Elementos, actividades y criterios para la identificación, comprensión y selección de conceptos en la indización analítica

Miguel Ángel Esteban Navarro
Universidad de Zaragoza

0.1. Resumen

Descripción de la naturaleza de los elementos léxico-conceptuales que se utilizan para la indización analítica de documentos textuales: los identificadores del contenido (materia, concepto y término), los empleados para la comprensión del significado de éstos (faceta y campo) y los usados para su expresión (términos de indización). A continuación, se exponen las actividades que se realizan para la identificación, comprensión, selección y expresión de las unidades de información que contiene el documento. Se presenta como la tarea central el diseño y uso de parrillas de indización construidas sobre la base del análisis de las facetas que caracterizan a un conjunto de materias relacionadas entre sí. Se discuten los criterios que ayudan a decidir que conceptos entre los identificados se deben representar, los cuales se presentan vinculados por parejas, condicionándose entre sí: exhaustividad y profundidad, especificidad y precisión, relevancia y pertinencia y coherencia y consistencia. Por último, se defiende, por extenso, la necesidad de mantener la indización analítica controlada por un lenguaje documental construido de acuerdo con principios científicos, como alternativa a los sistemas automatizados de indización por unitérminos para algunos tipos de documentos, porque su desaparición puede provocar el colapso del ciclo de información documental (Autor).

Palabras clave: Indización analítica. Concepto. Faceta. Análisis de facetas. Exhaustividad. Profundidad. Especificidad. Precisión. Relevancia. Pertinencia. Coherencia. Consistencia. Indización automática por unitérminos. Lenguajes documentales.

0.2. Abstract

Description of the nature of lexical and conceptual elements for the analytical indexing of textual documents: the identifiers of content (subject, concept and

term), the ones employed for the understanding of those (facet and field), and the others employed for its expression (indexing terms). Next, the activities for the identification, understanding, selection and expression of the information units of the document are exposed. The principal task is the design and the use of indexing grilles, developed from the faceted analysis of a group of subjects. The criteria for the determination of the concepts to represent are discussed, which are coupled in pairs but in relation between they: exhaustivity and depht, specificity and precision, relevance and aboutness, coherence and consistency. Finally, defence of the need of analytical indexing controlled by an indexing language builded with scientific principes, like option of the automatic uniterm indexing systems for some classes of documents, because its missing could crash the documentary information life cycle (Author).

Keywords: Analytical indexing. Concept. Facet. Faceted analysis. Exhaustivity. Depth. Specificity. Precision. Relevance. Aboutness. Coherence. Consistency. Automatic uniterm indexing. Indexing languages.

1. Concepto de indización

La indización consiste en un proceso destinado a identificar y a describir o caracterizar el contenido informativo de un documento mediante la extracción y la selección de las materias sobre las que versa (indización sintética) o de los conceptos presentes (indización analítica) para su expresión mediante unos términos de la lengua natural y su reunión en un índice, con objeto de permitir*a posteriori* la recuperación de ese documento del seno de una colección documental o de un conjunto de referencias documentales como respuesta a una demanda acerca del tipo de información que contiene. Es decir, la indización tiene como objetivo la representación del contenido de los documentos que forman parte de un conjunto para garantizar su eficaz recuperación durante el proceso de búsqueda en ese grupo, por lo que se trata de una técnica fundamental de la cadena documental, ya que afecta por igual a la fase de análisis y a la de salida. El concepto de indización se construye, por tanto, a partir del examen tanto de las actividades que se realizan durante el ejercicio de esta técnica, como de su función en un sistema de información documental.

El cumplimiento de esa función exige que la selección de los términos de indización durante el análisis de los documentos se efectúe siguiendo dos criterios: en primer lugar, como dominante, la adecuación al contenido informativo del documento, por lo que los términos de indización deben ser precisos, apropiados y exhaustivos; y, en segundo lugar, como secundaria pero también presente, las necesidades informativas del sistema documental en el que se realiza y a las que los documentos conservados deben dar satisfacción.

Asimismo, la expresión de los términos de indización se puede realizar de modo libre o haciendo uso únicamente de una lista de vocablos autorizados fijados *a priori*, con objeto de conseguir sobre la base del empleo de idénticas expresiones una coherencia entre los procesos de representación y de interrogación del catálogo, garantizando, de este modo, una adecuada recuperación de la información pertinente que contiene una colección. Es decir, se observa que se obtienen mejores resultados durante la recuperación cuando las demandas de los usuarios se pueden traducir a términos similares o iguales a los empleados durante la representación. La existencia de ese vocabulario auxiliar, denominado habitualmente lenguaje documental, también reduce el riesgo siempre presente de la subjetividad durante la indización y, por tanto, su negativa consecuencia de pérdida de coherencia cuando actúan múltiples indizadores en un mismo conjunto de documentos.

2. Elementos léxico-conceptuales de la indización

Centrándonos en el caso de la indización analítica de documentos textuales, para su ejercicio se emplean varios elementos léxico-conceptuales distribuidos en tres grupos, de acuerdo con su misión, que tienen su razón de ser en la propia naturaleza del texto: los identificadores del contenido, los que se emplean para la comprensión del significado y los que se utilizan para su expresión en un sistema documental.

2.1. Elementos para la identificación del contenido de los documentos

El contenido de un documento se compone de una serie de unidades de información dotadas de una mayor o menor complejidad interna y en profunda interrelación entre ellas, que se traducen en otros tantos niveles y elementos de información a identificar y representar durante el tratamiento documental. Básicamente, el contenido de un documento se puede reducir en un primer nivel a la presencia de una o varias materias, las cuales poseen una naturaleza compuesta y relacional formada por un tejido de conceptos que se expresan mediante la lengua en forma de términos. El documento se puede comparar a un objeto compuesto de materiales (materias) diversos formados por moléculas (conceptos) que se componen de átomos (términos) que a su vez poseen elementos atómicos (lexemas).

Por consiguiente, la identificación del contenido de un documento se reduce al descubrimiento de la naturaleza y las características concretas que adoptan en ese documento los elementos que componen cada uno de esos tres niveles de significado. Según el nivel de representación en el que se actúe se realizará un tipo u otro de indización: indización sintética si se detiene en las materias, indización

analítica si profundiza hasta los conceptos e indización automática por unitérminos si es un ordenador quien extrae términos.

La *materia* de un documento no se debe considerar una unidad simple, unidimensional y estable, sino que si consiste, de acuerdo con la norma UNE 50-121-91, en "cualquier concepto o combinación de conceptos que representa el tema de un documento» y el conjunto de sus relaciones, estamos ante una realidad compleja cuya plena identificación y comprensión exige atender durante su análisis a su capacidad para transmitir el conglomerado de conceptos que forman el contenido de un documento. Asimismo, la materia de un documento, según indica Langridge (1992, p. 74-75) parafraseando a Ranganathan, consiste en el asunto o los asuntos de los que trata ese documento más la forma del conocimiento desde la que son tratados; entendiendo por asunto aquello de lo que se ocupa un documento o, de modo más preciso, esa parte de la materia que adopta la expresión de fenómenos y relaciones. En definitiva, la materia de un documento no se puede considerar atendiendo únicamente a su naturaleza intrínseca, observada como algo ajeno al documento en el que se encuentra o encarna, sino que su plena comprensión exige, en primer lugar, desvelar la relación dinámica que los diversos asuntos, expresados por los conceptos, establecen dentro de un documento, ya que de esa relación se derivan modificaciones de significado; y, en segundo lugar, atender al ámbito disciplinar en que se sitúa el documento y la perspectiva desde la que el autor aborda la construcción del discurso.

