

ESTUDIO SOBRE LA EVALUACIÓN EN ORGANIZACIÓN Y REPRESENTACIÓN DEL CONOCIMIENTO

- Autores:** Antonio García Jiménez
Universidad Europea de Madrid
antonio.garcia@fcp.cin.uem.es
- Resumen:** Se presentan diversos estudios sobre la evaluación en la Organización y Representación del Conocimiento, en especial en la indización y en los tesauros. Este análisis se plantea tomando en consideración la perspectiva tradicional y otras perspectivas que pretenden objetivar y mejorar los procesos y métodos de trabajo relacionados con este dominio. Finalmente se revisa su influencia en la recuperación de información.
- Palabras:** Organización del Conocimiento; indización; tesauro; evaluación, recuperación de información.
- Abstract:** Different studies are presented about evaluation in Knowledge Organization, particularly in indexation and in thesauri. This analysis is achieved both from a traditional perspective and from new perspectives that try to make objective and to improve processes and methods related to this domain. Finally, the influence on information retrieval is reviewed.
- Keywords:** Knowledge Organization; indexation; thesaurus; evaluation; information retrieval.

1. Introducción

Al intentar precisar los diferentes modelos y métodos de evaluación tanto en el Análisis Documental de contenido (en concreto en la indización) como en los lenguajes documentales (en concreto en los tesauros), hay que tener presente la casi necesaria superación de las normas como único pilar de este tipo de operaciones e instrumentos. Se parte de la trascendencia metodológica y teórica que tiene la interdisciplinariedad en donde se sitúa este dominio, con vistas a entender y aprovechar todos los recursos que las disciplinas colaterales pueden aportar, así como de la creciente importancia de nuevos planteamientos en la forma de entender este campo y de la generalización de otros instrumentos de ayuda a la recuperación de información.

2. Evaluación y validación en el Análisis Documental

Si bien no se puede hablar en sentido estricto de una metodología de evaluación de la indización, sí es cierto que a lo largo de la tradición documentológica se han planteado diversas medidas sobre su efectividad. Así, Chaumier [1] recoge las siguientes medidas: la exhaustividad o profundidad de la indización, es decir, el término de descriptores afectados por el documento y la especificidad o pertinencia, y la coherencia, ya sea externa (la razón entre la unión del conjunto de conceptos utilizados por un analista A para describir un

documento y el conjunto propuesto por un analista B y su intersección) o interna (los resultados de la indización en dos momentos diferentes) y, finalmente, el tiempo de indización empleado.

En la misma línea, y mientras Lancaster [2] reconoce la consistencia y la calidad como medidas, van Slype [3] subraya, además de la profundidad y la coherencia, el tiempo de indización, la exhaustividad (que mide la calidad de la elección de los conceptos que son significativos para los usuarios), y la especificidad (que mide el grado en que los descriptores elegidos se corresponden con los conceptos vertidos en los documentos), que puede ser vertical, o lo que es lo mismo, cualquier descriptor se ha de ubicar en el mismo nivel de especificidad que el concepto, o puede ser horizontal (por la que un concepto compuesto debe incluirse en un sistema documental preferiblemente mediante un descriptor precoordinado).

Por su parte, Soergel [4], más allá de las ideas de relevancia, pertinencia y utilidad, también hace referencia, por un lado, a la exhaustividad, como el grado en el que los conceptos que indizan una entidad son cubiertos por los asignados a esa entidad y, por otro, a la especificidad, como el nivel genérico en el que los conceptos asignados a la entidad son expresados. También señala como criterios de calidad la corrección de la indización, en dos vertientes: la de la omisión de un descriptor o la de la introducción de un descriptor que no debiera aparecer. Finalmente, menciona la consistencia de la indización entendida como el acuerdo entre dos indizadores y el acuerdo existente entre dos sesiones de indización llevada a cabo por el mismo indizador.

Para poder completar esta recopilación, también nos debemos situar en un marco de trabajo distinto a los planteamientos que aquí se han recogido. En efecto, se trata de aquellos estudios que tienen como uno de sus objetivos precisar modelos de análisis o lectura documental [5] con vistas a la recuperación de la información, a través de los que se podría objetivar en mayor medida la indización. Así, García Jiménez [6] presenta un método de evaluación, que parte de los principios de evaluación del análisis de contenido, y que se prueba en función del modelo de casos de Cunha [7] (adaptado según los casos de Fillmore y Pottier) como modelo de análisis en el discurso periodístico. El cuadro de trabajo queda como sigue:

- **Validez semántica:** examina la capacidad del método de análisis para captar las ideas expuestas en el documento a través del *análisis logicista* [8], al objeto de percibir la capacidad de un modelo analítico para la detección de los argumentos de los textos. Por su parte, el contraste con el *esquema formal del discurso* en el que se sitúe servirá para observar su capacidad de desestructuración textual.
- **Validez pragmática:** delimita la capacidad que posee un modelo para dar respuesta a las demandas y necesidades de los usuarios, a través de las unidades de búsqueda por significación propias del discurso en el que se encuentre la aplicación.
- **Validez de la construcción:** que hace mención a las garantías que presenta el modelo en su diseño y aplicación gracias al estudio de: a) *Reacciones* (descripción del comportamiento del modelo analítico). Ofrece

ciertos indicadores capaces de dilucidar la pertinencia de un modelo de interrogación como vehículo de lectura documental; b) *Información ausente*: apartado de mayor dificultad de plasmación, en parte por la complejidad inherente a la categorización de la información ignorada, de lo que se deduce su menor peso en la evaluación final.

Por su parte, García Gutiérrez [9] en su intento de confirmar que el discurso periodístico es reducible a una tipología concreta de enunciados documentales con vistas a la formalización y a su reconocimiento artificial, propone una propuesta de evaluación. En resumen, ésta se despliega gracias a una muestra A, seleccionada mediante la aplicación de la técnica de la mitad sobre las unidades operativas elegidas, es decir, las macroproposiciones (una proposición que representa el conjunto de proposiciones de un texto y que es el resultado de la eliminación de algunas proposiciones y de la sustitución de proposiciones específicas por proposiciones supraordenadas) y a la verificación de los resultados al contrastarlos con una muestra B.

Por último, no se debe dejar de lado la posible manipulación procedente de la tarea indizadora. A decir de Chacón [10], y siempre según la naturaleza y las condiciones de la indización humana, la manipulación puede tener su origen en las diferentes formas de apreciación, sin olvidar su conexión con los lenguajes documentales. En definitiva, la manipulación puede ser voluntaria o involuntaria, puede venir de la mano de la falta o no de fidelidad al sentido y a la forma del texto, de las connotaciones implícitas, de los modelos culturales e ideológicos propios de cada sociedad, de la calidad de los descriptores en los lenguajes y de la capacidad de decisión del analista.

3. Evaluación en los lenguajes documentales: los tesauros

Los lenguajes documentales, dada la atención que se les ha prestado desde el ámbito de la práctica profesional, sí han sido objeto de una mayor preocupación teórico-práctica, de lo que se deriva una mayor inquietud por los problemas de evaluación. Así, según van Slype [11], es conveniente realizar el test del tesoro, que tiene por misión añadir descriptores y no-descriptores no insertados, modificar las estructuras jerárquicas y asociativas, así como eliminar la ambigüedad de los descriptores insuficientemente desarrollados. Tampoco olvida el estudio de oportunidad, que consiste, entre otros aspectos, en la definición del mercado, en la determinación de las características cualitativas y cuantitativas, en la localización de las fuentes terminológicas más adecuadas, en la delimitación del coste de operación, en la exposición del calendario de realización, etc. Por ejemplo, destaca su preocupación por el tamaño del tesoro, cuestión que depende del número de documentos del fondo documental, de la profundidad de la indización, del número medio de documentos que se darán como contestación a las consultas, etc. Desde su punto de vista, las características cualitativas se miden gracias a los siguientes indicadores:

- Ratio de precoordinación: razón entre el número de veces que aparecen las palabras no vacías que componen los términos de un descriptor y el número total de descriptores.
- Tasa de equivalencia: el número de no-descriptores/descriptores
- Tasa de enriquecimiento: el número de relaciones jerárquicas y asociativas/número total de descriptores
- Flexibilidad: la proporción de palabras significativas simples usadas para proponer los descriptores compuestos, y que también aparecen en el vocabulario como descriptores simples o no-descriptores.
- Conectividad: mide la razón existente entre el número de descriptores que incluyan como mínimo alguna relación de naturaleza semántica y el total de descriptores.

Lancaster [12] también menciona otras medidas como la idoneidad de introducción, la presentación o no de exposiciones alternativas, la adecuación del uso de las diferentes relaciones, el nivel de ambigüedad existente, la suficiencia en la especificada de los términos. Además, propone otras medidas:

- Ratio de reciprocidad: mide el grado en que los diferentes tipos de relaciones tienen reenvíos recíprocos. No obstante, no la entiende como adecuada, de la misma forma, que califica la razón de flexibilidad como una medida arbitraria.
- Definición: $(b-a)/b$, donde a es el número de descriptores que son ambiguos porque no presenta notas de aplicación, calificadores o relaciones jerárquicas que los definan y b es el número total de descriptores del vocabulario.
- El tamaño de los grupos de términos.

