

eISSN 2444-7986

DOI: <http://dx.doi.org/10.14201/orl201673.14814>

Artículo de revisión

## DESCRIPTORES Y PALABRAS CLAVE

### *Subject descriptors and keywords*

Beatriz MUÑOZ-MARTÍN

SACYL. Complejo Asistencial de Zamora. Biblioteca. Zamora. España

Correspondencia: [bmunozm@saludcastillayleon.es](mailto:bmunozm@saludcastillayleon.es)

Fecha de recepción: 3 de julio de 2016

Fecha de aceptación: 21 de julio de 2016

Fecha de Publicación: 13 de agosto de 2016

Conflicto de intereses: Los autores declaran no tener conflictos de intereses

Imágenes: Los autores declaran haber obtenido las imágenes con el permiso de los pacientes

Política de derechos y autoarchivo: se permite el autoarchivo de la versión post-print (SHERPA/RoMEO)

Licencia CC BY-NC-ND. Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional

© Universidad de Salamanca. Su comercialización está sujeta al permiso del editor

- RESUMEN** Introducción y objetivo: Es habitual que palabra clave y descriptor se usen indistintamente y esto, desde el punto de vista documental, no es correcto. El objetivo de esta revisión es intentar aclarar conceptos y orientar sobre su uso. Método: Revisión narrativa. Se establece las diferencias entre lenguaje natural y vocabulario controlado así como su aplicación y repercusión en la recuperación de la información. Resultados y conclusiones: La diferencia entre ambos conceptos es tan clara como necesaria a la hora de interrogar cualquier fuente de información. Las normas de publicación de la mayoría de las revista científicas biomédicas contribuyen, en cierta forma, a este uso inadecuado. La dificultad que entraña el uso del tesaurus para los autores se evidencia en los errores que éstos comenten en la selección de las palabras clave frente a los descriptores que finalmente se asignan a sus trabajos.
- PALABRAS CLAVE** descriptores; palabras clave; vocabulario controlado; lenguaje natural; recuperación de la información
- SUMMARY** Introduction and objective: The concepts of keyword and subject descriptor are usually exchanged, however it is wrong from a documentary standpoint. The objective of this report is to clarify concepts and guidance on their use. Method: Narrative review. The differences between natural language and controlled vocabulary as well as their implementation and impact on the retrieval of information was established. Results and conclusions: The distinction between the two concepts is clear and necessary to interrogate any source of information. Most of biomedical scientific journals contribute to this misuse. The difficulty for using the thesaurus by authors is clearly noticeable in the mistakes they made in the selection of keywords versus the assigned subject descriptors to the article in the end.
- KEYWORDS** subject descriptors; keywords; controlled vocabulary; natural language; information retrieval

## INTRODUCCIÓN

Desde la publicación de la primera edición, en 1979, de las actuales «*Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly work in Medical Journals*» [1] redactados por el ICMJE (*International Committee of Medical Journal Editors*), que popularmente conocemos como normativa o estilo Vancouver, son muchos los editores y editoriales de revistas biomédicas que han seguido estos requisitos técnicos como norma de presentación para los manuscritos enviado por los autores a sus revistas.

Dentro de estas recomendaciones, en el apartado «preparación del manuscrito», se especifica cada una de las partes de las que debe constar dicho manuscrito y el modo de consignarlas. Uno de esos apartados es «resumen y las palabras clave» (*abstract and keywords*). En la edición del año 79 se establecía que tras el resumen se registrarán de 3 a 10 «términos clave de indización» (*key indexing terms*) formados por palabras clave o frases cortas que ayudaran en la posterior indización del artículo. Se recomendaba que esos términos fueran extraídos de los *Medical Subject Heading (MeSH)* siempre que fuera posible.

Es, únicamente, en esta primera edición en que la se denominan a las palabras clave como «*key indexing terms*». En las sucesivas ediciones no se vuelve a recoger este término sino el de «*keywords*» en su sentido amplio, aunque se sigue aconsejando el uso del *MeSH* como herramienta para la identificación de esas palabras clave.

En la versión de 1988 se añade la opción del uso de «expresiones libres» cuando no se localice un término *MeSH* para la descripción del manuscrito y no se realizarán más cambios en este apartado hasta veinte años más tarde cuando, en la actualización del año 2008, se eliminan definitivamente de las recomendaciones el uso de las palabras clave de la preparación del manuscrito.

Actualmente las revistas en lengua inglesa de mayor impacto como *JAMA*, *New England Journal*, *Annals of Internal Medicine* o *The Lancet*, miembros oficiales del ICMJE, han seguido la línea más pura de estas recomendaciones y han eliminado el uso de las palabras clave de las instrucciones para sus autores.

En cambio otras revistas que también forman parte de la lista que siguen las directrices del ICMJE, han optado por continuar con su uso dentro de sus manuscritos. *Radiología*, *Revista Clínica Española* o *Gaceta Sanitaria*, son algunos ejemplos de revistas españolas.

