



ÍNDICE

AGRADECIMIENTOS	7
RESUMEN	9
INTRODUCCIÓN	11
CAPÍTULO 1. ESTADO DEL ARTE DE LOS RESÚMENES DOCUMENTALES	19
Orígenes de los resúmenes documentales	19
Las primeras publicaciones científicas:	
las revistas de resúmenes	22
Definiciones de resumen	27
Tipologías de resúmenes	34
Automatización del resumen. Tipologías	48
El hiperresumen	51
CAPÍTULO 2. EL HIPERTEXTO: SU IMPLICACIÓN	
EN RESÚMENES AUTOMÁTICOS	53
Definiciones asociadas al hipertexto	53
El prefijo hiper	53
Hipertexto	54

Multimedia	55
Hipermedia	55
Antecedentes y surgimiento del hipertexto	56
Características del hipertexto	58
Organización de la información en ambiente web	63
Ontologías	64
Taxonomías	65
Folksonomías y stackonomías	67
Topic Maps	68
Factores contextuales que intervienen en la elaboración de resúmenes	69
CAPÍTULO 3. TENDENCIAS EN LA ELABORACIÓN	
DE RESÚMENES AUTOMÁTICOS	75
Evolución de sistemas para el procesamiento de información en ambiente web	75
Métodos empleados en la elaboración de resúmenes automáticos	76
Procesamiento del lenguaje natural	78
Sistemas actuales para la elaboración de resúmenes automáticos: evolución y tendencias	82
Implicación de las minerías de texto, audio y video en los resúmenes automáticos	86
Aportes de la inteligencia artificial a los resúmenes automáticos	89
Experiencia cubana: sistema Corpus Miner	89
Sistemas que elaboran resúmenes mono y multidocumento	90
Hiperresúmenes: ¿utopía o realidad?	92
Perspectiva futura de los sistemas automáticos	93
BIBLIOGRAFÍA	97
ANEXO	111



AGRADECIMIENTOS

Ante todo a mi familia, y en particular a Javi y a mi tía Elvia.

A todo mi valioso equipo de trabajo, en especial a:

- Lourdes, por sus pinceladas de creatividad y buen ánimo, que me ayudaron a seguir adelante.
- Reina, Yosvany, Martica, los subdirectores docente y de I+D+I, por propiciarme las condiciones necesarias.
- A Marily.
- A Fernando Mendieta, Ananay y Elsa Wong por los arreglos técnicos.
- A los profesores de la Cátedra de Idiomas Paco, Niurka Hernández y Yusem, por la corrección de las traducciones.

Al Dr. José Llovera, casi de mi familia, por sus aportes y sabios consejos.

A la Dra. Ania Hernández Quintana y al MSc. Amed Leyva Mederos, por el entusiasmo que me impregnaron.

Al director del IBICT, Dr. Emir José Suaiden, por contribuir, con sus vastos conocimientos, a esta investigación.

A todos mis profesores y compañeros del Grupo 6 de la maestría en BCI y a su coordinador, el Dr. José Luis Rojas, por la posibilidad de momentos tan enriquecedores.

A los Drs. Alfonso López Yepes y José Antonio Moreira, por responder tan puntualmente a mis mensajes cuando les solicitaba información.

A todos, muchas gracias.



RESUMEN

El desarrollo científico técnico que tuvo lugar a partir del siglo XVII propició la aparición de las primeras revistas y servicios de resúmenes, primero en el continente europeo, hasta que a inicios del siglo XIX comienza a desarrollarse en los Estados Unidos. El siglo XX se caracterizó por un crecimiento significativo de la producción de publicaciones y servicios de resúmenes, convirtiéndose en las fuentes de referencia más solicitadas a nivel internacional, en su versión automatizada. Con el surgimiento de internet en la década del '60 surge una nueva modalidad de resúmenes, realizados de forma automática, que, al igual que los resúmenes elaborados manualmente, presentan serios problemas terminológicos y conceptuales. Aunque los sistemas para elaborar resúmenes automáticos han logrado avances significativos en cuanto a brevedad, legibilidad y precisión, están lejos de hacer comprensiones automáticas de los textos, de ahí que les falte coherencia y presenten ambigüedades lingüísticas e interpretativas. La perspectiva futura es que logren representar los conceptos desde diferentes dimensiones y analizar una realidad

determinada de forma evolutiva, pues las características del medio hipertextual exigen nuevas o renovadas formas representacionales adaptables a este entorno. Los hiperresúmenes son la perspectiva futura a la que se desea llegar, pero aún ni siquiera están bien definidos.