Un *concepto* es una representación mental de una parte de la realidad física o intelectual, que se construye a partir de la atribución, extracción y abstracción de una serie de caracteres comunes a un conjunto de objetos o fenómenos individuales, para formar una unidad de pensamiento que se expresa mediante un término compuesto de una o varias palabras u otro símbolo no lingüístico, como las expresiones gráficas (por ejemplo, una señal de tráfico). El concepto es tanto el resultado de un proceso de comprensión y representación de la realidad como el elemento básico para expresar y comunicar esa comprensión. Pero los conceptos no aparecen como entidades aisladas en el sistema cognitivo humano —y, por tanto, tampoco se muestran así en los documentos que conservan y transmiten el conocimiento—, sino que poseen una elevada capacidad intrínseca de relacionarse y de agruparse entre sí, formando, a modo de redes o de mosaicos de teselas, unos conjuntos denominados campos semánticos, que a su vez se articulan en torno a unos macroconceptos, donde las relaciones adoptan la forma de relaciones de significado. Un concepto consiste, básicamente, en una unidad de conocimiento estructurado que se materializa en un contexto concreto. Por consiguiente, el contenido de un concepto sólo se puede establecer de modo preciso mediante el descubrimiento de sus relaciones con otros conceptos en el seno de un campo.

Y por *término* se se entiende la representación o expresión lingüística de un concepto, que puede ser tanto una palabra como un grupo de palabras; entendiendo por palabra una secuencia de letras enmarcada por espacios en blanco que tiene un significado inteligible para el ser humano. No obstante, término y palabra no deben confundirse, ya que les distinguen dos caracteres. Primero, el término forma parte de un sistema de términos, que a su vez es la representación lingüística de un sistema de conceptos, perteneciente a la terminología de un campo del saber; es decir, el término expresa siempre un significado preciso y concreto en el marco de las relaciones que se establecen en el seno de cada lenguaje especializado. Y segundo, que se deriva de lo anterior, el término posee un grado superior de precisión o un contenido especial desconocido en la lengua general debido a que ha sufrido, para integrarse en el sistema al que pertenece, un proceso de normalización terminológica en su relación con el concepto que expresa; es decir, un concepto se expresa siempre mediante un único término y un término remite o expresa un único concepto, en el seno de un lenguaje especializado concreto. En conclusión, las palabras son propias de la lengua general y los términos de los lenguajes especializados; por consiguiente, la indización automática no se ocupa, en sentido estricto, de los términos, sino de las palabras, excepto cuando se realiza con el auxilio de un diccionario que indica los términos que se deben retener.

2.2. Elementos para la comprensión del significado de los documentos

Debido a la naturaleza compuesta y relacional del contenido de los elementos que representan y transmiten la información en un documento, la identificación de esta información no se obtiene directamente de la extracción independiente y sucesiva de materias simples, conceptos y términos. Esa operación exige, por el contrario, descubrir, ante todo, los conceptos presentes y el carácter de las relaciones que estos establecen entre sí dentro de un documento, para conseguir una correcta comprensión y precisión de su significado y el de las materias que forman mediante su combinación. Para ello se cuenta con el auxilio de otro elemento, la faceta, que permite descubrir las relaciones que mantienen entre sí los conceptos mediante la formulación de una serie de preguntas peculiares para el dominio disciplinar en que se sitúan la materia o las materias del documento y la forma de conocimiento mediante el que ese dominio se aborda. Esta técnica de indización se denomina análisis de facetas, la cual está estechamente vinculada con la teoría de los campos léxicos, semánticos y conceptuales. Por tanto, la fijación y el conocimiento de las facetas y los campos presentes en una disciplina y en un conjunto de materias facilitará la extracción y la comprensión de las unidades básicas del análisis de contenido.

Por *faceta* se entiende cada una de las diversas características semánticas o unidades de significado que se pueden distinguir dentro de una materia a partir del análisis de su significado, utilizando varios enfoques o perspectivas de análisis comunes para la disección de todas las materias que forman parte de una misma disciplina abordada desde una peculiar forma de conocimiento. Estas perspectivas de análisis o de disección del contenido de una materia también se denominan facetas. El contenido particular que adopta cada faceta en una materia se expresa mediante un concepto, de modo tal que en torno a cada una de las presentes se agrupa un conjunto de conceptos específicos para la categoría semántica que representa. Por ejemplo, según Vickery (1975, p. 181-189), la materia de documentos sobre Edafología se descubre y se expresa mediante la identificación y la reunión de una serie de conceptos que aluden al contenido concreto que adoptan en un documento cada una de las ocho facetas o categorías semánticas siguientes: tipo de suelo (pradera), estructura (granular), constituyentes (fósforo), propiedades (consistencia), procesos (nitrificación), operaciones (drenaje), técnicas de laboratorio (disolución) y generalidades (estabilidad).

Faceta es, por tanto, una noción bidimensional. Por una parte, se utiliza para denominar una característica intrínseca a la naturaleza de las materias que se encuentran en los documentos. Y, por otra parte, alude a las divisiones analíticas o categorías semánticas que emplea el análisis de facetas con objeto de descubrir el contenido concreto que adoptan esos caracteres en un materia, perimitiendo, por tanto, la comprensión de su significado y la identificación de los conceptos que indican la información que transmite un documento.

El origen de la noción de faceta y su primera construcción conceptual se encuentran en la obra de Ranganathan. Su nacimiento deriva de la observación y el análisis por el bibliotecario hindú de un fenómeno que ya hemos descrito: la naturaleza compuesta y relacional de las expresiones y nociones mediante las que el hombre representa la realidad y, por tanto, también de la información que transmite un documento expresada bajo la forma de materias. Para Ranganathan, el tema de un documento tiene una naturaleza compuesta, ya que se trata de un agregado de materias específicas donde cada cada una de las cuales es el tema principal tratado desde una perspectiva particular y formado por una composición más o menos compleja de conceptos simples. Pero la complejidad de estos conceptos no se limita a la diversidad y riqueza de los objetos físicos o abstractos a los que remiten, sino que también se basa en la existencia de una riqueza de relaciones entre ellos. Es decir, un tema no incluye solamente una multiplicidad de referentes físicos de las entidades clasificadas, con su correlato en la realidad, sino que por ser producto de una reflexión humana cada concepto y el conjunto de todos ellos remiten también a una serie diversa de elementos tales como agentes, funciones, procesos, operaciones, propiedades... introducidos por el autor del

documento y que, por consiguiente, se deben desvelar durante el tratamiento documental (Vickery, 1975, p. 8-10). La existencia de estos elementos, denominados facetas, se basa en el hecho de que el contenido de un concepto no se puede entender de modo aislado del resto, sino que la plena comprensión de su significado sólo es posible a partir de su agrupación con otros vecinos en un mismo campo semántico desvelando las relaciones que teje con ellos, así como las que establece el campo al que pertenece con la totalidad de los campos del universo.

