

Niveles de análisis documental de contenido

Blanca GIL URDICIAIN
Profesora de la EUBD Complutense

RESUMEN

Análisis de las características y funciones que desempeñan las operaciones de Análisis documental de contenido en el proceso de normalización y tratamiento de la información. En relación con la indización, se tienen en cuenta los sistemas tradicionales; en cuanto al segundo nivel de análisis, el resumen, se apunta la imprecisión terminológica existente en lo que a su categorización se refiere.

Palabras clave: Análisis documental / Indización / Resumen.

INTRODUCCION

El análisis constituye una fase fundamental del proceso documental. Se trata de una «operación o conjunto de operaciones destinadas a presentar el contenido de un documento de una forma distinta de la original, a fin de facilitar su consulta o su referencia en fase posterior»¹.

Las operaciones de entrada y salida de información en un sistema documental, estrechamente relacionadas, se articulan en base al Análisis de contenido de la documentación, que constituye una de las etapas claves de su proceso de normalización y tratamiento. Como medio de evitar solapamientos con otras disciplinas, me parece conveniente adoptar la denominación de Análisis Documental de Contenido (ADC) para describir las operaciones mediante las que, tras la lectura del documento o tras la inspección

¹ CHAUMIER, J. Les Techniques documentaires. París: PUF, 1979, p. 13.

de su título, índice, introducción, conclusiones y otras partes fundamentales, se determina el significado general del texto y se identifican y aíslan los conceptos representativos del contenido.

Hablar de análisis de contenido puede generar falsas interpretaciones por su posible acepción a otros análisis de este tipo, tales como el análisis semiótico o el análisis de contenido entendido como técnica de investigación. El análisis al que aquí nos referimos, se limita al ámbito documental, con las características que le son propias.

El ADC tropieza con una dificultad debido a que cada analista se vale de sus propios medios intelectuales para desarrollar la tarea, y dentro del marco de referencia constituido por su cultura, experiencias e intereses, factores que impiden la concordancia total entre los analistas. Acerca de este problema de la uniformidad escribe Courrier «dos indizadores elegirán para el mismo documento muy pocos descriptores idénticos, a veces menos del cincuenta por ciento. Eso ocurrirá también en el mismo indizador al operar en dos períodos diferentes»².

El lenguaje documental desempeña un rol mediador y normalizador que evita, en muchos casos, la arbitrariedad del indizador y, como consecuencia, facilita la tarea de recuperación, ya que proporciona puntos de acceso a la información de interés para el usuario.

Me parece oportuno apuntar la imprecisión terminológica que afecta al ADC, ya que bajo esta denominación se incluyen tanto tareas de análisis propiamente dicho, como de síntesis. El ADC comprende, en efecto una primera fase analítica, que corresponde a la captación de información, y una segunda, en la que se sintetizan los datos analizados. La terminología utilizada por la escuela soviética para describir el proceso: «análisis-síntesis» refleja sin ambigüedades los dos estadios que lo componen. Izquierdo Arroyo³ propone la denominación Tratamiento documental de Contenido (TDC) para evitar centrar todo el proceso de tratamiento en la operación de análisis.

El ADC puede llevarse a cabo en niveles de menor o mayor profundidad, en respuesta a las necesidades informativas del usuario. El primer nivel, la indización, atiende a la necesidad de identificar y localizar el documento, el segundo, el resumen, cumple con los anteriores objetivos y presenta además, de forma abreviada, al documento original.

Para realizar las tareas de indización y resumen con éxito es necesario examinar el contenido e identificar las estructuras del texto original; describir esas estructuras en mayor o menor profundidad dependiendo del ni-

² COURRIER, Y. Analyse et langage documentaires. *Documentaliste*, v. 13, n. 5-6 (1976), p. 182.

³ IZQUIERDO ARROYO, J.M. De la semiótica del texto a la semiótica documental. En Moreiro González, J.A. *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III/B.O.E., 1993, pp. 199-216.

vel de análisis a que sometamos al texto (descripción característica y descripción sustancial) y, por fin, representarias mediante un lenguaje documental en el caso de llevar a cabo una descripción característica del texto, o, en el segundo caso, mediante la redacción de un nuevo texto condensando al original.

La importancia de la lectura en ambos casos, es indiscutible ya que proporciona la única vía de acceso a la información contenida en el documento.

La anticipación o conocimientos previos sobre la materia, el dominio de las formas básicas de expresión del pensamiento, la capacidad de reconocer el significado de determinadas palabras señales del texto y el control de posibles defectos en la percepción debidos a hábitos, tales como las regresiones o la subvocalización, son, entre muchos otros, factores que benefician el eficaz diálogo autor-lector, necesario para la captación de las estructuras u organización interna del texto.

