

La rebelión de las formas: los aspectos formales del contenido como elemento de recuperación

Constança Espelt y Jesús Gascón

Universitat de Barcelona. Escola Universitària de Biblioteconomia i Documentació

Resumen

Las demandas de información de los usuarios y sus criterios para aceptar los resultados de la búsqueda no se basan únicamente en la materia tratada por un documento. El "cómo" se presenta esta materia es tanto o más decisivo a la hora de escoger. No obstante, los sistemas de recuperación de la información siguen basándose en un análisis de contenido centrado en los aspectos temáticos, dejando de lado los formales. El resultado es que la pertinencia de los resultados no suele ser la satisfactoria. Los autores presentan una tipología de elementos formales que se podrían tener en cuenta en un análisis de contenido más completo, ordenados según su grado de evidencia. Exponen algunos procedimientos habituales de indización que contemplan algunos de estos aspectos formales: las subdivisiones en lenguajes precoordinados, algunos campos de los registros de bases de datos bibliográficas, algunos buscadores de Internet que incorporan clasificaciones o los metadatos y formatos de codificación. Se valora la introducción de los elementos formales en la indización y la manera en que podría formalizarse: interfaces que permitan el filtrado acotando por estos atributos, ordenación de los resultados en base a ellos, etc. Finalmente, se exponen algunos casos que ejemplifican esta línea (Danmarks bibliotesskole, SUMS, bases de datos de imágenes, algunos catálogos, etc.).

Palabras Claves:

Recuperación de la información ; Análisis de contenido ; Indización de atributos ; Forma bibliográfica y aspectos formales ; Pertinencia de los resultados de la búsqueda bibliográfica ; Limitación de los resultados de la búsqueda ; Buscadores ; Metadatos ; Lenguajes precoordinados.

Abstract

Bibliographic reference requirements of users and their criteria for accepting search results are based not only on the subject matter dealt by the item. "How" is it presented is than important as the matter when users choose among different items. However, information retrieval systems are based on subject analysis focused on subject aspects, not formal ones. The consequence is that relevance of the answer is not satisfactory. The paper outlines a typology of formal elements to be considered in a more complete subject analysis, filed by their evidence level. Some usual procedures for indexing several of these formal aspects are described: precoordinate languages subdivisions, some fields in bibliographic databases, classifications in Internet search engines and metadata or codified formats. Some features of the introduction of formal elements in the indexing process are discussed, as well as its formulation: interfaces using these attributes as filters, answers filing based on attributes, etc. Finally, some cases of systems working in this trend are presented: Danmarks biblioteksskole catalog, SUMS, image databases, some catalogs, etc.

Keywords:

Information retrieval ; Subject analysis ; Attribute indexing ; Bibliographic form and formal aspects ; Bibliographic search answers relevance ; Search answers filtering ; Search engines classifications ; Metadata ; Precoordinate languages.

Introducción

La distinción entre materia y forma de los documentos es uno de los puntos fundamentales del tratamiento y de la recuperación de las materias. En todo tipo de centros y colecciones documentales, ya sean especializadas o enciclopédicas, destacan una serie de elementos formales que son de gran utilidad en funciones tan diversas como la distribución de los materiales en el espacio físico del centro o la rápida resolución de una consulta específica de información. Sin embargo la mayor parte de la atención se ha dedicado a la recuperación de los aspectos temáticos y relativamente poca a la de los aspectos formales.

Por lo que respecta a la fase del tratamiento, la norma ISO de métodos para el análisis de documentos [1], destinada al uso de sistemas de indización, constituye un claro ejemplo de esta falta de atención. Se describen los procedimientos para analizar temáticamente un documento, precisando la independencia del análisis respecto al sistema de indización a aplicar, pero no se hace mención específica de los elementos formales. Algunos manuales, como el de Brown [2] o el de Piggott [3], al tratar de esta primera etapa de la indización recomiendan la identificación de los elementos formales. Langridge [4] profundiza en esta línea y dedica una monografía al análisis de contenido, entendido como una fase inicial independiente de la representación de dicho contenido. En ella establece una primera distinción entre qué es un documento y de qué trata (*what is it - what is it about*), para añadir más adelante la forma bibliográfica externa, el nivel intelectual, la metodología aplicada, el punto de vista que adopta el autor y la audiencia a la que va dirigido un documento.

