



Temas de Biblioteconomía

El documento y sus clases

Análisis documental: indización y resumen

Autor: César Martín Gavilán

Fecha: 07/04/09

El concepto de documento y sus clases

Distinción entre dato, información y conocimiento

Datos, información y conocimiento se ubican respectivamente en tres niveles diferentes. En primer lugar, los datos se hallan en el territorio de lo físico. En segundo lugar, la información no es un conjunto especial de datos, sino que se sitúa en el territorio de lo conceptual: la información es el contenido semántico de los datos. Por último, el conocimiento, como estado cerebral o disposición neuronal, pertenece al territorio de lo mental.

- *Dato como soporte físico de la información:* podemos identificar los datos como acaecimientos físicos (pequeñas parcelas (o trozos) de la realidad) susceptibles de transportar asociada cierta información. Poseen una naturaleza material y pueden ser considerados como el soporte físico de la información. Cada uno de los enunciados (impresos o en soporte electrónico, dependiendo del formato o versión que esté utilizando el lector) que aparecen en este tema pueden ser considerados como datos. Es importante señalar algunas características de los datos de la mano de esta caracterización. Por un lado, al ser acaecimientos físicos, los datos son sencillos de capturar, estructurar, cuantificar o transferir. Por otro, que un mismo dato puede informar o no a un individuo dependiendo, como veremos a continuación, del bagaje previo de conocimiento del individuo. Además, los datos acostumbran a ser conjuntos de caracteres alfanuméricos materializados sobre un documento (físico o electrónico). Según esta caracterización, es posible abarcar sin tensión el sentido que se le da al concepto de dato en informática y telecomunicaciones: conjunto de caracteres asociados a un concepto. El conjunto de caracteres “35.879.987” respecto al concepto número del documento nacional de identidad (DNI), podría ser un ejemplo.

- *Información como contenido semántico del dato:* la información no posee una naturaleza física o material (como pasa en el caso de los datos), sino que pertenece al territorio de lo conceptual. Lo que permite que un dato transporte cierta información es la existencia de un código (o clave de codificación) que le asocia cierto contenido informativo. La información no depende de los receptores, sino de la clave de codificación que pone en relación el dato y aquello sobre lo que éste informa. El dato o conjunto de caracteres “Real Madrid – FC Barcelona: X” que aparecen en la página del periódico del lunes informa de que el Barça y el Real Madrid han empatado gracias a la existencia de un código (una clave de codificación) que le asocia ese contenido semántico al dato físico. Es importante distinguir entre dos fenómenos relacionados pero a la vez distintos: transportar información y adquirir información. Transportar información es una propiedad que poseen los datos gracias a la existencia de claves de codificación. Adquirir

información, en cambio, es una propiedad que poseen los individuos cuando son capaces de asimilar, a partir de su bagaje previo de conocimiento, la información que transporta un dato al interpretar ese dato a partir de la clave de codificación que se utiliza.

- *Conocimiento como estados mentales de un individuo a partir de la asimilación de información:* el conocimiento debe ser identificado con un tipo especial de estados mentales (o disposiciones neuronales) que posee un individuo y que presentan una serie de características propias. Por un lado, son estados mentales que adquiere el individuo a partir de un proceso de asimilación o metabolización de información. En este sentido el contenido semántico de esos estados mentales coincide con esa información asimilada. Y, por otro, actúan de guía en las acciones y la conducta de ese individuo. Sin embargo, las características del conocimiento no acaban aquí. El conocimiento, a diferencia de los datos y la información, se encuentra estrechamente relacionado con las acciones y las decisiones del sujeto que lo realiza; se llega incluso a poder evaluar ese conocimiento utilizando como indicativo esas acciones y decisiones. El conocimiento, además, es el factor crítico que permite la asimilación de nueva información, y por tanto la creación de nuevo conocimiento, por parte del sujeto que lo posee y suele verse reestructurado continuamente por las entradas de nueva información asimilada.

El concepto de documento como soporte de la información

Una vez esbozada la distinción entre los conceptos de dato, información y conocimiento, queda describir cuál es el papel que juega el concepto de documento dentro de todo este escenario.

En términos generales, se puede afirmar que los documentos siempre se han visto involucrados en la actividad intelectual del ser humano. Desde el principio de la historia del pensamiento, el hombre ha utilizado una serie de objetos o materiales donde poder plasmar y almacenar aquello que pensaba o sentía. Las pinturas rupestres, las tabletas de arcilla mesopotámicas, los muros de los edificios sagrados egipcios, los papiros, los pergaminos o, posteriormente, el papel, son claros ejemplos de tipo de objeto o materiales. En la actualidad, con el desarrollo de las tecnologías de la información y la comunicación, se está apostando cada día más por los formatos digitales para recoger la producción intelectual.

Para referirnos a todo este tipo de objetos o materiales utilizamos normalmente el término “documento”. O, dicho de otra manera, es posible identificar como documento todo aquel *soporte* donde se represente algún tipo de información. En este sentido, podemos incluir bajo el concepto de documento una hoja de papel escrito, un libro, una fotografía, una cinta de video, un DVD, un archivo creado con un procesador de textos, una base de datos o una página web.

Los documentos se caracterizan por poseer dos dimensiones: por un lado son algo físico y, por otro, contienen asociado un contenido informativo o información.