Para lograr una correcta representación del contenido de un documento, el indizador debe identificar con precisión los diversos términos que se usan para expresar los conceptos que aparecen en el documento y atribuirles un significado preciso. Esta doble operación se efectúa, cuando se sigue una técnica basada en el análisis de facetas, a partir del conocimiento de las relaciones léxicas y semánticas que los términos y los conceptos establecen, respectivamente, con otras unidades léxicas y conceptos dentro de una parcela más o menos reducida del conjunto del vocabulario, en cuyo seno se delimitan sus caracteres. Es decir, la plena comprensión del significado de los conceptos y la correcta identificación de los términos que los representan sólo se consigue en el marco de las agrupaciones en las que ambos se tienden a reunir en la lengua natural.

Detrás de esta hipótesis se encuentra una sólida teoría sobre cómo se relacionan y asocian las palabras y los conceptos, en definitiva, sobre como se articula la lengua. Esta teoría fue expuesta inicialmente por Ipsen en 1924 con su imagen del mosaico de la lengua, cuando afirmó, según García Gutiérrez (1990, p. 104-105): "las palabras autóctonas no están nunca solas en una lengua sino que se encuentran reunidas en grupos semánticos. Con ello no hacemos referencia sólo a un grupo etimológico o a palabras reunidas en torno a supuestas raíces, sino a aquellas cuyo contenido semántico objetivo se relaciona con otros contenidos (...) de tal manera que el grupo forma un campo estructurado en sí mismo, de contornos acoplados y las palabras quedan englobadas en una unidad semántica de orden superior". Trier precisó unos años después la tesis central de esta teoría con la afirmación de que "la exactitud de la comprensión de una palabra individual depende de la presencia psíquica de todo el campo y de su particular estructura" (ibidem, p. 106).

Guiados por esta convinción, los lingüistas han realizado numerosas investigaciones sobre la noción de *campo*, que han provocado el enriquecimiento de esa teoría relacional sobre la semántica del lenguaje y la presentación de una variada tipología de campos. Sin embargo, pese al evidente interés del estudio de los campos tanto para el indizador como para el gestor de lenguajes documentales, lo cierto es que apenas se ha prestado atención a esta teoría desde la Ciencia de la Información Documental. Si bien, por fortuna, algunos autores han comenzado a destacar en los últimos años el hecho de que la noción de campo se halla estre-

chamente emparentada con las nociones de faceta y de sistema de conceptos; apostando, por un parte, en su presentación como el sustrato teórico que legitima y permite la construcción de sistemas de conceptos, y, por otra parte, proponiendo su utilización como un instrumento que facilita el descubrimiento y la agrupación de los conceptos, con el consiguiente establecimiento de las relaciones pertinentes en un dominio concreto de la realidad.

La imagen de una granada ayuda a entender la función del campo en la indización analítica: la fruta granada sería el documento, donde se identificarían el grano con el concepto, la faceta con las partes de la granada, el campo con las subpartes de la granada en las que los granos (conceptos) se estructuran alrededor de un grano central (macroconcepto); si bien, en un documento únicamente están presentes de cada faceta un número reducido de los conceptos que tiene adscritos.

3. Etapas y tareas de la indización analítica

La indización analítica consta al igual que otras técnicas de representación documental, como la indización sintética y la clasificación, de dos etapas, compuestas de una serie de tareas y actividades: la fase de la idea, destinada a la identificación, comprensión, selección y extracción de los elementos de información que contiene el documento; y la fase verbal, centrada en la expresión del contenido mediante términos de indización y el establecimiento de relaciones entre estos en el seno de un índice. El detalle de las tareas y actividades que se realizan es el siguiente:

1. Fase de la idea: análisis conceptual del documento.

1.1. Identificación y comprensión de los conceptos. Actividades:

1.1.1. Examen del documento.

1.1.2. Complementariedad con otras técnicas documentales:

- Indización sintética: determinación de la materia.
- Resumen.

1.1.3. Análisis de facetas: diseño y uso de parrillas de indización.

1.2. Selección de conceptos. Se usan los siguientes criterios:

- Exhaustividad y profundidad.
- Especificidad y precisión.
- Relevancia y pertinencia.
- Coherencia y consistencia.

2. Fase verbal y notacional: expresión terminológica.

2.1. Conversión de los conceptos en términos de indización. Opciones:
- Términos de indización libres.
- Términos de indización controlados.

2.2. Construcción de la cadena de indización. Actividades:

2.2.1. Creación de un orden de citación de los términos de indización.

2.2.2. Establecimiento de enlaces sintácticos. Tipos:
- Yuxtaposición.
- Ponderación.
- Especificación: de puntos de vista, enlace o posición.
- Integración en un enunciado estructurado.
- Reenvío.

2.3. Integración de los términos de indización en índices estructurados.

4. Identificación y comprensión de los conceptos

La indización sintética de un documento no requiere un conocimiento profundo de la disciplina de la que trata, aunque este siempre sea deseable. Sin embargo, la indización analítica exige, si no ser especialista en la disciplina, sí tener, al menos, unos buenos conocimientos sobre sus objetivos, métodos y terminología. Evidentemente, hay campos como la indización de información periodística de actualidad de carácter general para los que puede ser suficiente disponer de una buena cultura general. Además, en cualquier caso, es necesario poseer unas características personales como curiosidad, capacidad de concentración, gusto por la lectura, poder de observación, facilidad y rapidez de comprensión lectora en los diversos niveles (palabra, frase, párrafo, capítulo y documento completo), elevado nivel de análisis y síntesis, minuciosidad y ganas de aprender.

Para facilitar la identificación y la comprensión de los conceptos es conveniente que esta tarea se realice si no a continuación de otras técnicas documentales como la indización sintética, la clasificación y el resumen, sí por lo menos de modo complementario a éstas. La indización sintética o determinación de la materia o las materias presentes en el documento ayuda a preveer los conceptos a identificar, a valorar la importancia de los presentes y a determinar la perspectiva desde la que se consideran (la forma del conocimiento). Por tanto, a pesar de que únicamente se realice una indización analítica, es conveniente extraer la materia de la que trata el documento aunque no sea con mucha precisión ni detalle, pudiéndose utilizar para ello los términos cabecera de un tesauro o los descriptores superiores en su jerarquía si se emplea este tipo de lenguaje documen-

tal como instrumento auxiliar de la indización. En cuanto al resumen, las ventajas que se obtienen de una indización efectuada no sobre un resumen pero sí tras realizar un resumen formal, sobre todo si es de tipo informativo, son evidentes, ya que de este modo se controla plenamente una actividad que de todos modos se efectúa aunque sea de modo inconsciente, y, por tanto, informal. Como mínimo, debería realizarse un resumen indicativo, pues con la inversión de un poco más de tiempo se mejora la indización y se ofrece, además, otro producto documental muy apreciado por el usuario.

No obstante, la tarea central y fundamental de la indización analítica es el diseño y uso de parrillas de indización construidas sobre la base del análisis de las facetas que caracterizan a un conjunto de materias relacionadas entre sí. Este tipo de indización es el resultado de la crítica que Ranganathan realizó, a partir de su reflexión sobre la naturaleza compuesta y relacional de la materia, al modo de representación y organización del conocimiento dominante en su tiempo, basado en la identificación del contenido de los documentos con una materia unidimensional para su posterior inclusión como una clase subordinada dentro de una estructura prefijada de acuerdo con un sistema de clasificación jerárquico, por lo que únicamente permite establecer relaciones de inclusión o de coincidencia entre los documentos de una colección. Como alternativa, Ranganathan propuso la faceta como la base de una nueva técnica de representación documental, denominada análisis de facetas (también conocida como indización analítico-sintética), que, en lugar de dedicarse a la extracción de materias simples, se centra en la identificación basada en en el análisis y la expresión sintética de los conceptos mediante los que se muestran la materia o las materias que componen el tema de un documento, a partir de la transformación de las facetas en preguntas presentadas bajo la forma de "parrillas de indización" peculiares para cada colección documental, para proceder a la posterior expresión de la materia mediante cadenas de términos de indización que indican las relaciones que se establecen entre los conceptos que la componen.