LA INDIZACION

La indización hace posible la entrada de documentos en el sistema documental, desempeña una función determinante en el momento de la recuperación, y uno de sus productos resultantes, el indizado, constituye la base para la creación de los lenguajes documentales.

Se trata de la fase de identificación de los conceptos que de forma explícita o implícita están presentes en el contenido del documento. Es una operación compleja, comprendida por varias fases; la norma ISO 5963-1985⁴ prescribe las siguientes:

- examen del documento y determinación de su contenido
- identificación y selección de los conceptos principales del contenido.
- selección de los términos de indización

Rowley⁵ cree necesario seguir los siguientes pasos para determinar el perfil del texto:

- familiarización con el contenido del documento.
- análisis y selección de las materias representativas del tema principal.
- elección de términos representativos del documento que expresen sin ambigüedad su contenido.

Chaumier⁶ por su parte identifica las fases de reconocimiento y extracción de los conceptos informativos y traducción de estos conceptos al lenguaje documental.

⁴ ISO 5963-1985. *Documentation. Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation*. Genève: ISO, 1985.

⁵ ROWLEY, J. *Abstracting and indexing*. London: Clive Bingley, 1988, p. 50.

⁶ CHAUMIER, J. *Análisis y lenguajes documentales: El tratamiento lingüístico de la información documental*. Barcelona: Mitre, 1986, p. 28.

La selección de términos de indización se hace en función de dos parámetros: exhaustividad y especificidad, también denominados profundidad y precisión, respectivamente. La exhaustividad está relacionada con el número de conceptos que se tienen en cuenta y que caracterizan el contenido de un documento. La especificidad está relacionada con la exactitud con que un concepto particular que aparece en un documento está representado por un término de indización., es decir, mide la proporción de términos extraídos de un documento que son realmente relevantes. Los dos parámetros están inversamente relacionados, de modo que el indizador habrá de establecer el equilibrio entre ambos; la balanza se inclinará en favor de uno u otro en función de las necesidades de profundidad o precisión determinadas por el usuario. Se pueden hallar los índices de exhaustividad y especificidad de la siguiente forma:

$$\text{Exhaustividad} = \frac{\text{N}^{\circ} \text{ de documentos recuperados y relevantes}}{\text{N}^{\circ} \text{ de documentos relevantes de la colección}}$$

$$\text{Especificidad} = \frac{\text{N}^{\circ} \text{ de documentos recuperados y relevantes}}{\text{N}^{\circ} \text{ de documentos recuperados}}$$

La mayoría de los especialistas dan por supuesta la necesidad de utilizar en todo sistema un lenguaje documental que garantice la extracción de los conceptos contenidos en el documento, ya sea de forma explícita o implícita. Una página más arriba mencionamos la referencia de Chaumier a este respecto, Van Slype⁷, por su parte, define la indización humana como la «operación que consiste en enumerar los conceptos sobre los que trata un documento y representarlos por medio de un lenguaje combinatorio».

Incluso Salton⁸, defensor del tratamiento de la información en lenguaje libre, escribe sobre la utilidad del lenguaje documental en el proceso de indización «Para obtener un alto nivel de exhaustividad en la indización, es probada la utilidad del lenguaje de indización, que provee una variedad de enfoques para cubrir un área temática dada. Para asegurar la precisión, debe utilizarse un sumamente específico lenguaje de indización y los términos deben llevar indicaciones adicionales de contenido, tales como ponderaciones del término e indicadores de relación con otros términos».

⁷ SLYPE, G. van. *Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 115.

⁸ SALTON, G. *Introduction to modern information retrieval*. New York, etc.: McGraw-Hill Publishing Company, 1983, p.160.

SISTEMAS DE INDIZACION

Existen tres métodos tradicionales de indización: por unitérminos, por descriptores y por materias, que se basan en la lectura lineal del texto y en la representación formal del contenido. Estos tres sistemas responden al análisis del texto a nivel de palabras, conceptos o temas, respectivamente.

La indización por temas fue la primera en utilizarse; realiza la normalización de los documentos en base a listas de encabezamientos de materias que permiten la caracterización a un nivel muy general. La recuperación por medio de este sistema se hace particularmente compleja cuando la información se presenta en soportes no tradicionales.