La relación entre los conceptos de documento y dato parece bien sencilla. Si el dato es el soporte físico de la información, el documento debe ser entendido como un conjunto de (un tipo especial de) datos.

Sobre la relación entre los conceptos de documento e información, diremos que si la información debe entenderse como el contenido semántico del dato derivado de una clave de codificación, el documento aparece como el *objeto material* donde se puede representar y materializar información. Esta representación y materialización permite explicar varias cosas. Por un lado, permite explicar cómo se puede transmitir la información: la información se representa (se asocia) en un documento a partir de un código y su transmisión se produce a partir de la propia transmisión material del documento. Y, por otro lado, permite también dejar claro por qué la conservación y almacenamiento del documento significa también la conservación y almacenamiento de la información que éste contiene. Sólo cabe analizar ese documento bajo la misma clave de codificación (o código) que se utilizó para asociarle ese contenido semántico concreto para poder recuperar esa información después del almacenamiento del documento.

Por último, en lo que se refiere a la articulación del concepto de documento frente al de conocimiento, entendido como estados mentales de un individuo construidos a partir de la asimilación de información y que rigen las acciones del propio sujeto. Frente a estos estados mentales, y a partir de su dimensión física y su capacidad de transportar información, el documento juega un papel muy importante: éste aparece como ese objeto material donde se puede representar y materializar esos estados mentales que residen exclusivamente en la cabeza de las personas. Y, al igual que pasaba en el caso de la información, esta representación y materialización permite explicar la transmisión y el almacenamiento de conocimiento a partir de la transmisión y almacenamiento de documentos.

En este sentido, por un lado, el conocimiento que reside en la cabeza de un individuo se plasma en un documento a partir de un código y su transmisión se produce a partir de la propia transmisión material del documento: cuando otro individuo es capaz de obtener la información asociada a ese documento transmitido y formar un nuevo estado mental a partir de la misma, podemos afirmar que se ha producido la *transmisión* de ese conocimiento. Por otro lado, a partir del mismo mecanismo, la conservación y almacenamiento del documento que se obtiene como fruto de la representación de un conocimiento concreto permite también la conservación y almacenamiento de ese conocimiento. Basta analizar ese documento bajo el mismo código que se

utilizó en la representación de esos estados mentales para poder recuperar la información asociada y crear nuevos estados mentales en otros individuos después del almacenamiento del documento. De esta manera ese conocimiento puede ser recuperado por cualquiera que lo necesite en el momento adecuado.

Clases de documentos

Para la ciencia de la documentación, el documento es a la vez medio y mensaje de información y conocimiento. De esta manera, el documento se caracteriza por una triple dimensión: el soporte físico o material, el mensaje informativo y la posibilidad de transmisión o difusión de este conocimiento. Esta triple dimensión que caracteriza al documento ha servido, a su vez, para establecer una tipología de los documentos en la que los estudiosos del campo de la documentación coinciden, más o menos, y que se puede resumir de la siguiente forma:

- Según el soporte material hablaríamos de documentos de papel (libros, revistas, folletos, etc.), material químico (películas), material magnético (cintas de vídeo, disquete de ordenador, casetes, etc.), soportes ópticos (CD-ROM, DVD, video-discos, etc.), etc.
- Según el mensaje informativo:
 - por la forma de expresión del contenido hablaríamos de documentos textuales (libros, revistas, etc.) y documentos no textuales (gráficos como mapas, planos, etc.; sonoros como cintas, discos, etc.; iconográficos como fotografías, carteles, etc.; audiovisuales como películas, vídeos, etc.; informáticos como programas de ordenador; tridimensionales como esculturas o juguetes; compuestos o multimedia cuando el documento combina varios contenidos mencionados anteriormente);
 - por la transformación del contenido hablaríamos de documentos primarios (originales, como libros y revistas, etc.), secundarios (que hacen referencia a los documentos primarios, como bibliografías, catálogos, índices, sumarios, boletines, bases de datos, etc.) y terciarios (tienen estructura formal de secundarios, pero contenido primario, como los diccionarios, enciclopedias, léxicos, tesauros, etc.).
- Según la posibilidad o forma de transmisión o difusión, desde el punto de vista social hablaríamos de documentos públicos, reservados e inéditos; y desde el punto de vista temporal hablaríamos de documento periódicos o no periódicos (de naturaleza monográfica).

En este tipo de clasificaciones tradicionales, no encontramos por ninguna parte términos como documento electrónico o digital. Será necesario, pues, ampliar esta tipología para dar cabida a los nuevos tipos de documentos surgidos

gracias a la revolución digital, a la tecnología hipertextual y la aparición de la Web y de Internet. Por ello debemos tener en cuenta:

- Nuevos materiales (documento electrónico/documento impreso): un documento electrónico difiere de un documento impreso en el material que lo conforma. Tablas de cera o arcilla, papiro, pergamino y papel han abierto paso a los discos y cintas magnéticas (casete, cinta de vídeo, disquete, disco duro de un ordenador, tarjetas de memoria, etc.) y a los discos ópticos (CD-ROM, DVD, etc.) que se imprimen y leen mediante láser sin que exista un contacto directo con el soporte. Ambos, documento impreso y documento electrónico, pueden contener el mismo texto, aunque cambia el soporte. Un documento electrónico es aquel cuyo contenido está en un soporte electrónico que, para su acceso, requiere una pantalla textual, una pantalla gráfica y/o unos dispositivos de lectura de audio, vídeo, etc. según el tipo de información que contenga. En algunos casos también se precisa la mediación de un ordenador (cuando la información está digitalizada) y en otros no (si se trata de información analógica).
- Nueva codificación de la información (documento digital/documento analógico): a menudo se identifica un documento electrónico con un documento digital, sin embargo, si hablamos con propiedad, no son la misma cosa. Un documento electrónico puede ser bien analógico, bien digital. Documentos electrónicos son, por ejemplo, una cinta de casete o una cinta de vídeo, que precisan de un dispositivo electrónico para su lectura, pero no son digitales. Lo que distingue un medio electrónico de un medio digital es, por una parte, la forma en que está codificada la información y, por otra, la necesaria mediación de un ordenador para decodificar esta información. En el caso de un documento digital, la información está codificada en bits, y para leer, visualizar o grabar la información se precisa de un dispositivo que transmita o grabe información codificada en bits. Al representarse digitalmente, los datos de entrada son convertidos en dígitos (0,1) inteligibles para la máquina y no para los sentidos humanos; y a la salida, otro dispositivo los convertirá en señales analógicas, inteligibles para los sentidos humanos. Un documento digital es, pues, aquél que contiene la información codificada en bits. De esta manera, las distintas morfologías de información (texto, imagen y sonido) se codifican en cadenas de ceros y unos. La dicotomía, pues, no sólo se establece entre documento impreso y documento electrónico, sino también y dentro de los propios documentos electrónicos, entre documento digital y documento analógico. Es claro que los medios analógicos (libros, periódicos, revistas, folletos, etc.) conviven hoy con los medios digitales a pesar de que se habla de que vivimos en la sociedad de la información entendiendo ésta como eminentemente digital, basada en el uso de los ordenadores y las redes de transmisión de datos. Existe un factor clave que favorece el avance de la información digital: la

producción, distribución, acceso y reusabilidad (esto es, la facilidad de selección, recuperación, reproducción y transmisión) es mucho más sencilla, accesible, rápida y barata que la de la información impresa.

- Nuevos tipos de documentos (documento virtual/documento real): Los documentos digitales pueden ser estáticos y persistentes o pueden ser generados dinámicamente y ser virtuales. Un documento virtual es un documento para el que existe un estado no persistente y en el que algunas o todas sus instancias se generan automáticamente en un lapso de tiempo. Un documento virtual puede entonces consistir de múltiples páginas, una visita guiada, applets de Java y puede tener o no tener enlaces (esto es, puede o no ser hipertextual). El contenido puede definirse mediante etiquetas o tags, una plantilla de documento, un programa informático, una consulta a una base de datos u otras muchas aplicaciones que sindicán contenidos. Los documentos virtuales han emergido al mismo tiempo que la interactividad y el deseo de obtener y recuperar documentos "a la carta", particularmente en la Web. Hasta el nacimiento de la WWW, el acceso a la información ha consistido en la recuperación de copias electrónicas de los documentos de un gran depósito de información relativamente estática. El término estático no se refiere aquí a la estabilidad de los documentos (que en la web se modifican y actualizan constantemente, cambian de lugar, desaparecen, etc.) sino a que el documento ya existe y tiene unos límites bien definidos. Lo que sucede ahora es que el acceso a la información se realiza a través de la manipulación de una gran colección de fuentes y recursos de información diferentes y que algunos de esos recursos son documentos y otros son procesos que crean documentos. El nacimiento de este nuevo tipo de documentos establece una nueva dicotomía entre los términos documento real (es el que se puede consultar sin ningún cambio, esto es, en su estado presente, actual) y el documento virtual, entendido como un documento que no es real, pero que contiene los datos específicos necesarios para producirlo. Se trata de una colección no organizada de bloques de información, junto con las especificaciones necesarias (herramientas y técnicas) que permiten la generación de un documento real.

Vemos entonces que la definición de documento debe ser lo más amplia posible, ya que tiene que integrar una gran variedad de soportes, formatos y distintas morfologías. La tipología de los documentos también se va ampliando a medida que surgen nuevas formas y tecnologías de lectura y escritura, nuevas formas de acceso y recuperación del documento, nuevas formas de estructurar la información y nuevos modos de interacción por parte del usuario. A la escritura manuscrita a través de diversos dispositivos (cuñas, pinceles, cañas, plumas, lápices, bolígrafos), le han seguido la tecnología de la imprenta, las máquinas de escribir y otros dispositivos electrónicos, los ordenadores y las redes telemáticas. Todos estos medios de escritura conviven hoy en el tiempo.

