y enfoques como los descritos en Barzilay y Elhadad(1999), Medelyan(2007) o Ercan y Cicekli

(2008), explotarlos para producir resúmenes. Vale la pena mencionar que poder identificar todas las entidades que están conectadas dentro de un documento o entre documentos, evita que los resúmenes sean comunes. *anáfora colgante* fenómeno, produciendo así resúmenes resultantes más coherentes (Elsner y Charniak 2008). Este fenómeno consiste en tener palabras en un texto (principalmente pronombres) sin su antecedente correcto incluido en el resumen. Por ejemplo, si un resumen contiene el pronombre "él", pero no se menciona su antecedente (por ejemplo, el presidente de España), esto conduciría a un resumen poco claro con este tipo específico de problema, lo que dificultaría la redacción del resumen, entender, o incluso incoherente. Para reducir este problema, algunos enfoques combinan el uso de la resolución de anáforas para ayudar a TS (Orăsan 2004; Mitkov et al. 2007). En estos enfoques, los documentos se procesan primero para resolver los pronombres anafóricos y luego se ejecuta un sistema de resumen para producir un resumen y determinar si un sistema de resolución de anáforas mejora o no la calidad del resumen. Debido al rendimiento moderado de este tipo de sistemas, esto es difícil de lograr y, contrariamente a la intuición, TS no mejora mucho. Sin embargo, en Orăsan (2007) se simuló un sistema ideal de resolución de anáforas, resolviendo manualmente las relaciones anafóricas en documentos científicos, y se comprobó que en tales situaciones los resultados de resumen mejoran notablemente. En Steinberger et al. (2007), se afirmó que la mejora asociada a TS, al utilizar un sistema de resolución de anáforas, no solo depende del menor rendimiento del resolutor de anáforas, sino también de la forma en que se utilizan las relaciones anafóricas. En consecuencia, utilizaron las relaciones anafóricas desde dos perspectivas diferentes: por un lado, para mejorar la calidad de los resúmenes, y por otro, para comprobar la coherencia de un resumen, una vez que éste ya estaba generado comprobando si las cadenas de correferencia del resumen son subcadenas de las identificadas en los documentos fuente.

2.1.5 Enfoques basados en aprendizaje automático

Los enfoques que se explican a continuación se basan en algoritmos de aprendizaje automático para producir resúmenes. Los primeros métodos de aprendizaje automático utilizados en TS incluyen clasificadores binarios (Kupiec et al. 1995), *Modelos ocultos de Markov* (Conroy y O'leary 2001; Schlesinger et al. 2002), y *bayesiano* métodos (Aone et al. 1998). Sin embargo, se puede utilizar una amplia gama de técnicas de aprendizaje automático para TS. suma neta (Svore et al. 2007) apuesta por la síntesis de un solo documento y produce extractos de documentos de noticias basados en redes neuronales, utilizando RankNet (Burgess et al. 2005) como algoritmo de aprendizaje para puntuar las frases y extraer las más importantes. Además de las funciones comunes basadas en palabras clave y la posición de la oración, un nuevo conjunto de funciones basado en Wikipedia y los registros de consultas también se utilizan de manera que, por ejemplo, las oraciones que contienen términos de consulta o entidades de Wikipedia contienen contenido importante. En Schilder y Kondadadi (2008), se presenta un resumen de varios documentos centrado en consultas, llamado FastSum, donde las oraciones se clasifican utilizando una técnica de aprendizaje automático llamada *Regresión de vectores de soporte* (RVS), y *Regresión de ángulo mínimo* para la selección de características. SVR se usó en resumen antes, en el enfoque descrito en Li et al. (2007), donde se utilizaron características de entidades basadas en palabras, frases, semánticas, así como en posiciones de oraciones o nombres para entrenar el clasificador automáticamente. Más adelante, se combinaron las características extraídas y luego se puntuaron las oraciones. En Wong et al. (2008), se presenta un enfoque de resumen extractivo, empleando métodos de aprendizaje supervisado y semisupervisado. Las características de la oración involucradas se agrupan en diferentes tipos: características de superficie, contenido, relevancia y evento que incluyen la posición de la oración, el número de palabras en una oración, el centroide y los términos de alta frecuencia, o la similitud entre oraciones, entre otros. En cuanto al enfoque supervisado, un

Máquinas de vectores soporte(SVM), mientras que para el *enfoque semisupervisado*, se utiliza un *SVM probabilístico* y un *Bayesiano ingenuo* clasificador están *co-entrenados para explotar datos no etiquetados*. La técnica *SVM* también se usó en *Fuentes et al.(2007)* para *detectar información relevante para ser incluida en un resumen centrado en la consulta*, donde se utilizaron *características estructurales, basadas en la cohesión y dependientes de la consulta para el entrenamiento*.

La ventaja de usar el *aprendizaje automático para TS* es que *permite probar fácilmente el rendimiento de una gran cantidad de características*, por ejemplo, *léxicas, sintácticas, estadísticas, etc.*, utilizando diferentes *paradigmas de aprendizaje automático para aprender cuáles son los más adecuados*. Sin embargo, *estos enfoques también necesitan un gran corpus de entrenamiento para poder obtener resultados concluyentes*. Por lo general, *el corpus consiste en un conjunto de resúmenes escritos por humanos, o documentos fuente anotados que contienen qué oraciones son importantes para el resumen y cuáles no*.

2.2 Conferencias y talleres relevantes

A finales de los años 90, la *Evaluación de resumen de texto TIPSTER₂(SUMMAC)* fue la *primera conferencia destinada a evaluar los sistemas de resumen automático*, donde se *probaron los resúmenes de texto en la clasificación de documentos y la respuesta a preguntas, con el fin de analizar si eran sustitutos adecuados para documentos completos*. Puede encontrar una explicación detallada de este foro de evaluación y cómo se evaluaron los resúmenes en *Mani et al.(2002)*. Colección de pruebas del Instituto Nacional de Informática para IR₃(*NTCIR*) también desarrolló una *serie de talleres de Desafíos de resumen de texto (TSC)*, que incluyeron *tareas de resumen en japonés* en 2001 (TSC), 2002 (TSC2) y 2003 (TSC3). Además de estas conferencias, las conferencias importantes que se enfocaron solo en TS fueron las *Conferencias de Comprensión de Documentos₄(DUC)* que se llevaron a cabo anualmente desde 2001 hasta 2007. En estas conferencias se propusieron *diferentes tareas a lo largo de los años*, teniendo en cuenta los *nuevos desafíos y requisitos para TS*, *obligando* también a los sistemas a ser *dinámicos y adaptables*. A lo largo de las ediciones de las jornadas DUC, *se puede observar cómo han ido evolucionando los sistemas de resumen*, así como las diferentes *metodologías de evaluación* que se han propuesto para evaluar los correspondientes resúmenes. Estos cambiaron de una evaluación manual completa, donde los evaluadores utilizaron el entorno de evaluación SEE₅ para facilitar la comparación del contenido de resúmenes automáticos y hechos por humanos, a una evaluación completamente automática del contenido utilizando ROUGE (*Lin 2004*) y Elementos Básicos (*Hovy et al. 2006*).

Las tareas involucradas en las *conferencias DUC* también cambiaron a lo largo de las ediciones, *comenzando al principio con el resumen genérico de un solo documento y continuando con el resumen de múltiples documentos centrado en la consulta*. En *Sobre et al.(2007)*. Más específicamente, la descripción general de algunas ediciones de DUC también se proporciona en *Over y Ligget (2002)* y *Maldita sea(2006)*. Este tipo de conferencias son muy útiles *para evaluar y comparar sistemas automáticos* y, al mismo tiempo, también *proporcionan un buen conjunto de corpus*, que comprende *documentos y resúmenes de modelos*, que están *disponibles gratuitamente bajo demanda*.⁶Desafortunadamente, debido a que todas las ediciones trabajaron bajo el dominio de newswire, *los datos tratan de un dominio único*. Desde 2008, las conferencias DUC ya no se organizan, porque tienen

2http://www-nlpir.nist.gov/related_projects/tipster_summac/.

3<http://research.nii.ac.jp/ntcir/outline/prop-en.html>.

4<http://www-nlpir.nist.gov/projects/duc/>.

5<http://www.isi.edu/licensed-sw/see/>.

6<http://www-nlpir.nist.gov/projects/duc/data.html>.

Forma parte de la **Conferencia de Análisis de Texto (TAC)**, dentro del cual se incluye una pista de resumen. En **TAC 2008**, se propusieron **dos tareas diferentes dentro de la vía TS**. La **primera** seguía la misma idea que la tarea de resumen de actualización en el DUC 2007, que consistía en **construir resúmenes que contenían información actualizada respecto a un conjunto determinado de documentos de noticias**, mientras que la **segunda** era una tarea piloto cuyo objetivo era **generar resúmenes de opinión**, de blogs. En el **TAC 2009** se **mantuvo la tarea de resumen de actualizaciones** pero, por otro lado, en lugar de la tarea de resumen de opiniones, se propuso una **nueva tarea relativa a la evaluación automática de resúmenes** (*Evaluación automática de resúmenes de pares*). El objetivo de esta tarea es **puntuar automáticamente un resumen para una métrica determinada que refleje el contenido del resumen**. En la actual edición del TAC (2010) se mantiene la tarea de evaluación pero, por el contrario, los **resúmenes de actualización se han convertido en resúmenes guiados**. La idea bajo este nuevo tipo de resúmenes es **alentar a los sistemas a usar un conocimiento semántico más profundo**, construyendo resúmenes que contengan información específica sobre diferentes aspectos de un tema. Por ejemplo, **si un conjunto de documentos trata sobre un accidente, es posible que nos interese información sobre cuándo ocurrió, por qué, dónde, etc.**

Mesa2muestra algunas características de las tareas involucradas en las conferencias mencionadas anteriormente. Las **palabras en negrita indican las novedades introducidas en las tareas de resumen a lo largo de los años**. En la primera conferencia de síntesis, SUMMAC (**Mani et al. 1999**), se evaluaron resúmenes de un solo documento centrados en consultas de documentos de noticias. Para ello se propusieron dos tareas de evaluación extrínseca y otra intrínseca. La evaluación extrínseca juzga la calidad del resumen en función de cómo afecta la finalización de alguna otra tarea, mientras que la evaluación intrínseca mide un resumen por sí solo. Por un lado, en la evaluación extrínseca, una *ad hoc* se sugirió una tarea en la que se evaluaran resúmenes indicativos con respecto a si permitían determinar rápidamente la relevancia de un documento enfocado en un tema específico. Además, también se propuso una tarea de categorización, cuyo objetivo era determinar si los resúmenes genéricos podían presentar efectivamente suficiente información para permitir que una persona categorizara correctamente un documento. Por otro lado, en cuanto a la tarea de evaluación intrínseca (tarea de pregunta-respuesta), el objetivo era medir el contenido de un resumen con respecto a qué grado contenía respuestas a varias preguntas relacionadas con el tema. Las tareas involucradas en los talleres TSC dentro de las conferencias NTCIR (**Fukushima y Okumura 2001; Okumura et al. 2004; Hirao et al. 2005**) también se ocupó de la evaluación utilizando métodos intrínsecos y extrínsecos. En el primer TSC se propusieron tres tareas, consistentes en producir resúmenes con una extensión específica. Las diferencias entre ellos eran que en la primera tarea, solo debían extraerse las oraciones importantes, mientras que en la segunda, los resúmenes generados automáticamente se comparaban con los elaborados por humanos. La tercera tarea involucró la evaluación extrínseca y los resúmenes se evaluaron con fines de recuperación de información. Esta tarea fue muy similar a la tarea *ad hoc* de la conferencia SUMMAC. El resumen de varios documentos se incluyó por primera vez en TSC2, en el que se definían tareas de un solo documento y de varios documentos. Sin embargo, desde entonces, el resumen multidocumento se convirtió en un tema central y, en consecuencia, en las siguientes conferencias (TSC3), las tareas ya no estaban dirigidas a producir resúmenes a partir de un solo documento de entrada. A medida que pasa el tiempo, los sistemas evolucionan y también lo hacen los requerimientos de los resúmenes, de acuerdo con las necesidades de la sociedad. Los cambios en los requisitos y sistemas de resumen a lo largo del tiempo también se pueden ver en las conferencias de DUC. Al principio, las tareas propuestas estaban dirigidas a producir resúmenes genéricos a partir de uno o varios documentos de entrada, pero al final se prestó más atención a los resúmenes centrados en la consulta y la novedad. Un aspecto a destacar fue el intento de realizar resúmenes multilingües entre las tareas propuestas estaban dirigidas a producir resúmenes genéricos a partir de uno o varios documentos de entrada, pero al final se prestó más atención a los resúmenes centrados en la consulta y la novedad. Un aspecto a destacar fue el intento de realizar resúmenes multilingües entre las tareas propuestas estaban dirigidas a producir resúmenes genéricos a partir de uno o varios documentos de entrada, pero al final se prestó más atención a los resúmenes centrados en la consulta y la novedad. Un aspecto a destacar fue el intento de realizar resúmenes multilingües entre

⁷<http://www.nist.gov/tac/>.

⁸<http://www.nist.gov/tac/2009/summarisation/aesop.09.guidelines.html>.

Tabla 2Características de la tarea de resumen para cada conferencia

Conferencia	Requisitos de la tarea de resumen
SUMMACa	Documento único, centrado en consultas, noticias
TSCb (NTCIR)	Centrado en consultas, genérico, noticias
TSC2 (NTCIR)	Soltero y multi-documento , genérico, noticias
TSC3 (NTCIR)	Multidocumento, genérico, noticias
DUC-01c	Único y multidocumento, genérico, noticias Único y
DUC-02	multidocumento, genérico, noticias
DUC-03	Multidocumento, centrado en la consulta , noticias
DUC-04	Documentos únicos y múltiples, orientados a temas, noticias, multilingüe
DUC-05	Multidocumento, centrado en consultas, noticias Multidocumento, centrado en
DUC-06	consultas, noticias Multidocumento, actualizar , centrado en la consulta, noticias
DUC-07	
TAC-08d	Multidocumento, actualización, centrado en consultas, basado en sentimientos , noticias & blogs
TAC-09	Multidocumento, actualización, centrado en consultas, noticias, evaluación Multi-documento,
TAC-10	guiado , centrado en la consulta, noticias, evaluación

una conferencia de evaluación de resúmenes de texto SUMMAC y tipos de resúmenes TIPSTER

b**TSC**Desafíos de resumen de texto

c**DUC**conferencias de comprensión de documentos

d**TAC**conferencias de análisis de texto

Inglés y chino en DUC (2004). Esta tarea consistió en producir resúmenes muy breves o breves en inglés a partir de un conjunto de documentos que previamente fueron traducidos automáticamente al inglés (su idioma original era el chino). El concepto de novedad e información novedosa se abordó por primera vez en la tarea de actualización de DUC (2007). El objetivo de esta tarea era generar un resumen a partir de un grupo de documentos relacionados, pero teniendo en cuenta que algunos de esos documentos ya habían sido leídos por los usuarios, por lo que la información contenida en ellos no necesitaba aparecer en el resumen. En cuanto al dominio de los documentos, las conferencias de la DUC también estuvieron enfocadas en los documentos de noticias. Por el contrario, además de los documentos noticiosos, se introdujo una fuente de información muy conocida en Internet, a saber, los blogs, a tratar,⁹Sin embargo, debido a la dificultad que entraña la propia tarea y el tipo de datos (blogs), la tarea de resumen de opinión quedó fuera del alcance del TAC (2009), manteniéndose únicamente la tarea de resumen de actualización e introduciendo una nueva tarea relativa a la evaluación automática de resúmenes. Finalmente, como se ha dicho, en TAC (2010) se mantiene la tarea de evaluación, mientras que se introduce un nuevo tipo de resúmenes, los resúmenes guiados.

En cuanto al número de grupos participantes, Tabla**3**muestra este número para la conferencia de cada edición. Vale la pena darse cuenta de cómo este número ha aumentado a lo largo de los años, lo que demuestra el creciente interés en el campo de investigación de TS. Como consecuencia, se están organizando talleres específicos centrados únicamente en TS dentro de conferencias importantes, como el Taller sobre extracción y resumen de información multilingüe y de fuentes múltiples.¹⁰(MIMIES), el Taller de Generación y síntesis de Lenguas¹¹(UCNLG+Suma), el Taller

⁹<http://www.nist.gov/tac/tracks/2008/summarisation/index.html>.

¹⁰<http://doremi.cs.helsinki.fi/mmies2/>.

¹¹<http://www.nltg.brighton.ac.uk/ucnlg/ucnlg09/>.

Tabla 3Número de grupos de participantes para cada conferencia

Nombre de la conferencia (Año)	Número de grupos de participantes
SUMMAC (1999)	dieciséis
TSC (2001)	9
TSC2 (2002)	8
TSC3 (2003)	9
DUC (2001)	15
DUC (2002)	17
DUC (2003)	21
DUC (2004)	22
DUC (2005)	31
DUC (2006)	34
DUC (2007)	32
TAC (2008)	38
TAC (2009)	39

en Resumen y presentación de resultados de búsqueda web (WSSP),¹²o el 1er Taller Internacional sobre Descubrimiento, Resumen y Uso de Clusterings Múltiples.¹³

3 Resumen del texto en el contexto actual

Esta sección se divide en dos subsecciones. Por un lado, en la Secc.3.1, y por otro lado, TS para nuevos escenarios también se explica brevemente en la Secc.3.2. Aparte de los tipos de resúmenes clásicos, como resúmenes de un solo documento o de varios documentos, genéricos o centrados en consultas, etc., existen varios tipos de resúmenes novedosos e interesantes, donde se persiguen objetivos específicos (por ejemplo, resúmenes basados en sentimientos). Además de este tipo de resúmenes, la aparición de nuevos escenarios también es interesante para TS. En lugar de realizar investigaciones en los mismos conjuntos de datos basados tradicionalmente en noticias o documentos científicos, en los últimos años se ha prestado mucha atención a dominios novedosos, como textos literarios, patentes o blogs. Por ello, también los abordamos en este apartado.