Paralelamente en el análisis de las demandas de los usuarios se ha reconocido también la importancia del papel que juega la forma en la definición de la demanda [5]. Las preguntas del usuario son con frecuencia muy vagas, pero no por ello debemos olvidar que él tiene una orientación sobre lo que quiere, en términos de calidad y cantidad, que en definitiva condicionará su posterior satisfacción. El objetivo de la entrevista de información es establecer qué tipo de información necesita el usuario. Además de acotar debidamente el tema de la consulta, esto implica determinar cuanta información se requiere, de qué nivel, su posible dependencia del contexto geográfico. Por lo tanto la selección de los documentos que responden adecuadamente a una demanda deberá tener en cuenta éstos y otros aspectos como la objetividad o el punto de vista que adopta el autor, el enfoque que recibe el contenido de la obra, la audiencia a la que va dirigida. En el contexto de búsquedas más específicas, Fidel [6] recoge la misma recomendación al defender una indización centrada en el usuario (*request-oriented indexing*) frente a la indización centrada en el documento (*document-oriented indexing*).

Sin embargo la extensa investigación en esta parcela parece eludir el núcleo de la cuestión. Se da un hecho similar a lo que ocurre con el análisis de contenido: una vez se ha puesto de manifiesto que se trata de una etapa esencial que determina la calidad del tratamiento del documento, se abandona su estudio para centrarse en los lenguajes de indización y clasificación y en la evaluación de los resultados de su aplicación. Así, mientras la bibliografía sobre la especificidad, las relaciones entre conceptos, los conceptos compuestos, y la adecuación de los lenguajes para plasmar con fidelidad la temática de los documentos es abundante, los estudios centrados en la forma han sido hasta hace poco más bien escasos.

Los avances tecnológicos actuales no han hecho grandes aportaciones en el tratamiento y recuperación de los aspectos formales. El optimismo generado por Internet parece tener únicamente en cuenta el principio de la disponibilidad de la información. Llevamos ya casi un siglo constatando el crecimiento exponencial de la documentación científico-técnica, casi una década inmersos en la sociedad de la información, y sin embargo se sigue avanzando más en la capacidad de almacenar cantidad y variedad de documentos que en la calidad, ya sea de la recuperación o de los documentos recuperados. Si antes teníamos dificultades para seleccionar lo que buscábamos en lo que vendría a ser el colmado de la esquina, ahora nos encontramos perdidos en la inmensidad de una gran superficie comercial, donde encontramos de todo, y muy especialmente basura y productos inútiles.

Elementos formales a tener en cuenta

Antes de pasar a analizar hasta que punto las técnicas de representación del contenido de los documentos contemplan la inclusión de aspectos formales, es imprescindible concretar a qué elementos hace referencia esta denominación genérica. Algunos de ellos se reflejan en el análisis de contenido habitual en catálogos y bases de datos: la materia, la forma bibliográfica externa, la forma interna y la disciplina (no siempre) y las concreciones de la materia en un ámbito geográfico o cronológico. Otros no se identifican en el análisis porque, sencillamente, nos resulta imposible presuponerlos en la globalidad de los usuarios: serán aspectos importantes para la adecuación del documento a la demanda, pero dependerán del bagaje particular de cada usuario. No podremos, pues, incluirlos en el análisis documental; como mucho, podremos tenerlos en cuenta durante el intercambio de información que tendremos con el usuario cuando éste plantea qué es lo que quiere. Así, aspectos como la novedad de los contenidos del documento o el conocimiento previo de la materia serán diferentes para cada usuario, como también lo serán su respuesta emocional a elementos como la personalidad del autor, sus preferencias y gustos, el conocimiento de otros documentos relacionados con los encontrados, etc. Es un terreno en donde el bibliotecario no puede adentrarse hasta conocer al usuario. Por lo tanto, aunque autores como Barry [7] los consideren entre los criterios que influyen en la adecuación de la respuesta a la demanda, en este trabajo no los tendremos en cuenta.