Los sistemas de gestión de hipertextos, los lenguajes hipertextuales, los lenguajes estructurados y la Web han incrementado la escala de los sistemas de recuperación de la información y han introducido nuevas nociones y nuevos tipos de documentos. Podemos, pues, ampliar la tipología clásica de los documentos incluyendo nuevos tipos como los siguientes:

- Según el soporte material hablaremos de documentos físicos (soportes tradicionales como el papel o la película; soportes magnéticos como el videocasete o el disco duro; y soportes ópticos como el CD o el DVD) y virtuales (en línea, alojados en la memoria de servidores remotos).
- Según el mensaje informativo:
 - por la forma de expresión del contenido hablaremos de documentos reales y documentos virtuales (tanto hipertextuales como conceptuales)
 - por la posibilidad de transformación del contenido por parte del usuario hablaremos de documentos estáticos y documentos dinámicos (que permiten la interactividad del usuario y pueden generarse mediante diferentes aplicaciones).
 - por su forma de codificación hablaremos de documentos analógicos (su paradigma lo constituye, por una parte, el libro impreso, pero también objetos como la microforma, la cinta de video o el registro fonográfico.) y documentos digitales.
- Según el tipo o posibilidad de acceso:
 - por el tipo de acceso al documento hablaremos de documentos locales (el acceso se realiza a través de un soporte físico como es el disco duro del ordenador o a través de algún periférico) o documentos remotos (el acceso se realiza en línea a través de una red).
 - por la posibilidad de acceso, transmisión y difusión, desde el punto de vista social hablaremos de documentos públicos y documentos privados; desde el punto de vista temporal hablaremos de documentos permanentes y de documentos transitorios; y desde el punto de vista espacial hablaremos de documentos en línea (acceso a través de Internet o una intranet) y documentos fuera de línea.
- Según el tipo de autor, hablaremos de documentos de autoría individual (personales, institucionales y corporativos) y documentos de autoría colectiva (tanto participativos, caso de los blogs, wikis y demás redes sociales, como no participativos).

Y aunque esta lista no pretende ser exhaustiva, sirve para mostrar cómo la tecnología digital e Internet han dado paso a nuevas concepciones teóricas y

prácticas del documento y a la necesidad de establecer nuevas tipologías documentales.

Análisis documental

El análisis documental se puede definir como el conjunto de operaciones (unas técnicas y otras intelectuales) que se realizan para representar tanto la forma como el contenido de documentos primarios, generando de esta forma otros documentos secundarios cuyo objetivo no es otro que facilitar al usuario la identificación precisa y recuperación posterior de los documentos primarios representados. Por ejemplo, son fruto del análisis documental las construcciones descriptivas estructuradas y normalizadas que creamos para proporcionar al usuario información bibliográfica de documentos a los que se puede acceder de alguna forma en o desde nuestra biblioteca: es decir, los registros bibliográficos de nuestro catálogo son el producto de un análisis documental. Desde el punto de vista de la recuperación de la información, estos documentos secundarios son tremendamente útiles y poseen ventajas evidentes frente a los propios documentos primarios (siempre y cuando el análisis documental se haya realizado respetando normas y convenciones, claro está): son más manejables, facilitan la identificación del documento original y establecen vías de acceso a su información.

Muchos y muy diversos pueden ser los métodos de análisis utilizados para representar los documentos. La mayor o menor profundidad del análisis dependerá en gran medida de aspectos como:

- Tipo de usuarios y necesidades de información más o menos especializadas
- Tipo de servicios que se ofrecen (no es lo mismo una biblioteca pública que una biblioteca universitaria)
- Tipo de documentos a analizar (libros, artículos de revistas, literatura gris, prensa, legislación...)

En cualquier caso, el concepto de análisis documental abarca siempre la descripción formal (externo) y la descripción del contenido (interno) del documento.

- El análisis documental como descripción formal: este proceso, más técnico que intelectual, tiene como objetivo solucionar el problema de la identificación de los documentos primarios para facilitar su localización o acceso posterior. Su campo de actuación se concreta en dos operaciones distintas:

- la descripción bibliográfica: entendida como la recolección y presentación organizada de los elementos o datos identificativos del documento. Por ejemplo, de forma normalizada siguiendo las áreas descriptivas de la ISBD.
 - la catalogación: entendida como la elección de los puntos de acceso formales que guiarán el camino hacia la descripción bibliográfica. Por ejemplo, de forma normalizada siguiendo los capítulos 14-16 de las RC (o la segunda parte de las AACR2).
- El análisis documental como descripción del contenido: se trata de una operación más intelectual que técnica y, por lo tanto, más compleja, que pasa necesariamente por la identificación del contenido de los documentos, producto de la cual se crean puntos de acceso semánticos o resúmenes, que no sólo son unos eficaces intermediarios entre los documentos primarios y sus posibles usuarios, sino que a veces desempeñan misiones de anticipo o incluso sustitutos de los documentos que representan. Las operaciones de análisis documental interno son de tres tipos:
 - Clasificación: operación de síntesis que trata de discernir el contenido fundamental de los documentos para formalizarlos y representarlos mediante la ayuda de un lenguaje artificial preestablecido.
 - Indización: operación de análisis que trata de retener las ideas más representativas para vincularlas a unos términos de indización adecuados, bien procedentes del lenguaje natural o documental.
 - Resumen: operación de análisis y síntesis que trata de representar de forma abreviada el contenido del documento primario, constituyendo en sí mismo una entidad documental autónoma.

Dejando de lado la clasificación, en este tema nos centraremos en la indización y el resumen.

La indización

La UNESCO define indización como un proceso que consiste en describir y caracterizar un documento con la ayuda de representaciones de los conceptos contenidos en dicho documento. Para Neet indizar es “analizar los documentos y aislar en la riqueza del lenguaje natural empleado por los autores, todos los conceptos esenciales o pertinentes que deben ser retenidos en previsión de búsquedas posteriores”.