INTRODUCCIÓN

Los cambios científico-tecnológicos de finales del siglo XIX han propiciado el aumento excesivo de la producción documentaria a escala internacional. Para paliar esta situación se han creado mecanismos para el control bibliográfico de parte de estos materiales, pero más que eso se necesita que las nuevas formas de organización y representación de la información que van surgiendo sean capaces de sintetizar los altos volúmenes informativos que día a día se producen y cuya tendencia es la integración de estructuras, formas y formatos, donde se incluyen nuevos aditamentos como voz, imagen, sonido y movimiento.

Una de las herramientas más utilizadas como solución a la incapacidad de procesar manualmente toda la información que existe en determinada rama del saber es el resumen, ya que permite obtener información abreviada sobre el contenido de los documentos, minimizando el tiempo dedicado a la búsqueda y recuperación de información.

La existencia del resumen documental es una necesidad para cualquier profesional, ya que evita tener que revisar numerosos documentos para encontrar los que realmente sean pertinentes a sus necesidades informativas.

Además, facilita la indización y la transferencia de información desde el punto de vista del idioma (por lo general todas las investigaciones prestigiosas tienen un resumen en inglés). Las revistas de resúmenes constituyen uno de los medios por excelencia de presentación de este producto, y brindan información, retrospectiva o reciente, sobre las publicaciones de cualquier especialidad. Una de las tendencias en este campo es su automatización, convirtiéndose en un instrumento eficaz para la indización automática, porque *"el resumen sigue siendo el metadato de más complejidad y relevancia informativa"* (Lancaster y Pinto Molina, 2001: 134).

En las últimas cuatro décadas el resumen documental ha sido objeto de estudio fundamentalmente en los países de habla inglesa (Inglaterra y Estados Unidos). Otros países que han avanzado en el tema son Holanda, Alemania, Francia, la antigua URSS y Canadá. Entre los países de habla hispana se destaca España en las primeras posiciones, muestra de ello son los trabajos de autores de la talla de María Pinto Molina y José Antonio Moreiro, por sólo citar a los dos que más se han pronunciado sobre el tema. La mayoría de las investigaciones que se han realizado en este sentido han estado enfocadas al estudio del resumen como práctica, descuidando otros contextos, como el lingüístico, cuestión que analiza profundamente Pinto Molina. Latinoamérica ha sido una de las regiones que ha experimentado menor desarrollo en el tema.

La experiencia de mi país, Cuba, en esta actividad data de la década del '70, aunque ha sido un tema débilmente abordado por especialistas de la profesión desde el punto de vista epistemológico, prevaleciendo su estudio desde la óptica de los servicios de información. En este sentido pueden citarse los trabajos de diploma "Estudio sistematizado del comportamiento de los resúmenes", de María de los A. Lamazares (1988); "Tendencias que se perfilan en la elaboración de publicaciones de resúmenes cubanas: tipología, estructura y representación", de Lissette Guerra Jamud (1989); "Elaboración de resúmenes en algunas instituciones informativas cubanas: apuntes, caracterización y análisis", de las autoras Mylen Pombo Tanda y Yamilet Trelles Mutis (1993), entre otras. Todas estas investigaciones, en el período comprendido entre 1987 y 1993, fueron tutoradas por Sonia Gironelly Pérez.

Una de las aristas más estudiadas, pero a la vez sumamente polémica, es la relacionada con la tipología de los resúmenes, figurando entre los más mencionados los informativos, indicativos, mixtos, de autor, estructurados, entre otros. Según estudio de Lissette Guerra (1989) en su tesis de licenciatura, resulta evidente que en la década de los '80 los resúmenes informativos eran los predominantes en Cuba y las fuentes más empleadas para su elaboración estaban en los idiomas español y ruso.