3.1 Nuevos tipos de resúmenes

En lo que resta de esta subsección, se describirán diferentes tipos de resúmenes que han aparecido recientemente. Vale la pena enfatizar el hecho de que los tipos de resúmenes que se describen a continuación se enfocan principalmente en las necesidades del usuario o intentan manejar de manera eficiente grandes cantidades de información. Respecto a los primeros, hemos seleccionado resúmenes personalizados, actualizados y basados en sentimientos, ya que su objetivo común es realizar un resumen cuyo contenido viene determinado directamente por los requisitos del usuario (el usuario debe delimitar qué tipo de información le interesa). Con respecto a estos últimos, se analizan las encuestas y los resúmenes abstractivos, porque representan dos buenos ejemplos de resúmenes que deben construirse empleando

¹²<http://www.wssp.info/2009.html>.
¹³<http://eecs.oregonstate.edu/research/multiclust/>.

técnicas que van más allá de la mera concatenación de oraciones, y actualmente parece que esa es la tendencia y objetivo final de los sistemas TS.

3.1.1 *Resúmenes personalizados*

El objetivo de este tipo de resúmenes es proporcionar un resumen que contenga la información específica que le interesa a un usuario. Esto significa que diferentes usuarios pueden tener diferentes necesidades, por lo que los sistemas de resúmenes tienen que determinar el perfil del usuario antes de seleccionar la información relevante que será incluirse en el resumen final.

En (Agnihotri et al. 2005) el perfil del usuario se determina mediante un método de mapeo estadístico a partir de los rasgos de personalidad de los usuarios identificados mediante pruebas de personalidad al contenido. El análisis realizado sobre 59 usuarios mostró que solo algunos rasgos como el género y solo algunas características (por ejemplo, el texto) fueron de ayuda para personalizar el resumen. El principal inconveniente de este enfoque son los datos limitados que se utilizan para la configuración experimental y, debido a la dificultad de la tarea, los experimentos en otro entorno son muy difíciles de ampliar o replicar. Respecto también a los resúmenes personalizados, en Díaz y Gervas(2007) se propone un enfoque para producir resúmenes de noticias que contengan información relevante para un perfil de usuario determinado. La idea es seleccionar aquellas oraciones que son más relevantes para un modelo de usuario dado. Esto se hace calculando la similitud entre el modelo de usuario de un individuo concreto y cada una de las frases del documento, por lo que dependiendo de qué parte del modelo se elija para calcular la similitud, se pueden obtener varios resúmenes personalizados posibles. El modelo de usuario está determinado por la combinación de características de dominio específicas, un conjunto de palabras clave, que refleja las necesidades de información que no cambian con el tiempo, y un nivel de retroalimentación de relevancia, que tiene en cuenta los cambios proporcionados por los usuarios. comentario. El extenso conjunto de experimentos realizados demostró la idoneidad de este tipo de resúmenes,

En Kumar et al.(2008), se generan resúmenes personalizados en función del área de especialización y los intereses personales de un usuario. Con este fin, se desarrolla un modelo de antecedentes de usuario tomando como base la información que se encuentra en Internet con respecto a una persona, como su página Web personal, o publicaciones en línea. Una vez identificado el perfil de usuario, se determina la relevancia de las frases del documento en función de este perfil. Se proponen dos funciones de puntuación, una para información genérica y otra específica para el usuario. El primero se basa en la frecuencia de los términos para extraer las oraciones genéricas más relevantes, mientras que el segundo calcula la probabilidad de que las oraciones genéricas contengan también información específica del usuario. Luego, en la etapa de generación de resumen, se seleccionan y extraen las oraciones mejor clasificadas. Aunque este enfoque es muy interesante, su principal dificultad es que la desambiguación de la entidad del nombre se debe realizar cuando se busca información específica de una persona en Internet. Esto sería esencial porque puede suceder que las personas compartan el mismo nombre pero sean totalmente diferentes (por ejemplo, George Bush podría referirse tanto al presidente de EE. UU. de 1989 a 1993 como al presidente de EE. UU. de 2001 a 2009 (su hijo)).

En (Berkovski et al. 2008), se lleva a cabo una evaluación preliminar del usuario para evaluar diferentes aspectos de las actitudes de los usuarios hacia los TS personalizados. Se sugieren tres experimentos con el propósito de analizar este tema. En primer lugar, se evalúa si la personalización de los resúmenes tiene o no el efecto deseado en los usuarios. Luego, se analiza el impacto de la longitud de los resúmenes. Finalmente, se valora el grado de fidelidad entre los resúmenes personalizados y los documentos originales. Las conclusiones derivadas del análisis son muy preliminares. Se demuestra que cuanto más información personalizada contenga un resumen, mejor será preferido. También se afirma que los usuarios prefieren resúmenes ni demasiado cortos ni demasiado largos; sin embargo, no se dan pistas sobre cuál debería ser la longitud óptima.

3.1.2 Actualizar resúmenes

El resumen de actualizaciones intenta generar resúmenes teniendo en cuenta que los usuarios tienen un conocimiento previo del tema sobre el que quieren leer, por lo que solo están interesados en los eventos más recientes relacionados con ese tema. Este tipo de resúmenes surgió gracias a la tarea propuesta en las conferencias DUC y TAC.

Para generar resúmenes de actualización, el enfoque descrito en [Sweeney et al.\(2008\)](#) consiste en incorporar novedad a los resúmenes, minimizando la superposición de contenido entre una oración de resumen y una potencial candidata. En [Witte et al.\(2007\)](#), [Bellemare et al. \(2008\)](#), [Li et al.\(2009\)](#) los resúmenes de actualización se construyen sobre la base de estructuras de datos de gráficos de conglomerados, que se basan en el contexto y en el conjunto de documentos que se van a resumir. Se propone un esquema de clasificación de oraciones según la superposición entre las oraciones de los grupos y el contexto, por lo que finalmente se establecen rangos y se generan resúmenes seleccionando oraciones de cada rango. El enfoque sugerido en [Li et al.\(2008\)](#), define el concepto de historia (aquellos documentos ya conocidos por un lector), e introduce un nuevo tipo de características (características de filtrado), que reflejan que el sumario se resume con su historia. Por lo tanto, en tales casos, las funciones de filtrado se pueden calcular a través de dos métricas de similitud diferentes para excluir aquellas oraciones que son similares a la historia. Una de estas métricas de similitud se basa en la fórmula del coseno de la distancia, mientras que la otra toma en consideración unigramas, bigramas y funciones sintácticas de las palabras y las combina linealmente para obtener finalmente una métrica de similitud.

Los algoritmos de aprendizaje automático también se aprovechan para generar resúmenes de actualización. [Schilder et al.\(2008\)](#) se basa en el sistema de resumen FastSum ([Schilder y Kondadadi 2008](#)) que utiliza SVM, pero también se tienen en cuenta las características relacionadas con información nueva y antigua, como entidades nuevas/antiguas, frecuencia de documentos/palabras nuevas/antiguas. Tales características penalizan oraciones que son similares a las de las previamente seleccionadas. Además, en [Fischer et al.\(2009\)](#), las métricas de similitud se usan como características dentro de un paradigma de aprendizaje automático supervisado, un clasificador de perceptrones, en lugar de usarse para clasificar oraciones directamente. Junto con estas características, también se emplea un segmentador de discurso para determinar las posibles unidades suborales que también se incluirán en el resumen. Ambos enfoques obtuvieron buenos resultados, manteniéndose en la mitad del ranking entre todos los participantes.

En enfoques recientes, como [Liu et al.\(2009\)](#) y [Nastase et al.\(2009\)](#), la información de fondo está tomada de artículos de Wikipedia. Por un lado, en el enfoque anterior, Wikipedia se usa para producir un resumen, tomando el primer párrafo de la entrada relacionada con un tema, y luego calculando la similitud entre las posibles oraciones de resumen con las que ya están contenidas en el texto basado en Wikipedia. Las oraciones con menor similitud serán seleccionadas para el resumen de actualización. Por otro lado, este último utiliza Wikipedia para recuperar conceptos que se discuten en el conjunto de documentos a resumir. De esta forma, es posible predecir qué conceptos tienen más probabilidades de encontrarse en resúmenes bien formados. Además del conocimiento de Wikipedia, este enfoque también realiza la compresión de oraciones para dar al resumen final una naturaleza abstracta.

3.1.3 Resúmenes basados en opiniones

En los últimos años, la subjetividad que aparece en los documentos ha dado lugar a un nuevo tipo de resúmenes emergentes: los resúmenes basados en sentimientos, que deben tener en cuenta el sentimiento que una persona tiene hacia un tema, producto, lugar, servicio, etc. En consecuencia, TS y Sentiment Analysis (SA), también conocido como minería de opiniones, deben combinarse para

producir este tipo de resúmenes. SA proporciona el sentimiento asociado a un documento en diferentes niveles (documento, fragmento, oración o incluso nivel de palabra) (Pang y Lee 2008), mientras que TS identifica las partes más relevantes de un documento y construye a partir de ellas un fragmento de texto coherente (el resumen). En cuanto a los resúmenes basados en sentimientos, las opiniones deben ser detectadas y clasificadas primero, según su subjetividad (si una oración es objetiva o subjetiva, por ejemplo), y luego según su polaridad (positiva, negativa o neutra). Más adelante, TS se encarga de determinar qué oraciones se incluirán en el resumen y generar el resumen final. Los sistemas de resumen basados en sentimientos que participaron en el *Tarea piloto de resumen de opinión* de la conferencia TAC 2008, como Conroy y Schlesinger (2008), El et al. (2008), Balahur et al. (2008), o Bossard et al. (2008) sigue estos pasos.

Sin embargo, fuera del alcance de la competencia TAC, también se pueden encontrar otros enfoques interesantes. por ejemplo, en Beineke et al. (2004) se utilizan algoritmos de aprendizaje automático para determinar qué oraciones deben pertenecer a un resumen, después de identificar posibles tramos de texto de opinión. Las características que resultaron útiles para ubicar citas de opinión dentro de un texto incluyeron la ubicación dentro del párrafo y el documento, y el tipo de palabras que contenían. Del mismo modo, en Zhuang et al. (2006) se identifican las características relevantes (por ejemplo, guión, actores para una película) y las palabras de opinión y su polaridad (ya sea un sentimiento positivo o negativo) y luego, después de identificar todos los pares válidos de característica-opinión, se produce un resumen, pero centrándose únicamente en las reseñas de películas. Normalmente, las reseñas en línea contienen también calificaciones numéricas que los usuarios otorgan cuando brindan una opinión personal sobre un producto o servicio. El enfoque descrito en Titov y McDonald (2008) propuso una *Sentimiento multiaspecto* modelo. Este modelo estadístico utiliza clasificaciones de aspecto para descubrir los temas correspondientes y extraer fragmentos de texto. Además, en Lerman y McDonald (2009), se sugiere un enfoque para producir resúmenes contrastivos en el dominio de las opiniones de los consumidores. El resumen contrastivo se refiere al problema de generar un resumen para dos entidades con el fin de resaltar sus diferencias, por ejemplo, los sentimientos de diferentes personas sobre varios productos. Para producir este tipo de resúmenes adaptan el *Coincidencia de aspecto de sentimiento* modelo descrito en Lerman et al. (2009), diseñado originalmente para generar resúmenes basados en la opinión de un solo producto. Este modelo determina qué oraciones extraer comparando el sentimiento promedio de una oración con respecto al sentimiento promedio de la entidad específica, seleccionando así las más cercanas.

3.1.4 Resúmenes de encuestas

Este tipo de resúmenes tiene como objetivo proporcionar una visión general de un tema o entidad en particular. Por lo general, son más largos que cortos, porque intentan capturar los hechos más importantes relacionados con una persona, por ejemplo. A continuación, nos enfocamos en resúmenes biográficos, resúmenes de encuestas y artículos de Wikipedia como tres de los tipos de TS más recientes que se pueden categorizar dentro de este grupo.

El desafío de producir resúmenes a partir de biografías se presentó en Zhou et al. (2004). La idea detrás del resumen biográfico de varios documentos es producir un texto que contenga los aspectos más relevantes de una persona específica, respondiendo preguntas, como “¿Quién es Barack Obama?” por ejemplo. Para realizar esta tarea, se utilizan varios algoritmos de aprendizaje automático (*bayesiana ingenua*, SVM y *Árboles de decisión*) para clasificar oraciones. Además, la información redundante se elimina en una etapa posterior. Un sistema de resumen biográfico similar que utiliza también técnicas de aprendizaje automático se describe en Biadys et al. (2008). La diferencia con el anterior es que se usa un clasificador binario para discriminar entre oraciones biográficas y no biográficas, y luego se entrena un modelo de regresión SVM para reordenar oraciones biográficas extraídas usando la Wikipedia como corpus. La etapa final de este enfoque es emplear una heurística de reescritura para crear los resúmenes finales.

Otro enfoque interesante para este tipo de TS es el uso de citas de artículos. En [Kan et al.\(2002\)](#) se demostró que a partir de las entradas bibliográficas era posible producir un resumen indicativo. La idea principal detrás de esta suposición es que dichas entradas contienen información tanto informativa como indicativa, por ejemplo, detalles sobre el recurso o metadatos, como el autor o el propósito del artículo). En su investigación, se desarrolla un gran corpus anotado (2000 entradas anotadas) para tales fines. Siguiendo la idea de generar resúmenes a partir de esta información de entrada, en [Qazvinian y Radev\(2008\)](#) las citas se analizan para producir un resumen de un solo documento a partir de artículos científicos. El objetivo final es generar resúmenes sobre un tema específico. Además, el trabajo descrito en [Mohamed et al.\(2009\)](#) aborda este tema y, en consecuencia, presenta algunos experimentos preliminares de la utilidad del texto de cita para generar automáticamente encuestas técnicas. Se utilizan tres tipos de entrada (artículos completos, resúmenes y textos de citas), y se tienen en cuenta los sistemas de resumen ya existentes para crear tales encuestas, por ejemplo, Lex-Rank ([Erkan y Radev 2004](#)), recortadora ([Zajic et al. 2007](#)) y C-RR y C-LexRank ([Qazvinian y Radev 2008](#)). Entre las conclusiones extraídas de los experimentos, se demostró que la creación de encuestas técnicas de documentos múltiples se beneficia considerablemente de los textos de citas.

A diferencia de estos enfoques, ([Sauper y Barzilay 2009](#)) sugieren la creación automática de artículos de Wikipedia utilizando plantillas específicas de dominio que se inducen a partir de documentos generados por humanos. Para producir tales artículos, se emplea un motor de búsqueda para recuperar documentos relacionados con un tema, que se consideran como entrada del proceso de resumen. Siguiendo la misma estructura de un artículo de Wikipedia, la información apropiada para cada sección está determinada por técnicas de aprendizaje automático, extractos de entrenamiento basados en cuán representativos son para un tema seleccionado.

3.1.5 Resúmenes abstractivos

Para simplificar el problema del resumen, la mayoría de los enfoques siguen un paradigma extractivo, generando las oraciones más relevantes de un documento o documentos sin realizar ningún cambio. Aunque es un método muy extendido, los resúmenes resultantes suelen presentar varios problemas con respecto a su calidad, como la falta de coherencia o “*anáfora colgante*”. El paradigma abstractivo puede solucionar estas limitaciones, ya que intenta producir nuevos textos a partir de los fragmentos de información o conceptos identificados como relevantes. A pesar de no ser un tema novedoso, en los últimos años se ha incentivado la investigación en abstracción, debido a que la información se repite a lo largo de los documentos, y se requieren formas específicas de transmitir y presentar la información. Métodos como la compresión de oraciones ([Zajic et al. 2007](#)), fusión de oraciones ([Barzilay y McKeown 2005](#)) o generación de lenguaje natural ([Radev y McKeown 1998](#)) se han aplicado tradicionalmente para la generación de resúmenes.

Además, se han realizado varios análisis para comprender cómo los humanos resumen ([Jing y McKeown 2000](#); [Jin 2002](#)). En consecuencia, se analizan las operaciones básicas para transformar la información fuente en información resumida. Por ejemplo, ([Hasler 20007](#)) afirma que la técnica que hacen los humanos es copiar y pegar el mismo material presente en los documentos fuente. Sin embargo, se aplican algunos cambios leves en la mayoría de los casos, y se identifican dos tipos de operaciones, atómicas y complejas, que involucran eliminación, inserción, reemplazo, reordenación o fusión (las dos primeras son operaciones atómicas mientras que las tres últimas son complejas). De la evaluación realizada en términos de coherencia, los resultados mostraron que el 78% de los resúmenes fueron más coherentes que los extractos. En ([Fiszman et al. 2004](#)) se propone un enfoque para generar resúmenes a partir de documentos biomédicos. La idea principal es identificar

predicados semánticos usando SemRep¹⁴ luego producir el resumen de forma esquemática. El proceso de resumen comprende la identificación de dichos predicados y la conectividad entre ellos. Más adelante, la novedad y la prominencia de cada predicción se calcula en función de los recuentos de frecuencia de términos. Saggion(2009) sugiere un enfoque novedoso para combinar diferentes fragmentos de información que han sido extraídos de uno o más documentos. A partir de un vocabulario predefinido (p. ej. *dirigirse, indicar, para reportar*, etc.) el algoritmo es capaz de decidir cuál de estas expresiones es más apropiada para una oración, dependiendo del contenido y del resumen parcial generado. La motivación de esta investigación es estudiar en qué medida la adición de información adicional que no está presente en los documentos fuente es útil y beneficia el proceso de abstracción. Usando técnicas de aprendizaje automático y experimentando con diferentes tipos de clasificadores (por ejemplo, *Árboles de decisión*), los resultados mostraron que el mejor clasificador, basado en características de resumen, incluyendo información lingüística, semántica, cohesiva, discursiva o posicional, es capaz de predecir correctamente el 60% de los casos.