Con base en la anterior definición, el término indización debe entenderse como un proceso que consiste en conceptualizar el contenido del documento y traducirlo a un lenguaje normalizado como descriptores e identificadores, o a un lenguaje libre expresado en palabras claves, con la finalidad de crear

índices, los cuales son generados en la actualidad por los sistemas de gestión de las bases de datos.

Este proceso supone un análisis del contenido del documento, basado en el conocimiento de la disciplina, la estructura del documento, las necesidades del usuario y el vocabulario normalizado que se utiliza. De la fidelidad, consistencia y especificidad en la asignación de términos, depende la pertinencia y precisión en la recuperación de la información, ya que el lenguaje del análisis es el puente entre la información que contiene el documento y el usuario que consulta la base. Tratándose de un proceso subjetivo, la posibilidad de reflejar con exactitud el contenido de un documento es algo no siempre factible. Ensayos sobre la coincidencia en la asignación de términos a un mismo documento entre varios documentalistas han demostrado que, aún siendo expertos en el campo de estudio, la concordancia en la terminología de la indización apenas alcanza el 40% de promedio. Esto refleja el difícil papel del profesional de la documentación en la complicada y costosa tarea de interlocutor entre el contenido y naturaleza del documento y la necesidad del usuario, sobre todo en la actualidad, con este cambiante escenario de las TICs que hace más complicado si cabe el proceso del análisis documental.

Los principales factores que afectan a una buena indización son:

- **Especificidad:** nos referimos al nivel de concreción que deben tener los descriptores en relación con el tema del documento. Así un trabajo sobre impuestos concretos como el IRPF o el IVA deberá ser analizado por dichos términos, siendo demasiado genérico la utilización del concepto "sistema fiscal". ¿Hasta qué punto debe ser específica una indización? Hasta el mismo que el tema del documento lo exija. Esto no significa que en algunos casos, y siempre que las necesidades de información de los usuarios así lo recomienden, podamos añadir un término genérico.
- **Exhaustividad:** relacionado con el concepto anterior, la exhaustividad implica reflejar todos los temas y conceptos relevantes. Es difícil valorar este aspecto ya que, aunque una gran cantidad de entradas puede facilitar la recuperación de un documento, muchas veces este factor entra en contradicción con la pertinencia del documento hallado.
- **Pertinencia:** es el elemento fundamental a tener en cuenta para una buena indización. La pertinencia de un descriptor representa su grado de adecuación al concepto expresado en el texto original. Un buen descriptor debe ser siempre un punto preciso y fiel de acceso y responder a las expectativas de búsqueda del usuario. El indizador debe siempre ponerse en el lugar del especialista que cuando busca por ese término que hemos utilizado va a encontrarse con el documento que se está analizando.

- **Coherencia:** la coherencia en la indización se refiere al grado de concordancia, uniformidad entre indizadores o un mismo indizador. Una medida de coherencia puede hallarse a través del promedio de descriptores comunes. La coherencia depende del número de descriptores asignados, del vocabulario utilizado, de la terminología del documento y, en general, de la capacidad de conceptualizar del responsable de la indización.

Según la norma UNE 50121:1990 Métodos para el análisis de documentos, determinación de su contenido y la selección de los términos de indización (equivalente a la ISO 5963:1985), la indización consiste en tres etapas, que tienden a solaparse en la práctica:

- Examen del documento y determinación del contenido a partir del título, resumen, sumario, introducción, conclusiones, notas,...
- Identificar y selección de los conceptos teniendo en cuenta el nivel de indización escogido
- Traducción de los principales conceptos escogidos al lenguaje de indización

Hay diversos sistemas de indización, entre los que se encuentran:

- **Indización por palabras clave (lenguaje libre):** mediante colección no ordenada (sólo en orden alfabético) de palabras significativas o no vacías (es decir, todo aquello que no son artículos, conjunciones, pronombres, preposiciones, numerales, ciertos verbos y adverbios), extraídas automáticamente por un ordenador a partir del título, del resumen y, cada vez más a menudo, del texto completo de los documentos registrados dentro un determinado sistema documental.
- **Indización por descriptores libres (lenguaje libre):** mediante colección no ordenada (sólo en orden alfabético) de conceptos destacados, por un proceso intelectual, a partir de los documentos registrados dentro un sistema documental determinado. Estos conceptos son expresados por palabras o expresiones extraídas de los documentos, o propuestos por los documentalistas, sin verificar si existen en una lista establecida a priori
- **Indización por autoridades (lenguaje controlado):** mediante colección no ordenada (sólo en orden alfabético) de conceptos destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro un sistema documental determinado. Estos conceptos son expresados por palabras y expresiones extraídas de una lista finita hecha a priori; sólo los términos que hay en la lista pueden ser utilizados para indizar los documentos y realizar consultas.
- **Indización por tesauros (lenguaje controlado):** mediante lista estructurada de conceptos destinados a representar de manera unívoca el contenido de

los documentos y de las consultas dentro un sistema documental determinado. Los términos forman una lista finita establecida a priori. Sólo los términos que están en la lista pueden ser empleados para indizar los documentos y realizar consultas. Se ordena generalmente alfabéticamente, pero acepta otras formas de presentación. Las relaciones de equivalencia, jerárquicas y asociativas entre los términos están más trabajadas que en las listas de encabezamientos de materia. Son ámbitos temáticos especializados. Su elaboración exige un estudio previo del área temática a cubrir. Se basa en la postcoordinación.