Quedan aún aquellas características inherentes al documento que, sin quedar recogidas en los análisis de contenido habituales, pueden ayudar, en gran manera a delimitar el resultado de la búsqueda: la lengua utilizada, la claridad de exposición, la actualidad del tratamiento del tema, el lugar donde trabaja el autor, la audiencia a la que se dirige, etc. Algunos de estos aspectos pueden encontrarse en la descripción resultante del análisis formal del documento. Otros afectan más a su contenido, pero no tanto a la entidad de la materia o del documento (qué materia es, qué es el documento, cuál su disciplina, etc.) sino a sus atributos (cómo es esta materia, cómo se trata, cómo se expresa). Se trata, por tanto, tal como propone Green [8], de añadir a la indización temática (*subject indexing*) una indización de atributos (*attribute indexing*).

Aunque una gran parte de los elementos están muy interrelacionados, podríamos clasificar estos aspectos en diferentes grupos, en función de la dificultad o inmediatez que implica su detección y formulación, resultando una secuencia que parte de los más evidentes para llegar a los más encubiertos (ver el cuadro siguiente).

<i>Elementos detectables en una primera observación del documento</i>	<i>Elementos condicionados por la estructuración interna del contenido</i>	<i>Elementos condicionados por el autor</i>	<i>Elementos obtenidos a partir de la comparación entre documentos</i>
Forma externa	Forma interna	Punto de vista	Calidad y validez de la información
Lengua	Disciplina	Escuela, discípulos	Carácter único de la información
Actualidad	Género	Afiliación	Carácter polémico o consensuado del tratamiento del tema
Legibilidad física	Nivel intelectual Audiencia Alcance o nivel de detalle Claridad de la exposición Tangibilidad Metodología aplicada al objeto de estudio Espacio y tiempo Contexto social	Intencionalidad Resultados o conclusiones positivos o negativos	

Elementos formales que se pueden considerar en el análisis del documento

Algunos serán "evidentes" en una primera observación del documento: la materia, la forma externa en que se presenta, el lenguaje (ya referido a información textual como musical, gráfica, simbólica, etc.), las características físicas que hacen que la información del documento sea más o menos fácil de asimilar (desde la legibilidad de la tipografía a la inclusión de elementos icónicos, ventanas, etc. que faciliten el uso de un programa informático) o la actualidad de la información ofrecida, de importancia relativa en función del área de conocimiento tratada (es un dato, éste, que puede quedar reflejado en la fecha de publicación del documento o, en caso de que ésta no coincidiera con la fecha de creación de la obra, en una nota, como parte del análisis formal).

En un segundo nivel encontramos elementos que ya requieren un examen más detallado del contenido del documento en tanto que están relacionadas con la estructuración interna de su contenido y con aspectos menos evidentes, en general. Aunque en algunos casos sean fácilmente identificables, podríamos incluir aquí la forma interna y la disciplina dentro de la cuál se trata la materia, o el género a que pertenece (en obras de creación artística y literaria, se entiende).

La audiencia sólo se refleja en la indización de casos muy concretos (algunos libros infantiles, por ejemplo). Su inclusión en el análisis documental agilizaría la recuperación de documentos pertinentes. Se refiere al tipo de público para el que ha sido pensado el documento (aficionados, legos en la materia, colectivos sociales o profesionales, por ejemplo), no tanto a la experiencia ni a los conocimientos previos que requiere su asimilación (nivel intelectual). El alcance de la obra correspondería a la cantidad y al nivel de detalle de la información que proporciona, a la cobertura que da a más o menos aspectos de la materia tratada. Ligada a estos aspectos podríamos situar la capacidad del autor de hacerse inteligible, ya sea por su claridad de exposición o por el uso de una terminología más o menos consensuada.