La generación de lenguaje natural (NLG) también se aplica para producir resúmenes abstractos. En (Yu et al. 2007) se producen resúmenes muy breves a partir de grandes colecciones de datos numéricos. Los datos se presentan en forma de tablas y se genera un nuevo texto para describir los hechos que dichos datos representan. En primer lugar, los datos deben analizarse y comprenderse antes de generar las descripciones. En particular, para el último paso se utiliza un módulo NLG, que tiene en cuenta específicamente tres tipos de información a generar: antecedentes, descripción general y patrones más significativos encontrados en la recopilación de datos. Belz(2008) también propusieron un enfoque TS basado en NLG para generar pronósticos meteorológicos automáticamente, pero centrándose principalmente en la etapa NLG.

Otros enfoques abstractos se basan en el uso de plantillas para estructurar la información que se ha identificado previamente, por ejemplo, utilizando un sistema de extracción de información. Kumar et al.(2009) intenta generar informes a partir de la información del evento almacenada en bases de datos de diferentes dominios (biomédico, deportivo, etc.). Los resúmenes escritos por humanos se utilizan para determinar la información que se incluirá en un resumen, donde se generan algunas plantillas y se identifican en los textos los patrones para completar dichas plantillas. Del mismo modo, en Carenini y Cheung(2008) también se identifican patrones, pero dado que el objetivo es generar resúmenes contrastivos, los marcadores discursivos que indican contraste como *“aunque”, “sin embargo”,* etc. también se agregan para que el resumen suene más natural.

3.2 Nuevos escenarios para resúmenes de texto

Aunque la mayor parte del trabajo en TS se ha centrado tradicionalmente en las noticias (Gotti et al. 2007; Nenkova et al. 2005; Nenkova 2005), documentos científicos (Jaoua y Hamadou 2003; Teufel y Moens 2002), o incluso documentos legales (Saravanan et al. 2006; Cesarano et al. 2007), estos no son los únicos escenarios en los que se han probado los enfoques TS. A continuación, se describen varios escenarios nuevos en los que también se ha aplicado TS. Del análisis de TS en tales escenarios, vale la pena destacar el hecho de que, aunque la naturaleza de los documentos es totalmente diferente entre dominios, y puede parecer que cada dominio necesitaría una forma diferente de abordar el proceso de TS, en la práctica, las técnicas empleadas no experimentan grandes cambios. Se puede ver que las características estadísticas o posicionales son las preferidas. En algunos casos, se agrega vocabulario específico, pero generalmente se realizan cambios menores para adaptar los enfoques de TS a otros escenarios. Por lo tanto, puede surgir la siguiente discusión: *“¿Sería mejor desarrollar sistemas genéricos para una amplia gama de escenarios, aunque con un rendimiento moderado, o construir sistemas muy específicos que pudieran obtener un mayor rendimiento?”*. Sobre

¹⁴<http://skr.nlm.nih.gov/papers/index.shtml#Sem\discretionary-Rep>.

Por otro lado, dado que algunas características, como la frecuencia de términos o la frecuencia de documentos inversos, pueden considerarse independientes del dominio, un enfoque apropiado podría ser combinar este tipo de características con las específicas del dominio dentro del mismo proceso TS. Esto permitiría que, para cada dominio, el proceso pudiera beneficiarse de cuestiones como el vocabulario específico o la estructura de los documentos, aumentando así su desempeño en el escenario específico respecto a un TS genérico por sí solo.

3.2.1 Texto literario

Intenta resumir textos literarios, ya sean cuentos (Kazantseva 2006) o textos más largos, es decir, libros (Mihalcea y Ceylan 2007) también se han abordado en los últimos años. En Mihalcea y Ceylan (2007), se analizan las dificultades del TS cuando se dirige a la síntesis de libros, construyendo un benchmark, donde se apunta específicamente a la evaluación de las síntesis de libros. Además, también se abordan varias técnicas para el resumen de libros, por ejemplo, la segmentación del texto, lo que sugiere un enfoque de resumen basado en el sistema MEAD ya existente (Radev et al. 2001), con algunos cambios particulares, para adaptar el sistema a la síntesis de documentos extensos. Más adelante, en Ceilán y Mihalcea (2009), se generan dos tipos de resúmenes: resúmenes objetivos e interpretativos. El primero contiene información sobre los eventos que ocurren en los libros y su trama, mientras que el segundo intenta capturar las ideas y pensamientos del autor. A partir de este análisis, se encuentra que aproximadamente el 48% de los resúmenes objetivos pueden reconstruirse mediante operaciones de cortar y pegar a partir del documento original. Sin embargo, para los resúmenes interpretativos, este número se reduce a solo el 25%. Para cuentos (Kazantseva 2006), se generan resúmenes indicativos para ayudar al usuario a decidir si lee o no el documento completo. La información relevante a incluir se determina en función de los rasgos lingüísticos, como los tiempos gramaticales, las expresiones temporales, la voz, el significado del verbo y el tipo de discurso (directo o indirecto).

3.2.2 Reivindicaciones de patentes

El estilo particular de escritura de las patentes, aunque difícil de procesar debido al tipo de lenguaje empleado, también ha sido objeto de TS. Un enfoque interesante que funciona a aproximadamente el 60% (medida F) se puede encontrar en Mille y Wanner (2008) donde se desarrolla un sistema de síntesis multilingüe para español, inglés y francés aprovechando la estructura de las reivindicaciones de patentes y empleando características discursivas y semánticas así como patrones de dependencia para el proceso de síntesis. También realizan simplificaciones lingüísticas para dar a los resúmenes resultantes un carácter abstractivo. En Tseng et al. (2007). Con respecto a TS, el proceso extractivo clasifica las oraciones en función de la frecuencia de las palabras clave, la similitud con el título del reclamo de patente y las palabras clave que contiene. Además, se consideran las características posicionales. Finalmente, todas estas características se combinan de forma lineal y las oraciones de mayor peso, hasta la longitud deseada, se seleccionan para formar el resumen final. Las frases clave también se identifican y utilizan como característica para determinar la relevancia de una oración en Trapey y Trapey (2008). Además, se emplean técnicas de agrupamiento para obtener la densidad de información de una oración. Sin embargo, a diferencia de los enfoques anteriores, la novedad de este enfoque se basa en la incorporación de características específicas del dominio basadas en frases y oraciones temáticas de un documento de patente determinado. Trapey et al. (2009) amplían el trabajo previo con conocimiento ontológico para recuperar las palabras clave y frases específicas del dominio utilizando jerarquías de conceptos y relaciones semánticas. Los resultados se evalúan en términos de relaciones de comprensión y retención.

Por un lado, el ratio de compresión es la relación del número de palabras entre el resumen y su documento original. Por otro lado, el índice de retención indica el valor promedio del índice de recuperación y el índice de precisión. Los resultados muestran que la mejor relación de compresión es del 20%, que es la línea de última generación (Morris et al. 1992), y el uso de ontologías mejora la tasa de retención.

3.2.3 Leyenda de imágenes

La necesidad de producir descripciones breves de imágenes también puede verse como un problema de TS, donde se produce un resumen a partir de un conjunto de documentos relacionados que se refieren a una imagen anotada con información geográfica. En (Deschacht y Moens 2007) se producen leyendas de imágenes utilizando la información asociada relacionada con una imagen. Aprovechan el contexto inmediato de la imagen para extraer dicha información, por ejemplo, texto en etiquetas HTML. Su objetivo principal es detectar y clasificar correctamente las entidades que aparecen en las imágenes y luego calcular la prominencia de dicha entidad con el objetivo final de producir una breve anotación para la imagen. Similarmente, Feng y Lapata (2008) sugieren un modelo de anotación de imágenes que es capaz de aprender leyendas de imágenes a partir de documentos auxiliares y anotaciones ruidosas. Los documentos auxiliares son muy útiles para esta tarea ya que pueden brindar información importante relacionada con la imagen, permitiendo así generar descripciones más precisas de la imagen. Aker y Gaizauskas (2009) proponen un enfoque basado en modelos lingüísticos (n-gramas) para generar resúmenes de 250 palabras para pies de imagen utilizando los corpus descritos en Aker y Gaizauskas (2010). Los resultados obtenidos son muy esperanzadores, siendo mejorados posteriormente mediante patrones de dependencia (Ahmet y Gaizauskas 2010), que se comportaron muy cerca de los obtenidos utilizando el primer párrafo de los artículos de Wikipedia como línea base de resumen. Además, en Plaza et al. (2010), dos enfoques de TS basados en características estadísticas (frecuencia de términos y longitud de sintagmas nominales) y gráficos semánticos que utilizan conceptos de WordNet, también se prueban dentro del mismo corpus. Los resultados para ambos enfoques fueron aceptables, obteniendo alrededor de un 10% de recuerdo según la métrica ROUGE-SU4, y mejorando el enfoque de modelos de lenguaje propuesto originalmente en Aker y Gaizauskas (2009).

3.2.4 Géneros textuales de la Web 2.0

Resúmenes de nuevos géneros textuales, como blogs (Balahur-Dobrescu et al. 2009; Lloret et al. 2009), reseñas (Balahur y Montoyo 2008; Zhuang et al. 2006) o hilos (Zajic et al. 2008; Balahur et al. 2009) también se puede encontrar en la literatura. Las técnicas de resumen utilizadas dentro de estos enfoques están en la línea de las presentadas en la Secc. 3.1.3, siendo fundamental la integración de técnicas de análisis de sentimiento para generar resúmenes a partir de estos nuevos géneros nacidos con la Web social. Centrándonos en TS, en Balahur et al. (2009), las técnicas empleadas se basan en recuentos de frecuencia de términos, mientras que en Balahur-Dobrescu et al. (2009) los resúmenes se generan utilizando *Análisis semántico latente*. Otros enfoques, por ejemplo Zhuang et al. (2006) simplemente confían en la salida del sistema SA para agrupar oraciones según su polaridad sin preocuparse por ninguna técnica TS.

4 Combinación de resúmenes de texto con sistemas inteligentes

El objetivo de esta Sección es presentar cómo los resúmenes pueden ayudar a otros sistemas, analizando así la aplicabilidad de TS dentro de otros sistemas inteligentes. Esto se puede considerar como

una forma de evaluar indirectamente los resúmenes, también conocida como evaluación extrínseca. El rápido crecimiento de la información conduce a un aumento en el tiempo, cuando se quiere tratarla de manera eficiente. Por lo tanto, los resúmenes pueden ser una buena manera de permitir que los sistemas dediquen menos tiempo de procesamiento, si se utilizan en lugar del documento completo. Además, al mismo tiempo, los resúmenes pueden ser adecuados para eliminar información ruidosa, manteniendo así solo la realmente importante. Estos aspectos se derivan de la propia definición de resumen, donde un resumen es una representación breve pero precisa del contenido de un documento o de un conjunto de ellos. La razón por la cual esta Sección, aunque relacionada de alguna manera con la evaluación, es independiente de la Sección 5 es que aquí nos interesa más profundizar y destacar la utilidad de TS para otros sistemas, como los que se explican a continuación. En particular, nos enfocamos en la recuperación de información, respuesta a preguntas y clasificación de textos.

4.1 Combinación de resúmenes de texto con recuperación de información

El objetivo de la Recuperación de información (IR) es encontrar material (generalmente documentos) de naturaleza no estructurada (generalmente texto) que satisfaga una necesidad de información dentro de grandes colecciones (generalmente almacenadas en computadoras) (Manning et al. 2008). TS se ha combinado con IR desde una doble perspectiva. Por un lado, varios enfoques utilizan resúmenes para beneficiar a IR, por ejemplo, en la etapa de indexación, mejorando el tiempo de recuperación de documentos y su rendimiento. Por otro lado, otros enfoques toman como entrada para TS la salida de los sistemas IR (es decir, los documentos recuperados por el sistema IR) y, en algunos casos, se proporcionan resúmenes en lugar de fragmentos tradicionales como salida del sistema IR. por ejemplo, en Kan y Klavan(2002), los resúmenes se emplean para presentar una visualización alternativa de los documentos provenientes de un marco IR estándar. Además, se analiza enKaisser et al.(2008), concluyendo que la longitud preferida por los usuarios depende del tipo de consulta. Sin embargo, el enfoque más común es combinar IR y TS de la siguiente manera: primero se recuperan los documentos relacionados con un tema y luego se genera un resumen que tiene en cuenta estos documentos. Por lo tanto, IR ayuda a recopilar solo los documentos relevantes para una consulta, mientras que TS selecciona la información más importante de ellos.radev y ventilador(2000) propone un resúmenes de documentos múltiples independiente del dominio, que genera resúmenes a partir de los resultados de búsqueda en la Web. Del mismo modo, DULCE (Steinberger et al. 2008) se basa en un motor de búsqueda para recuperar documentos relevantes para una consulta de la Web, y luego técnicas de resumen basadas en*Análisis semántico latente* se utilizan para identificar y extraer las frases más importantes de los documentos recuperados utilizando, al mismo tiempo, la similitud del coseno para evitar la redundancia en los resúmenes finales. El sistema QCS (Dunlavy et al. 2007) también integra un módulo IR pero, en lugar de recuperar documentos directamente de Internet, lo hace a partir de una colección de documentos estáticos. Una vez que se han recuperado los documentos relevantes, el sistema los agrupa según su tema principal y, finalmente, se produce un resumen para cada grupo. El proceso de resumen se realiza en dos pasos. En primer lugar, se genera un resumen de un solo documento para cada grupo de documentos y, luego, esas oraciones de resumen extraídas se tienen en cuenta para producir el resumen final. La forma en que se seleccionan las oraciones para que formen parte de los resúmenes es usando un*Modelo oculto de Markov*, calculando la probabilidad de una oración con respecto a si es una buena oración de resumen o no.

Se han llevado a cabo menos investigaciones para analizar cómo los resúmenes de texto pueden ser beneficiosos para RI. EnSakai y Spark-Jones(2001) se comprobó que los resúmenes genéricos con una tasa de compresión que oscilaba entre el 10 y el 30% eran los más apropiados para la etapa de indexación en RI, concluyendo que un índice resumido era tan efectivo como el índice de texto completo, para búsquedas orientadas a la precisión. EnSzlávik et al.(2006), si el resumen era útil en la recuperación XML interactiva fue

investigado, proporcionando así resúmenes de elementos XML para permitir a los usuarios navegar y juzgar documentos XML más fácilmente.

4.2 Combinación de resumen de texto con respuesta a preguntas

La respuesta a preguntas (QA) tiene como objetivo responder automáticamente preguntas, ya sean simples o complejas, planteadas en lenguaje natural (Strzalkowski y Harabagiu 2007). En la literatura se pueden encontrar investigaciones específicas en las que se han integrado diferentes enfoques de TS en un sistema de control de calidad. El enfoque mostrado en Mori et al. (2004) analiza la efectividad de las firmas de temas en el contexto de resumen de control de calidad de varios documentos para un tipo particular de preguntas. Los resúmenes generados eran sobre personas y contenían la respuesta a la pregunta. “¿Quién es X?”, donde X es una persona. Se encontró que, aunque las firmas de temas pudieron capturar información enfatizada en los textos fuente correspondientes, esto no fue suficiente, ya que los resúmenes hechos por humanos también contenían algunos detalles que se mencionaron, a pesar de no enfatizarse.

Un enfoque interesante para el control de calidad se presenta en Demner-Fushman y Lin (2006), donde se combinan técnicas de TS e IR para dar respuesta a preguntas pertenecientes al dominio médico. Preguntas como “¿Cuál es el mejor tratamiento farmacológico para X?” se abordan identificando primero los medicamentos a partir de un conjunto de citas y luego agrupando los resúmenes correspondientes, de modo que se pueda producir un breve resumen extractivo para cada uno de ellos. Los resúmenes se generan al mostrar el título del resumen, la intervención principal y la oración con la puntuación más alta, que se determina mediante técnicas de aprendizaje automático supervisado. También en el ámbito médico, el sistema BioSquash (Shi et al. 2007) resume múltiples documentos biomédicos que responden a una pregunta específica. El sistema, basado en un resumen genérico, tiene cuatro componentes principales. El *Anotador* El módulo anota los documentos y la pregunta con información sintáctica y semántica superficial, y luego las relaciones entre los conceptos en los documentos y las preguntas están determinadas por el *Similitud de conceptos* módulo. Los módulos restantes, el *Extractory* y *Editomódulos*, se centran en la selección de contenido y la legibilidad lingüística, respectivamente. El resultado final es un resumen fluido relevante para una pregunta relacionada con el dominio biomédico.