- Indización por listas de encabezamientos de materia (lenguaje controlado): mediante lista ordenada alfabéticamente, y no estructurada conceptualmente, de aquellos términos que deben representar de forma unívoca el contenido de los documentos. Los términos forman parte de una lista establecida a priori. Sólo los términos que aparecen en esta lista pueden utilizarse para indizar. Se emplea con fondos generales o enciclopédicos. Son útiles tanto para búsquedas manuales como automatizadas. Se basa en la precoordinación. El encabezamiento de materia representa un término del documento mediante una serie de términos del lenguaje natural y relacionados entre sí (encabezamiento y subencabezamientos).

La indización automatizada

La generación de índices en una base de datos con las palabras extraídas del título, resumen y palabras claves del autor, fue el primer método utilizado para indizar documentos en bases de datos bibliográficas. En la extracción de ocurrencias los sistemas informáticos han mostrado una enorme eficacia. Pero no olvidemos ésta es sólo la primera parte del proceso de indización, aunque realmente la mayor parte de los llamados motores de búsqueda se quedan en este primer paso y generan índices a partir de la simple extracción de palabras.

El paso siguiente es la asignación, que implica, o debería implicar, la traducción de las palabras extraídas en términos que describen el contenido. Los algoritmos que se usan para este proceso, llamados en inglés de *assignment indexing* son varios: en algunos casos el sistema posee un tesoro, que funciona para el control terminológico y que dispone de conjuntos de términos que permiten vincular los términos extraídos con los del lenguaje controlado.

La presencia de grandes masas textuales en la Web y la difusión de los lenguajes de marcado ha producido una modificación en las perspectivas de trabajo. A partir de ese momento, con la complementariedad de formas de extracción y asignación en ambiente Web, podemos decir que nos encontramos ya frente a lo que se conocen como sistemas asistidos de indización automática (Computer Assisted Indexing systems, CAI).

Los CAI se caracterizan por su capacidad de manipular documentos extrayendo palabras significativas del título, resumen y palabras claves asignadas por los autores, o del texto del documento, transformando las palabras extraídas mediante algún procedimiento de análisis, en una asignación de términos que pretende representar el contenido. Por tanto, en los CAI hay que considerar cuáles son los procedimientos de extracción y de análisis. Metodológicamente utilizan diferentes enfoques: análisis semántico, análisis conceptual con base estadística y análisis en base a reglas de reconocimiento, entre otros.

Resumen

Antes de introducirnos en las características de los resúmenes y las metodologías para prepararlos, parece aconsejable aclarar qué entendemos por *resumen*. Un resumen es una representación abreviada y exacta de un documento, en la medida de lo posible utilizando el propio lenguaje del autor, en que los puntos contenidos se exponen en el mismo orden que en el original, sin agregar interpretación o crítica alguna, y sin identificación de quien escribió el resumen. Debe constituir una entidad en sí mismo y ser fácilmente comprensible sin hacer referencia al documento original.

Propósitos del resumen

Sin lugar a dudas, la gran disponibilidad de bases de datos bibliográficas en línea ha cambiado y ampliado el papel de los resúmenes en el pasado. Hoy en día los resúmenes contenidos en las bases de datos pueden examinarse con la modalidad de “búsqueda en texto libre”, lo cual permite al usuario realizar búsquedas en línea por cualquier término que se considere apropiado. El sistema recupera referencias de todos los documentos cuyos resúmenes contenga esos términos o combinación de ellos. Este proceso se realiza sin el uso de un lenguaje de indización específico. Considerando esta opción, cada día más frecuente, de búsqueda en texto libre, la necesidad de poner especial cuidado en la construcción del resumen, tanto en sus aspectos de contenido temático como formales (palabras adecuadas, correcta formación de frases, etc.), se hace evidente.

Un resumen puede prepararse con diferentes propósitos. Es importante distinguir estos propósitos pues ellos son determinantes para decidir el tipo de resumen a utilizar, es decir, si se debe condensar la información de los documentos o sólo indicar de qué tratar. Podemos distinguir entre propósitos generales y específicos.

Propósitos generales:

- Ayudar a decidir a un lector interesado en el tema de un documento si le conviene o no leerlo completo.
- Proporcionar al lector interesado parcialmente en el documento, el máximo de información posible, a fin de evitarle una lectura total.
- Permitir realizar búsquedas a través de un sistema informático.

Propósitos específicos:

- Indicar la naturaleza de la información contenida en el documento original.
- Sugerir la conveniencia de obtener el documento para un estudio intensivo.
- Proporcionar al lector un medio rápido de mantenerse informado sobre los avances en su campo de interés.
- Guía al lector a una lectura más inteligente del trabajo original, destacando de él los aspectos más significativos.