Otros aspectos interesantes, presentes en muchas demandas, concretan algún elemento de la materia o de la forma en que se ha tratado: la tangibilidad (el poder saber previamente si el documento referenciado contiene datos concretos y constatados o si especula sobre la materia en un plano teórico), los métodos y procedimientos presentes en el trabajo que se han aplicado al objeto estudiado (es frecuente encontrarse con preguntas como "He de hacer un estudio sociológico utilizando tal método; ¿me podría proporcionar otros estudios, sobre lo que sea, que también lo utilicen?" o lo mismo en trabajos científico-técnicos).

Algunos de ellos, como la concreción en el espacio y el tiempo han sido reconocidas tradicionalmente como categorías esenciales en el análisis de contenido de algunas áreas de conocimiento. Además del ámbito geográfico concreto también podría interesar, en algunos casos, la especificación del ambiente o entorno en que se sitúa la obra (por ejemplo, el mar, la montaña, etc., de forma genérica).

El contexto social podría relacionarse con aspectos como el entorno ya comentado o el punto de vista del autor. Un documento que trate la misma materia, en el mismo ámbito geográfico y cronológico, puede variar muchísimo dependiendo del contexto social (el grupo social estudiado, la ideología política, religiosa, etc.) en que se sitúa y que condiciona el resultado final: un estudio sobre la industrialización de un lugar concreto cambiará si el objeto de estudio es la burguesía industrial o si lo son los obreros. Los aspectos contextuales pueden aumentar notablemente la exhaustividad de la indización y, no obstante, no han sido suficientemente explotados, con excepciones como las bases de datos de imágenes, ya sean iconográficas (de obras de arte) ya de fotografías o vídeos con objetivos periodísticos.

El tercer nivel corresponde a características que vienen dadas por la personalidad del autor. Mientras que algunas pueden ser más o menos fáciles de determinar, como la tendencia o escuela de pensamiento en que se inscribe o el lugar en que trabaja (ya sea un organismo concreto, ya un sector laboral genérico), otras pueden encontrarse muy encubiertas y requerirán un examen crítico y casi entre líneas de la obra en cuestión. Por su propia naturaleza, son aspectos que, aunque puedan expresarse abiertamente, también pueden ser subliminales y responden a una actitud personal que el mismo autor quiere mantener velada. El punto de vista, la intencionalidad (que no tiene por qué coincidir con el efecto causado: en una novela o una película, el autor quizás ha querido transmitir un estado de ánimo concreto y, en cambio, al leerla o verla, dependiendo de muchos factores, la interpretamos de otra forma) o el mensaje que el autor quiera transmitir. Incluimos aquí la constatación de cómo son las conclusiones o resultados a que llega el autor en lo que se refiere al tema tratado, en tanto que será el resultado de la actitud del autor ante la cuestión. No se trata tanto de valorarlas sino de señalar si son negativas o positivas: pueden responder a

demandas del tipo "Necesito estudios sobre la influencia del tabaco en la morbilidad femenina en que se demuestre que efectivamente están relacionados", por ejemplo.

Finalmente, en un último nivel, encontramos aspectos que en mayor o menor grado dependen de la relación de la información que da el documento con la de otros. El autor puede plantear el trabajo como una polémica o como una posición que consensúe diferentes tendencias en la disciplina tratada; puede ser un trabajo que aporte una información única en su campo o, por el contrario, puede resultar fácil encontrar otras obras sobre el mismo tema. Más delicada es la determinación de la calidad y la validez del trabajo. Si bien para ello nos podríamos apoyar en otras fuentes que refrendaran las tesis que expone, la evaluación sólo tendría credibilidad si partimos de un conocimiento de la materia tratada. Aunque a primera vista, muchas veces, se pueda dilucidar si un trabajo tiene consistencia o no (por la metodología, las conclusiones, etc.), la familiaridad con la materia facilitaría mucho las cosas. Por otro lado, la validez está sujeta a los vaivenes del tiempo: una investigación válida hoy puede no serlo mañana, a la luz de un nuevo descubrimiento, independientemente de que tenga una calidad indiscutible. O a la inversa.