El procedimiento de unitérminos consiste en la consideración de las palabras, no de las materias componentes del texto. Este método, ideado por Taube, ofrece la posibilidad de profundizar en el análisis del texto, aunque al excluir del mismo tan sólo los términos vacíos, genera un alto nivel de exhaustividad que provoca distorsiones en la comunicación. El sistema de unitérminos presenta otro inconveniente de carácter sintáctico en el momento de la recuperación, ya que al utilizar términos simples, pueden producirse falsas combinaciones entre ellos, provocando la recuperación de innumerables documentos no pertinentes.

A este tipo de indización siguió el procedimiento de caracterización por medio de descriptores, que la norma ISO define como «término usado para representar un concepto dado, conocido también como «término preferente»».

Mooers fue el primero en referirse al descriptor en relación con la indización de documentos, abandonando la descomposición morfológica estricta y la indización a nivel de palabra. Este sistema, al posibilitar la utilización de más de un término en la descripción, elimina la ambigüedad y alteraciones de sentido. Se basa en los conceptos, lo que dota de mayor flexibilidad a la tarea de indizar y permite hacer frente a cuestiones específicas formuladas por el usuario.

Con estos métodos tradicionales, tanto si se trata de indización manual como automática, el procedimiento se realiza mediante la segmentación del texto en términos significativos del contenido. Ni semántica ni sintaxis son suficientes para llevar a cabo el análisis del texto con exactitud, lo hacen por aproximación; bien es cierto que eliminan ambigüedad, pero para profundizar en el contenido del documento se precisa trabajar a nivel de contexto.

Algunas líneas de investigación, representadas en España por García Gutiérrez¹⁰, proponen métodos de análisis basados en la significación, en

⁹ ORGANIZACION INTERNACIONAL DE NORMALIZACION: *ISO 5963-1985: Documentation. Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection de termes d'indexation*. Genève: ISO, 1985, p. 3.

¹⁰ GARCIA GUTIERREZ, A. *Análisis documental del discurso periodístico*. Madrid: CTD, 1992, pp.136-154.

los que el significante es considerado como simple medio de acceso al significado. Se trata de métodos experimentales que subrayan la importancia del contexto en el análisis y que, aunque son útiles para la interrogación del texto, aún no han demostrado su validez para la representación formal de los resultados del análisis.

LA OPERACION DE RESUMIR

Esta tarea, que tiene analogías con la indización, constituye el segundo nivel del análisis documental de contenido.

En nuestro idioma se utiliza el término resumen tanto para denominar a la operación como al producto resultante, lo que puede generar falsas interpretaciones. Otros idiomas, como el inglés, tienen recursos gramaticales que obvian esta dificultad. En España, algunos teóricos recurren al uso del verbo para referirse a la operación y al sustantivo cuando hacen alusión al producto de la misma. También se habla de condensación y de descripción sustancial para hacer mención a la operación de resumir.

Por resumen se entiende la representación condensada del contenido de un documento, presentado preferentemente en un único párrafo de longitud determinada, y no incluyendo elementos críticos personales de quien lo redacta.

----- Entre los fines que se persiguen con la operación de resumir, son destacables los siguientes:

Su objetivo primordial es el de permitir la rápida y precisa identificación de los puntos esenciales del texto original, de manera que el usuario pueda decidir, sin pérdida de tiempo, si le interesa o no su lectura completa. Sirve además el resumen para superar barreras lingüísticas, ya que, como recomienda Unesco, se redacta en la lengua utilizada en el servicio o centro de documentación donde es elaborado y se traduce a uno de los idiomas de más generalizada difusión. Gran cantidad de resúmenes ofrecen de este modo la oportunidad de conocer al menos la información esencial presente en los documentos primarios.

Destacaremos también aquí el papel que desempeña este producto documental como auxiliar de indización, tanto manual como automática. Dado que en su seno figuran los temas esenciales descritos en el texto original, indizar en base al resumen supone la garantía de dar con los términos pertinentes bien sea extraídos de forma manual, como por los métodos lingüísticos de los que se sirve la indización automática.

Ahora bien, el resumen sólo cumplirá con los fines señalados anteriormente si en el momento de su redacción se tienen presentes principios como la entropía, pertinencia y relevancia, es decir, si transmitimos la mayor información posible utilizando el menor número posible de signos, si trasladamos convenientemente la información que contiene el documen-

to original y si expresamos sin ambigüedades el mensaje del documento primario.

CATEGORIZACION DEL RESUMEN

El resumen puede ser llevado a cabo por un documentalista o analista o por el propio autor del documento original. Uno y otro tipo de resumen tienen ventajas e inconvenientes. La ventaja del resumen escrito por el autor reside en su perfecto conocimiento de los temas abordados. Como inconveniente, se le puede achacar el exceso de subjetividad al caracterizar su propia obra.