En términos generales, la evaluación de los tesauros [13] se despliega alrededor de la distribución temática de los tesauros, de la calidad informativa de la presentación, de la consistencia sintagmática de los descriptores, de la estructura semántica (agrupaciones de tipo jerárquico y asociativo), así como de la valoración estética. A nuestro entender, estas medidas no se detienen suficientemente en el necesario estudio, si se desea una evaluación completa, de las vías lógicas, léxicas y sociales que subyacen en la construcción de cualquier tesoro, de las que se podrían derivar posibles sesgos en un sistema documental. En realidad, se buscan fórmulas que garanticen que un tesoro determinado es capaz de representar correctamente el contenido de los documentos, de ofrecer un marco de referencia para el análisis y de colaborar con el buen funcionamiento del sistema documental, siempre en relación con el uso.

En consecuencia, se entiende que el método de evaluación debe ofrecer fórmulas explícitas de validación orientadas a la naturalidad, la flexibilización y la superación, cuando sea pertinente, del orden jerárquico. A tal efecto, García Jiménez [14] presenta una alternativa basada en la aplicación de modelos conceptuales como fórmulas de evaluación/construcción de los tesauros, ya que se entiende que pueden ofrecer nuevas perspectivas en las relaciones de los tesauros y en la configuración de los campos conceptuales, de lo que se

deduce la capacidad para la detección de la lógica interna, los defectos y las omisiones en un lenguaje documental. Los modelos estudiados fueron los siguientes:

- Los principios universales de los campos léxicos en los sistemas documentales [15].
- El modelo ampliado de las 5W (con la forma estructurante ampliada siguiente: quién/por quién; qué/cómo; por/para qué/quién; dónde; cuándo).
- El modelo de casos como potencial conjunto de categorías capaz de organizar léxicos en cualquier ámbito del saber.
- Las categorías del análisis de contenido [16] como modelo de testado (materia, tema o asunto; dirección o punto de vista; criterio o valor; autoridad, fuente o prueba; objetivo o meta; medio o método; actores; características personales y rasgos de carácter; origen; destino; forma de la afirmación; intensidad; mecanismo o recurso).
- Modelo de relaciones vectoriales para la construcción de nuevos lenguajes documentales [17] mediante modos estructurantes y vectores: vector modal, procesual, adjetival, actancial e instrumental.
- Los rasgos estructurantes de dos grupos diferentes de categorías facetadas presentadas por Aitchison y Gilchrist [18], basadas bien en la tradición documentológica, bien en campos o disciplinas concretas.

Sin entrar a explicar el modo de aplicación, de este intento se pretende extraer modelos capaces de excluir el plano intuitivo como principio de trabajo, aunque en cierta medida, se sitúen al margen del aspecto normativo. La aplicación se realiza a través de una organización vectorial completada por una perspectiva enunciativa y lingüística en dos niveles: la aplicación de las categorías de cada modelo y el estudio del entramado relacional de cada uno de los descriptores seleccionados como muestra. La evaluación se realiza a partir de los siguientes criterios:

- **Fiabilidad de una categoría única.** Estima el solapamiento entre categorías dentro un mismo modelo. No obstante, queda pendiente dilucidar la importancia de este criterio en la construcción y evaluación de lenguajes documentales.
- **Fiabilidad condicional.** Detecta el valor suficiente de cada una de las categorías empleadas.
- **Validez de la construcción.** Mide la suficiencia de cada modelo en su aplicación. Permite verificar el estado de un tesoro además de sacar a la superficie, en gran medida, las bases subyacentes de construcción. Como aspecto negativo, se detecta la escasa descripción, la multiaplicación y su ineficacia para precisar la pertinencia de los términos.
 - **Contextualidad y no contradicción.** Mide la idoneidad de un modelo para determinar la pertinencia de todos sus elementos.
 - **Reciprocidad.** Precisa la capacidad de un modelo para formar ejes estructurales y permitir la bidireccionalidad de las relaciones.
 - **Organización formal.** Mide la aptitud de un modelo para precisar si un tesoro se presenta de forma ordenada y útil. Se observa que este

apartado depende de criterios de uso y supera la capacidad de organización y decisión que en general incorporan los modelos de relaciones.

- **Homogeneidad.** Delimita la capacidad para la detección de lagunas conceptuales en el tesoro. Su funcionamiento requiere en muchos casos fuentes de conocimiento externas, pero ofrece expectativas positivas en la mayoría de modelos analizados.