Finalmente, el sistema QAAS (Torres-Moreno et al. 2009) ha resultado de la integración de un sistema TS con un sistema QA. En este enfoque, un resumen genérico de múltiples documentos de varias tasas de compresión se combina con un sistema de control de calidad, lo que permite reducir el espacio de búsqueda de documentos, en comparación con cuando se utiliza todo el documento. Los resultados obtenidos muestran que el número de respuestas correctas devueltas por el sistema combinado aumenta. El resumen genérico también se utiliza para identificar zonas textuales informativas en documentos. Sin embargo, se identifican las limitaciones del uso de resúmenes genéricos para el control de calidad y, por esta razón, el sistema de resúmenes genéricos se adapta a uno centrado en consultas mediante la expansión de consultas y la reevaluación de las oraciones seleccionadas por los resúmenes genéricos de acuerdo con los términos de la pregunta.

Aparte de estos enfoques, también se puede encontrar en la literatura la combinación inversa, QA aplicada a TS. En Mori et al. (2005), se utiliza un motor de control de calidad para ayudar a determinar la importancia de las oraciones, que se calcula en función de las puntuaciones producidas por el sistema de control de calidad y un conjunto de consultas. El objetivo final es generar un resumen, y para ello finalmente se integra el motor QA en un resumen multidocumento genérico.

4.3 Combinar resumen de texto con clasificación de texto

La clasificación de texto (TC), también conocida como categorización de texto, tiene como objetivo clasificar automáticamente un conjunto de documentos en categorías de un conjunto predefinido (Sebastián 2002). En Ker y Chen

(2000), las funciones de resumen (es decir, la posición y la frecuencia de las palabras) se utilizan para categorizar las noticias según diferentes categorías (por ejemplo, "dinero"), alcanzando el 82 % del valor de precisión. Tomar resúmenes de texto en lugar de documentos completos es el enfoque sugerido en Shen et al. (2004), bajo el supuesto de que pueden ser un buen filtro de ruido. Dado que las páginas web contienen demasiada información irrelevante que puede ser perjudicial para TC, los resúmenes pueden extraer la información más importante, produciendo un nuevo texto, que luego se utiliza con fines de clasificación. Los enfoques TS utilizados se basan en la frecuencia de los términos y *Análisis semántico latente*. Realizan una gran experimentación con más de 150.000 páginas web y 64 categorías. Los resultados obtenidos muestran que el algoritmo de clasificación basado en resúmenes propuesto mejora aproximadamente un 8,8% respecto a los documentos completos. La misma idea y conjunto de datos se analiza en Shen et al. (2007) donde, además, se estudia la tasa de compresión óptima para los resúmenes. Los resúmenes del 20 y 30% alcanzan los mejores resultados. Sin embargo, en el rango de 10 a 40%, se demuestra que los resúmenes de texto pueden mejorar el rendimiento de la clasificación hasta cierto punto en documentos completos.

La tarea de inferencia de calificación puede verse como un tipo particular de TC. Su objetivo es identificar la evaluación del autor de una entidad, producto, servicio, etc. con respecto a una escala ordinal basada en su evaluación textual de la entidad (Pang y Lee 2005). Por lo tanto, puede considerarse como un problema de clasificación de opiniones. Por lo general, las opiniones se clasifican con respecto a dos o tres dimensiones, subjetivas frente a objetivas, o positivas frente a neutrales frente a negativas, respectivamente (Wilson et al. 2005). Sin embargo, es frecuente encontrar textos en los que los usuarios dan una puntuación, en función de cuánto les ha gustado o no un producto, película, restaurante, hotel, servicio, etc., lo que normalmente va asociado a una escala de valoración (1 = peor, ... 5 = mejor). Al abordar esta tarea con documentos breves (reseñas que contienen como máximo tres oraciones), el proceso de clasificación obtiene buenos resultados (74%) (Saggion y Funk 2009). Por el contrario, solo alcanza el 32% cuando se trata de textos más extensos. En consecuencia, y tomando como base el supuesto antes mencionado sobre la idoneidad del TS para filtrar el ruido, en Lloret et al. (2010) y Saggion et al. (2010), se lleva a cabo un conjunto preliminar de experimentos para predecir la calificación de una reseña utilizando resúmenes de texto en lugar de documentos completos. Estos experimentos comprenden el análisis de una amplia gama de tipos de resumen de diferentes tasas de compresión. En particular, se estudian resúmenes genéricos, centrados en consultas y basados en sentimientos, junto con varios tipos de líneas de base, incluidas la primera y la última oración de una reseña. Aunque se afirma que los resúmenes centrados en consultas y basados en sentimientos pueden ser más apropiados para la tarea de inferencia de calificación, este análisis tiene dos limitaciones principales. Por un lado, esta tarea es muy compleja, ya que se trabaja con una granularidad muy fina y, por ejemplo, las diferencias entre un texto de 4 y uno de 5 pueden ser muy sutiles. Por otro lado,

5 Evaluación de resumen de texto

Los métodos para evaluar sistemas se pueden clasificar en términos generales en dos grandes categorías: *intrínseca* o *extrínseca* (Spärck-Jones y Galliers 1996). En el contexto de TS, el primero evalúa un resumen en sí mismo, por ejemplo según su contenido de información, mientras que el segundo se enfoca en probar la efectividad de un sistema de resumen en otras aplicaciones (por ejemplo, IR).

En cuanto a la evaluación intrínseca, existen diferentes métodos que se pueden tener en cuenta para evaluar un resumen. De acuerdo a maní (2001a), se puede distinguir entre evaluar la *calidad* el *informatividad* de un resumen. Aparte de estos dos, se propone otro, *fidelidad a la fuente*, que determina la informatividad resumida en el contexto

del documento fuente, es decir, si el resumen contiene los mismos o similares conceptos relevantes que tiene el documento fuente. El problema con este método es cómo dar cuenta de los conceptos relevantes en el documento fuente. Sin embargo, a pesar de tener diferentes metodologías intrínsecas para evaluar resúmenes, el enfoque más común es evaluar la informatividad del resumen comparando su contenido con un modelo humano, que se considera un resumen de referencia. Siguiendo esta idea, se han desarrollado varios métodos que se analizan en la Secc.5.1. Estos métodos se centran en la cantidad de información reflejada en el resumen, es decir, su contenido. Sin embargo, debido a la subjetividad inherente asociada a los resúmenes, no es posible construir un patrón oro justo para decidir que cualquier resumen automático que no sea similar a este no sería un buen resumen. La concepción de lo que es un buen resumen varía mucho entre distintas personas, y también depende de para qué está destinado el resumen. Este y otros problemas relacionados con este tipo de evaluación también se explican en la Secc.5.3. Además, otros enfoques de evaluación se preocupan más por una evaluación cualitativa, cuyo objetivo es evaluar la calidad de un resumen con respecto a diferentes criterios, como la gramaticalidad o el enfoque. Aunque estos problemas son muy difíciles de automatizar, en los últimos años han surgido varios enfoques que intentan proporcionar este tipo de evaluación y se describirán en la Secc.5.2.

Con respecto a los métodos de evaluación extrínseca, se han propuesto varios escenarios como métodos de superficie para la evaluación resumida inspirados en diferentes disciplinas (Hassel 2007). Ejemplos de estos escenarios son: *El juego de Shannon*, cuyo objetivo es cuantificar el contenido de la información adivinando tokens, de modo que se pueda recrear el documento original; *El juego de preguntas*, en el que se pone a prueba la comprensión del lector del resumen y la capacidad de transmitir los conceptos principales; *El juego de clasificación*, que consiste en determinar la categoría ya sea para documentos originales o para resúmenes, y medir la correspondencia entre ellos; y *Asociación de palabras clave*, en el que se proporciona una lista de palabras clave y el objetivo de la tarea es verificar si los resúmenes contienen tales palabras o no. También enmaní (2001a) se describen diferentes métodos de evaluación extrínseca, como *evaluación de relevancia*, en el que se pide a los sujetos que determinen la relevancia de un tema, ya sea en un resumen o en un documento fuente; o *comprensión lectora*, que consiste en responder a pruebas de opción múltiple, habiendo leído el resumen o el documento completo. Además, la aplicación de TS en otro tipo de sistemas, como IR, QA o TC descritos anteriormente en la Secc.4 también podría considerarse como una forma de evaluar extrínsecamente los resúmenes. Por lo tanto, si los resúmenes son capaces de ayudar en otras tareas, se puede suponer que son lo suficientemente buenos para ser utilizados para mejorar otras aplicaciones, aunque pueden estar lejos de ser perfectos con respecto a cuestiones como la coherencia.

5.1 Evaluación de la informatividad

Se ha hecho un gran esfuerzo para desarrollar herramientas de evaluación que se puedan utilizar para evaluar el carácter informativo de un resumen en diferentes niveles. Las conocidas métricas de recuperación de información, precisión, recuperación y medida F (Van Rijsbergen 1981) también se han adaptado a TS. La memoria evalúa qué porción de las oraciones seleccionadas por un humano también son identificadas por un sistema, mientras que la precisión es la fracción de estas oraciones identificadas por un sistema que son correctas (Nenkova 2006). F-measure es una combinación de precisión y recuperación. Sin embargo, al usar estas métricas, es posible que dos extractos igualmente buenos se juzguen de manera muy diferente. *Utilidad relativa* (Radev y Tam 2003) se propuso como otra métrica para la evaluación, en la que varios jueces clasifican cada oración en la entrada con una puntuación, dándoles un valor entre

0 a 10, con respecto a su idoneidad para su inclusión en un resumen. Por lo tanto, cuanto más alto se clasificaron las oraciones, más adecuadas fueron para un resumen.

A la luz de esto, una métrica de resumen basada en información, llamada *puntaje factoides*, se sugiere en Teufel y Halteren (2004), por lo que los resúmenes automáticos se evalúan en función de factoides (unidades atómicas de información que representan el significado de una oración). La idea es utilizar varios resúmenes de referencia como patrón oro y medir la superposición de información entre ellos. El método de la pirámide (Passoneau 2010) sigue una filosofía similar. Su objetivo es identificar información con el mismo significado en diferentes resúmenes escritos por humanos, lo que se denomina *Resumen de unidades de contenido* (UCE). Cada SCU tiene un peso en función del número de evaluadores humanos, que expresaron la misma información, y estos pesos siguen una distribución específica, lo que permite diferenciar los contenidos importantes de los menos importantes. Para evitar la laboriosa tarea de anotar manualmente todas las SCU, se sugiere intentar realizar esto automáticamente en Fuentes et al. (2005). Sin embargo, el esfuerzo manual necesario para detectar factoides o SCU a lo largo de una colección de resúmenes sigue siendo una de las principales desventajas de ambos métodos.

Inspirado en BLEU (Papineni et al. 2002), una herramienta para evaluar automáticamente la salida de un sistema de traducción automática, se desarrolla una herramienta para evaluar automáticamente un resumen en Lin (2004). Su nombre, ROUGE, significa *Suplente orientado a la recuperación para la evaluación de Gisting*, y obtiene precisión, recuperación y medida F.¹⁵ Esta herramienta se basa en la co-ocurrencia de n-gramas, y la idea detrás de esto es comparar el contenido de un resumen con uno o más resúmenes de referencia y ver la cantidad de n-gramas de palabras que todos tienen en común. Se pueden calcular diferentes tipos de n-gramas, como unigramas (ROUGE-1), bigramas (ROUGE-2) o la subsecuencia común más larga (ROUGE-L). La hipótesis de este método es que dos textos que tienen un significado similar, también deben compartir palabras o frases similares.

Bajo el supuesto de que la mejor métrica de similitud debe ser la que mejor discrimine entre resúmenes manuales y generados automáticamente, QARLA, un marco de evaluación descrito en Amigó et al. (2005), es sugerido. Con un conjunto de resúmenes de referencia, un conjunto de resúmenes automáticos y un conjunto de métricas de similitud, el marco proporciona los siguientes tipos de medidas: QUEEN, que es una estimación de la calidad de un resumen automático; KING, como estimación de la calidad de una métrica de similitud; y por último, JACK, como medida para indicar la fiabilidad del conjunto de resúmenes automáticos. En este enfoque, se utiliza un número total de 59 métricas de similitud diferentes, incluidas métricas de recuerdo, precisión, longitud y frecuencia de oraciones y distribución gramatical. Las métricas de recuperación y precisión se basan en el método de superposición de n-gramas de ROUGE, que combina varias opciones de n-gramas con diferentes pasos de preprocesamiento.

Además, con el fin de abordar y mejorar los inconvenientes derivados de la comparación de palabras fijas de n-gramas, se propone un nuevo marco de evaluación en Hovy et al. (2006). La idea subyacente de este método es dividir una oración en unidades muy pequeñas de contenido, llamadas *Elementos básicos dieciséis* (BE), que se definen como tríos de palabras que constan de un encabezamiento y un modificador o argumento, con su relación con el encabezamiento (*cabeza | modificador | relación*). Su objetivo es permitir una mayor flexibilidad para la coincidencia de diferentes expresiones equivalentes. Más adelante, para abordar las deficiencias de expresiones diferentes pero igualmente buenas, ParaEva (Zhou et al. 2006) es desarrollado. Su objetivo es proporcionar un método de evaluación de resumen, facilitando la detección de coincidencias de paráfrasis. La detección de paráfrasis se realiza de acuerdo con una estrategia de tres niveles. Primero, paráfrasis de varias palabras entre frases en los resúmenes de referencia,

¹⁵ La última versión, ROUGE-1.5.5, nos permite obtener precisión, recuperación y medida F. Las versiones anteriores solo calculaban el valor de recuperación.

dieciséis <http://www.isi.edu/publications/licensed-sw/BE/index.html>.

así como en los resúmenes automáticos se identifican. Luego, para aquellos fragmentos que no coincidían, el método intentaba encontrar sinónimos entre palabras sueltas y, si esto también fallaba, finalmente se realizaba una coincidencia léxica simple.

Ensalvado(2007), las gramáticas de texto se emplean para evaluar automáticamente los resúmenes de texto. Una gramática de texto es una forma de describir una estructura de texto válida de manera formal (Van Dijk 1972), y considera la estructura superficial y profunda a través de las relaciones entre oraciones (microestructuras) y la estructura del texto como un todo (macroestructura), respectivamente. Bajo el supuesto de que la superposición de vocabulario no es suficiente para medir la informatividad de un resumen, este enfoque se basa en una lista de proposiciones previamente identificadas, y los humanos tienen que decidir si cada proposición es relevante o no y establecer varios grupos para enfrentar el problema de cuantificar la informatividad de cada proposición. Luego, se proponen tres puntajes, basados en *informatividad* (cuántas proposiciones están presentes en el resumen), *desinformación* (se detectan declaraciones engañosas del resumen) y *gramaticalidad* (que se relaciona con la corrección de las oraciones en base a problemas ortográficos o gramaticales, así como problemas de coherencia). La aplicación de este método en resúmenes escritos por humanos y automáticos muestra que los resúmenes humanos obtienen puntajes más altos que los automáticos, como era de esperar. Sin embargo, el principal inconveniente de este método es que, aunque los resultados muestran que los resúmenes humanos son mejores que los automáticos, no es posible saber qué tan bien se correlacionaría con la evaluación humana. Además, se requiere la intervención humana para identificar proposiciones y evaluar la cantidad de información errónea y falta de gramaticalidad que tienen los resúmenes, lo cual es muy costoso y requiere mucho tiempo y no sería fácilmente escalable.

*AutoSummESP*¹⁷(Giannakopoulos et al. 2008b) es otro método automático desarrollado recientemente que ha demostrado tener una alta correlación con los juicios humanos. Este método se diferencia de los demás en tres aspectos principales: (1) el tipo de información estadística extraída; (2) la representación elegida para esta información extraída, y (3) el método utilizado para calcular la similitud entre los resúmenes. Aquí, la comparación entre resúmenes se lleva a cabo construyendo primero gráficos de caracteres de n-gramas y luego comparando sus representaciones para establecer un grado de similitud entre los gráficos. Además, su metodología es independiente del idioma, por lo que se espera que funcione también en otros idiomas.

Owczarzak(2009) sugieren DEPEVAL(summ), que es una métrica basada en dependencia. La idea aquí es similar a Elementos Básicos, y consiste en comparar los triples de dependencia extraídos de los resúmenes automáticos contra los de los resúmenes del modelo. La principal diferencia con Basic Elements es el analizador utilizado. Mientras que Basic Elements utiliza Minipar,¹⁸ DEPEVAL(summ) se prueba con diferentes analizadores, por ejemplo, el analizador Charniak.¹⁹

GEMS (Modelado generativo para la evaluación de resúmenes) (Katragada 2010) sugiere el uso de términos de firma para analizar cómo se capturan en los resúmenes automáticos. Los términos de la firma se calculan sobre la base de etiquetas de parte del discurso, como sustantivos o verbos; términos de consulta y resúmenes de términos de referencia. La distribución de los términos de la firma se calcula en el documento fuente y luego se obtiene la probabilidad de que un resumen esté sesgado hacia dichos términos de la firma.

Por otro lado, se han propuesto diferentes metodologías de evaluación específicamente para otros idiomas además del inglés, como por ejemplo para el chino utilizando el recurso HowNet. How-Net²⁰ es un sistema de conocimiento electrónico para inglés y chino, y se diferencia de otros

¹⁷ RESUMEN AUTOMático Evaluación basada en gráficos N-gram:<http://www.iit.demokritos.gr/~ggianna/>.

¹⁸<http://webdocs.cs.ualberta.ca/~lindek/minipar.htm>.