Representación disponible y recuperable

Los procedimientos habituales de representación del contenido en catálogos y bases de datos han tenido en cuenta, a veces, algunos de estos elementos. Los lenguajes de indización precoordinados, como los basados en los LCSH, por ejemplo, incluyen, además de las subdivisiones geográficas y cronológicas con las que se cubren los aspectos de espacio y tiempo, algunas subdivisiones que responden a formas externas ("-- Bibliography", "--Dictionaries"), internas ("--History" o "--Criticism and interpretation"), audiencia ("-- Children books", "-- Popular works"). En contados casos queda reflejado el género, como en las obras literarias que tratan un evento o personaje en particular, mediante la subdivisión correspondiente, o en algunos casos de antologías de obras literarias o en las obras musicales, en que se utiliza un encabezamiento referido al género. Los aspectos formales que cubren son, por tanto, unos pocos y sólo en determinadas ocasiones.

Siguiendo las mismas directrices, las propuestas del OCLC/LC Fiction Project [9] para la indización sistemática de algunos atributos (género, materia tratada, personajes y ambientes) de las obras de creación artística y literaria han originado, en el caso del género, listas de encabezamientos para su control. Para los otros atributos se aplican subdivisiones ("-- Fiction", "-- Drama", etc.) bajo los encabezamientos de persona, lugar, materia: "Peasants -- Fiction" o "Divorce -- Tales". Su aplicación, sin embargo, está lejos de haberse normalizado.

Los valores y campos de los formatos MARC incluyen códigos para designar algunas formas externas e internas, la audiencia, la lengua (incluso detallando la lengua original, si se trata de una traducción), el ámbito geográfico y, muy vagamente, el género. Pensados para la recuperación de la información del registro, no se han aprovechado lo suficiente. No tendría que ser excesivamente complicado elaborar interfaces que facilitaran al usuario la recuperación de esos códigos mediante términos del lenguaje natural y, sin embargo, pocos sistemas automatizados lo permiten.

Lo mismo ha pasado con la recuperación de lenguajes de clasificación como la CDU. Pocos programas son capaces de gestionarlos debidamente.

De hecho, en la mayoría de los catálogos se puede constatar que los procesos de reconversión no han ido acompañados de una reflexión suficientemente profunda sobre los cambios que introducían en la recuperación de la información. De aquí que una vez implantada la aplicación informática se haya seguido trabajando con los mismos criterios que con el catálogo en fichas, desaprovechando las nuevas posibilidades de acceso que se habían conseguido. Quizás, en estos casos, hubiera valido la pena plantearse, o como mínimo considerar sus costes y sus beneficios, no ya una "reindización retrospectiva" sino una adaptación del sistema de indización al entorno automatizado, aplicable a los registros introducidos a partir del momento de la reconversión. O, como mínimo, partiendo de la posibilidad de la búsqueda por palabras clave, se podrían haber formalizado algunos elementos de la descripción bibliográfica, sobre todo en la zona de notas, que hubieran podido servir para suplir algunas de las carencias de los lenguajes de indización utilizados.

En el ámbito de las bases de datos bibliográficas, la capacidad de recuperación de elementos formales varía mucho de una a otra. En algunas, como Medline, la indización por materia se complementa con etiquetas que permiten acotar la búsqueda, por ejemplo, a algunos tipos de documento, metodología seguida o de estudio. En otras, los resúmenes detallan algunos aspectos formales que no se reflejan en los lenguajes controlados utilizados.

Algunos buscadores de Internet ofrecen alternativas a la búsqueda por palabra clave, en forma de clasificaciones por grandes clases temáticas: Yahoo, Infoseek, Olé, etc. En algunas de ellas no se tiene en cuenta únicamente las materias, sino aspectos como la audiencia (las clases de "Ocio" o de "Webs en catalán, vasco, gallego" de Olé, por ejemplo). La importancia que se le da al ámbito geográfico también es digna de ser remarcada. No se trata de clases excluyentes, sino que el mismo recurso puede incluirse en categorías diferentes. Por otra parte, son clasificaciones altamente dinámicas que van cambiando a medida que se hace necesario.