Con el auge de las tecnologías de la información y las comunicaciones comienzan a sobresalir otros tipos, amparados por los cambios sustanciales que tuvo la información científica y tecnológica derivados de la aplicación de la informática y las telecomunicaciones, tales como los hiperresúmenes y los resúmenes automatizados, surgidos con la explosión y desarrollo de los sistemas de recuperación de información en el ciberespacio, pero que aún están lejos de solucionar los graves problemas de la recuperación de información y mucho menos el fenómeno de la *infoxicación*, debido a que carecen de técnicas léxico-sintácticas complejas para realizar la selección, lo que repercute en las acciones lógico-semánticas de la interpretación y en los procesos de producción documentaria que tienen un amplio sentido pragmático.

La presente investigación realiza un estudio histórico-evolutivo del comportamiento de los resúmenes documentales, desde sus orígenes hasta la actualidad, siguiendo una óptica conceptual general, que parte del análisis de las definiciones de resúmenes documentales, sus tipologías y especificidades, entre otros aspectos que buscan develar tendencias contemporáneas en la elaboración de resúmenes.

En este sentido, se pretende que la exposición sistematizada de estas cuestiones constituya la base para que, en un futuro cercano, otros estudiosos continúen esta trayectoria, y demuestren, de forma más categórica, si tiene sentido la existencia de tantas definiciones y tipos de resúmenes, muchas de las cuales, aunque con nombres distintos, hacen referencia a un mismo fenómeno y en su mayoría no se centran en los aspectos medulares de este tipo de representación. La comprensión de esta parte tan elemental (tradicional) constituye una base insoslayable para el posterior estudio de los resúmenes documentales automatizados.

Todo lo anterior con el propósito de lograr mayor claridad en el tema y contribuir, además, a la elaboración y edición de un texto básico para la enseñanza superior en la carrera Bibliotecología y Ciencia de la Información (BCI) y profesionales de la especialidad en general.

También, además de actualizar la temática en cuestión, se trascienden las fronteras de lo tradicional hasta adentrarse en un entorno completamente nuevo, pues en la medida en que las tecnologías han irrumpido en la forma de elaborar los resúmenes documentales, el procesador se ha visto obligado a modificar sus modelos mentales y ampliar sus competencias profesionales, con vista a lograr resúmenes de mayor calidad en entornos extremadamente dinámicos.

Por otra parte, esta investigación constituirá un basamento teórico importante para estudios posteriores relacionados con el tema, sobre todo en el área latinoamericana, cuya contribución a este segmento del conocimiento

ha sido escasa y donde la localización y recuperación de información actualizada y evaluada sobre la cuestión se hace difícil.

Tomando como base los planteamientos anteriores, se elaboró la siguiente **pregunta de investigación**:

¿Cuáles rasgos han caracterizado, desde una mirada histórica, la evolución de la producción de resúmenes documentales, desde sus orígenes hasta la actualidad?

Objetivo general

Estudiar el comportamiento evolutivo de la producción de resúmenes documentales desde sus orígenes hasta la actualidad.

Objetivos específicos

1. Estudiar las tipologías de resúmenes y sus complejidades.
2. Caracterizar la plataforma hipertextual y su implicación en resúmenes automáticos.
3. Identificar tendencias en la elaboración de resúmenes automáticos.

Métodos utilizados

Para el desarrollo de la presente investigación, de corte teórico-conceptual, se emplearon los métodos que a continuación se describen:

- Análisis documental clásico: desde la óptica del análisis de contenido, permite seleccionar las ideas más relevantes de un documento con el propósito de expresar su contenido sin ambigüedades con vista a la recuperación de la información. Desde una mirada formal o externa, identifica los elementos objetivos que distinguen a los documentos, como autor, título, fecha de publicación, editorial, etc. El proceso de análisis requirió un cuestionamiento crítico e inteligente de una realidad determinada, unir teorías dispersas. El análisis *"es una operación intelectual que posibilita descomponer mentalmente un todo complejo en sus partes y cualidades, en sus múltiples relaciones y componentes"* (Pérez Rodríguez et al., 1996: 70).
- Histórico-lógico: permite conocer el fenómeno estudiado desde los orígenes hasta la actualidad, lo cual favorece el conocimiento de su historia, evolución, tendencias, regularidades y características esenciales (Pérez Rodríguez et al., 1996: 75).

Para la consecución de este trabajo se realizaron las siguientes tareas, según el orden en que aparecen:

- Se localizó la información existente y disponible sobre los resúmenes documentales.
- Se clasificaron todos los documentos identificados con el propósito de seleccionar los más pertinentes para los propósitos de la investigación.
- Se analizó el contenido de los documentos seleccionados, para detectar patrones, tendencias, convergencias y contradicciones teóricas sobre el tema.
- Se analizaron, críticamente, los resultados develados en el paso anterior, con el propósito de comprender la realidad analizada.

