



# La investigación en indización y resumen en España

• ISIDORO GIL LEIVA

## 1. Introducción

En primer lugar, juzgamos necesario explicar de un modo breve qué es y para qué sirve la indización y el resumen con el fin de contextualizar el presente trabajo. La indización y el resumen son, sin duda, las tareas más complejas, interdisciplinarias y subjetivas de todo el proceso de análisis de los documentos. Son labores complejas porque requieren para su ejecución una buena formación general, una sólida base en Documentación, y en la materia a resumir o indizar. Son interdisciplinarias puesto que se precisan conocimientos en Terminología, Lingüística, y Documentación, entre otras. Y subjetivas porque raramente dos profesionales alcanzan una idéntica indización o resumen para el mismo documento.

En el proceso documental la indización y el resumen tienen por fin el análisis del contenido de los documentos.

### 1.1. Delimitación, etapas y principios de la indización

La indización tiene como principal objetivo el almacenamiento y recuperación de la información. En la gestión documental la indización interviene en dos momentos diferentes, dando lugar a lo que podríamos denominar dos vertientes enfrentadas. La primera es la indización de los documentos que tiene por misión el almacenamiento de los mismos, generalmente en bases de datos, una vez que han sido analizados y representados sus conceptos esenciales a través de términos de indización. Y en segundo lugar, la indización de las preguntas para iniciar la recuperación de los documentos depositados en dichas bases de datos.

El proceso de la indización se lleva a cabo en dos momentos: a) análisis de los documentos y preguntas para la selección de sus conceptos esenciales explícitos e implícitos, b) la conversión de los términos

seleccionados en lenguaje natural a un lenguaje, generalmente controlado, por medio de herramientas como vocabularios controlados, listas de encabezamientos de materia o tesauros. De este modo, consideramos la indización como la llave para el almacenamiento y recuperación de la información.

Los principios que intervienen en la indización son:

- **Exhaustividad:** relativa al número de conceptos que caracterizan el contenido íntegro de un documento y no a la cantidad de términos asignados
- **Especificidad:** precisión a la hora de detectar el tema o asuntos principales que trata un documento
- **Corrección:** relacionada con la omisión de términos correctos o inclusión de términos innecesarios
- **Consistencia:** grado de coincidencia a la hora de presentar los términos de indización entre uno o varios indizadores.

Las bibliotecas analizan los documentos para asignar encabezamientos y subencabezamientos por medio de las listas de encabezamientos de materia. Por estas razones es preciso que los bibliotecarios tengan presente en cada momento los principios mencionados anteriormente, así como otros más específicos sobre los que se asientan estas listas con el fin de obtener una buena indización. Estos principios básicos son:

- **Síntesis:** reducción del contenido de un documento al menor número posible de materias.
- **Lingüístico:** empleo del lenguaje habitual y en el orden natural del idioma, prevaleciendo el sustantivo frente al adjetivo como elemento inicial, aunque hay excepciones.
- **Economía:** asignación del mínimo número de encabezamientos, por lo general de uno a tres para una biblioteca pública.

En cambio, en bibliotecas especializadas, centros de documentación y bases de datos para representar el contenido de los documentos suele trabajarse con descriptores incluidos en un vocabulario controlado o



en un tesoro. Los descriptores son expresiones menos complejas que los encabezamientos de materias, no condicionan tanto el almacenamiento y recuperación de la información y no hay que asignar un número más o menos preestablecido.

## 1.2. Delimitación, principios y tipología del resumen

Al comienzo señalamos que el resumen es la otra operación que constituye el análisis del contenido documental. El principal objetivo de un resumen es reproducir el contenido esencial de un documento de forma condensada para que un usuario pueda decidir sobre la conveniencia de consultar el documento completo. Los principios caracterizadores de los resúmenes son los siguientes, según Moreiro González [1993]:

- **la entropía:** busca el mayor grado de comunicación con el menor número de signos
- **la pertinencia:** adecuada transferencia informativa del documento original al resumen o incluso para la localización de los usuarios adecuados a los que va dirigido dicho producto
- **la coherencia:** la correcta distribución de la información a lo largo del resumen
- **la corrección lingüística:** búsqueda de la corrección ortográfica, la gramaticalidad y univocidad en la estructura y significación de las oraciones
- **el estilo:** utilización de mensajes concisos, exactos y directos para conseguir una comunicación fácil, clara y rápida
- **el distanciamiento tolerable:** que transcurra el menor tiempo posible entre la publicación del original y la llegada del resumen al usuario.