Por ejemplo, los indizadores de la base de datos Pascal (Ménillet, 1992) proponen la distinción de las siguientes categorías para el análisis de documentación médica: enfermedad (malformación), término anatómico (arteria pulmonar), finalidad del estudio (diagnóstico, etiología, determinismo genético), materia de estudio (prematuro), tipo de estudio (estudio comparativo), fenómeno fisiológico (presión arterial), técnica de exploración (arteriografía), tratamiento (vasodilatador) y lugar geográfico. En cambio, si se indiza documentación de actualidad se deben utilizar las facetas bajo las que se articula el esquema comunicativo básico para la redacción del *lead* (delantera) de una noticia, los 7 W: sobre quién se informa (*who*), sobre qué hecho se informa (*what*), lugar dónde acaece el hecho sobre el que se informa (*where*), cuándo ha sucedido el hecho (*when*), por qué se

ha producido ese hecho (*why*), cómo ha sucedido (*how*) y para qué ha sucedido o que consecuencias ha tenido (*what for*).

En caso de que no exista un estudio previo sobre las facetas operativas en una disciplina pero se desee hacer uso de esta técnica, se puede utilizar como alternativa para el diseño de parrillas de indización de las pautas que propone la norma UNE 50-121-91 para ayudar a identificar los conceptos presentes en un documento:

a) ¿Trata el documento de algún objeto sometido a una acción?

b) ¿Contiene algún concepto activo? (por ejemplo, una acción, un procedimiento, etc.).

c) ¿Se ve afectado el objeto por la acción identificada?

d) ¿Trata del agente causante de la acción?

e) ¿Se describen los medios para llevar a cabo la acción? (por ejemplo, instrumentos, técnicas o métodos especiales).

f) ¿Existen factores considerados en un medio o lugar particular?

g) ¿Se identifican variables dependientes o independientes?

h) ¿Se trata el tema desde un punto de vista particular no asociado normalmente a ese campo? (por ejemplo, estudio de la religión desde un punto de vista sociológico).

Se pueden utilizar, por tanto, la siguientes facetas: concepto aglutinante, objeto, acción, procedimiento, causa, consecuencia, medios de procedimiento, factor temporal, contexto espacial, variables dependientes, variables independientes, enfoque singular, etc.

Durante el análisis de contenido se suelen adscribir a cada faceta un determinado número de conceptos peculiares que, a su vez, no aparecen vinculados con el resto de las facetas; es decir, en torno a cada faceta se agrupan un conjunto de conceptos específicos para la categoría que representa. Asimismo, el grupo de conceptos vinculado a una faceta está dotado de una estructura interna, determinada por los campos semánticos operantes en la lengua, ya que el significado de los conceptos dentro de cada faceta está condicionado por las relaciones semánticas que estos establecen entre sí dentro de los marcos naturales de agrupación de la lengua natural: campos léxicos que remiten a campos conceptuales. Este hecho ha permitido diseñar unos lenguajes documentales auxiliares del análisis de facetas para documentación específica: primero, las clasificaciones facetadas y, más adelante, el tesauro facetado.

En una clasificación facetada los conceptos no se disponen siguiendo una única escala de subordinación o árbol de conocimiento, sino que se reúnen de

acuerdo con la categoría semántica a la que pertenecen, también conocidas como facetas o divisiones analíticas, y dentro de cada uno de los grupos creados se jerarquizan generando unas estructuras arborescentes autónomas y paralelas; es decir, cada criterio de división da lugar a su propio árbol clasificatorio de entidades. Por consiguiente, la expresión de la materia de un documento utilizando una clasificación facetada no se limita a su integración en una única clase, sino que la notación por la que se representa es el resultado de la combinación y síntesis de acuerdo con una serie de reglas preestablecidas de los códigos que indican las clases de cada jerarquía a las que pertenecen los diversos componentes de la materia, identificados a partir de su análisis desde la perspectiva de cada una de las dimensiones semánticas contempladas para la estructuración del sistema de clasificación utilizado como auxiliar. Debido a la combinación del análisis y la síntesis durante el proceso de identificación y representación de la materia, las clasificaciones facetadas se denominan también clasificaciones analítico-sintéticas. Este tipo de sistema de clasificación dió origen en la década de los años sesenta a los tesauros gracias a su combinación con la técnica de representación basada en los unitérminos.

5. Selección de los conceptos

La selección y extracción de los conceptos que representan el contenido de un documento debe seguir unos criterios que ayuden a decidir que conceptos entre los presentes deben retenerse durante la indización y cuáles desecharse. Como indica la norma UNE 50-121-91: "el indizador no tiene necesariamente que utilizar como términos de indización todos los conceptos identificados durante el examen del documento". Estos criterios se presentan vinculados por parejas, donde un criterio condiciona al otro: exhaustividad y profundidad, especificidad y precisión, relevancia y pertinencia y coherencia y consistencia. Además, las cuatro parejas se encuentran relacionadas entre sí de tal modo, que no se puede aplicar una sin tener en cuenta el resto. Asimismo, la aplicación de esos criterios para la selección de conceptos está determinada en todo caso por la presencia de cuatro factores: el carácter del documento a indizar, la finalidad con que se van a utilizar los términos de indización (por ejemplo, imprimir índices alfabéticos o crear una base de datos bibliográfica), el tipo de contenido y de objetivo de cada documento, y las características del usuario dominante en el sistema de información en que se realiza la indización.

5.1. Exhaustividad y profundidad

La exhaustividad en la identificación y la selección de conceptos durante la indización está "relacionada con el número de nociones que se tienen en cuenta, y que caracterizan el contenido íntegro de un documento" (norma UNE 50-121-

91). Se trata de recoger todos aquellos conceptos que cumplan dos condiciones: en primer lugar, que representen información realmente presente en el documento; y, en segundo lugar, que esta información sea significativa. Es decir, se debe seleccionar conceptos que tengan a la vez carga informativa o importancia en el documento y valor potencial (pertinencia) para los usuarios del sistema de información.

La exhaustividad trata de conseguir un compromiso entre no dejar parcelas del contenido de un documento sin representar y extraer información presente pero de escasa utilidad para los usuarios. La norma UNE 50-121-91 asume este compromiso cuando apunta que "La cobertura de la indización no debe interpretarse de una forma demasiado estricta. Hay que tener en cuenta que términos de indización creados inicialmente para un grupo de usuarios (por ejemplo, científicos y técnicos) pueden ser utilizados por otros grupos de usuarios (por ejemplo, economistas). Se aconseja que los indizadores de literatura científica y técnica tengan presente otros aspectos del tema, en particular los sociales y los económicos"; para más adelante matizar: "El principal criterio de selección de conceptos debe ser su valor potencial como elemento de expresión del tema del documento para su recuperación. En la selección de conceptos, el indizador debe tener en mente las preguntas que se pueden hacer al sistema de información, en la medida en que dichas preguntas se pueden conocer. En efecto, este criterio constituye la principal función de la indización".