La introducción de metadatos en los recursos electrónicos podría ser la solución más viable para filtrar los resultados de las búsquedas en Internet [10]. La variedad de datos (metadatos) que se han utilizado en diferentes proyectos y formatos es suficientemente grande como para poder buscar información a partir, casi, de cualquier aspecto tanto temático como formal (nivel intelectual, etc.). Precisamente esta diversidad de formatos y la falta de uniformidad resultante es, quizás, el gran pero de este recurso de recuperación de información.

Incluir aspectos formales en los sistemas de recuperación de la información: ¿por qué no?

El objetivo de la ampliación de los elementos formales en la representación de los documentos es ayudar a mejorar la satisfacción del usuario ante los resultados. Dicha optimización puede provenir de dos vías: facilitar la recuperación de un mayor número de documentos o restringir los resultados de acuerdo con su adecuación a la demanda. Aunque tradicionalmente la eficacia de la búsqueda se haya asociado al clásico equilibrio entre las medidas de índice de recuperación e índice de precisión, en la actualidad la ampliación y variedad de los contextos de búsqueda hacen inclinar claramente la balanza hacia el índice de precisión.

La pertinencia de los documentos recuperados será pues, el primer paso hacia la satisfacción del usuario ante los resultados. Decimos un primer paso porque hablar de pertinencia cuando los resultados de la búsqueda son miles de documentos no tiene ningún sentido, la satisfacción de las necesidades del usuario requiere un número abordable de documentos, aún en el hipotético caso de que todos los documentos recuperados fueran temáticamente pertinentes, se impondría una selección en la cuál nos pueden ser de gran utilidad la valoración de los aspectos formales. Si bien la representación de algunos de los elementos conlleva una ampliación de la recuperación (disciplina, forma interna), en su gran mayoría se caracterizan por su poder de actuar como filtro dentro de un conjunto de documentos pertinentes desde el punto de vista temático. Así lo manifiesta R. Green [8] al asignar a la indización de los atributos la función de aumentar la precisión, frente a la de aumentar la recuperación correspondiente principalmente a la indización temática.

Como consecuencia de este potencial los elementos formales deberían adquirir un mayor protagonismo en las fases inicial y final de la búsqueda. Los sistemas de categorías, ya sea como instrumento para ayudar al usuario a definir los límites de su demanda o bien para situar la búsqueda en el contexto más acorde con sus exigencias, deberían ampliar la tendencia actual a presentar estructuras alternativas. López-Huertas [11] propone categorías que respondan a aspectos subjetivos, refiriéndose concretamente a la intencionalidad y al propósito de la búsqueda. La ampliación puede ser más generalizada para incluir también aspectos de metodología, tangibilidad, alcance o punto de vista.

El filtro podría aplicarse igualmente en la fase final de presentación de los resultados. En la mayoría de las interfaces la presentación y el tipo de ordenación de los resultados parecen un detalle sin importancia, sin embargo la visualización condiciona totalmente la comprensión, y por lo tanto la valoración de la pertinencia. Los criterios de ordenación del elevado número de documentos obtenidos mediante las técnicas de recuperación de coincidencia parcial habituales no pueden limitarse a los clásicos autor, título, fecha de aparición, ni tampoco a una supuesta pertinencia, en base a cálculos probabilísticos de frecuencia de aparición de los términos de

búsqueda. La facilidad de obtención del documento, la legibilidad física, la claridad de exposición, la validez de las fuentes son aspectos que pueden ser especialmente útiles en el momento de la visualización. Cuando las búsquedas producen resultados superiores a cien títulos, por mencionar una cifra a partir de la cuál sería práctica nuestra propuesta, la disponibilidad de varias presentaciones de los resultados sin tener que repetir la búsqueda favorecería la selección de los títulos más pertinentes, y por lo tanto contribuiría a una mayor satisfacción del usuario.