Tras este estudio se observó como aspectos negativos la pérdida de cierto nivel de control sobre los núcleos conceptuales, las lagunas de los modelos, la necesidad de fórmulas de representación (de organización y enunciativa), la falta de profundización sobre la base léxica y sobre el aspecto de uso y definitorio, así como la excesiva implicación del factor humano. Como aspectos positivos, siempre en el contexto de un marco tecnológico concreto, destacan la objetivación y explicitación de los procesos, el cumplimiento satisfactorio del objetivo de naturalidad, y la capacidad de los modelos para plantear relaciones entre descriptores que, en muchos casos, quedarían sin resolver.

En último lugar, en la misma línea de estudios que dan importancia a otros modos de entender los lenguajes documentales [19], recogemos la aportación realizada conforme los denominados lenguajes epistemográficos [20]. Estos lenguajes son constructos cognitivos que funcionan como una base de conocimientos con fines documentales. En este contexto, se recomienda la comparación con otros tesauros y otros lenguajes documentales del mismo dominio, el análisis del número de entradas, la observación de la economía de signos junto a la poscoordinación como principios que deben prevalecer, el estudio del nivel de uso de todos los componentes estudiados, de la preponderancia del eje sintagmático frente al nivel eje paradigmático, de la capacidad de aproximación a la lengua natural, del grado en que las asociaciones posibles son cubiertas, de la validez del software empleado, así como el análisis de la base léxica y de la estructura y composición de las categorías. Así pues, se establecen dos tipos de análisis:

1) **Análisis sobre la base léxica:** composición del vocabulario; tipos de descriptores empleados; extensión del vocabulario; precisión de los descriptores; número de no-descriptores; nivel de aclaración de los descriptores y de uso de las notas de alcance; calidad en el planteamiento de las polijerarquías; igualdad en el tratamiento de los descriptores.

2) **Análisis de la estructura y las categorías:** solapamientos e insuficiencias en las facetas (o macrocategorías, en términos de lenguajes epistemográficos), entendidas como el ángulo más genérico y abstracto de denominación y exclusivamente de organización de los campos conceptuales; la coherencia en la adscripción de términos en los diversos campos conceptuales; nivel de cobertura de las macrocategorías; utilidad y funcionalidad de los diferentes componentes estructurales (en especial las relaciones horizontales (TR)); coherencia en la estructura; precisión de las notas aclaratorias para la delimitación de las coberturas de las distintas categorías; facilidad para la

percepción de las estructuras jerárquicas y asociativas; detección de lagunas en las relaciones; normalización de las convenciones empleadas.

5. La indización, los tesauros y la recuperación de información

Otra forma de evaluación, y que en muchas ocasiones ha sido discriminada en el estudio de la indización y de los lenguajes documentales, es su incidencia en la recuperación de información. Según Soergel [21], la exhaustividad de la indización puede producir un aumento en la tasa de *recall* y también en el porcentaje de información correctamente descartada. Mientras, los efectos de una indización específica en una búsqueda concreta dependen de la corrección de la indización puesto que unos descriptores específicos hacen más vulnerable a la búsqueda a los errores de indización. Por su parte, los efectos de una indización específica en una búsqueda amplia dependen de la capacidad del sistema de búsqueda. Por otra parte, la corrección de la indización influye de forma determinante en la recuperación. De hecho, la tasa de *recall* está vinculada al grado en el que la indización es completa. Por último, la consistencia sólo afecta indirectamente a la recuperación y sobre todo en su vinculación con la corrección de la indización.

Para Lancaster [22], la falta de especificidad tiene efectos en la precisión, si bien no necesariamente en la exhaustividad. En realidad, la especificidad del vocabulario es el factor que puede afectar en mayor medida a la precisión de un sistema de recuperación. De hecho, se ha demostrado que los vocabularios con un alto grado de especificidad pueden provocar tanto la disminución de la exhaustividad como un aumento de la precisión en términos de recuperación de información. Y un tesoro con menos términos específicos dará lugar a una alta exhaustividad pero a una escasa precisión. Por último, señala que la presencia de coordinaciones falsas y de relaciones incorrectas provocarán una disminución en la exhaustividad.

A decir de algunos autores [23], los términos relacionados (TR), que son menos comprendidos en la construcción de tesauros y menos empleados en la recuperación de información frente a otros tipos de relaciones, como las jerárquicas, deben asentarse como objeto de estudio en su relación con la recuperación de información, pese a ser evidente el problema del potencial incremento del ruido en la recuperación, sobre todo si no presentan una correcta y pertinente definición.