### ¿QUÉ ES UN PALABRA CLAVE?, Y ¿UN DESCRIPTOR?

Es habitual que palabra clave y descriptor se usen indistintamente y esto, desde el punto de vista documental, no es del todo correcto.

Una palabra clave es una palabra o frase corta significativa que describen el contenido de un trabajo en lenguaje natural, el mismo que se utiliza en la comunicación humana. Son términos libres y variados que dependen de la riqueza del vocabulario de quien los utilice.

Frente a esta variabilidad y ambigüedad del lenguaje natural (género, número, sinonimia, polisemia), se encuentra el lenguaje controlado, que surge como herramienta en la indización y en la recuperación de la información. Lo componen términos normalizados dentro de un lenguaje documental llamados descriptores que representan de manera unívoca un concepto [2]. Su función principal es la de homogenizar los términos empleados para una búsqueda más precisa.

Es el documentalista quien, en el proceso de tratamiento documental al que se somete el artículo para ser recogido en una base de datos, otorga esos términos normalizados en función de la temática (contenido) del documento, con el objetivo de facilitar su recuperación. Es lo que conocemos como indización (o indexación si se trata de un recurso informático).

Los descriptores están recogidos en el tesoro, definido por la norma ISO 2788-1986 como «un vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento». Los tesauros relacionan los términos en lenguaje natural (palabras clave) con los normalizados (descriptores).

En las bases de datos biosanitarias el tesoro más conocido es el *MeSH (Medical Subject*

*Headings*), utilizado por la *National Library of Medicine* para todas sus bases de datos. Es el que recomienda el *ICMJE* en sus normas. En esta línea, y con el propósito de permitir el uso de una terminología común en tres idiomas (portugués, español e inglés), *BIREME* desarrolló a partir del tesoro *MeSH*, los «Descriptores en Ciencias de la Salud (DeCS)», que permiten la indización y recuperación en las bases de datos que conforman la Biblioteca Virtual en Salud (BVS).

#### DIFERENTES CONCEPTOS, DIFERENTES USOS

La diferencia entre ambos conceptos es tan clara como necesaria a la hora de interrogar cualquier fuente de información.

Si nos encontramos frente a una base de datos su propia estructura indicará qué debemos utilizar.

Recordemos que las bases de datos están formadas por una sucesión de registros (*records*) que corresponden a las referencias de los artículos indizados.

Los registros son representaciones que contienen la información básica para proceder a la identificación y recuperación de esos artículos.

Cada referencia se estructura en campos (*fields*) que contienen información relativa a la descripción e identificación del documento. Los campos varían de una base de datos a otra o, incluso, si la versión es gratuita o de pago.

Los campos más comunes suelen ser: autor/es, título del artículo, resumen, descriptores, título de la revista, año de publicación, volumen de la revista y las páginas, aunque puede contener muchos otros.

La estructuración de los registros en campos hace que sea más sencilla y eficaz la gestión y recuperación de la información. Si los registros recogieran el texto completo del artículo, tal cual, como si fuera una fotocopia, cualquier palabra, independientemente de su significado tendría el mismo valor de cara a la recuperación, lo que provocaría una gran imprecisión en los resultados de la búsqueda. Existen dos tipos de campos que conforman el registro: los que ayudan a identificar el documento (autores, revista, año, volumen, número, etc.) y los que describen su contenido: el título, el resumen, las palabras clave

(en las bases de datos que incluya ese campo) y los descriptores.

De entre estos últimos, el título, el resumen y las palabras clave utilizan un lenguaje libre (han sido redactados por el propio autor), mientras que el campo descriptores está basado en un lenguaje controlado, asignado a posteriori al manuscrito durante el proceso de indización.

Como consecuencia, la búsqueda de contenido mediante descriptores únicamente tiene sentido en recursos que contengan ese campo y un tesoro que ayude a su normalización y equivalencia con los términos naturales.

Las palabras clave, en cambio, como términos libres, se emplearán para la recuperación de información en el título, resumen o en otros campos que la propia base de datos haya elaborado para la búsqueda de contenido en lenguaje natural. Es el caso de la *Web of Science* o *Scopus*, entre otras, que han incluido campos de búsqueda de palabras clave dadas por los autores. En estos casos, lenguaje controlado y natural resultarían complementarios para una búsqueda de información pertinente y exhaustiva.

Otro enfoque requiere la búsqueda en motores, metabuscadores y portales. Estas herramientas no disponen de tesoro, luego los descriptores (términos normalizados) y las palabras clave (términos libres) serán tratados del mismo modo: como palabras libres que aparezcan en el texto. No existen campos normalizados, la recuperación de la información se basa en coincidencias bajo una indización automatizada [3].