En cuanto a su tipología, según este mismo autor, encontramos resúmenes por la mediación informativa que realizan, por la descodificación del texto integro, de acuerdo con el mediador, por el número de fuentes analizadas, por el modo de difusión y por la calidad de las fuentes.

## 2. Tendencias en la investigación sobre indización y resumen en España

En primer lugar, diremos que, salvo alguna excepción, la indización y el resumen han sido poco estudiados por los investigadores y profesionales españoles en Biblioteconomía y Documentación. En cambio, a principios de los noventa proliferaron trabajos que trataban aspectos del análisis del contenido documental desde la generalidad. Para comprobar el estado de la investigación en indización y resumen nada mejor que revisar la bibliografía publicada en España. Las fuentes consultadas fueron publicaciones periódicas, monografías, libros de compilación y tesis doctorales sobre Biblioteconomía y Documentación.

Las investigaciones en indización han abundado más que sobre el resumen. Pero esta circunstancia se ha producido tanto a nivel nacional como internacional. En nuestro país mostraremos a continuación algunos datos que confirman esta aseveración. Y para avalar esta idea con respecto al exterior, simplemente hay

que revisar publicaciones periódicas norteamericanas, como *Journal of Documentation*, *Journal of the American Society for Information Science* o *Information Processing & Management*; publicaciones francesas como *Documentaliste-Sciences de l'Information* o brasileñas como *Ciencia da Informação*.

## 2.1. La investigación en resumen

Comencemos apuntando que Sánchez Casabón y García Marco [1995] analizaron la producción bibliográfica española sobre análisis de contenido documental y lenguajes documentales entre los años 1982 y 1994 (en concreto setenta y cinco artículos). Para ello tomaron los trabajos publicados en destacadas revistas sobre Biblioteconomía y Documentación de nuestro país (1). En este trabajo concluyeron en cuanto al reparto temático que más del 50% de los asuntos elegidos por los autores analizados trata de las clasificaciones, los tesauros y los resúmenes, temas abordados en la mayoría de las ocasiones con una amplia visión. El resumen acapara el 14,5% de los trabajos analizados. En cambio, el de la indización está incluido en un grupo de temas como la catalogación, autoridades o lingüística computacional que, entre todos, acaparan tan sólo el 9,8% de la bibliografía estudiada.

La investigación sobre el resumen ha sido escasa en España hasta el momento, a excepción de varios investigadores. Nos referimos a J.A. Moreiro y M. Pinto Molina, ambos profesores de Biblioteconomía y Documentación de las Universidades Carlos III y Granada respectivamente.

Para comprobar el estado de la investigación sobre el resumen procedimos del siguiente modo. En abril de 1997 consultamos la Base de Datos ISOC del Consejo Superior de Investigaciones Científicas con el fin de localizar los artículos publicados sobre este tema en las principales revistas españolas sobre Biblioteconomía y Documentación. Si al resultado de esta búsqueda unimos la bibliografía localizada manualmente contabilizamos un total de diecisiete trabajos publicados, de los cuales quince son artículos y sólo dos libros dedicados íntegramente al resumen. Los autores de estas monografías son los citados Moreiro González y Pinto Molina.

Por último, realizamos una búsqueda en la Base de datos Teseo, del Ministerio de Educación y Ciencia. Esta base de datos recoge las tesis doctorales presentadas en Universidades españolas desde 1976 hasta la actualidad. La respuesta fue cero Tesis doctorales sobre el resumen.

En definitiva, los resultados quedan extractados en la siguiente tabla:

	Artículos	Monografías	Tesis doct.	Total
Resumen	15	2	0	17

Los temas abordados en los trabajos sobre el resumen son la metodología para su elaboración, tipología, servicios y revistas de resúmenes. Mientras que apenas se ha escrito sobre



su automatización.