Una interficie donde además de la posibilidad de visualización de la lista de resultados se ofreciera un menú mediante el cual se pudieran filtrar teniendo en cuenta aspectos formales como los comentados no parece que sea técnicamente difícil: se trata de ligar los códigos ya existentes (de metadatos, de formato bibliográfico, etc.) a una expresión inteligible por el usuario y que, al activarla, limite la visualización de los resultados únicamente a aquellos que cumplan el argumento requerido.

En el terreno de la aplicación práctica puede objetarse la dificultad de la identificación de los elementos formales así como de su representación. La subjetividad de la valoración de algunos de ellos, la falta de durabilidad de factores como la validez, la actualidad o el prestigio de las fuentes, el elevado coste de la identificación de todos estos elementos, por la inversión de tiempo que supone, la falta de listas de descriptores que permitan normalizar su expresión y asegurar un nivel de coherencia en su asignación, son factores a tener en cuenta, pero en todo caso no constituyen una barrera insalvable.

En primer lugar, la necesidad de complementar la recuperación temática mediante atributos formales parece suficientemente avalada por el marco teórico que aporta el paradigma cognitivo y las conclusiones de estudios de satisfacción de usuarios. Por otra parte, la existencia de experiencias en campos temáticos acotados puede ilustrar su funcionamiento y servir de base para una aplicación más general.

En indización de imágenes y de obras de creación literaria las iniciativas han sido múltiples, entre ellas destacamos el sistema de clasificación multidimensional desarrollado por Pejtersen [12]. Este sistema, aplicado en los catálogos de la Danmarks biblioteksskole y del Nordiske boghus projekt, incorpora junto al tema, la situación en el contexto temporal y social, la experiencia emocional que el autor quiere transmitir, y la accesibilidad en base a las características físicas del documento y al grado de dificultad del lenguaje utilizado. Recientemente dicho sistema y el mencionado OCLC/LC Fiction Project han servido de base para una propuesta de indización en bibliotecas especializadas de Buenos Aires [13]. Situándonos en áreas propiamente especializadas y multidisciplinarias, Beghtol [14] se refiere a la conveniencia de aplicar técnicas de clasificación a los elementos formales, y menciona un sistema diseñado para indizar de forma exhaustiva los métodos aplicados en la investigación del crimen y la violencia, ámbito que incluye disciplinas tan diversas como derecho, criminología, sociología, psicología, medicina y biología. Podemos añadir una experiencia similar más cercana, como la del catálogo de la biblioteca de la Fundació CIDOB, en Barcelona, especializada en relaciones internacionales, cooperación y desarrollo. En él, sobre las demandas de sus usuarios, se han ido introduciendo descriptores referidos a la forma del documento ("Cronología", "Informe", etc.), el nivel intelectual ("Material básico" de nivel divulgativo o de iniciación), la audiencia ("Material didáctico", para educadores), la metodología de trabajo ("Estudio comparado") o la tangibilidad ("Estadísticas", "Datos básicos"), entre otros. La búsqueda combinada de estos descriptores formales y los de materia da unos resultados de un alto nivel de pertinencia.

Por último, a nivel global, merece destacarse el prototipo System for Universal Media Searching [15] que, como indica el subtítulo Framework for a Meta-Data Digital Reference Room, tiene el objetivo de ofrecer los recursos de referencia y organización de materiales de las principales bibliotecas. Proporciona diez instrumentos básicos de navegación en cuyo desarrollo se ha recurrido con frecuencia al uso de los aspectos formales. La inclusión de "How?" y "Why?" en las preguntas que ayudan al usuario a definir la búsqueda, un sistema de categorías basado exclusivamente en la intención de la búsqueda, la posibilidad de definir las preferencias de acceso del usuario ajustando el sistema al nivel educativo, la identificación de dos grupos de niveles de conocimiento para distinguir entre objetos de conocimiento (clasificaciones, bibliografías, enciclopedias...) e interpretaciones de este conocimiento, son tan sólo algunas opciones indicativas del alcance de este prototipo.