Por otra parte, un resumen puede ser preparado con uno o más de los siguientes fines:

- Ser publicado en una revista de resúmenes
- Acompañar al documento original
- Ser incluido en un boletín de alerta
- Formar parte del registro bibliográfico y ser almacenado en una base de datos

Características del resumen

Las siguientes son algunas de las principales características que debe tener un resumen:

- a) Un resumen debe ser claro, metódico, ordenado. Esto implica un uso exacto y apropiado de las palabras. La presentación de los conceptos o temas debe conservar el orden del documento. El resumen debe decir sólo lo esencial, sin usar un estilo narrativo, pero tampoco telegráfico.
- b) Debe ser preciso, vale decir, debemos evitar el uso de términos ambiguos y de frases coloquiales.
- c) El resumen debe ser conciso. Es algo más que preciso y breve. Ser conciso significa rechazar lo inútil y lo superfluo; evitar los epítetos y la repetición de términos equivalentes. Significa dar a una frase la plenitud de su sentido con mínimas palabras.

d) Debe ser fiel reflejo del documento original. El resumen debe ser una imagen reducida del documento original, conteniendo sus elementos esenciales, esto es, resultados, técnicas o métodos, y conclusiones.

e) Necesita ser autosuficiente. Debe serlo en el sentido de tener una entidad propia. Esto significa que, acompañado de una correcta cita bibliográfica, puede existir por sí mismo, sin estar inserto en o acompañado al original.

f) Fácilmente comprensible por sí mismo. En un resumen nunca debe hacerse referencia al documento, pues el resumen en sí es una abreviación del original. Por lo tanto, un resumen no debe “contar” de qué trata el documento, como si se tratara de hechos ajenos. No debe emplearse frases como “El autor dice...” o “Este artículo trata de...”, etc.

La autoría del resumen

Los resúmenes pueden ser preparados por los autores del documento primario, por especialistas en el tema, o por profesionales (documentalistas) que manejan esta técnica. Cada alternativa tiene sus ventajas y desventajas. Sin embargo, es importante decir que, por lo menos en el caso de los artículos científicos, es una responsabilidad del autor el acompañar su escrito con un resumen. En todo caso, la preparación de buenos resúmenes exige a quien desea realizar este trabajo, cuatro condiciones indispensables: el conocimiento de tema, el dominio de ciertas técnicas, un muy buen manejo del idioma y una cierta capacidad.

- Autor: sin duda, el autor tiene el mejor conocimiento de los contenidos del documento. Además, el hecho de que él prepare un resumen al momento que presente su trabajo, significará un gran ahorro para los servicios de resúmenes. Sin embargo, una desventaja es la probable falta de objetividad del autor respecto a su propio trabajo, lo que le puede llevar a escribir un resumen que “mejore” el propio documento.
- Expertos: generalmente, los expertos en un tema suelen llegar a escribir excelentes resúmenes tras recibir un cierto entrenamiento. Su mayor problema es la dificultad para abstenerse de emitir juicios de valor respecto al documento. Sin lugar a dudas, sometidos a cierta metodología y disciplina, pueden escribir mejores resúmenes que los autores.
- Documentalistas: son personas que, sin ser especialistas en el tema, están entrenados para preparar resúmenes objetivos, no críticos, correctos en la forma y la extensión. Su entrenamiento enfatiza la inclusión de palabras claves en el texto, lo que facilita la búsqueda en texto completo. Sin embargo, estas personas pueden tener serias dificultades al tratarse de temas complejos o muy especializados.

Tipos de resúmenes

Distinguiremos básicamente dos tipos de resúmenes: indicativos e informativos. Una variación de los indicativos son los llamados resúmenes “integrados” y la combinación de las técnicas de los resúmenes indicativos e informativos da lugar a la tipología híbrida del resumen “informativo-indicativo”.

- Resumen indicativo: el resumen indicativo, llamado también “descriptivo” es aquél que sólo describe brevemente el contenido del documento, sin incluir aspectos metodológicos, ni las conclusiones del autor. Sólo intenta ayudar al usuario a comprender el alcance del documento original, sin entregarle una cuenta detallada de su contenido. Su extensión suele ser entre 25 y 50 palabras.
- Resumen integrado: esta modalidad de resumen indicativo tiene la particularidad que el resumen en sí está formado por una serie de palabras claves, conectadas por conjunciones, preposiciones, formas verbales, etc.
- Resumen informativo: el resumen informativo incluye información sobre propósito del trabajo, hechos claves, datos, observaciones, metodologías, hallazgos, conclusiones. El resumen informativo ha sido comparado con una hoja a la que se le quita toda la pulpa dejando sólo la nervadura. Un perfecto esqueleto estructural, a partir del cual es perfectamente posible establecer su apariencia original. Idealmente, el resumen informativo es una versión abreviada, pero completa e inteligible del contenido significativo y conclusiones de un documento. Su extensión suele ser entre 200 y 350 palabras. En general, los resúmenes informativos deben ser preparados por personas especialistas en el tema, o por el autor, siempre que hayan recibido entrenamiento en la metodología de preparación de resúmenes. El resumen informativo puede dividirse en cuatro secciones:
 - Alcance y propósito
 - Métodos empleados
 - Resultados obtenidos
 - Conclusiones o interpretaciones de los resultados

El mantenimiento de este orden es importante, la información debe entregarse en este orden dentro de párrafos individuales. La primera parte del resumen (alcance y propósito) es de particular importancia. Si está bien escrita, puede ser concluyente para que el lector sepa si el documento es de su interés. El resto del resumen, aunque esencial, pierde valor sin este párrafo inicial. En esta parte del resumen, es útil para el lector encontrar expresiones como: “breve” “exhaustivo”, “teórico”, “experimental”, “de campo”, etc., que le señalan la forma de tratamiento o enfoque del tema. La segunda sección (métodos empleados) necesita ser explícita, incluyendo a veces, equipos y materiales empleados, especialmente si se trata de una

metodología novedosa (aspecto este último que sólo puede detectar un especialista en el tema). La tercera sección (resultados obtenidos) se refiere a los hallazgos concretos. La cuarta sección (conclusiones) incluye las recomendaciones y es realmente importante, pues las conclusiones suelen reflejar el documento entero.