En cuanto a la primera dimensión del principio de exhaustividad, la selección de conceptos que transmitan información real, se pueden dar algunos consejos para su aplicación. Así, Van Slype (1996, p. 116-117) aconseja rechazar tres tipos de nociones: las nociones que están relacionadas con las expuestas en el documento pero de las que el autor afirma que no va a tratar, las nociones que sólo aparecen a título de ejemplo y las nociones que aparecen sin contener una información suficiente o completa para interesar al usuario. Asimismo, recomienda retener "las nociones explícitamente presentes sobre las que se aporta información susceptible de responder a una necesidad del sistema documental" (ibidem, p. 117).

En cuanto a la segunda dimensión del principio de exhaustividad, la selección de información útil para el usuario, su cumplimiento exige del indizador que reflexione tanto sobre los conceptos que son más apropiados para el carácter del grupo de usuarios al que sirve como sobre las preguntas habituales que este grupo puede hacer al sistema de información. La importancia de este principio es tal, que el indizador deberá identificar y seleccionar también "las nociones implícitamente contenidas dentro del documento que, si bien no son designadas como tales por el autor, son tratadas suficientemente en detalle como para interesar oca-

sionalmente a los usuarios" (ibidem, p. 117). Asimismo, podrá modificar incluso tanto el procedimiento de indización como el lenguaje de indización auxiliar que usa, si bien sin distorsionar su estructura y lógica de construcción, tras comprobar a lo largo de sucesivas búsquedas que el sistema no responde adecuadamente a las necesidades y las características de los usuarios (norma UNE 50-121-91).

¿Cuál es el grado de exhaustividad deseable? ¿cómo medir la exhaustividad de una indización? La respuesta a estos dos interrogantes está relacionada con la presencia de un segundo criterio presente en la selección de los conceptos identificados durante la indización, complementario al de la exhaustividad: la profundidad.

Algunos autores entienden por profundidad, e incluso por exhaustividad, el número de términos de indización que se utilizan para representar el contenido de un documento, partiendo del principio de que la indización será más profunda cuantos más términos se utilicen. Así, según Chaumier (187, p. 32), "la profundidad de indización es el término de descriptores afectados por el documento", situándose una profundidad media en una franja de ocho a doce descriptores por documento. En la misma línea, la profundidad de la indización alcanza para Pinto (1991, p. 130) diversos niveles cuyos límites se encuentran en el número de descriptores: indización genérica o superficial cuando únicamente se destacan los temas principales, indización intermedia o descripción del conjunto de los temas mediante un máximo de diez términos de indización e indización en profundidad o indicación de todos los temas del documento mediante más de diez términos. El documentalista encuentra impuestos estos niveles por factores tales como los efectivos y la cualificación del personal, el volumen y la naturaleza de la información a indizar, el sistema de almacenamiento y recuperación que se utiliza y los medios económicos con que se cuenta (ibidem, p. 130).

No obstante, no consideramos adecuada esta forma de abordar la profundidad, ya que si se atiende a los caracteres del principio de exhaustividad, el número de términos de indización que se pueden asignar a un documento debe ser resultado de una tensión dialéctica entre la cantidad de información que este contenga y las supuestas necesidades de los usuarios que prevea el documentalista. La limitación arbitraria de los términos de indización "puede conducir a una pérdida de objetividad en la indización y a una deformación de la información que se podrá utilizar en la recuperación" (norma UNE 50-121-91). Porque como advierte Van Slype (1991, p. 123), una indización poco exhaustiva provocará que no se recuperen documentos pertinentes, aumentando los silencios del sistema documental a las preguntas de los usuarios, pero una exhaustividad demasiado elevada permitirá, en cambio, recuperar documentos no pertinentes, disminuyendo la precisión de las respuestas.

Por consiguiente, la medición de la exhaustividad de una indización se debe efectuar desde la perspectiva de la adecuación del vocabulario de indización para la recuperación. Y esa adecuación depende tanto de la amplitud terminológica de los términos de indización elegidos para expresar los conceptos como de la existencia de un lenguaje documental para su control.

La amplitud terminológica de un término de indización puede medirse, según Maron (1979), en dos niveles: el intensional, basado en la amplitud semántica de su significado, y el extensional, relativo al número de documentos que puede indizar adecuadamente ese término. Estos niveles no guardan ninguna proporción, pues el mayor grado de generalidad de un término respecto a otro no supone, necesariamente, que aquel pueda abarcar un mayor número de documentos que éste durante la indización. Así, de acuerdo con la vertiente extensional de un término, un documento estará indizado con mayor profundidad que otro, cuanto más términos de indización reciba; sin embargo, de acuerdo con la otra vertiente, "aquel documento indizado con términos más específicos estará intensionalmente indizado con más profundidad que el que contenga términos más genéricos", con independencia del número de términos que se le asigne. Asimismo, este hecho está relacionado con la presencia de un lenguaje documental, de modo que la amplitud intensional crece y, por tanto, se logra una mayor exhaustividad, cuando se emplea un tesauro específico para el ámbito del conocimiento al que pertenece el documento, que si se usa un tesauro no específico o una lista de descriptores libres.

En definitiva, la determinación de la profundidad de la indización no se basa en el número de términos de indización empleados, sino que está ligada con el grado de profundidad semántica que poseen esos términos de indización en el lenguaje de recuperación que resulta de una indización libre o en el tesauro que se utiliza en una indización controlada. Este fenómeno se conoce como el control de la amplitud terminológica intensional de las unidades de representación; el cual exige tener presente durante la indización el principio de especificidad, de tal modo que entre profundidad y especificidad se establece una relación directamente proporcional, ya que con el uso de términos específicos se obtiene mayor profundidad y con el empleo de términos más generales, aunque estos sean más numerosos, se obtiene menor profundidad. Este hecho conjura el riesgo de que una profundidad demasiado elevada aumente el ruido durante la búsqueda o que excesivamente baja provoque silencio, cuando la exhaustividad se entiende exclusivamente en clave de número de términos de indización, es decir, cuando se confunde con la amplitud terminológica extensional.

5.2. Especificidad y precisión

La especificidad en la selección y expresión de los conceptos durante la indización consiste en reducir al máximo posible la amplitud del significado de los términos, eliminando la presencia de términos generales como unidades básicas de representación; es decir, se trata de lograr la mayor concreción posible en la utilización de los términos, con objeto de representar efectivamente los conceptos presentes en el documento. La especificidad se reduce cuando un concepto se representa mediante un término que tiene un significado más general. De ahí que, como indica la norma UNE 50-121-91, "la especificidad está relacionada con la exactitud con que un concepto particular que aparece en un documento está representado por un término de indización".

La exactitud o precisión consiste en la representación o expresión lingüística del concepto presente en un documento por el término más adecuado; lo cual exige para su logro efectivo que un concepto se exprese siempre y únicamente mediante un término y un término de indización en el sistema de información remita exclusivamente a un concepto. A su vez, conseguir esto depende, en gran medida, de la existencia de un lenguaje documental como auxiliar de la indización y de la calidad de ese lenguaje.

Asimismo, cuando la indización se efectúa con el auxilio de un tesauro podemos distinguir, según van Slype (1991, p. 123), dos tipos de especificidad: una vertical y otra horizontal. La especificidad vertical significa que el descriptor elegido debe situarse en el mismo nivel de especificidad que el concepto representado o, por defecto, en el nivel jerárquico inmediatamente superior en el tesauro, pues de este modo se aumenta la precisión en la formulación de la ecuación de búsqueda. Por ejemplo, el concepto "vaca frisona" se indizará por "vaca frisona", en tanto que exista en el tesauro, y no por "vaca". La especificidad horizontal consiste en que un concepto compuesto debe traducirse mediante un descriptor precoordinado, si existe en el tesauro, antes que por la asociación de descriptores simples o de dos descriptores candidatos, ya que de este modo se disminuye el riesgo de efectuar falsas coordinaciones durante la búsqueda. Por ejemplo, el concepto "cultivos de huerta" se indizará por el descriptor "horticultura" antes que por los descriptores "cultivo" y "huerta".