La bibliografía se presenta según las indicaciones del *estilo Harvard* y se acota al final de la tesis. Las referencias dentro del texto igualmente se registrarán por dicho estilo.

La estructura capitular de la investigación quedó de la siguiente manera:

Capítulo 1: Estado del arte de los resúmenes documentales

Resumen

El gran interés de bibliotecarios y monjes de la antigüedad de conservar las colecciones y posteriormente el aumento de las publicaciones, propició el surgimiento de los primeros antecedentes de los resúmenes documentales, cuyo fin era identificar y recuperar los materiales bibliográficos, destacando lo esencial del contenido de los documentos. En los siglos siguientes el resumen se vio favorecido por la invención de la imprenta y la Revolución Científica, dando paso a las primeras revistas de resúmenes en el continente europeo y a los servicios de resúmenes. El desarrollo de internet propició el surgimiento de resúmenes elaborados automáticamente, aunque sus resultados no han sido muy satisfactorios. Los resúmenes documentales, desde sus orígenes, han presentado serios problemas terminológicos y conceptuales, los cuales se evidencian en las definiciones y tipologías presentadas. Las normas actuales para su elaboración abordan superficialmente los aspectos del contenido de los resúmenes, excluyendo tipos documentales del medio tradicional y a las nuevas representaciones hipertextuales. Los hiperresúmenes, también procedentes del medio electrónico, son la perspectiva futura a la que se desea llegar, pero aún ni siquiera están bien definidos.

Capítulo 2: El hipertexto: su implicación en resúmenes automáticos

Resumen

El medio hipertextual presenta características que lo distinguen del impreso, como la pérdida de la linealidad discursiva, debido a la presencia de una estructura asociativa que reproduce la estructura de la memoria humana; la combinación de texto con imágenes fijas y en movimiento, secuencias de video; inestabilidad de los contenidos; etc. Estos rasgos exigen nuevas o renovadas formas representacionales adaptables a este entorno. En este contexto surgen las ontologías, de gran utilidad en la elaboración de resúmenes automáticos debido a que asignan un significado unívoco a cada término en dependencia del dominio donde se esté utilizando, reduciendo la ambigüedad y mejorando la calidad de los resúmenes. Las taxonomías son una excelente herramienta clasificatoria para dominios específicos, establecen relaciones de equivalencia y jerarquía de contenidos. Cada categoría tiene un significado unívoco, lo que contribuye a que los sistemas de sumarización relacionen y recuperen fácilmente informaciones similares dispersas en la Web. En este sentido desempeña un rol fundamental la norma internacional Topic Maps, que indica cómo representar conceptos y sus relaciones, normalizando los términos empleados de manera inequívoca. Debido a todas estas facilidades han contribuido enormemente en la elaboración de resúmenes automáticos, pues en la medida en que cada concepto tenga asociado un único significado dentro de un contexto determinado, los sistemas tendrán más posibilidades de elaborar resúmenes de calidad.

Capítulo 3: Tendencias en la elaboración de resúmenes automáticos

Resumen

Muchos métodos y técnicas empleados en la actualidad para la elaboración de resúmenes automáticos tienen su origen en la década del '50, pero se han ido enriqueciendo con la contribución de las mineras de texto, imagen y video, la inteligencia artificial y el procesamiento del lenguaje natural. Algunos sistemas incorporan bases de conocimientos que reproducen el conocimiento de expertos. A pesar de que estos sistemas han logrado avances significativos en cuanto a brevedad, legibilidad y precisión, aún están lejos de hacer comprensiones automáticas de los textos, de ahí que les falte coherencia y presenten

ambigüedades lingüísticas e interpretativas. La perspectiva futura es que logren hacer razonamientos e inferencias, combinando métodos y técnicas provenientes de diferentes áreas del conocimiento, con el fin de representar los conceptos desde diferentes dimensiones y analizar una realidad determinada de forma evolutiva y no de forma estática. Hasta el momento, los últimos adelantos provienen de la ingeniería y la informática y no del dominio de la ciencia de la información, que prácticamente ha permanecido al margen.