¹⁹<ftp://ftp.cs.brown.edu/pub/nlparser/>.

²⁰<http://www.keenage.com/>.

Tabla 4Tipos de métodos de evaluación de la informatividad

Enfoque	Automático	Semiautomático
Utilidad relativa		-
Puntuación factoides		-
método de la pirámide		-
COLORETE	-	
QARLA	-	
SER	-	
gramáticas de texto		-
ParaEval	-	
AutoSummESP	-	
DEPEVAL(suma)	-	
GEMAS	-	
Evaluación de HowNet.	-	

bases de datos léxicas existentes, por ejemplo WordNet,²¹ en la forma en que se calcula la similitud de las palabras. Además, HowNet proporciona información más rica y cada concepto se representa sin ambigüedades por su definición y vínculos de asociación con otros conceptos. Es un recurso bien conocido para el idioma chino y se ha aplicado a muchos enfoques. En (Wang et al. 2008) se propone un enfoque para evaluar resúmenes basado en HowNet. A pesar de que este método también se basa en estadísticas de co-ocurrencia de n-gramas, su principal novedad es el uso de HowNet para calcular la similitud de palabras, por lo que también se pueden tener en cuenta los sinónimos. Además, también se muestra que también se podrían detectar en cierta medida algunas métricas de calidad, como la concisión o el ordenamiento de secuencias.

Mesa4 representa los métodos antes mencionados distinguiendo entre métodos automáticos y semiautomáticos. Este último se referirá a aquellos métodos, como Pyramid, que necesitan algún tipo de anotaciones humanas. Vale la pena enfatizar el hecho de que todos los métodos automáticos, excepto el que se basa en gramáticas de texto, se basan en resúmenes escritos por humanos de referencia que se utilizan para evaluar los automáticos, lo que puede generar algunos problemas si el resumen automático difiere mucho de lo humano-escrito con respecto al vocabulario que contiene.

La investigación sobre la evaluación automática del contenido de los resúmenes ha recibido una atención especial gracias a la pista AESOP propuesta por primera vez en el TAC 2009, que fomenta la investigación sobre este tema. En el contexto de esta tarea se propusieron versiones mejoradas de los enfoques antes mencionados, tales como Giannakopoulos y Karkaletsis(2009) o Conroy et al.(2009), que experimentan con diferentes enfoques basados en las ideas de ROUGE. Debido al éxito en la participación de la tarea, se considera nuevamente para la nueva edición de la conferencia (TAC 2010).

5.2 Evaluación de la calidad

En términos generales, una deficiencia de los métodos de evaluación existentes es que solo evalúan la calidad de un resumen según su contenido, y no tienen en cuenta otros aspectos importantes, como la coherencia o la no redundancia. La evaluación de la calidad de un resumen teniendo en cuenta otras cuestiones distintas a su carácter informativo, siempre ha estado en la mente de los investigadores. El objetivo del Protocolo FAN descrito en Minel et al.(1997) fue evaluar la calidad de un resumen independientemente de

²¹<http://wordnet.princeton.edu/>.

el texto fuente y la información que contiene. Se propusieron cuatro criterios: (1) número de anáforas desprovistas de referentes; (2) ruptura de segmentos textuales; (3) presencia de oraciones tautológicas; y (4) legibilidad del resumen. Todos estos criterios fueron evaluados manualmente por dos jurados. A la luz de esto, también se propuso otro protocolo: el Protocolo MLUCE. La idea detrás de este protocolo es permitir que los usuarios potenciales evalúen los resúmenes, dependiendo de para qué querían el resumen. Por ejemplo, si quieren el resumen solo para decidir leer todo el documento o no, o por el contrario, que sirva como síntesis del documento fuente. Nuevamente, la evaluación se lleva a cabo manualmente como lo es para el Protocolo FAN. Los intentos de evaluar la indicatividad y la aceptabilidad también se han abordado en [Saggion y Lapalme\(2000\)](#). El primero mide si el resumen es capaz de extraer los temas del documento, mientras que el segundo determina si una oración seleccionada por un sistema de resumen es adecuada en comparación con lo que los humanos habrían seleccionado, por lo que se necesita la intervención humana para evaluar este criterio.

Más recientemente, en [Conroy y dang\(2008\)](#), se aborda la necesidad de disponer de herramientas que evalúen contenidos así como otros aspectos lingüísticos. En las conferencias DUC y TAC, los resúmenes se evalúan con respecto a cinco preguntas de calidad lingüística (*gramaticalidad, no redundancia, Claridad referencial, Enfocar, y Estructura y Coherencia*) que no impliquen ninguna comparación con un resumen de referencia. Este tipo de evaluación es realizada manualmente por evaluadores humanos expertos, quienes califican la calidad de un resumen de acuerdo con una escala de cinco puntos. Además, en la literatura podemos encontrar algunos estudios para predecir la calidad del texto a través del análisis de varios factores de legibilidad ([Pitler y Nenkova 2008](#)). La idea aquí es analizar la calidad de un texto por medio de diferentes criterios que incluyen vocabulario, sintaxis o discurso, para dar cuenta de la correlación entre esos factores y las calificaciones de legibilidad humana recopiladas previamente. Cada criterio se modela de una manera diferente, por ejemplo, el vocabulario se representa en términos de unigramas y la sintaxis se modela a través de características, como el número promedio de frases nominales o el número promedio de frases verbales. Los resultados muestran que al combinar todos los factores de legibilidad propuestos, la predicción obtiene una precisión cercana al 90%, por lo que esta idea podría aplicarse y extenderse para evaluar la calidad de un resumen. Otros enfoques, como los presentados en [Barzilay y Lapata\(2005\)](#), [Lapata y Barzilay\(2005\)](#), o [Hasler\(2008\)](#) se enfocan en modelar la coherencia local desde diferentes perspectivas, con el propósito de representar y medir la coherencia del texto. Los enfoques van desde el *Teoría de centrado* ([Grosz et al. 1995](#)) al desarrollo de modelos sintácticos y semánticos, para capturar la distribución de entidades en el texto, o el grado de conectividad entre oraciones, respectivamente. Los intentos de evaluar automáticamente la gramaticalidad de un resumen se han explorado en [Vadlapudi y Katragadda\(2010b\)](#). Los modelos de N-gramas, en particular los unigramas, bigramas, trigramas y la subsecuencia común más larga se utilizan para capturar este aspecto. Además, este problema se considera como un problema de clasificación, donde las oraciones de resumen se clasifican en clases sobre la base de su aceptabilidad. El parámetro de aceptabilidad se estima mediante trigramas. Los métodos propuestos se evalúan de la misma manera que se evaluaron los resúmenes en las conferencias DUC o TAC. Los resultados obtenidos se correlacionan bien (85% como máximo) con respecto a las evaluaciones manuales ya existentes. Además, en [Vadlapudi y Katragadda\(2010a\)](#), también se investigan aspectos de estructura y coherencia a partir de cadenas léxicas y la relación semántica de dos entidades. Los resultados que utilizan este enfoque alcanzan un 70% utilizando la correlación de Spearman.

5.3 Limitaciones de los métodos de evaluación existentes

Aunque los métodos antes mencionados realmente ayudan a evaluar los resúmenes automáticos, hay varias preguntas relacionadas con la evaluación que aún quedan sin resolver. Por lo tanto, algunos

Se pueden abordar las deficiencias de la tarea de evaluación en general, así como algunos problemas específicos que tienen las herramientas de evaluación.

En primer lugar, tener una colección de documentos y sus correspondientes resúmenes es una tarea muy costosa. Por un lado, diferentes humanos escribirían diferentes resúmenes, unos con más abstracción, otros sintetizando más la información, y así sucesivamente. Esto significa que es posible tener varios resúmenes válidos, aunque diferentes en contenido. Figura 1 muestra tres ejemplos diferentes de resúmenes (A, B y C). Además, su correspondiente documento original se puede ver en la Fig. 2. Estos documentos hacen referencia al documento AP880911-0016 tomado del corpus DUC 2002. Vale la pena señalar que los resúmenes A y B, en la parte superior de la figura, respectivamente, son producidos por humanos expertos, mientras que el resumen C se genera automáticamente utilizando un sistema que extrae las oraciones más relevantes de un texto, empleando la vinculación textual para eliminar la redundancia, y la frecuencia de palabras como esquema de ponderación, dando más importancia a aquellas oraciones que contienen palabras con alta frecuencia (Lloret et al. 2008). En consecuencia, los resúmenes A y B tienen un carácter abstracto, mientras que el resumen producido automáticamente se ha elaborado siguiendo un enfoque extractivo. Como puede verse en los ejemplos dados, ninguno de ellos es idéntico, aunque contienen información similar. Parte del vocabulario se comparte entre todos los resúmenes (*"tormenta tropical Gilbert en el este del Caribe"*), pero otros hechos se expresan con palabras diferentes (*"Puerto Rico emitió una alerta de inundación para Puerto Rico"*, *"Florida se esperan inundaciones en Puerto Rico"*, *"Gilbert trajo inundaciones costeras [...] a la costa sur de Puerto Rico"*). Para evaluar el contenido de cada resumen, calculamos las puntuaciones de ROUGE tomando como referencia otro resumen escrito por humanos diferente a los que se muestran en la Fig. 1. En cuanto al valor de recuerdo para ROUGE-1, se obtiene un 46% para el resumen automático, mientras que los resultados para los resúmenes escritos por humanos son 53 y 49%, respectivamente. Aunque estos valores son muy cercanos, significa que el primer resumen humano sería el mejor, pero como se puede ver, todos los resúmenes contienen información relevante. Además, puede ocurrir que si estos resúmenes fueran evaluados manualmente por diferentes personas, los resultados podrían variar, debido a la subjetividad inherente al proceso de evaluación de resúmenes. Por lo tanto, es un asunto muy delicado decidir cuál de estos resúmenes es el mejor y cómo podemos tomar tal decisión. En consecuencia, elegir resúmenes escritos por humanos como estándar de oro puede no ser la solución óptima para el problema de evaluación de resúmenes.

Con respecto a los ejemplos anteriores, la dificultad aumenta cuando un resumen producido siguiendo un enfoque extractivo tiene que ser comparado con otro generado por uno abstractivo. Diferentes estudios han demostrado que, si los humanos tuvieran que decidir qué oraciones de los documentos son más relevantes para pertenecer al resumen final para producir extractos, también estarían en desacuerdo sobre qué oraciones representan mejor el contenido de un documento. Por lo tanto, el bajo acuerdo entre humanos es un problema cuando se utilizan sus resúmenes para comparar. En Donaway et al. (2000) se mostró la varianza en el valor de recordación, dependiendo de qué resumen humano se tomó como referencia para la comparación con el automático. Este problema también se planteó en maní (2001b). Además, la equivalencia semántica entre diferentes sustantivos, por ejemplo mediante sinonimia, o expresiones, cuando hay varias formas de expresar una misma idea, es otro inconveniente de la evaluación (Nenkova 2006), porque la mayoría de los métodos solo realizan un análisis superficial y no tienen en cuenta el significado semántico de las frases. En cuanto a los métodos automáticos, se realizaron algunas críticas a la herramienta ROUGE. Por un lado, en Sjöbergh (2007) se abordó el hecho de que se podría producir fácilmente un resumen para obtener puntajes ROUGE altos. Para probar esta hipótesis, se desarrolló un método de resumen simple, utilizando una estrategia de selección de palabras codiciosa, y aunque los resúmenes generados no eran buenos desde el punto de vista humano, obtuvieron

Tropical Storm Gilbert in the eastern Caribbean strengthened into a hurricane Saturday night. The National Hurricane Center in Miami reported its position at 2 a.m. Sunday to be about 140 miles south of Puerto Rico and 200 miles southeast of Santo Domingo. It is moving westward at 15mph with a broad area of cloudiness and heavy weather with sustained winds of 75mph gusting to 92mph. The Dominican Republic's Civil Defense alerted that country's heavily populated south coast and the National Weather Service in San Juan, Puerto Rico issued a flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday.

Hurricane Gilbert is moving toward the Dominican Republic, where the residents of the south coast, especially the Barahona Province, have been alerted to prepare for heavy rains, and high winds and seas. Tropical Storm Gilbert formed in the eastern Caribbean and became a hurricane on Saturday night. By 2 a.m. Sunday it was about 200 miles southeast of Santo Domingo and moving westward at 15 mph with winds of 75 mph. Flooding is expected in Puerto Rico and the Virgin Islands. The second hurricane of the season, Florence, is now over the southern United States and downgraded to a tropical storm.

Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas. Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night. Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds and up to 12 feet feet to Puerto Rico's south coast. San Juan, on the north coast, had heavy rains and gusts Saturday, but they subsided during the night. On Saturday, Hurricane Florence was downgraded to a tropical storm and its remnants pushed inland from the U.S. Gulf Coast.

Figura 1 Ejemplos de resúmenes humanos y automáticos (resúmenes A, B y C, respectivamente)

Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas. The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph. "There is no need for alarm," Civil Defense Director Eugenio Cabral said in a television alert shortly before midnight Saturday. Cabral said residents of the province of Barahona should closely follow Gilbert's movement. An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo. Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night. The National Hurricane Center in Miami reported its position at 2 a.m. Sunday at latitude 16.1 north, longitude 67.5 west, about 140 miles south of Ponce, Puerto Rico, and 200 miles southeast of Santo Domingo. The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westward at 15 mph with a "broad area of cloudiness and heavy weather" rotating around the center of the storm. The weather service issued a flash flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday. Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds and up to 12 feet feet to Puerto Rico's south coast. There were no reports of casualties. San Juan, on the north coast, had heavy rains and gusts Saturday, but they subsided during the night. On Saturday, Hurricane Florence was downgraded to a tropical storm and its remnants pushed inland from the U.S. Gulf Coast. Residents returned home, happy to find little damage from 80 mph winds and sheets of rain. Florence, the sixth named storm of the 1988 Atlantic storm season, was the second hurricane. The first, Debby, reached minimal hurricane strength briefly before hitting the Mexican coast last month.

Figura 2 Ejemplo de un documento fuente

buenos resultados en algunos valores de ROUGE, por ejemplo, una puntuación de recuerdo del 41 % para ROUGE-1.²² Por otro lado, la correlación entre ROUGE y los resúmenes humanos resultó ser más baja de lo que se afirmaba, especialmente en otros tipos de resúmenes, por ejemplo en

²² Aunque este valor pueda parecer poco elevado, es bastante bueno en relación con el estado del arte.

resumen del discurso (Liu y Liu 2008). En cuanto al método de Elementos Básicos, su principal inconveniente es que utiliza algunos módulos de preprocesamiento dependientes del lenguaje para parsear y cortar, por lo que no sería muy fácil aplicarlo a otros lenguajes, especialmente si carecen de dichos recursos.

Con respecto a la evaluación de la calidad, los métodos sugeridos aún se encuentran en sus primeros pasos. Realizar este tipo de evaluación es realmente difícil y desarrollar una buena métrica que se correlacione bien con la evaluación humana no es un tema trivial. Normalmente, se pide a los jueces expertos que evalúen los resúmenes utilizando una escala de 5 puntos, que comprende valores cualitativos con respecto a una pregunta o afirmación. Por ejemplo, en la evaluación realizada en DUC 2001, las preguntas tenían esta forma:

¿En qué medida el resumen dice lo mismo una y otra vez?

1. Bastante; la mayoría de las oraciones son repetitivas
2. Más de la mitad del texto es repetitivo
3. Algunas repeticiones
4. Repeticiones menores
5. Ninguno; el resumen no tiene información repetida

que significa exactamente “*alguna repetición*” o “*repetición menor*”? Sería posible mapear en una escala cuantitativa la primera y la última afirmación, por ejemplo, ninguna significaría ninguna repetición en absoluto, pero los límites entre las del medio son muy sutiles. Además, declaraciones como las del ejemplo contienen un grado de subjetividad que no es posible capturar automáticamente. Todos estos problemas hacen que la tarea de evaluar la calidad de un resumen sea realmente desafiante y difícil de abordar desde un punto de vista automático.

6 Direcciones futuras en el resumen de texto

Desde finales de los años 50, las metodologías y sistemas de resumen han experimentado grandes avances. Se desarrollan nuevos enfoques, teniendo en cuenta también aspectos lingüísticos en algunos casos, lo que permite que un resumen automático sea algo más que una mera unión de frases. Además, estos avances han dado lugar a nuevos tipos de resúmenes (por ejemplo, actualización o resúmenes personalizados) y nuevos escenarios donde los resúmenes juegan un papel crucial (por ejemplo, reivindicaciones de patentes, blogs, reseñas). Sin embargo, todavía hay mucho margen de mejora, especialmente debido a la gran cantidad de datos disponibles en diferentes formatos, y el rápido desarrollo de la tecnología, que trae nuevos desafíos para este campo de investigación, como multi-documento, multi- resumen lingüístico o multimedia. Otro tema desafiante es la evaluación, que plantea las siguientes preguntas: “¿Es posible evaluar un resumen de manera objetiva?”, “¿Es justo comparar los resúmenes automáticos con los escritos por humanos?”, “¿Cómo podría evaluarse automáticamente la calidad de un resumen?”.

