

PALABRAS CLAVE, DESCRIPTORES Y RECUPERACIÓN DE LA INFORMACIÓN

Sr. Director:

Existe una confusión generalizada entre los usuarios no expertos de los sistemas documentales, sobre el significado y grado de precisión de las denominadas palabras clave. Confusión que tiene su raíz en los dos momentos en que un investigador entra en contacto con un sistema documental: 1) cuando debe identificar las palabras clave a su artículo, generalmente siguiendo las normas para presentación de originales de la revista a que piensa enviarlo, y 2) cuando realiza una búsqueda bibliográfica, introduciendo en el sistema un número determinado de palabras clave. Naturalmente, el investigador que quiera recuperar su propio artículo solicitará al sistema las mismas palabras clave que él introdujo, y, lo que puede parecer paradójico, cuanto más imperfecto y menos desarrollado esté el sistema más posibilidades tiene de obtener una respuesta satisfactoria a su pregunta; es decir, a la solicitud formulada en los términos que él estimó más pertinentes.

La mayoría de los manuales de estilo, al referirse a las palabras clave se limitan a unas pocas, aunque sin duda útiles, recomendaciones sobre el número aconsejable de las mismas y la conveniencia de consultar índices (los de la propia revista a la que se piensa enviar el original, Index Medicus, Índice Médico Español, etc.). Prácticos consejos, sin duda, pero que no aclaran suficientemente a los autores las diferencias que existen entre un vocabulario controlado (tesauro de descriptores), y uno que no lo está (lista de palabras clave); tanto en lo que se refiere a su elaboración y mantenimiento, como a su posterior explotación por bases de datos documentales. Los conceptos de precisión y exhaustividad de una búsqueda bibliográfica están íntimamente relacionados con esta cuestión.

Una palabra clave es una palabra significativa (no vacía, es decir, no puede ser una partícula gramatical, ni preposiciones, conjunciones, artículos, pronombres, adverbios, etc.). Una palabra clave puede pertenecer por tanto a cualquier categoría gramatical. Puede ser un nombre (común o propio), puede ser un adjetivo, e incluso puede ser un verbo; y aparecer en cualquier forma gramatical. En resumidas cuentas, para el sistema documental son palabras clave todas las que conforman la referencia bibliográfica, con independencia del campo en que se encuentren. Veamos algunas definiciones. Para la Asociación Americana de Bibliotecas, una palabra clave es *toda palabra significativa en el resumen, título o texto de una obra que se utiliza como descriptor*¹. Y en términos parecidos define Martínez de Sousa en su «Diccionario de bibliología»: *Palabra o grupo de palabras escogidas del título o el texto de un documento para caracterizar el contenido y facilitar su localización, especialmente empleada en los índices*².

Por su parte, el «Manual de estilo para publicaciones biomédicas» de *Medicina Clínica* las define como *términos esenciales dentro del artículo y que se destinan a emplearse como coordenadas de búsqueda bibliográfica*³.

A la hora de indizar un documento, la aplicación de un descriptor va a depender del análisis de contenido del artículo y su nivel de profundidad, y aunque nadie puede pretender un conocimiento de su contenido mejor que su autor, los problemas en la representación de ese contenido en un sistema de información y sus posibilidades de recuperación por el mismo hacen de la indización una operación más compleja. No se trata sólo de representar un documento por medio de un conjunto de palabras clave, que casi siempre serán acertadas, sino de garantizar su recuperación; y las palabras clave, por sí solas, son muchas veces insuficientes. Pero para que un documento pueda ser recuperado, previamente debe haber sido almacenado adecuadamente. Es decir, debe ocupar un lugar lógico (generalmente más de uno) en la gran masa de documentos de un determinado sistema de información o base de datos⁴. Cuando esto se produce, es decir, cuando el contenido de un documento se confronta con una lista establecida previamente, lista en cuya elaboración se han tenido en cuenta los problemas de sinonimia y polisemia que caracterizan al lenguaje natural, y cuya finalidad es garantizar la recuperación del mismo, puede hablarse con propiedad de descriptores.