No obstante, puede preferirse el uso de términos de indización más generales cuando el concepto que represente está poco desarrollado en el documento o cuando el documentalista "considere que un exceso de especificidad puede actuar de forma negativa sobre el sistema de indización" (norma UNE 50-121-91), por ejemplo, inflacionando el número de descriptores con términos poco significativos dentro del sistema documental desde la perspectiva de sus objetivos.

5.3. Relevancia y pertinencia

La relevancia de los conceptos identificados y seleccionados depende de la adecuación de los términos de indización que los expresan tanto para transmitir el contenido de los documentos mejor que otros términos, como para facilitar la posterior formulación de demandas al sistema de información. El principio de relevancia de los términos de indización propuestos se encuentra vinculado, por tanto, con el de pertinencia; es decir, con la capacidad de los términos de indización para conectar con las necesidades de los usuarios y la peculiaridad de su vocabulario, facilitando, por tanto, la comunicación de estos con el sistema de información. En resumen, un término de indización debe adecuarse, de acuerdo con los principios de relevancia y de pertinencia, tanto al contenido del documento como a las necesidades y el vocabulario de los usuarios.

Ambos principios se pueden entender, por consiguiente, como una extensión o desarrollo del principio de exhaustividad. Además, el principio de relevancia se encuentra estrechamente relacionado con la caracterización de la profundidad desde la perspectiva de la amplitud terminológica de los términos de indización, ya que la distinción entre intensión y extensión de un término, propuesta por Maron, se basa en la idea de la relevancia documental a partir de la consideración de un documento como "una colección de unos cuantos conceptos relevantes" (Landry y Rusch, 1968. Ref. García Gutiérrez, 1984, p. 123).

5.4. Coherencia y consistencia

El principio de coherencia consiste que en un sistema de información varios indizadores o un mismo indizador en momentos distintos asignen idénticos términos de indización a un mismo documento. Este principio también se encuentra estrechamente vinculado al de exhaustividad; así, la plena coherencia se logra cuando "los términos de indización asignados a un documento y el nivel de exhaustividad conseguido son idénticos en cualquier indizador" (norma UNE 50-121-91).

Lograr la coherencia en la indización exige obligatoriamente la presencia de un lenguaje documental que ayude a identificar y representar los conceptos presentes en un documento. Pero este lenguaje documental debe poseer como una de sus principales características la consistencia, es decir, que el alcance del léxico empleado se circunscriba a significados previamente seleccionados y semánticamente controlados, para evitar confusiones y variaciones en el empleo de los vocablos derivadas de la ambigüedad semántica de la lengua natural. En definitiva, cada término en el lenguaje documental debe ajustarse, deliberadamente, a un único significado mediante la eliminación de sinónimos y cuasisinónimos, lo cual se logra mediante el respeto del principio de univocidad durante su construcción (Esteban, 1997, p. 138-140).

Evidentemente, pese a disponer de un lenguaje documental de tal tipo, una coherencia total es una meta inalcanzable debido a los diversos factores que intervienen en el proceso de indización, destacando por su dificultad de control los relativos al carácter intelectual y personal del proceso. Los indizadores suelen percibir de forma diferente el contenido real del documento, la parte de ese contenido susceptible de responder a las necesidades de los usarios, los conceptos más importantes y los términos de indización elegidos para su representación (Van Slype, 1991, p. 32). Y tampoco se debe olvidar el peso de otros factores actuantes en la coherencia de la indización —si bien más fácilmente controlables—, como la existencia y calidad de un manual de indización, la formación recibida por los documentalistas, su meticulosidad, etc.

Por consiguiente, como recoge la norma UNE 50-121-91, la coherencia sólo es posible en una "situación ideal". Sin embargo, no por eso un sistema de información ha de abandonar la aspiración de conseguir la máxima coherencia, máxime porque se trata también de una necesidad ineludible cuando la información se intercambia entre diferentes centros miembros de una misma red.

6. Expresión de los conceptos

Una vez seleccionados los conceptos a representar de acuerdo con los criterios expuestos, llega el momento de convertirlos en términos de indización. Esta traducción es ante todo un proceso de control de vocabulario, que puede efectuarse de dos formas: de modo espontáneo a partir del documento o mediante el auxilio de un lenguaje documental. Asimismo, esas unidades de información se pueden expresar mediante códigos numéricos, alfabéticos o alfanuméricos, cuyo significado se encuentra en un lenguaje documental auxiliar, o mediante palabras claves extraidas de la lengua natural cuya morfología y semántica pueden estar normalizadas si también se usa un lenguaje documental.

Habitualmente, los indizadores recurren a la palabra clave, por lo que es fácil confundirla con la noción de término de indización, sin embargo, aquella es únicamente un tipo concreto de esta clase de término. Así, la norma UNE 50-121-91 define término de indización como "La representación de un concepto en forma de un término derivado del lenguaje natural, preferiblemente un sustantivo simple o compuesto; o de un código de clasificación. Un término de indización puede constar de más de una palabra. En un lenguaje de indización controlado, un término se designa como descriptor o no-descriptor". Por consiguiente, cuando el término de indización se ofrece de modo codificado se denomina código o notación y cuando se presenta en lengua natural se denomina palabra clave. De acuerdo con el grado de condensación, el nivel de control y el tipo de lenguaje documental que se emplee para su normalización, se distinguen varias clases de palabras claves: unitérmino (palabra clave no normalizada), encabezamiento de

materia (palabra clave normalizada mediante una lista de autoridades o de encabezamientos de materia) y descriptor (palabra clave normalizada mediante un tesauro), sobre las que se establecen otros tantos sistemas de indización.

El control de vocabulario de los términos de indización libres o palabras claves no normalizadas se caracteriza por un control mínimo en el ámbito semántico y entre escaso o elevado en el campo morfológico. El grado de control morfológico depende de si el documentalista designa los conceptos tal y como los encuentra en el documento o si los enuncia él mismo sometiéndolos a una cierta normalización: por ejemplo, transformando las formas verbales y adjetivas en formas nominales ("medida" en vez de "medir"; "longitud" en lugar de "largo") o respetando ciertas convenciones respecto al género y el número.

El índice que resulta de la agrupación de todos las palabras es un lenguaje documental de estructura muy simple: una lista alfabética de términos sin ningún tipo de relaciones semánticas. Este lenguaje documental puede adoptar dos formas: lista de palabras clave cuando no se ha efectuado ningún tipo de control morfológico y semántico durante la indización; o lista de descriptores libres, si sobre una lista de palabras clave se realiza cierto grado de control de la forma y del significado de sus términos durante y después de la indización. Cuando una lista de palabras claves evoluciona a lista de descriptores libres, ya se inicia, en sentido estricto, un proceso de control, pues en posteriores indizaciones el documentalista debera comparar y controlar los términos extraídos con las designaciones normalizadas que figuran en la lista.