- Resumen informativo-indicativo: conjuga los niveles del resumen indicativo y del informativo. Las partes centrales son representadas de forma informativa, mientras que aquellos aspectos de menor trascendencia pero útiles son tratados indicativamente.

También podemos añadir dos tipos de resúmenes menos frecuentes: el analítico y el crítico. El resumen analítico supone mayor detalle informativo (aunque no excede de 500 palabras) y es poco habitual porque requiere bastante tiempo para su elaboración. El resumen crítico incluye una valoración sobre la profundidad y alcance del trabajo, metodología y contribución al desarrollo del conocimiento, y es poco habitual porque requiere que el analista sea un experto en la materia objeto de investigación.

Pautas y recomendaciones para resúmenes

Además de las pautas que cada servicio de resúmenes establece, existe normativa internacional para la confección de resúmenes, como la ANSI/NISO Z39.14:1997 (R2002), Guidelines for Abstracts; la “Guía para la preparación de resúmenes” de la UNESCO (1968); o la UNE 50103:1990, Preparación de Resúmenes (equivalente a la ISO 214:1976). Esta última norma, en concreto, establece las siguientes recomendaciones para la preparación de resúmenes informativos:

a) Propósito. Debe indicarse los principales objetivos y alcance del estudio, o bien las razones por las cuales se escribió el documento, a menos que ello se desprenda claramente del título. Sólo debe hacerse referencia a publicaciones anteriores si éstas constituyen parte esencial del propósito del documento.

b) Metodología. Las técnicas o métodos utilizados deben describirse sólo en la medida necesaria para su comprensión, pero identificando claramente las técnicas nuevas, describiendo sus principios metodológicos básicos, el alcance de su aplicación y el grado de exactitud que puede lograr. Si corresponde, debe señalarse las fuentes de información (encuestas, revisiones bibliográficas, etc.).

c) Resultados. Los resultados deben describirse en la forma más precisa e informativa posible. Pueden ser de carácter experimental o teórico, o consistir en información reunida, en relaciones o correlaciones advertidas, en efectos observados, etc. Si no es posible incluir todos los resultados

novedosos, debe darse prioridad a lo siguiente: verificación de nuevos hechos, resultados nuevos de valor a largo plazo, descubrimientos significativos que contradicen teorías anteriores, o que el autor destaca como especialmente pertinentes en relación con un problema práctico.

d) Conclusiones. Debe describirse el alcance de los resultados, especialmente en cuanto al propósito de la investigación. Las conclusiones pueden ir en relación con: recomendaciones, evaluaciones, aplicaciones, sugerencias, nuevas relaciones e hipótesis aceptadas o rechazadas.

Metodología

El procedimiento utilizado para la preparación de resúmenes, parte con el análisis conceptual. El propósito de este análisis es extraer ciertas nociones básicas del contenido temático de un documento. El análisis se realiza revisando el documento entero y leyendo cuidadosamente secciones seleccionadas: la tabla de contenido, la introducción, los párrafos iniciales de los capítulos, las conclusiones, etc. Una vez que el documentalista se ha familiarizado con el documento, comienza a escribir el resumen; procura extraer ideas o pensamientos del contexto, y trata de reunirlos armoniosamente, con la misma intención ideada por el autor.

Es importante saber qué tomar y qué descartar: por lo tanto, es evidentemente importante el conocimiento que la persona tenga del tema tratado en el documento. Hay que destacar, sin embargo, que aunque el conocimiento del tema es esencial, igualmente determinante en la preparación de un buen resumen es un hábil manejo del idioma y una gran capacidad de análisis y de síntesis. Y aunque la persona sea experta en este tipo de trabajo, es muy recomendable que siempre prepare un borrador para luego pulirlo y limpiarlo de palabras superfluas, cuidando al mismo tiempo la redacción.

BIBLIOGRAFÍA

Mario Pérez-Montoro Gutiérrez. "El documento como dato, conocimiento e información". En: *Tradumàtica : traducció i tecnologies de la informació i la comunicació*, Núm.2 (2003)

María Jesús Lamarca Lapuente. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*.

<http://www.hipertexto.info/documentos/document.htm>

Mela Bosch. "La indización asistida para el manejo de conocimiento en organizaciones frente al tratamiento documental en bibliotecas." En: *Biblios*. Año 7, No.23, Ene – Mar. 2006

María Texia Iglesias Maturana. "Guía para la preparación de resúmenes". En: *Serie bbliotecología y gestión de información*, N° 41, octubre 2008