El objetivo de este trabajo ha sido realizar una investigación sobre el estado del arte en ST, centrándonos especialmente en los nuevos tipos de resúmenes y escenarios planteados en los últimos años. Sin embargo, con el fin de proporcionar algunos antecedentes básicos sobre este campo de investigación, también se realizó una breve revisión de enfoques de resumen conocidos, agrupando estos enfoques en diferentes grupos en lo que respecta a las técnicas o algoritmos empleados. Además, se describieron los foros y conferencias internacionales sobre ST que se han realizado a lo largo de los años, con el fin de brindar una perspectiva histórica de este campo de investigación, destacando cómo ha evolucionado y la atención que la comunidad investigadora le ha prestado. . Es muy interesante ver el impacto de TS en la investigación comunitaria, ya que se puede concluir que, a pesar de su antigüedad y dificultad, atrae cada vez a más investigadores. Un

Una parte importante de esta revisión bibliográfica trata sobre el estado del arte actual del TS, donde se puede apreciar cómo se ha ido adaptando a los requerimientos de la nueva sociedad. Esto ha dado como resultado nuevos tipos de resúmenes y nuevos escenarios en los que la síntesis puede ser de gran ayuda. Además, también se analiza la combinación de TS con otros sistemas inteligentes, en concreto, IR, QA o TC, ya que los resúmenes de texto pueden mejorar el rendimiento de otros sistemas, siendo muy apropiado usarlos en combinación con otras aplicaciones. Finalmente, una encuesta en TS no puede evitar el problema de la evaluación y, en consecuencia, también se presta especial atención a este tema. Aunque existen varias métricas y métodos de evaluación, que pueden evaluar la información de un resumen (si el resumen contiene la información correcta), o su calidad (si la información está claramente expresada y es legible), todavía hay algunos aspectos desafiantes, que permiten mucho margen de mejora con respecto a este tema. Sin embargo, es esencial analizar los métodos de evaluación actuales y sus limitaciones, con el fin de dar algunas pistas para futuras evaluaciones.

El análisis realizado en este trabajo nos permite tener información básica sobre el pasado de los TS, el estado del arte actual y las posibles tendencias a futuro. En lo que respecta a los enfoques de TS, vale la pena mencionar que a lo largo de los años, los enfoques existentes están cambiando. Por ejemplo, se proponen nuevos algoritmos de aprendizaje automático para abordar el TS; sin embargo, las características utilizadas no varían demasiado con respecto a las ya existentes (por ejemplo, frecuencia del término, parte del discurso, posición). Lo que parece estar cambiando rápidamente a lo largo de los años son los tipos de resúmenes, ya que la sociedad tiene que adaptarse a los nuevos requisitos de los usuarios. Considerando que al principio la síntesis genérica y de un solo documento eran uno de los tipos más importantes, actualmente la síntesis multidocumento e incluso multilingüe ha cobrado gran importancia debido a la gran cantidad de información que tenemos que manejar. Esto también se puede comprobar en los foros internacionales dedicados a los TS, en los que, año tras año, se actualizan las pistas propuestas. Además, en los últimos años existe una tendencia hacia la generación de resúmenes con fines específicos, como los resúmenes basados en sentimientos o personalizados. Vale la pena recalcar que la generación de resúmenes también está cobrando mucha importancia, y en la medida en que mejoren las técnicas existentes actualmente, es posible emplearlas para generar resúmenes cercanos a los escritos por humanos. Además, también van en aumento nuevos escenarios distintos a los tradicionales, como el newswire, o los artículos científicos,

La evaluación de un resumen, ya sea automático o escrito por humanos, es un tema delicado, debido a la subjetividad inherente asociada al proceso. En este artículo, presentamos dos tipos diferentes de evaluación, intrínseca y extrínseca, dependiendo de si queremos evaluar el resumen en sí o su desempeño para cumplir con los objetivos de otras aplicaciones. La evaluación que realizan los métodos y herramientas actuales es mayoritariamente intrínseca. Además, se puede distinguir entre informatividad y evaluación de la calidad. Como se puede observar, la mayoría de las herramientas automáticas actuales evalúan el contenido de un resumen, y solo unos pocos enfoques intentan determinar si el resumen generado es de alta calidad o no, en función de diferentes criterios, como la gramaticalidad o la coherencia. Los métodos existentes tienen algunas limitaciones, en el sentido de que se basaron principalmente en la superposición de vocabulario entre un resumen automático y un modelo. Por otro lado, en cuanto a la evaluación de la calidad, en los últimos años han surgido novedosos enfoques para automatizar este proceso, lo que supone un paso más hacia este campo de investigación, ya que este tipo de evaluación ha sido realizada de forma totalmente manual por jueces expertos, por lo que lejos. Aunque todavía son muy preliminares, representan un buen punto de partida para abordar este tema. Para enfrentar con éxito el problema de las escalas cualitativas, primero debemos investigar métodos capaces de mapear aspectos cualitativos. lo que supone un paso más hacia este campo de investigación, ya que hasta el momento este tipo de evaluación ha sido realizada de forma totalmente manual por jueces expertos. Aunque todavía son muy preliminares, representan un buen punto de partida para abordar este tema. Para enfrentar con éxito el problema de las escalas cualitativas, primero debemos investigar métodos capaces de mapear aspectos cualitativos.

(por ejemplo, repeticiones menores) en cuantitativas (por ejemplo, *¿Cuántas repeticiones (en número) deben estar presentes para ser consideradas menores? ¿10% del texto?, ¿5% de las entidades nombradas?*).

Se pueden extraer algunas conclusiones sobre las tendencias de TS para el futuro. Como decíamos anteriormente, los requerimientos de la sociedad cambian y la información crece a un ritmo exponencial, lo que obliga a TS a adaptarse a las nuevas necesidades. Para los próximos años, la síntesis multidocumental y multilingüe será fundamental, ya que la misma información puede aparecer en un elevado número de documentos pero también en diferentes idiomas. Y cabe mencionar que esta información tiene que ser presentada de forma coherente y yendo más allá de la concatenación de oraciones. Por tanto, los paradigmas abstractivos o al menos híbridos se convertirán en uno de los principales retos a resolver. Los enfoques híbridos serían capaces de identificar y seleccionar fragmentos relevantes de información y luego fusionar, comprimir o eliminar dicha información para generar nueva información resumida. Como consecuencia, es posible tener en cuenta los beneficios de los enfoques extractivo y abstractivo juntos. Además, dado que hoy en día los usuarios juegan un papel crucial en Internet, los resúmenes basados en sentimientos, resúmenes personalizados y actualizados también serán muy importantes, porque un resumen debe proporcionar la información exacta que un usuario requiere. Cómo presentar dicha información es otro tema que se está volviendo cada vez más importante en el sentido de que, tradicionalmente, la entrada y la salida de un sistema TS es texto. Sin embargo, esta tendencia está cambiando y podemos encontrar muchos enfoques que resumen otros tipos de entrada, como reuniones o videos, o producir la salida en un formato diferente al texto. Los resúmenes basados en opiniones, resúmenes personalizados y actualizados también serán muy importantes, porque un resumen debe proporcionar la información exacta que requiere un usuario. Cómo presentar dicha información es otro tema que se está volviendo cada vez más importante en el sentido de que, tradicionalmente, la entrada y la salida de un sistema TS es texto. Sin embargo, esta tendencia está cambiando y podemos encontrar muchos enfoques que resumen otros tipos de entrada, como reuniones o videos, o producir la salida en un formato diferente al texto.²³ Esto consistiría, por ejemplo, en tomar texto como entrada, pero por el contrario presentar el resumen en otro formato, por ejemplo, mediante estadísticas, tablas, gráficos, escalas de calificación visual, etc., lo que permitiría a los usuarios visualizar los resultados de forma inmediata, y tal vez encontrar la información que les interesa más rápidamente. Además, estas representaciones visuales también podrían complementarse con resúmenes de texto. En cuanto a la evaluación, los principales temas a enfocar son la evaluación intrínseca, la realización de investigaciones en métodos novedosos y formas de evaluar el resumen, tanto de acuerdo con la información que contiene como con la forma en que se presenta. Debido a que el proceso de evaluación contiene un alto grado de subjetividad, aún no sabemos si el proceso podría ser bastante automatizado, porque primero se debe definir un buen criterio para distinguir qué considerar relevante y qué no. Lo mismo ocurre con la evaluación de la calidad de un resumen. Si bien se establecen algunas pautas para llevar a cabo esta tarea, siempre existe un margen de subjetividad, que genera resultados diferentes cuando el mismo resumen es evaluado por dos humanos. Aparte de esto, en la medida en que los métodos semánticos progresen, será más factible dar cuenta de expresiones equivalentes. Esto ayudará a los métodos automáticos, ya que sería posible determinar si el resumen contiene contenido apropiado o no. que generan resultados diferentes cuando el mismo resumen es evaluado por dos humanos. Aparte de esto, en la medida en que los métodos semánticos progresen, será más factible dar cuenta de expresiones equivalentes. Esto ayudará a los métodos automáticos, ya que sería posible determinar si el resumen contiene contenido apropiado o no. que generan resultados diferentes cuando el mismo resumen es evaluado por dos humanos. Aparte de esto, en la medida en que los métodos semánticos progresen, será más factible dar cuenta de expresiones equivalentes. Esto ayudará a los métodos automáticos, ya que sería posible determinar si el resumen contiene contenido apropiado o no.

Finalmente, a pesar de tener más de 50 años, TS sigue vivo con un gran interés entre la comunidad investigadora. De hecho, este campo de investigación es muy dinámico, ya que se adapta continuamente a las nuevas necesidades y desafíos. Aunque el rendimiento de TS aún es moderado y los resúmenes generados están lejos de ser perfectos, se puede ver cómo la combinación con otros sistemas conduce a la mejora del rendimiento general del sistema combinado, lo que ayuda a desarrollar sistemas aún más inteligentes. La evaluación de un resumen presenta también grandes retos, que se están abordando desde hace años, y se obtienen mejoras respecto a enfoques anteriores. Todas las posibilidades que ofrece TS junto con la amplia aplicación que tiene en el mundo real, lo convierten en un campo de investigación interesante para investigar,

²³ En este documento, solo nos enfocamos en el texto y, en consecuencia, aunque existen otros formatos de entrada/salida, estaban fuera del alcance de este documento.

Expresiones de gratitud Esta investigación ha sido financiada por el proyecto PROMETEO "Desarrollo de Técnicas Inteligentes e Interactivas de Minería de Textos" (2009/119) y el proyecto referencia ACOMP/2010/286 de la Generalitat Valenciana, así como por el Gobierno de España (subvención no .TIN2009-13391-C04-01). Elena Lloret está financiada por el programa FPI (BES-2007-16268) del Ministerio de Ciencia e Innovación de España, en el marco del proyecto TEXT-MESS (TIN2006-15265-C06-01).

Referencias

- Agnihotri L, Kender JR, Dimitrova N, Zimmerman J (2005) Estudio de usuario para generar resumen personalizado perfiles de maría. En: Actas de la conferencia internacional IEEE sobre multimedia y exposición (ICME). págs. 1094–1097
- Ahmet A, Gaizauskas R (2010) Generación de descripciones de imágenes utilizando patrones relacionales de dependencia. En: **Pro-actas de la 48ª reunión anual de la asociación de lingüística computacional**
- Aker A, Gaizauskas R (2009) Generación de resumen para imágenes referenciadas a topónimos usando lenguaje de tipo de objeto. modelos de calibre. En: Actas de la conferencia internacional sobre avances recientes en el procesamiento del lenguaje natural (RANLP-2009)
- Aker A, Gaizauskas R (2010) Resúmenes de modelos para imágenes relacionadas con OPTlocalización. En: Actas de lenguaje recursos y evaluación
- Amigó E, Gonzalo J, Peñas A, Verdejo F (2005) QARLA: un marco para la evaluación del resumen de textos sistemas de zación. En: ACL '05: actas de la 43.ª reunión anual de la asociación de lingüística computacional. págs. 280–289
- Ando R, Boguraev B, Byrd R, Neff M (2005) Resumen de documentos múltiples habilitado para visualización por Iter-Reescalamiento de Residual ativo. Nat Lang Eng 11 (1): 67–86
- Angheluta R, Busser RD, Francine Moens M (2002) The use of topic segmentation for automatic summarization En: Actas del taller posterior a la conferencia ACL-2002 sobre resumen automático. págs. 66–70
- Aone C, Okurowski ME, Gorlinsky J (1998) Resumen entrenable y escalable usando NLP robusto y máquina aprendizaje. En: Actas de la 36.ª reunión anual de la asociación de lingüística computacional y la 17.ª conferencia internacional sobre lingüística computacional, vol 1. pp 62–66
- Azzam S, Humphreys K, Gaizauskas R (1999) Uso de cadenas de correferencia para el resumen de texto. En: Proceder-ings del taller ACL'99 sobre correferencia y sus aplicaciones
- Balahur A, Montoyo A (2008) Extracción y resumen de opiniones basadas en características multilingües de clientes revisiones de tomer. En: Actas de la 13ª conferencia internacional sobre aplicaciones del lenguaje natural a los sistemas de información. págs. 345–346
- Balahur A, Lloret E, Ferrández O, Montoyo A, Palomar M, Muñoz R (2008) La participación del equipo de la DLSIUAEs pación en las pistas TAC 2008. En: Actas de la conferencia de análisis de texto (TAC)
- Balahur A, Lloret E, Boldrini E, Montoyo A, Palomar M, Martinez-Barco P (2009) Resumen de hilos en blogs que utilizan la polaridad de opinión. En: Actas del taller internacional sobre eventos en tipos de texto emergentes (eETT). págs. 5–13
- Balahur-Dobrescu A, Kabadjov M, Steinberger J, Steinberger R, Montoyo A (2009) Resumen de opiniones en hilos de blogs. En: Actas de la conferencia de Asia Pacífico sobre lenguaje, información y computación. págs. 606–613
- Baldwin B, Morton TS (1998) Resumen dinámico basado en correferencias. En: Actas de la tercera con-Conferencia sobre métodos empíricos en el procesamiento del lenguaje natural (EMNLP-3)
- Barzilay R, Elhadad M (1999) Uso de cadenas léxicas para resúmenes de texto. En: Avances en texto automático resumen págs. 111–122
- Barzilay R, Lapata M (2005) Modelando la coherencia local: un enfoque basado en entidades. En: Actas de la 43. reunión anual de la asociación de lingüística computacional (ACL'05). págs. 141–148
- Barzilay R, McKeown KR (2005) Fusión de oraciones para resúmenes de noticias en múltiples documentos. lingüista informático 31(3):297–328
- Beineke P, Hastie T, Manning C, Vaithyanathan S (2004) Una exploración del resumen de sentimientos. En: Actas del simposio de primavera AAAI sobre la exploración de la actitud y el afecto en el texto: teorías y aplicaciones
- Bellemare S, Bergler S, Witte R (2008) ERSS en TAC 2008. En: Actas de la conferencia de análisis de texto (TAC)
- Belz A (2008) Generación automática de textos de pronóstico del tiempo utilizando generación probabilística integral Modelos de espacio de ción. Nat Lang Eng 14 (4): 431–455

- Berkovsky S, Baldwin T, Zukerman I (2008) Resumen de texto personalizado basado en aspectos. En: Actas de la 5ª conferencia internacional sobre hipermedia adaptativa y sistemas adaptativos basados en la web. pp 267–270 Biadys F, Hirschberg J, Filatova E (2008) Un enfoque no supervisado para la producción de biografías usando Wikipedia. En: Actas de ACL-08: HLT. págs. 807–815
- Boguraev BK, Neff MS (2000) Segmentación del discurso como ayuda al resumen de documentos. En: Actas de la 33ª conferencia internacional de Hawái sobre ciencias de sistemas, vol 3. p 3004
- Bossard A, Générèux M, Poibeau T (2008) Descripción de los sistemas LIPN en TAC 2008: resumen información y opiniones. En: Actas de la conferencia de análisis de texto (TAC)
- Branny E (2007) Evaluación resumida automática basada en gramáticas de texto. Inf. dígito J 8(3).<http://revistas.tdl.org/jodi/article/viewArticle/232>
- Burges C, Shaked T, Renshaw E, Lazier A, Deeds M, Hamilton N, Hullender G (2005) Aprender a clasificar usando descenso de gradiente. En: Actas de la 22ª conferencia internacional sobre aprendizaje automático. pp 89–96
- Carenini G, Cheung JCK (2008) Resumen extractivo vs. resumen abstracto basado en NLG del texto evaluativo: el efecto de la polémica del corpus. En: Actas de la quinta conferencia internacional de generación de lenguaje natural, ACL 2008. pp 33–40
- Cesarano C, Mazzeo A, Picariello A (2007) Un sistema para la similitud resumen-documento en el dominio notarial. Taller internacional sobre la aplicación del sistema experto en bases de datos: 254–258
- Ceylan H, Mihalcea R (2009) La descomposición de los resúmenes de libros escritos por humanos. En: Actas de la 10ª conferencia internacional sobre lingüística computacional y procesamiento inteligente de textos (CICLing '09). págs. 582–593
- Cole R (ed) (1997) Estudio del estado del arte en la tecnología del lenguaje humano. Prensa de la Universidad de Cambridge, cambridge
- Conroy J, Schlesinger J (2008) CLASSY en TAC 2008 Métricas. En: Actas de la conferencia de análisis de texto (TAC)
- Conroy JM, Dang HT (2008) Cuidado con la brecha: peligros de divorciar las evaluaciones del contenido resumido de la línea. calidad guística. En: Actas de la 22ª conferencia internacional sobre lingüística computacional (Coling 2008). págs. 145–152
- Conroy JM, O'leary DP (2001) Resumen de texto a través de modelos ocultos de Markov. En: SIGIR '01: actas de la 24ª conferencia internacional anual ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 406–407
- Conroy JM, Schlesinger JD, O'Leary DP (2009) CLASSY 2009: resumen y métricas. En: Actas de la conferencia de análisis de texto (TAC)
- Cristea D, Postolache O, Pistol I (2005) Resumen a través de la estructura del discurso. En: Actas de la lingüística computacional y procesamiento inteligente de textos, 6ª Conferencia Internacional (CICLing 2005). págs. 632–644
- Cunha ID, Fernández S, Velázquez-Morales P, Vivaldi J, SanJuan E, Moreno JMT (2007) A new hybrid summarizer basado en modelo de espacio vectorial, física estadística y lingüística. En: MICAI 2007: avances en inteligencia artificial. págs. 872–882
- Dang HT (2006) Overview of DUC 2006. En: El taller de comprensión de documentos (presentado en la *HLT/NA-LCA*). Brooklyn, Nueva York, Estados Unidos
- Demner-Fushman D, Lin J (2006) Extracción de respuestas, agrupamiento semántico y resumen extractivo para respuesta a preguntas clínicas. En: Actas de la 21.ª conferencia internacional sobre lingüística computacional y la 44.ª reunión anual de la asociación de lingüística computacional. pp 841–848 Deschacht K, Moens MF (2007) Análisis de texto para la anotación automática de imágenes. En: Actas del 45 reunión anual de la asociación de lingüística computacional. págs. 1000–1007
- Díaz A, Gervás P (2007) Resumen personalizado basado en modelo de usuario. Gestión de procesos de información 43(6):1715–1734
- Donaway RL, Drummey KW, Mather LA (2000) Una comparación de clasificaciones producidas por evaluación de resumen medidas de uación. En: Actas del taller NAACL-ANLP 2000 sobre resumen automático. págs. 69–78
- Dunlavy DM, O'Leary DP, Conroy JM, Schlesinger JD (2007) QCS: un sistema para consultar, agrupar y resumen de documentos. Gestión de procesos de información 43 (6): 1588–1605
- Edmundson HP (1969) Nuevos métodos de extracción automática. En: Mani I, Maybury M (eds) Avances en resumen automático de texto. págs. 23–42
- Elsner M, Charniak E (2008) Modelado de coherencia inspirado en la coreferencia. En: Procedimientos de ACL-08: HLT, papeles cortos. págs. 41–44
- Ercan G, Cicekli I (2008) Modelado de temas basado en la cohesión léxica para resúmenes. En: Actas de la IX Congreso Internacional de Lingüística Computacional y Procesamiento Inteligente de Textos. pp 582–592 Erkan G, Radev DR (2004) LexRank: centralidad léxica basada en gráficos como prominencia en el resumen de texto. J Artif Intel Res (JAIR) 22: 457–479