La International Organization for Standardization (ISO 5963 *Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*, equivalente de una norma española, UNE 50-121)⁵ define descriptor como *aquel término usado siempre, en la indización, para representar un concepto dado, conocido también como «término preferente»*, y lo relaciona con los términos no preferentes, que no son otra cosa que los sinónimos o cuasi sinónimos del anterior. La palabra clave persigue el mismo objetivo. La diferencia estriba en que no siempre se usa el mismo término, ya que el autor no tiene por qué conocer el término preferente y puede optar por cualquier sinónimo que juzgue más conveniente. Las palabras clave no están destinadas, aunque se empleen con frecuencia, a la indización de un documento, sino a su recuperación a través de los índices de la propia revista y de las bases de datos que las consideren como puntos de acceso. *Debemos tener siempre presente que las palabras clave sirven para buscar y ser buscados*⁶. Basta comparar los descriptores de una referencia bibliográfica contenida en una base de datos que utiliza el tesauro, con las palabras clave que el autor asignó al artículo en cuestión. Coincidirán en ocasiones, pero no siempre. Y en cualquier caso, el número de descriptores asignado por la base de datos para representar el documento, es sensi-

blemente mayor a las palabras clave que utilizó el autor. La razón está en que el autor practicó, consciente o inconscientemente, una operación reductora, condensando el contenido de su artículo en las tres, seis u ocho palabras clave que suelen aconsejar las normas de publicación de la mayoría de las revistas biomédicas. Para la recuperación de ese mismo trabajo, y ésta es la razón de ser de los descriptores, es necesario que su número sea muy superior. En cierto modo podría decirse que el autor busca la especificidad, es decir, la mayor exactitud posible en la representación del contenido, mientras que el indizador persigue la exhaustividad, es decir, la identificación de todos los conceptos de un documento susceptibles de ser interrogados⁷. En la norma antes citada se dice textualmente: *El principal criterio de selección de conceptos debe ser su valor potencial como elemento de expresión del tema del documento para su recuperación*⁵.

Digamos, a modo de conclusión, que del mismo modo que la práctica del resumen se ha generalizado y las bases de datos

lo reproducen íntegro, son ya muchas las revistas que han abandonado la práctica de las palabras clave. En la determinación del contenido de un artículo, cuyo objetivo último es la asignación de un número determinado de descriptores al documento que garanticen su recuperación, los indizadores tienen en cuenta el título y el resumen preferentemente, de los que extraen términos que concuerden con los utilizados en el tesoro de descriptores de la base de datos. Las palabras clave, cuando el artículo las contiene, les servirán sin duda de orientación, pero sólo podrán ser asignadas al documento después de un estricto control de vocabulario que resuelva, como en el caso del tesoro, todos los problemas de sinonimia, polisemia y ambigüedad semántica a que puedan dar lugar. Control de vocabulario exigible a una base de datos, pero en ningún caso a los autores.

Manuel Arranz

*Instituto Valenciano de Estudios en Salud Pública
Valencia*

Bibliografía

1. American Library Association. *Glosario ALA de Bibliotecología y Ciencias de la Información*. Madrid: Díaz de Santos, 1988: 190.
2. Martínez de Sousa J. *Diccionario de bibliología y ciencias afines*, 2ª ed. Madrid: Fundación Germán Sánchez Ruipérez, 1993: 661.
3. Medicina Clínica. *Manual de estilo: publicaciones biomédicas*. Barcelona: Doyma, 1993: 53.
4. Chaumier J. *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Mitre, 1986: 28.
5. AENOR. *Documentación. Tomo 2. Normas fundamentales: recopilación de normas USE*. Madrid: AENOR, 1994: 153-9.
6. Lafuente PJ, Bacaicoa E, Apellaniz A. Importancia de las palabras clave en las publicaciones científicas. *Gac Sanit* 1994; 8: 50-1.
7. Van Slype G. *Los lenguajes de indización. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1993: 162.