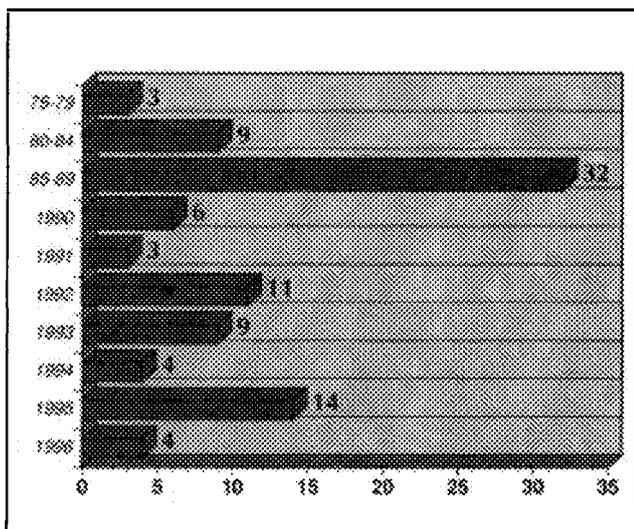
## 2.2. La investigación en indización

Como ya dijimos, la investigación en indización ha sido más abundante. El concepto de indización lo utilizamos en un sentido amplio ya que comprende tanto los aspectos teóricos, como los productos y las herramientas para la indización. Para comprobar esta afirmación consultamos de nuevo la Base de datos ISOC. Con las referencias proporcionadas y las localizadas de modo manual, obtuvimos un total de noventa y cinco trabajos sobre indización, de los cuales cuatro fueron libros, y otros cuatro tesis doctorales, todas sobre la construcción de tesauros.

Por tanto, si esquematizamos los datos quedan así:

	Artículos	Monografías	Tesis doct.	Total
Indización	87	4	4	95

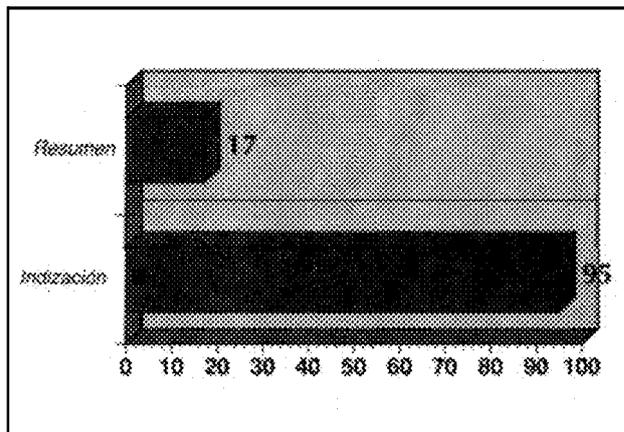
En la mayoría de los textos sobre indización el tema estudiado ha sido la elaboración de tesauros, seguido de la metodología y enseñanza de la indización. Y en menor medida, la indización de materiales no textuales, la calidad de la indización y las técnicas y tecnologías aplicadas a este proceso. En el siguiente gráfico se muestra la evolución ocurrida desde 1975 a 1996 en los trabajos relativos a la indización.



La observación de este gráfico muestra un aumento progresivo en la producción bibliográfica desde mitad de los años setenta hasta finales de los ochenta. A partir de aquí se suceden recesos y crecimientos alternativos. El aumento en el quinquenio 85-89 se debe a la publicación de un número importante de trabajos sobre elaboración de terminologías y tesauros. Después, destacan 1992 y 1995, y éste último porque en ese año se publicaron las Actas del I Encuentro de ISKO-España celebrado en 1993. Estas Actas recogen once artículos relativos a la indización, ya que una parte importante de este Congreso se dedicó a experiencias docentes en lenguajes docu-

mentales en las Diplomaturas de Biblioteconomía y Documentación españolas.

Las diferencias en publicación de trabajos en indización y resumen pueden contemplarse perfectamente con el gráfico dispuesto a continuación.



Las causas por las que se ha trabajado más en indización que en resumen no son fáciles de establecer y de enumerar. Pero, en cualquier caso, baste mencionar que, por un lado, la indización se utiliza tanto en las bibliotecas (adjudicación de encabezamientos de materia) como en los centros de documentación (asignación de descriptores). En cambio, la elaboración de resúmenes no es una tarea común de las bibliotecas, a no ser que sea especializada, y además el resumen tampoco es una labor extensamente difundida en los centros de documentación españoles. Y posiblemente, la razón de esta situación es la falta de personal en los centros, por lo que con la indización se cubre la necesidad mínima de tener debidamente almacenados los documentos para su recuperación en cualquier momento. Además, la indización se realiza con la asistencia de una serie de herramientas (vocabularios controlados, listas de encabezamientos de materia y tesauros) a las que los profesionales e investigadores les han dedicado bastantes trabajos tanto de elaboración como de evaluación.