- Fan J, Gao Y, Luo H, Keim DA, Li Z (2008) Un enfoque novedoso para habilitar el resumen semántico y visual de imágenes. *rización para la búsqueda exploratoria de imágenes*. En: MIR '08: actas de la 1ra conferencia internacional ACM sobre recuperación de información multimedia. págs. 358–365
- Fellbaum C (1998) WordNet: una base de datos léxica electrónica. La prensa del MIT, Cambridge
- Feng Y, Lapata M (2008) Anotación automática de imágenes utilizando información de texto auxiliar. En: Actas de ACL-08: HLT. págs. 272–280
- Filatova E, Hatzivassiloglou V (2004) Resumen extractivo basado en eventos. En: Marie-Francine Moens SS (ed) *El resumen del texto se bifurca: actas del taller ACL-04*. págs. 104–111
- Fisher S, Dunlop A, Roark B, Chen Y, Burmeister J (2009) Resumen de OHSU y sistemas de vinculación de entidades. En: Actas de la conferencia de análisis de texto (TAC)
- Fiszman M, Rindflesch TC, Kilicoglu H (2004) Resumen de abstracción para gestionar la biomedicina literatura de investigación. En: Moldovan D, Girju R (eds) *HLT-NAACL 2004: taller sobre semántica léxica computacional*. págs. 76–83
- Fuentes M, González E, Ferrés D, Rodríguez H (2005) QASUM-TALP en DUC 2005 evaluado automáticamente con una métrica basada en pirámide. En: *El taller de comprensión de documentos (presentado en la HLT/EMNLP reunión anual)*
- Fuentes M, Alfonseca E, Rodríguez H (2007) Máquinas de vectores de soporte para resumen centrado en consultas capacitados y evaluados en datos piramidales. En: Actas de la 45.ª reunión anual de la asociación de lingüística computacional actas del volumen complementario de las sesiones de demostración y póster. pp 57–60
- Fukushima T, Okumura M (2001) Desafío de resumen de texto: evaluación de resumen de texto en NTCIR taller 2. En: Actas de la segunda reunión del taller NTCIR sobre la evaluación de la recuperación y el resumen de textos en chino y japonés. págs. 9–13
- Giannakopoulos G, Karkaletsis V (2009) GRÁFICOS N-GRAM: representación de documentos y conjuntos de documentos en evaluación resumida del sistema. En: Actas de la conferencia de análisis de texto (TAC) Giannakopoulos G, Karkaletsis V, Vouros G (2008a) Prueba del uso de gráficos de n-gramas en resumen subtares. En: Actas de la conferencia de análisis de texto (TAC)
- Giannakopoulos G, Karkaletsis V, Vouros G, Stamatopoulos P (2008) Evaluación del sistema de resumen Revisado: gráficos de N-gramas. *ACM Trans Speech Lang Process* 5(3):1–39
- Goldstein J, Mittal V, Carbonell J, Kantrowitz M (2000) Resumen de documentos múltiples por extracción de oraciones En: Taller NAACL-ANLP 2000 sobre Resumen automático. págs. 40–48
- Gonçalves PN, Rino L, Vieira R (2008) Resumiendo y refiriendo: hacia extractos cohesivos. En: DocEng '08: actas del octavo simposio de ACM sobre ingeniería de documentos. págs. 253–256
- Gotti F, Lapalme G, Nerima L, Wehrli E (2007) GOFASUM: un resumen simbólico para DUC. En el taller de comprensión de documentos (presentado en la HLT/NAACL)
- Grosz BJ, Weinstein S, Joshi AK (1995) Centrado: un marco para modelar la coherencia local de Discurso. *Lingüista informático* 21 (2): 203–225
- Harabagiu S, Lacatusu F (2005) Temas temáticos para resúmenes de múltiples documentos. En: SIGIR '05: actas de la 28ª conferencia internacional anual ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 202–209
- Hasler L (2007) De extractos a resúmenes: operaciones de producción de resúmenes humanos para resúmenes asistidos por computadora marización En: Actas del taller RANLP 2007 sobre procesamiento del lenguaje asistido por computadora (CALP). págs. 11–18
- Hasler L (2008) Teoría de centrado para la evaluación de la coherencia en resúmenes asistidos por computadora. En: (ELRA) ELRA (ed) Actas de la sexta conferencia internacional sobre recursos lingüísticos y evaluación (LREC'08) Hassel M (2007) Resumen de texto automático fácil de usar y portátil. Tesis doctoral, Departamento de Numeración Análisis ical y Ciencias de la Computación, Instituto Real de Tecnología
- He L, Sanocki E, Gupta A, Grudin J (1999) Auto-resumen de presentaciones de audio y vídeo. En: MULTIMEDIA '99: actas de la séptima conferencia internacional ACM sobre multimedia (Parte 1). págs. 489–498
- He T, Chen J, Gui Z, Li F (2008) CCNU en TAC 2008: proceder sobre el uso del método semántico para automatización resumen En: Actas de la conferencia de análisis de texto (TAC)
- Hearst MA (1997) TextTiling: segmentación de texto en pasajes de subtemas de varios párrafos. *lingüista informático* 23(1):33–64
- Hirao T, Okumura M, Fukushima T, Nanba H (2005) Desafío 3 de resumen de texto: evaluación de resumen de texto uación en el taller NTCIR 4. En: Actas del cuarto taller NTCIR sobre investigación en tecnologías de acceso a la información, recuperación de información, respuesta a preguntas y resumen. págs. 407–411
- Hovy E, Lin CY (1999) Resumen de texto multilingüe automatizado y su evaluación. Reporte técnico Instituto de Ciencias de la Información, Universidad del Sur de California
- Hovy E, Lin CY, Zhou L, Fukumoto J (2006) Evaluación de resumen automatizada con elementos básicos. En: Actas de la 5ª conferencia internacional sobre recursos lingüísticos y evaluación (LREC)

- Jaoua M, Hamadou AB (2003) Resumen automático de texto de artículos científicos basado en la clasificación de la Población del extracto. En: Actas de lingüística computacional y procesamiento inteligente de textos, 4ª conferencia internacional. págs. 623-634
- Jing H (2002) Uso de modelos ocultos de Markov para descomponer resúmenes escritos por humanos. *lingüista informático* 28(4):527-543
- Jing H, McKeown KR (2000) Resumen de texto basado en cortar y pegar. En: Actas del 1er Norte Capítulo estadounidense de la conferencia de la asociación de lingüística computacional. pp 178-185 Kaisser M, Hearst MA, Lowe JB (2008) Mejora de la calidad de los resultados de búsqueda mediante la personalización de la longitud de los resúmenes. En: Actas de ACL-08: HLT. págs. 701-709
- Kan MY, Klavans JL (2002) Uso de técnicas bibliotecarias en resúmenes automáticos de texto para información recuperación. En: JCDL '02: actas de la segunda conferencia conjunta ACM/IEEE-CS sobre bibliotecas digitales. págs. 36-45
- Kan MY, Klavans JL, McKeown KR (2002) Uso de la bibliografía anotada como recurso para resumen En: Actas de la conferencia sobre evaluación y recursos lingüísticos. pp 1746-1752
- Katragadda R (2010) GEMS: modelado generativo para la evaluación de resúmenes. En: Actas de la XI Congreso Internacional de Lingüística Computacional y Procesamiento Inteligente de Textos, CICLing. págs. 724-735
- Kazantseva A (2006) Un enfoque para resumir cuentos. En: Actas del trabajo de investigación del estudiante-tienda en la 11ª conferencia del capítulo europeo de la asociación de lingüística computacional. págs. 55-62
- Ker SJ, Chen JN (2000) Una categorización de texto basada en la técnica de resumen. En: Actas de la Taller ACL-2000 sobre avances recientes en procesamiento de lenguaje natural y recuperación de información. págs. 79-83
- Khan AU, Khan S, Mahmood W (2005) MRST: una nueva técnica para el resumen de información. En el segunda conferencia mundial de enformatika, WEC'05. págs. 249-252
- Kumar C, Pingali P, Varma V (2008) Generación de resúmenes personalizados utilizando documentos web disponibles públicamente. En: Actas de la conferencia internacional IEEE/WIC/ACM de 2008 sobre inteligencia web y la conferencia internacional sobre tecnología de agentes inteligentes. págs. 103-106
- Kumar M, Das D, Agarwal S, Rudnicky A (2009) Resumen de eventos no textuales mediante la aplicación de aprender a generar lenguaje basado en plantillas. En: Actas del taller de 2009 sobre generación y resumen de idiomas (UCNLG+ Sum 2009). págs. 67-71
- Kuo JJ, Chen HH (2008) Generación de resúmenes de documentos múltiples: uso de palabras informativas y de eventos. MCA Proceso de información sobre idiomas transasiáticos (TALIP) 7(1):1-23
- Kupiec J, Pedersen J, Chen F (1995) Un resumen de documentos entrenables. En: SIGIR '95: Actas de la 18ª conferencia anual internacional ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 68-73
- Lapata M, Barzilay R (2005) Evaluación automática de la coherencia del texto: modelos y representaciones. En: Pro-actas de la 19ª conferencia internacional conjunta sobre inteligencia artificial. pp 1085-1090 Lerman K, McDonald R (2009) Resumen contrastivo: un experimento con reseñas de consumidores. En: Pro-actas de las tecnologías del lenguaje humano: la conferencia anual de 2009 del capítulo norteamericano de la asociación de lingüística computacional, volumen complementario: artículos breves. pp 113-116 Lerman K, Blair-Goldensohn S, McDonald R (2009) Resumen de sentimientos: evaluación y aprendizaje del usuario preferencias En: Actas de la 12.ª conferencia del capítulo europeo de la ACL (EACL 2009). págs. 514-522
- Li S, Ouyang Y, Wang W, Sun B (2007) Resumen de varios documentos mediante regresión de vectores de soporte. En: El taller de comprensión de documentos (presentado en la HLT/NAACL). Rochester, Nueva York, EE. UU. Li S, Wan W, Wang C (2008) Tarea de resumen de actualización de TAC 2008 de ICL. En: Actas del análisis del texto conferencia (TAC)
- Li S, Wang W, Zhang Y (2009) Resumen de actualización de Tac 2009 de icl. En: Actas del análisis del texto conferencia (TAC)
- Lin CY (2004) ROUGE: un paquete para la evaluación automática de resúmenes. En: Actas de texto ACL taller de resúmenes. págs. 74-81
- Lin CY, Hovy E (2000) La adquisición automatizada de firmas de temas para resúmenes de texto. En: Actas de la 18ª conferencia sobre lingüística computacional. págs. 495-501
- Liu F, Liu Y (2008) Correlación entre ROUGE y evaluación humana de resúmenes de reuniones extractivas. En: Actas de ACL-08: HLT, artículos breves. págs. 201-204
- Liu M, Yu B, Fang F, Sun H (2009) Tarea de resumen de actualización de TAC 2009 de WUST. En: Actas de la conferencia de análisis de texto (TAC)
- Lloret E, Palomar M (2009) Una combinación gradual de características para construir sistemas de resumen automático. En: Actas de la 12ª conferencia internacional sobre texto, habla y diálogo (TSD). págs. 16-23

- Lloret E, Ferrández O, Muñoz R, Palomar M (2008) Un enfoque de resumen de texto bajo la influencia de vinculación textual. En: Actas del quinto taller internacional sobre procesamiento del lenguaje natural y ciencia cognitiva (NLPCS 2008). págs. 22–31
- Lloret E, Balahur A, Palomar M, Montoyo A (2009) Hacia la construcción de un resumen de opinión competitiva sistema: retos y claves. En: Actas de la NAACL. Taller de Investigación Estudiantil y Consorcio de Doctorado. págs. 72–77
- Lloret E, Saggion H, Palomar M (2010) Experimentos sobre clasificación de opinión basada en resumen. En: Procedimientos del taller NAACL HLT 2010 sobre enfoques computacionales para el análisis y la generación de emociones en el texto. págs. 107–115
- Luhn HP (1958) La creación automática de resúmenes de literatura. En: Avances en el resumen automático de textos. págs. 15–22
- Mani I (2001) Resumen automático. John Benjamins Publishing Co. Ámsterdam, Filadelfia, EE. UU. Mani I (2001b) Evaluación de resumen: una descripción general. En: Actas del capítulo norteamericano de la Asociación de Lingüística Computacional (NAACL). Taller de Resumen Automático Mani I, Maybury MT (1999) Avances en resumen automático de textos. The MIT Press, Cambridge Mani I, House D, Klein G, Hirschman L, Firmin T, Sundheim B (1999) El texto de TIPSTER SUMMAC resume evaluación de marización. En: Actas de la novena conferencia sobre el capítulo europeo de la asociación de lingüística computacional. págs. 77–85
- Mani I, Klein G, House D, Hirschman L, Firmin T, Sundheim B (2002) SUMMAC: resumen de un texto evaluación. *Nat Lang Inglés* 8 (1): 43–68
- Mann WC, Thompson SA (1988) Teoría de la estructura retórica: Hacia una teoría funcional de la organización del texto. *Texto* 8(3):243–281
- Manning CD, Raghavan P, Schtze H (2008) Introducción a la recuperación de información. Prensa de la Universidad de Cambridge, Nueva York, Nueva York, EE. UU.
- Marcu D (1999) Los árboles de discurso son buenos indicadores de la importancia en el texto. En: Avances en texto automático resumen págs. 123–136
- McCargar V (2005) Enfoques estadísticos para el resumen automático de texto. *Bull Am Soc Inf Sci Technol* 30(4):21–25
- Medelyan O (2007) Computación de cadenas léxicas con agrupamiento de grafos. En: Actas del estudiante ACL 2007 taller de investigación págs. 85–90
- Mihalcea R (2004) Algoritmos de clasificación basados en gráficos para la extracción de oraciones, aplicados al resumen de texto. En: Actas de la ACL 2004 sobre carteles interactivos y sesiones de demostración. pág. 20
- Mihalcea R, Ceylan H (2007) Exploraciones en el resumen automático de libros. En: Actas de la junta conferencia sobre métodos empíricos en el procesamiento del lenguaje natural y el aprendizaje computacional del lenguaje natural (EMNLP-CoNLL). págs. 380–389
- Mille S, Wanner L (2008) Resumen multilingüe en la práctica: el caso de las reivindicaciones de patentes. En: Actas de la 12ª Conferencia de la Asociación Europea de Traducción Automática. págs. 120–129
- Minel JL, Nugier S, Piat G (1997) ¿Cómo apreciar la calidad del resumen automático de texto? Ejemplos de Protocolos FAN y MLUCe y sus resultados en SERAPHIN. En: Actas del taller de resumen de texto escalable inteligente en conjunto con el capítulo europeo de la asociación de lingüística computacional (EACL). págs. 25–30
- Mitkov R, Evans R, Orasan C, Ha LA, Pekar V (2007) Resolución de anáfora: en qué medida ayuda aplicaciones de la PNL? En: Actas del sexto coloquio de anáfora discursiva y resolución anáfora. págs. 179–190
- Mohammad S, Dorr B, Egan M, Hassan A, Muthukrishnan P, Qazvinian V, Radev D, Zajic D (2009) Usando citas para generar levantamientos de paradigmas científicos. En: Actas de las tecnologías del lenguaje humano: la conferencia anual de 2009 del capítulo norteamericano de la asociación de lingüística computacional. págs. 584–592
- Mori T (2002) Relación de ganancia de información como peso del término: el caso del resumen de los resultados de IR. En: Actas de la 19ª conferencia internacional de lingüística computacional. págs. 1–7
- Mori T, Nozawa M, Asada Y (2004) Resumen de múltiples documentos centrado en respuestas múltiples usando un cuestionario-motor de respuesta En: COLING '04: actas de la 20ª conferencia internacional sobre lingüística computacional. págs. 439–445
- Mori T, Nozawa M, Asada Y (2005) Resumen de múltiples documentos centrado en respuestas múltiples utilizando una pregunta-motor de respuesta *ACM Trans Asian Lang Inf Process (TALIP)* 4(3):305–320
- Morris AH, Kasper GM, Adams DA (1992) El efecto y las limitaciones de la condensación automática de texto en Rendimiento de comprensión de lectura. *Res. del sistema de inf.* 3(1):17–35
- Nastase V, Milne D, Filippova K (2009) Resumiendo con conocimiento enciclopédico. En: Actas de la conferencia de análisis de texto (TAC)