En cambio, los términos de indización controlados por un lenguaje documental diseñado *a priori* se caracterizan por un elevado control en los campos morfológico, semántico y sintáctico, por igual. No obstante, se logra un grado mayor de control cuando se utiliza un tesauro de descriptores en lugar de una lista de autoridades, debido a que la riqueza de sus relaciones semánticas y su grado de especificidad se revelan unos eficaces auxiliares para la selección de los términos de indización. Las peculiaridades del uso de cada uno de estos lenguajes de indización han sido descritas con gran claridad por Van Slype (1992), donde también se describen las actividades de que consta la tarea final de la indización analítica: la construcción de la cadena de indización (ibidem, p. 119-122).

7. La función del lenguaje documental

En los últimos años, con el desarrollo de los sistemas automatizados de indización por unitérminos, la indización analítica controlada ha perdido parte de su prestigio e incluso se cuestiona su conveniencia universal para todo tipo de documentos textuales. Sin embargo, en nuestra opinión, esta no debería desaparecer, pues, por una parte, la indización analítica todavía presenta grandes ventajas, y,

por otra parte, si se acepta lo anterior, únicamente el uso de un lenguaje documental construido de acuerdo con unos principios científicos garantiza la correcta elaboración de las parrillas de indización y la adecuada utilización de los criterios que se deben seguir durante la selección conceptual (Esteban, 1997).

Para disponer de sólidos argumentos en favor de la defensa de indización analítica controlada, consideramos necesario recordar, previamente, como preámbulo, cual es la función de un lenguaje documental en un sistema de información. El problema al que los lenguajes documentales se presentan como solución es tan viejo como la invención de la escritura. Por un lado, alguien desea encontrar información sobre un tema que le preocupa; por otro, existe un número inmenso de documentos pero no se pueden encontrar si no se conoce previamente el autor o el título. Surge, por tanto, la necesidad de inventar un sistema de búsqueda capaz de indicar al demandante todos los documentos que pueden ser de su interés. Para ello se siguen cuatro etapas: comprensión, extracción y selección del contenido de cada documento presente en una colección, expresión de ese contenido en una fórmula de indización para su almacenamiento junto con las restantes en una lista ordenada con referencias al lugar donde se conserva el documento, expresión del tema de búsqueda y hallazgo de los documentos mediante el índice. Toda la eficacia del sistema descansa en la comparación de dos fórmulas, la del indizador y la del recuperador. Como ambos utilizan espontáneamente su vocabulario usual para concebir y expresar una materia, la solución más simple consiste en emplear la lengua natural para construir las fórmulas. Este es el sistema que sigue el índice analítico de un libro.

Pero, desafortunadamente, esta solución tan simple muestra pronto los límites de su eficacia: los términos que se utilizan para expresar una misma materia en lenguaje natural por el indizador y el buscador no siempre son coincidentes, por lo que convierten la comparación en inoperante. Las razones son bien conocidas: los idiomas de los protagonistas pueden ser diferentes (multilingüismo), cada uno concibe y formula sus ideas en función de su cultura y su léxico (factor individual) e incluso cuando estos factores de divergencia no existen, la sinonimia y la polisemia provocan inevitablemente variantes y ambigüedades en la enunciación de una materia.

¿Cómo preservar en la expresión de una materia la similitud del sentido a través de la variedad y la incertidumbre del lenguaje natural? La solución se ha encontrado en la invención de un lenguaje artificial, denominado lenguaje documental, que se caracteriza fundamentalmente por dos rasgos: por contener todos los términos autorizados para la expresión de las materias, asegurando de este modo una correspondencia total entre una materia y su enunciado (Maniez, 1993, p. 150), y por poseer una simplicidad de organización y una economía de signos donde cada elemento aporta la mayor cantidad posible de información al tiempo

que evita la redundancia no suministrando la información trasmitida por sus compañeros (Chaumier, 1977, p. 132). De este modo, el lenguaje documental desempeña un papel central como lenguaje intermediario entre una pregunta y una oferta de información, permitiendo ajustar exactamente la respuesta a la demanda sin confusión ni omisión.

Sin embargo, una materia es inseparable de su contexto y el interés por un tema suele implicar el interés por otros cercanos. Por ello, se considera conveniente que un lenguaje documental permita también al demandante modular su pregunta navegando entre la expresión estricta de una materia y expresiones vecinas o más amplias. Esta utilidad secundaria se suele denominar "función zoom" y exige que los lenguajes documentales contengan diversas relaciones semánticas entre sus términos.

En definitiva, del uso de los lenguajes documentales emana un diálogo comunicativo que complementa en un segundo nivel a la comunicación documental primera entre el creador de un documento y su lector (Chaumier, 1977, p. 126-148). La principal característica de este diálogo es la disminución de la ambigüedad en la emisión y la recepción de los mensajes, debido a que la existencia de un lenguaje documental proporciona un vocabulario para la demanda compatible con el utilizado durante el análisis del contenido de los documentos, ayuda a ordenar los términos durante la interrogación y permite focalizar o ampliar la ecuación de búsqueda, mejorando, de este modo, los resultados de la recuperación. Asimismo, un uso correcto del lenguaje documental durante el tratamiento y la recuperación de la información eliminan el silencio (la información demanda no recuperada) y el ruido (la información recibida no solicitada).

Este es el objetivo básico de los lenguajes documentales y legitimador de su existencia: la representación y organización de la información documental de un modo preciso y consistente con el fin de garantizar, por una parte, la adecuación de las representaciones y las relaciones que propone entre los documentos de un fondo a la naturaleza de la información conservada en ellos, y, por otra parte, la coherencia tanto de las representaciones efectuadas por diferentes documentalistas como entre el proceso de tratamiento y el de recuperación de la información. La función de los lenguajes documentales en la cadena documental no se limita, por tanto, a controlar el proceso de indización y a hacer posible la representación, sino que su presencia se siente, sobre todo, durante la formulación de las preguntas a un sistema de información.

Por último, señalar que los elementos para la identificación, comprensión y representación del contenido de los documentos también se utilizan para la construcción de los lenguajes documentales, ya que los sistemas de representación y organización deben estar formados por la misma materia que compone la infor-

mación documental para ser unos auxiliares efectivos durante los procesos de tratamiento y recuperación de información.

8. La indización analítica controlada frente a las tecnologías de la información

La reflexión sobre los fundamentos y la práctica de la indización analítica y de los lenguajes documentales no está hoy de moda, así como la defensa de su necesidad. En el mundo de la Internet y del tratamiento automático de la información no parece haber hueco para una de las prácticas documentales más importantes y un tema de investigación tradicional de la Documentación, a la que tanto ha ayudado a consolidar como técnica y como disciplina científica, así como a darle prestigio social. Pero eso no ha de significar que esa práctica y esa investigación se deban abandonar, pese a su actual minusvaloración en favor de las tecnologías de la información.

En nuestra opinión, todavía existe la necesidad de continuar investigando en el campo de la indización analítica y de seguir utilizado esta técnica para el tratamiento de documentación científico-técnica en particular y la especializada en general, allá donde la complejidad y la especificidad de los documentos y las necesidades de los usuarios exija un firme y eficaz control del contenido de una colección documental. Se trata de un ámbito donde todavía hay mucho que hacer, sobre todo en lo que afecta al diseño de programas de aplicación de sus principios y de control de su calidad, pues todavía son muy abundantes los problemas de ruidos, silencios e imprecisiones durante la recuperación documental en muchas bases de datos.