## 3. Posibles líneas de investigación en indización y resumen

En este trabajo solamente se ha revisado la investigación en indización y resumen en España y hemos comprobado que en general las publicaciones concernientes a estos temas han sido escasas. Pero si ampliamos esta visión podemos afirmar que la investigación en Biblioteconomía y Documentación en nuestro país es un hecho relativamente reciente con respecto a países como Francia, Inglaterra, Brasil y más aún Estados Unidos. Y esto ha provocado que todavía existan bastantes parcelas donde apenas se ha indagado o incluso permanezcan sin cubrir.

Parte de esta situación podría paliarse si los documentalistas y bibliotecarios iniciaran investigacio-



nes individuales o en grupos de trabajo, de menor o mayor envergadura, encaminadas a descubrir realidades que por su proximidad conocen bien. Otros temas más complejos, se podrían abarcar con la elaboración de tesis doctorales desde programas de doctorado ofertados por las Facultades de Biblioteconomía y Documentación españolas. En definitiva, algunos temas sobre los que sería de interés iniciar, o en algunos casos ampliar, su estudio son los siguientes:

- Nivel de implantación del resumen como tarea común en los centros de documentación. Si es escaso, cuáles son las razones y en qué podría beneficiar a los usuarios-clientes potenciales en el caso de su generalización.
- Disciplinas y ciencias que intervienen en el proceso de resumir, así como la relación entre proceso cognitivo y esta tarea.
- Propuestas dirigidas a la automatización de la indización y el resumen.
- Aportaciones del procesamiento del lenguaje natural en la automatización de estos procesos.
- Estudio de los posicionamientos de profesionales e investigadores de Biblioteconomía y Documentación con respecto a la automatización del resumen y de la indización.
- Calidad de la indización en las bibliotecas y centros de documentación.
- Elaboración de tesauros para determinadas áreas del conocimiento que carecen de estas herramientas.

#### 4. Conclusiones

La revisión de la bibliografía española sobre la indización y el resumen que hemos localizado permite concluir que:

1. El resumen ha sido menos trabajado que la indización por los investigadores y profesionales españoles del área de Biblioteconomía y Documentación.
2. Los aspectos más tratados en las publicaciones sobre el resumen son la metodología y tipología, mientras que en la indización han sido la elaboración de tesauros, y en menor medida asuntos terminológicos, los teóricos y su enseñanza.
3. El número de trabajos dirigidos a la automatización de estas áreas ha sido escaso.
4. Para lograr un mayor desarrollo en cuanto al nivel de investigación en el ámbito de Biblioteconomía y Documentación sería deseable un mayor esfuerzo por parte de profesionales individuales, desde grupos de trabajo o desde programas de doctorado de Facultades de Biblioteconomía y Documentación. Esto propiciará, sin duda alguna, el avance tan esperado y necesario de nuestra área.

\* **Isidoro Gil Lelva** es Profesor de la Facultad de Biblioteconomía y Documentación de Murcia.

#### Referencias

- MOREIRO GONZÁLEZ, J.A.: *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III. 1993
- SÁNCHEZ CASAHÓN, A.I.; GARCÍA MARCO, J.F.: *La investigación sobre análisis de contenido y los lenguajes documentales en las publicaciones periódicas españolas de información y documentación (1982-1994)*. En: *Revista Española de Documentación Científica*, 18, 2, 1995, p. 155-171

#### NOTA:

(1) *Boletín de la Asociación Andaluza de Bibliotecarios*, *Boletín de la ANABAD*, *Ciencias de la Documentación*, *Cuadernos de ADAB*, *Documentación de las Ciencias de la Información*, *ITEM*: *Revista de Biblioteconomía y Documentación*, *Revista de Documentación de las Ciencias de la Información* y *Revista Española de Documentación Científica*.

P  
U  
B  
L  
I  
C  
I  
D  
A  
D