- Nenkova A (2005) Resumen automático de texto de noticias: lecciones aprendidas del documento bajo conferencia permanente. En: Actas de la Asociación Estadounidense de Inteligencia Artificial (AAAI). págs. 1436-1441
- Nenkova A (2006) Evaluación de resúmenes para texto y discurso: problemas y enfoques. En: INTERSPEECH-2006, artículo 2079-Wed1WeS.1
- Nenkova A, Siddharthan A, McKeown K (2005) Aprendizaje automático del estado cognitivo para documentos múltiples resumen de noticias. En: HLT '05: actas de la conferencia sobre tecnología del lenguaje humano y métodos empíricos en el procesamiento del lenguaje natural. págs. 241-248
- Neto JL, Santos A, Kaestner CAA, Freitas AA (2000) Generación de resúmenes de texto a través de la importancia relativa de temas. En: IBERAMIA-SBIA '00: Actas de la conferencia internacional conjunta, 7ª Conferencia Iberoamericana de IA. págs. 300-309
- Okumura M, Fukushima T, Nanba H, Hirao T (2004) Desafío de resumen de texto 2 resumen de texto evaluación en NTCIR taller 3. Foro SIGIR 38(1):29-38
- Orăsan C (2004) La influencia de los pronombres personales para el resumen automático de artículos científicos. En: Actas del coloquio anáfora discursiva y resolución anáfora. págs. 127-132
- Orăsan C (2007) Resolución de anáforas pronominales para resúmenes de texto. En: Actas de los avances recientes sobre el procesamiento del lenguaje natural. págs. 430-436
- Orăsan C (2009) Evaluación comparativa de los métodos de ponderación de términos para el resumen automático. Cuant J Lingüista 16 (1): 67-95
- Orăsan C, Pekar V, Hasler L (2004) Una comparación de los métodos de resumen basados en la estimación de la especificidad del término En: Actas de la cuarta conferencia internacional sobre recursos lingüísticos y evaluación (LREC2004). págs. 1037-1041. Disponible en: <http://clg.wlv.ac.uk/papers/orasan-04a.pdf>
- Over P, Liggett W (2002) Introducción a DUC: una evaluación intrínseca del resumen de texto de noticias genéricas sistemas En: El taller de comprensión de documentos
- Over P, Dang H, Harman D (2007) DUC en contexto. Gestión de procesos de información 43(6):1506-1520
- Owczarzak K (2009) DEPEVAL(summ): evaluación basada en dependencias para resúmenes automáticos. En: Pro-actas de la conferencia conjunta de la 47.ª reunión anual de la ACL y la 4.ª conferencia conjunta internacional sobre procesamiento del lenguaje natural de la AFNLP. págs. 190-198
- Pang B, Lee L (2005) Ver estrellas: explotar las relaciones de clase para la categorización de sentimientos con respecto a escalas de calificación. En: Actas de la asociación de lingüística computacional. pp 115-124 Pang B, Lee L (2008) Minería de opiniones y análisis de sentimiento. Found Trends Inf Retr 2(1-2):1-135 Papineni K, Roukos S, Ward T, Zhu WJ (2002) BLEU: un método para la evaluación automática de la traducción automática En: Actas de la 40ª reunión anual de la asociación de lingüística computacional. pp 311-318 Passonneau RJ (2010) Evaluación formal y funcional del método de la pirámide para la evaluación resumida del contenido uación Nat Lang Eng 16 (2): 107-131
- Pitler E, Nenkova A (2008) Revisando la legibilidad: un marco unificado para predecir la calidad del texto. En: Pro-actas de la conferencia de 2008 sobre métodos empíricos en el procesamiento del lenguaje natural. pp 186-195
- Plaza L, Díaz A, Gervás P (2008) Concept-graph based biomedical automatic Summarization Using Ontologies En: Coling 2008: Actas del tercer taller de gráficos de texto sobre algoritmos basados en gráficos para el procesamiento del lenguaje natural. págs. 53-56
- Plaza L, Lloret E, Aker A (2010) Mejora de los subtítulos automáticos de imágenes mediante técnicas de resumen de texto. En: Actas de la 13.ª conferencia internacional sobre texto, voz y diálogo (TSD) Qazvinian V, Radev DR (2008) Resumen de artículos científicos mediante redes de resumen de citas. En: Pro-actas de la 22ª conferencia internacional sobre lingüística computacional (Coling 2008). pp 689-696 Radev DR, Fan W (2000) Resumen automático de listas de aciertos de motores de búsqueda. En: Actas de la ACL-2000 taller sobre avances recientes en el procesamiento del lenguaje natural y recuperación de información. pp 99-109 Radev DR, McKeown KR (1998) Generación de resúmenes en lenguaje natural a partir de múltiples fuentes en línea. Lingüista informático 24 (3): 470-500
- Radev DR, Tam D (2003) Evaluación de resumen utilizando utilidad relativa. En: CIKM '03: Actas de la XII congreso internacional sobre gestión de la información y el conocimiento. págs. 508-511
- Radev DR, Blair-Goldensohn S, Zhang Z (2001) Experimentos en resúmenes de documentos únicos y múltiples utilizando MEAD. En: Primera conferencia de comprensión de documentos. págs. 1-7
- Radev DR, Hovy E, McKeown K (2002) Introducción al número especial sobre resúmenes. lingüista informático 28(4):399-408
- Saggion H (2008) Resumen automático: una descripción general. Revue française de linguistique appliquée XIII(1). págs. 63-81
- Saggion H (2009) Un algoritmo de clasificación para predecir la estructura de los resúmenes. En: Actas de la 2009 taller sobre generación y síntesis de lenguaje (UCNLG+Sum 2009). pp 31-38 Saggion H, Funk A (2009) Extracción de opiniones y hechos para Business Intelligence. RNTI E-17:119-146

- Saggion H, Lapalme G (2000) Análisis selectivo para la abstracción automática: evaluación de la indicatividad y aceptabilidad. En: Procedimientos de acceso a la información multimedia basada en contenido (RIAIO). pp 747-764
- Saggion H, Lloret E, Palomar M (2010) Uso de resúmenes de texto para predecir escalas de calificación. En: Actas de el primer taller sobre enfoques computacionales para la subjetividad y el análisis de sentimientos (WASSA) Sakai T, Sparck-Jones K (2001) Resúmenes genéricos para la indexación en la recuperación de información. En: SIGIR '01: actas de la 24ª conferencia internacional anual ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 190-198
- Saravanan M, Ravindran B, Raman S (2006) Mejora del resumen de documentos legales mediante modelos gráficos. En: Procedimientos de conocimiento legal y sistemas de información—JURIX 2006: la 19ª conferencia anual sobre conocimiento legal y sistemas de información. págs. 51-60
- Sauper C, Barzilay R (2009) Generación automática de artículos de wikipedia: un enfoque consciente de la estructura. En: Actas de la conferencia conjunta de la 47.ª reunión anual de la ACL y la 4.ª conferencia internacional conjunta sobre el procesamiento del lenguaje natural de la AFNLP. págs. 208-216
- Schilder F, Kondadadi R (2008) FastSum: resumen rápido y preciso de varios documentos basado en consultas. En: Actas de ACL-08: HLT, artículos breves. págs. 205-208
- Schilder F, Kondadadi R, Leidner JL, Conrad JG (2008) Thomson reuters en TAC 2008: filtrado agresivo con FastSum para actualización y resumen de opiniones. En: Actas de la conferencia de análisis de texto (TAC)
- Schlesinger JD, Okurowski ME, Conroy JM, O'Leary DP, Taylor A, Hobbs J, Wilson H (2002) El rendimiento de la máquina en el contexto del rendimiento humano para el resumen de varios documentos. En: Actas del taller DUC 2002 sobre resumen de texto
- Sebastiani F (2002) Aprendizaje automático en categorización de texto automatizada. ACM Comput Surv 34(1):1-47
- Shen D, Chen Z, Yang Q, Zeng HJ, Zhang B, Lu Y, Ma WY (2004) Clasificación de páginas web a través de resúmenes. En: SIGIR '04: actas de la 27.ª conferencia internacional anual ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 242-249
- Shen D, Yang Q, Chen Z (2007) Reducción de ruido mediante resumen para la clasificación de páginas web. Información Gestión de procesos 43(6):1735-1747
- Shi Z, Melli G, Wang Y, Liu Y, Gu B, Kashani MM, Sarkar A, Popowich F (2007) Respuesta a preguntas resumen de múltiples documentos biomédicos. En: CAI '07: actas de la 20ª conferencia de la sociedad canadiense para estudios computacionales de inteligencia sobre avances en inteligencia artificial. págs. 284-295
- Sjöbergh J (2007) Las versiones anteriores del sistema de evaluación de resumen ROUGEeval eran más fáciles de engañar. Gestión de procesos de información 43(6):1500-1505
- Spärck Jones K (1999) Resumen automático: factores y direcciones. En: Avances en resumen automático de texto marización págs. 1-14
- Spärck Jones K (2007) Resumen automático: el estado del arte. Inf Process Manag 43(6):1449-1481
- Spärck-Jones K, Galliers JR (eds) (1996) Evaluación de los sistemas de procesamiento del lenguaje natural, un análisis y revisión, notas de conferencias en informática, vol 1083. Springer, Berlín
- Steinberger J, Poesio M, Kabadjov MA, Ježek K (2007) Dos usos de la resolución anáfora en resumen. Gestión de procesos de información 43(6):1663-1680
- Steinberger J, Jezek K, Sloup M (2008) Resumen de temas web. En: Actas de la 12ª edición internacional Congreso sobre edición electrónica. págs. 322-334
- Strzalkowski T, Harabagiu S (2007) Avances en la respuesta a preguntas de dominio abierto (texto, habla y lenguaje) Tecnología). Springer-Verlag New York, Inc., Secaucus, Nueva Jersey, EE. UU.
- Sun JT, Shen D, Zeng HJ, Yang Q, Lu Y, Chen Z (2005) Resumen de páginas web utilizando datos de clics. En: SIGIR '05: actas de la 28.ª conferencia internacional anual ACM SIGIR sobre investigación y desarrollo en recuperación de información. págs. 194-201
- Svore KM, Vanderwende L, Burges CJ (2007) Mejorar el resumen de un solo documento al combinar RankNet y fuentes de terceros. En: Actas de la conferencia conjunta sobre métodos empíricos en el procesamiento del lenguaje natural y el aprendizaje computacional del lenguaje natural (EMNLP-CoNLL). pp 448-457
- Sweeney S, Crestani F, Losada DE (2008) Muéstrame más: Resumen de longitud incremental usando novedad detección. Gestión de procesos de información 44(2):663-686
- Szlávik Z, Tombros A, Lalmas M (2006) Investigación del uso del resumen para la recuperación interactiva de XML. En: SAC '06: Actas del simposio ACM 2006 sobre computación aplicada. pp 1068-1072
- Teng Z, Liu Y, Ren F, Tsuchiya S, Ren F (2008) Resumen de un documento único basado en la identificación del tema local tificación y frecuencia de palabras. En: MICAI '08: actas de la séptima conferencia internacional mexicana sobre inteligencia artificial 2008. págs. 37-41. <http://dx.doi.org/10.1109/MICAI.2008.12>
- Teufel S, Halteren Hv (2004) Evaluación del contenido de la información mediante análisis factoides: anotación humana y estabilidad En: Actas de la conferencia de 2004 sobre métodos empíricos en el procesamiento del lenguaje natural. págs. 419-426

- Teufel S, Moens M (2002) Resumen de artículos científicos: experimentos con relevancia y estadística retórica. *tus.. Comput Linguist* 28(4):409-445
- Titov I, McDonald R (2008) Un modelo conjunto de calificaciones de texto y aspectos para resumir sentimientos. En: *Pro-actas de ACL-08: HLT*. págs. 308-316
- Torres-Moreno JM, St-Onge PL, Gagnon M, El-Bze M, Bellot P (2009) Sistema de resumen automático junto con un sistema de preguntas y respuestas (QAAS). *Lenguaje informático de noticias de PNL*. <http://arxiv.org/abs/0905.2990v1>
- Trappey A, Trappey C, Wu CY (2009) Resumen automático de documentos de patente para conocimiento colaborativo. *Sistemas y servicios de borde. J Syst Sci Syst Eng* 18(1):71-94
- Trappey AJC, Trappey CV (2008) Un método de gestión del conocimiento de I+D para el resumen de documentos de patentes. *rización Sistema de datos de gestión industrial* 108(2):245-257
- Tseng YH, Lin CJ, Lin YI (2007) Técnicas de minería de textos para el análisis de patentes. *Gestión de procesos de información* 43(5):1216-1247
- Vadlapudi R, Katragadda R (2010a) Sobre la evaluación automatizada de la legibilidad de los resúmenes: captura de gramática, maticidad, enfoque, estructura y coherencia. En: *Actas del taller de investigación estudiantil NAACL HLT 2010*. págs. 7-12
- Vadlapudi R, Katragadda R (2010b) Evaluación cuantitativa de la gramaticalidad de los resúmenes. En: *Proceder-ings de la 11ª conferencia internacional sobre lingüística computacional y procesamiento inteligente de textos, CICLing*. págs. 736-747
- Van Dijk TA (1972) Algunos aspectos de las gramáticas de texto. *Un estudio en Lingüística Teórica y Poética*, La Haya-parís, Mouton
- Van Rijsbergen CJ (1981) *Recuperación de información*. Elsevier, Ámsterdam
- Wan X, Yang J, Xiao J (2007) Hacia un enfoque unificado basado en gráficos de afinidad para varios documentos múltiples resúmenes. En: *Actas de la 11ª conferencia europea*. págs. 297-308
- Wang C, Long L, Li L (2008) Evaluación basada en HowNet para el resumen de texto en chino. En: *Actas de la conferencia internacional sobre procesamiento de lenguaje natural e ingeniería de software*. pp 82-87 Wilson T, Wiebe J, Hoffmann P (2005) Reconociendo la polaridad contextual en el análisis de sentimiento a nivel de frase. En: *Actas de los métodos empíricos en el procesamiento del lenguaje natural*. págs. 347-354
- Witte R, Krestel R, Bergler S (2007) Generating update resúmenes para DUC 2007. En: *The document under-taller de pie (presentado en el HLT/NAACL)*
- Wong KF, Wu M, Li W (2008) Resumen extractivo mediante aprendizaje supervisado y semisupervisado. En: *Actas de la 22ª conferencia internacional sobre lingüística computacional (Coling 2008)*. págs. 985-992
- Yu J, Reiter E, Hunter J, Mellish C (2007) Elección del contenido de resúmenes textuales de grandes series temporales. *Conjuntos de datos. Nat Lang Inglés* 13 (1): 25-49
- Zajic D, Dorr BJ, Lin J, Schwartz R (2007) Reducción de candidatos múltiples: compresión de oraciones como herramienta para tareas de resumen de documentos. *Gestión de procesos de información* 43 (6): 1549-1570
- Zajic DM, Dorr BJ, Lin J (2008) Técnicas de resumen de un solo documento y de varios documentos para correo electrónico. *subprocesos que usan compresión de oraciones. Gestión de procesos de información* 44(4): 1600-1610
- Zechner K, Waibel A (2000) DiaSumm: resumen flexible de diálogos espontáneos en libre dominios. En: *Actas de la 18ª conferencia sobre lingüística computacional*. págs. 968-974
- Zhou L, Ticea M, Hovy E (2004) Resumen biográfico de varios documentos. En: *Actas de la conferencia sobre métodos empíricos en el procesamiento del lenguaje natural*. págs. 434-441
- Zhou L, Lin CY, Munteanu DS, Hovy E (2006) ParaEval: uso de paráfrasis para evaluar resúmenes de forma automática. *calladamente En: Actas de la conferencia de tecnología del lenguaje humano/Asociación Norteamericana de Lingüística Computacional*. págs. 447-454
- Zhuang L, Jing F, Zhu XY (2006) Minería y resumen de reseñas de películas. En: *CIKM '06: actas de la 15ª conferencia internacional ACM sobre gestión de la información y el conocimiento*. págs. 43-50