Esto no ha de interpretarse como una defensa de que el análisis de contenido y la gestión y el uso de los lenguajes documentales se puedan entender en la actualidad al margen de las tecnologías de la información. Porque los nuevos sistemas de captación y almacenamiento de información, los ordenadores que procesan los datos y las redes de transmisión de mensajes documentales no son exclusivamente unas herramientas al servicio de esa técnica y los lenguajes documentales, sino que modelan el nuevo entorno en el que estos se desarrollan y, por tanto, al que se deben adaptar.

El influjo de estas tecnologías actúa en dos direcciones:

- Por una parte, la expansión del ordenador ha contribuido al éxito del tesauro ya que al permitir utilizar la lógica booleana facilita el uso de sistemas postcoordinados y, además, permite almacenar las relaciones semánticas de un tesauro e integrarlas en el módulo de búsqueda. El resultado es que tanto el mantenimiento del tesauro como la reducción o ampliación de los términos seleccionados durante la búsqueda pueden ser gestionados auto-

máticamente sin que el usuario deba conocer la estructura del lenguaje o el mecanismo de las operaciones. El lenguaje documental se convierte en "transparente" sin perder por ello su categoría de elemento esencial del sistema de información.

- Pero, por otra parte, el ordenador también puede ser una alternativa al empleo de lenguajes documentales tanto durante la indización como durante la recuperación, con la proliferación de programas de indización e incluso de clasificación automáticos y la promoción de las estrategias de recuperación en texto libre. Sin embargo, un análisis más detallado de este hecho permite matizar esta conclusión.

En el caso de la indización automática hay que distinguir dos situaciones. El fenómeno más frecuente es la presencia de un programa de indización como complemento de la indización humana, donde la intervención del ordenador se limita a los títulos y eventualmente a los resúmenes para completar los términos controlados por el indizador. Un ejemplo corriente es el índice de las grandes bases de datos, que destinado a una función de puntero se contenta con seleccionar las palabras significativas del texto mediante la técnica sumaria pero eficaz de la eliminación de las palabras vacías, previamente decididas por un hombre. La segunda realidad es la total sustitución de la indización humana por la máquina. Todavía en fase de investigación, no sólo persigue la selección de los conceptos importantes sino también traducirlos a términos no ambiguos; lo cual exige introducir un léxico de sinónimos y parentescos semánticos, en otras palabras, un tesauro. Es, por tanto, erróneo pensar que los avances de la indización automática condenarán a los lenguajes documentales; más que de desaparición, se debería pensar en utilización compartida entre el indizador y el diseñador de programas informáticos. Sin embargo, el aumento exponencial de los documentos ofrecidos a los usuarios con el desarrollo de los servidores de información electrónica accesibles por la red Internet ha provocado que una especie del primer tipo de programas de indización, denominada robots o motores de búsqueda, sean los responsables absolutos sin intervención humana de la representación documental en este contexto.

En cuanto a la recuperación de la información, el avance de las técnicas informáticas está modificando la relación del usuario con los lenguajes documentales. El usuario ha adquirido la costumbre de interrogar por sí mismo las bases de datos en lenguaje libre en todo el texto o en campo predeterminados, ignorando la existencia del lenguaje documental que ha auxiliado la selección de los términos de indización. La comodidad de recuperaciones deficientes pero rápidas, confiando desmesuradamente en la propia pericia, vence a la alternativa de aprender el uso correcto de un lenguaje documental. La ignorancia animada por el deslumbramiento tecnológico promociona incluso la idea de que el princi-

pal obstáculo entre los documentos y un usuario es el lenguaje documental que actúa como mediador, desconociendo el objeto, el motivo y los resultados de esa función de mediación. Además, hay que considerar el hecho económico de que el coste informático de almacenar texto completo baja continuamente y el coste intelectual de indizar sube. Los límites actuales de la indización automática son evidentes. Como resultado de todos estos factores, cada vez existen más colecciones documentales sin indizar ni clasificar, descargando los productores de las bases de datos a los usuarios una parte cada vez más importante de las actividades de representación y organización documental con la incorporación de programas que permiten que el usuario índice los documentos en el momento de hacer la búsqueda.

Los documentos se almacenan en la base de datos en texto completo, sin palabras claves. El usuario introduce las palabras que definen su ecuación de búsqueda sin utilizar operadores lógicos. El ordenador revisa todos los textos y selecciona aquellos en los que aparecen los términos demandados, ordenados de mayor a menor relevancia según el número de veces que esas palabras están presentes mediante un sencillo cálculo. Sin embargo, como el resultado de la búsqueda no permite discriminar cuales son los documentos más pertinentes y relevantes a la consulta entre los obtenidos, se intenta corregir esta deficiencia mediante la técnica de la ponderación.

Polisemia, sinonimia, redundancia, conceptos que no se expresan en un texto por sus términos... son fenómenos ignorados por estos sistemas, que no evitan sus negativas consecuencias durante la búsqueda. Exhaustividad, especificidad, pertinencia, relevancia son indicadores de calidad sistemáticamente ignorados por programas que se ofrecen como el gran avance tecnológico de la documentación. En la actualidad, las ventajas de seguir una estrategia de recuperación que combine la búsqueda a través de índices controlados por un lenguaje documental con la interrogación por texto libre quedan eliminadas en favor del imperio de este último sistema. Lo cual nos invita a plantear si la evidente crisis de los lenguajes documentales no amenaza incluso con quebrar el modelo del ciclo de la información documental. ¿Un futuro de la gestión de la información documental sin lenguajes documentales no representaría un riesgo de retorno a un pasado predocumental?

9. Referencias

Chaumier, Jacques (1977). L'analyse documentaire : le traitement linguistique de l'information documentaire. París : Entreprise Moderne d'Edition, 1977.

Chaumier, Jacques (1987). Análisis y lenguajes documentales. Barcelona : Mitre, 1987.

Esteban Navarro, Miguel Ángel (1997). Construcción y mantenimiento de clasificaciones documentales. // Pinto Molina, María (ed.). Manual de clasificación documental. Madrid : Síntesis, 1997, p. 131-174.

García Gutiérrez, Antonio Luis (1984). Lingüística Documental. Madrid : Paraninfo, 1984.

García Gutiérrez, Antonio Luis (1990). Estructura lingüística de la documentación. Murcia : Universidad, 1990.

Langridge, Derek W. (1992). Classification. Londres : Bowker-Saur, 1992.

Maniez, Jacques (1993). Los lenguajes documentales de clasificación. Madrid : Fundación Germán Sánchez Ruipérez ; Pirámide, 1993.

Maron, M. E. Depht of Indexing. // Journal of the American Society for Information Science, 30 : 3 (july 1979) 220-228.

Ménillet, Dominique (1992). Grilles d'indexation et de préindexation. L'exemple de Pascal. // Documentaliste. 29 : 4-5 (1992) 183-190.

Pinto Molina, María (1991). Análisis documental: fundamentos y procedimientos. Madrid : EUDEMA, 1991.

Landry, B. C. ; Rusch, J. E. (1968). Toward a Theory of Indexing. // Proceedings of the ASIS. 1968.

Slype, Georges van (1991). Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales. Madrid : Fundación Germán Sánchez Ruipérez ; Pirámide, 1991.

UNE 50-121-91. Documentación. Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización. Norma equivalente a la norma ISO 5963:1985. Madrid : AENOR, 1991.

Vickery, Brian C. (1975). Classification and indexing in science. Londres : Butterworths Scientific Publications, 1975.