Conclusiones

Mejorar la recuperación por materias es una asignatura pendiente demasiado generalizada, pero no por ello debería dejar de ser uno de los objetivos prioritarios de cualquier sistema de información. Desde el plano de la realidad de las bibliotecas y la industria de la información, la excusa suele ser que se trata de una empresa quijotesca, que requeriría una dedicación excesiva del personal. De hecho, es evidente que la tecnología no constituye el único obstáculo, por no decir que no es óbice para llevar a cabo el intento. La rapidez, la facilidad, la simplicidad y, por qué no, la uniformidad son cualidades importantes en las interfaces, pero no se pueden obviar las limitaciones del resultado que proporcionan. En algunos casos, la optimización equivaldría a sacar un mayor rendimiento al tratamiento que se hace de los documentos: se trata únicamente de que las interfaces de búsqueda pongan al alcance del usuario lo que permanece oculto e inútil, bajo codificaciones diversas. Por lo que respecta al resto, para alcanzar logros significativos es imprescindible explorar nuevas vías sin perder de vista el objetivo: la recuperación de información pertinente a una demanda concreta de un usuario concreto, con la "circunstancia" que implica. Con un nuevo enfoque, no parece utópico que las actuales técnicas de indexación automática permitan recuperar por atributos tales como el nivel intelectual o la tangibilidad. En todo caso, una indexación en que además de las materias se tuviesen en cuenta sus atributos aumentaría la calidad de los sistemas de información y facilitaría la recuperación, incrementando el nivel de satisfacción del usuario.

Bibliografía

- [1] ISO. *Methods for examining documents, determining their subjects, and selecting indexing terms*. ISO 5963:1985.
- [2] Publicada en español como: AENOR. *Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indexación*. UNE 50-121-91.
- [3] Brown, A. G. *An introduction to subject indexing*. 2nd ed. London: Clive Bingley, 1982.
- [4] Piggott, Mary. *A topography of cataloguing showing the most important landmarks, communications and perilous places*. London: Library Association, 1988.
- [5] Langridge, D.W. *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989.
- [6] Katz, William A. *Introduction to reference work*. 7th ed. New York: McGraw-Hill, 1997.
- [7] Fidel, Raya. User-centered indexing. *Journal of the American Society of Information Science*, 45 (8), 1994, 572-576.
- [8] Barry, Carol L. User-defined relevance criteria: An exploratory study. *Journal of the American Society of Information Science*, 45 (3), 1994, 149-159.
- [9] Green, Rebecca. The role of relational structures in indexing for the humanities. *Knowledge organization*, 24 (2), 1997, 72-83.
- [10] Down, N. Subject access to individual works of fiction: participating in the OCLC/LC Fiction Project. *Cataloging and classification quarterly*, 20 (2), 1995, 61-69.
- [11] Dempsey, Lorcan; Heery, Rachel. Metadata: a current view of practice and issues. *Journal of documentation*, 54 (2), 1998, 145-171.
- [12] López-Huertas Pérez, M^a José. La representación del usuario en la recuperación de la información. *Jornadas Españolas de Documentación VI*, 521-526.
- [13] Pejtersen, Annelise Mark; Austin, Jutta. Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria. *Journal of Documentation*, 39 (4), 1983, 230-246; 40 (1), 1984, 25-35.
- [14] Pené, Mónica; Martínez, Ana María. Descripción de contenido de las obras de ficción en los catálogos en línea. *Revista argentina de bibliotecología* 1, 1998, 7-18.
- [15] Beghtol, Clare. Knowledge domains: Multidisciplinary and bibliographic classification systems. *Knowledge organization* 25, (1/2), 1998, p. 1-12.
- [16] Veltman, Kim H. *SUMS (System for Universal Media Searching) demo homepage* [en línea]. <http://www.sumscorp.com/index1.html> [consultado el 24 de febrero de 1999].