La ventaja de los resúmenes escritos por documentalistas reside en su experiencia práctica y su conocimiento de las necesidades del usuario. Presenta como inconveniente su desconocimiento o conocimiento parcial de materias especializadas.

Aparte de esta tipificación en función del sujeto ejecutor del resumen, existe otra en base a la carga informativa. La escuela de pensamiento más generalizada distingue, según este criterio, dos tipos de resúmenes: informativo e indicativo. Rowley¹¹ y Pinto¹², entre otros, identifican además el indicativo-informativo, que sería aquel que se redacta en parte en un estilo informativo, tratando en estilo indicativo los aspectos menos relevantes del texto.

El resumen informativo presenta la mayor cantidad posible de información cuantitativa o cualitativa aportada por el autor del documento. Está indicado especialmente para la condensación del contenido de documentos científicos y técnicos. El resumen informativo debe indicar la estructura del contenido significativo del documento y presentará información relativa a los objetivos, la metodología experimental propuesta, los resultados obtenidos y las conclusiones del autor.

El resumen indicativo se limita a señalar el tema de estudio. Puede ser de gran utilidad en el tratamiento de material audiovisual y de documentación presentada en soportes informáticos, o en microformas, ya que permite la descripción de esos documentos, cuyo tratamiento no puede limitarse a la cumplimentación del tradicional asiento bibliográfico.

En cuanto a la extensión del resumen, considero que su longitud no depende ni de las dimensiones del documento a analizar ni del carácter indicativo o informativo del resumen, sino que responderá a la exhaustividad del análisis.

¹¹ ROWLEY, J. Types of abstracts. Epigr. 2.3. de *Abstracting and indexing*. London: Clive Bingley, 1988, pp. 14-15.

¹² PINTO MOLINA, M. *El resumen documental: Principios y métodos*. Madrid: Fundación Germán Sánchez Ruipérez, 1992, pp. 249-251.

No voy a tratar aquí del estilo, metodología, normalización y otros aspectos relativos a la descripción sustancial de documentos, dada la extensión que ello conllevaría; solamente señalaré, para terminar, la gran importancia que tiene el cuidado del estilo y contenido de los resúmenes especialmente cuando se incluyen en una base de datos, en donde, en muchos casos, las palabras que lo componen son utilizadas para realizar búsquedas en texto libre, pasando a convertirse en una importante herramienta de recuperación.

Apuntaré, por último, un hecho que se da demasiado a menudo en las pruebas selectivas para el ingreso en la Escala de Ayudantes de Archivos y Bibliotecas. La imprecisión en cuanto a la categorización del resumen documental genera conflictos a las personas que pretenden pasar determinadas pruebas de oposición, ya que en las bases de las convocatorias se pide la redacción de un resumen analítico, además de uno informativo y otro indicativo. Consultadas las fuentes¹⁵ que me ha sido posible, todas coinciden en identificar el resumen analítico con aquel, de carácter informativo, redactado por el propio autor del texto original. Dado que son muchos los años en que se vienen convocando plazas para el mencionado cuerpo, para las que se exigen este tipo de pruebas, me gustaría desde aquí urgir la necesidad de normalización en la denominación de los resúmenes documentales para evitar tanto la falta de rigor científico, como los malos ratos de los opositores que no saben qué hacer cuando se les pide llevar a cabo un resumen que, por principio, no pueden redactar.

BIBLIOGRAFIA

- CLEVELAND, D.B. & CLEVELAND, A.D. *Introduction to indexing and abstracting*. 2th ed. Englewood, Col.: Libraries Unlimited, 1990, pp. 163-167.
- MOREIRO, J.A. *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlo III/B.O.E., 1993, pp. 162-167.
- NEET, H.E. *A la recherche du mot clé: Analyse documentaire et indexation alphabétique*. Genève: Les Editions IES, 1989, p. 95.
- PINTO MOLINA, M. *El resumen documental: principios y métodos*. Madrid: Fundación Germán Sánchez Ruipérez, 1992, p. 253.
- ROWLEY, J. *Abstracting and indexing*. 2th ed. London: Clive Bingley, 1990, pp. 14-18.
- UNESCO. *Guía para la redacción de artículos científicos destinados a la publicación*. Paris: Unesco-Programa General de Información y UNISIST, 1983.

¹⁵ AFNOR (ASSOCIATION FRANÇAISE DE NORMALISATION). *Documentation: recommandations aux auteurs des articles scientifiques et techniques pour la rédaction des résumés: NF Z 44-004*, Paris: 1984.