6. Conclusiones

El estudio de las diferentes operaciones e instrumentos que aquí se han tratado no pueden dejar de lado las implicaciones procedentes de su complejidad y de la dificultad a la hora de alcanzar un grado de uniformidad completamente satisfactorio. A la luz de lo anteriormente expuesto, varias son las ideas dignas de resaltar. En primer lugar, la Organización del Conocimiento requiere investigar los cambios que el contexto hipermedia está produciendo en el Análisis Documental y en los lenguajes documentales. En segundo lugar, se

debe ampliar los métodos reseñados a otras variantes de instrumentos de ayuda a la recuperación de información, así como a los sistemas clasificatorios, con especial atención a su generalización en Internet. Finalmente, se considera como un objeto de estudio relevante la unificación de los diversos planteamientos de evaluación, siempre que no se les haga depender de una perspectiva previa determinada.

Bibliografía

- [1] Chaumier, Jacques. *Análisis y lenguajes documentales. El tratamiento lingüístico de la información documental*. Madrid, Mitre, 1986.
- [2] Lancaster, F. W. *Indexing and Abstracting in Theory and Practice*. The Library Association, 1991.
- [3] [11] Slype, Georges van. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Salamanca, Fundación Germán Sánchez Ruipérez; Madrid, Pirámide, 1991.
- [4] [21] Soergel, Dagobert. *Indexing and Retrieval Performance: The Logical Evidence*, JASIS, 1994, 45(8), p. 589-599.
- [5] Pinto, María y Gálvez, Carmen. *Análisis Documental de contenido. Procesamiento de información*. Madrid, Síntesis, 1996.
- [6] [14] García Jiménez, Antonio. *Metodología de validación del Análisis Documental y de los lenguajes documentales en el discurso periodístico*. Madrid, Universidad Complutense de Madrid, 1995 (Tesis doctoral).
- [7] Cunha, Isabel. *Do mito à análise documentária*. Sao Paulo, Edusp, 1990.
- [8] Gardin, J. C. *Le calcul et la raison: essais sur la formalisation du discours savant*. París, L'Ecole des Hautes Études en Sciences Sociales, 1991.
- [9] García Gutiérrez, Antonio. *Procedimientos de Análisis Documental Automático. Estudio de Caso*. Sevilla, Instituto Andaluz del Patrimonio Histórico, 1996.
- [10] Chacón Gutiérrez, Inmaculada. *La fiabilidad del proceso documental del discurso periodístico como fuente de información: análisis, método y valoración*. Madrid, Universidad Complutense de Madrid, 1994. (Tesis doctoral)
- [12] [22] Lancaster, F. W. *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington, Information Resources Press, 1992.
- [13] Álvaro Bermejo, Concepción; Villagrà Rubio, Angel y Sorli Rojo, Ángela. Desarrollo de lenguajes documentales formalizados en lengua española: II. Evaluación de los tesauros disponibles en lengua española. *Rev. Esp. Doc. Cient.* 1989, vol 12, nº 3, p. 283-305.
- [15] [17] García Gutiérrez, Antonio. *Estructura lingüística de la documentación: teoría y método*. Murcia, Universidad, Secretariado de Publicaciones, 1990.
- [16] López-Aranguren, Eduardo. El análisis de contenido, en García Ferrando, M.; Ibañez J. y Alvira, F.(comp.), *El análisis de la realidad social. Métodos y técnicas de investigación*, 2ª ed. Madrid, Alianza, 1989, p. 383-414.
- [18] Aitchison J. y Gilchrist A. *Thesaurus construction: a practical manual*, 2nd. ed. rep. London, Aslib, 1990.
- [19] López-Huertas, María José. *Potencialidad evolutiva del tesoro: Hacia una base de conocimiento experto*, (Actas del IV Congreso ISKO-España

EOCONSID'99. Granada, 22-24 de abril de 1999), Granada, Isko-Facultad de Biblioteconomía y Documentación, 1999, p. 133-140.

[20] García Gutiérrez, Antonio. *Principios de lenguaje epistemográfico: la representación del conocimiento sobre Patrimonio Histórico Andaluz*. Sevilla, Instituto Andaluz del Patrimonio Histórico, 1998.

[23] Alani, Harith; Jones, Christopher, and Tudhope, Douglas. Associative and Spatial Relationships in Thesaurus-Based Retrieval, *Research and Advanced Technology for Digital Libraries (4th European Conference, ECDL 2000) Proceedings*, (Lectures notes in computer science, vol. 1923), Lisboa, Springer, 2000, p. 45-58.