Cuando el usuario interroga a un motor de búsqueda, teclea unas palabras que estima relevantes para su consulta (palabra clave). De acuerdo al comportamiento de los buscadores, la página de resultados le devolverá un *ranking* de documentos ordenados tanto por la cantidad de palabras coincidentes con el criterio de búsqueda que aparezcan en el documento como por el número de veces que aparecen, de tal forma que las palabras clave que busca el usuario y las palabras clave que el autor ha dado a su trabajo, deben coincidir. Y será mucho más probable que el usuario realice su búsqueda en términos libres, que conozca y maneje, que el que acuda a un tesoro para localizar el descriptor más adecuado [4].

AUTORES Y EDITORES: 100% PALABRAS CLAVE

La intención de todo autor cuando publica su obra es que su trabajo sea consultado y en el mejor de los casos, citado. Ése es el verdadero reconocimiento de su investigación. Pero para ello es condición indispensable que sea recuperado de entre otros trabajos similares [4].

Solicitar a los autores que asignen palabras clave a sus manuscritos cuando son remitidos a una revista para su publicación, es un valor añadido en la recuperación de sus trabajos. Como hemos visto, existen bases de datos que las indexan y las convierten en otro campo de búsqueda más en lenguaje libre para el usuario.

Si, además, su artículo va a ser publicado en una revista indizada en una base de datos con tesauro, sabe que a su trabajo se le asignarán, posteriormente, unos descriptores que harán aún más efectiva la recuperación.

El problema estriba cuando se exige a los autores que esas palabras clave (que por definición son términos del lenguaje natural) sean extraídas de un tesauro (vocabulario normalizado). Entonces ambos conceptos pierden su significado absoluto, se nombran indistintamente y comienzan a usarse de forma errónea.

Estudios como los de Gil-Leiva y Alonso-Arroyo, Rockelle y Ansari [5-7], evidencian el alto porcentaje de errores o inexactitudes cometidos por los autores en la selección de las palabras clave frente a los descriptores que finalmente se asignaron a dichos artículos, debido, principalmente, a la dificultad que entraña el uso del tesauro en personas no expertas.

No es de extrañar, por tanto, que el *ICMJE* haya eliminado de sus recomendaciones la asignación por parte de los autores de las palabras clave (entendidas como descriptores, ya que se recomendaba el uso del tesauro *MeSH* para su normalización), puesto que éstas parecen servir más de guía al indizador que como elemento real de recuperación de contenido para el usuario [5-7]. Pretender utilizar un vocabulario controlado en lo que se convertirá, tras la indización, en un campo de lenguaje libre en una base de datos, es inoperante y absurdo.

Consecuentemente a lo expuesto, cuando una publicación no esté recogida en una base de datos y a sus artículos no se le asignen descriptores, o cuando se trate de trabajos en

forma de *post*, *blog* o estén publicados en una página web, será responsabilidad de los autores dotarlos de palabras clave significativas (y no descriptores) que representen el contenido para lograr un buen posicionamiento en los motores de búsqueda y recolectores, teniendo presente que:

- El ranking, principalmente, se efectúa por número coincidencias de la palabra y concepto buscado;
- que el usuario nunca va a recurrir a un tesauro para consultar en un motor de búsqueda.

Una mala elección de esas palabras clave podría dificultar la recuperación y difusión de los trabajos, puesto que se trata, como hemos visto, de una herramienta de doble vía: de quienes escriben y de quienes buscan la información [8]. Los trabajos que estén bien indexados en los buscadores serán más fáciles de encontrar y con más probabilidad de ser citados. Incluso autores como Macías-Alegre [9], estiman que «(...) esta responsabilidad podría asumirse desde los consejos editoriales por especialistas que afinen los términos clave y eviten errores».

La opción más consecuente de las publicaciones que se encuentren en esta tesitura, es la de encaminar sus esfuerzos para ser incluidas en un repositorio o usar un sistema de administración y publicación de revistas línea. Ambas alternativas facilitan al editor, y por inclusión, al autor, el gran obstáculo que supone la falta de visibilidad en la red porque permiten no sólo la posibilidad de gestionar los metadatos, si no la facilidad de ser recolectados por agentes externos.

Los metadatos permiten la descripción del documento y su indización en los motores de búsqueda y su accesibilidad desde múltiples fuentes [10]. Recogen elementos sobre el formato y la identificación del documento; elementos sobre información relativa a la propiedad intelectual del recurso y elementos relacionados con el contenido. Es en este último donde se recogen los metadatos sobre la materia y palabras clave suministradas por el autor en su trabajo y de los que se servirán los motores de búsqueda para crear sus índices.

*Revista ORL* es un claro ejemplo de buen posicionamiento web mediante el uso de metadatos. Es una publicación que ha sabido adaptarse y evolucionar con las alternativas

que tenía a su alcance sin perder, en palabras de su Director «sus criterios de calidad, de rigor formal y de contenido» [11].

#### BIBLIOGRAFÍA

1. International Committee of Medical Journal Editors [Internet]. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals [acceso el 1 de junio de 2016]. Disponible en: <http://www.icmje.org/>.
2. Martín-Gavilán C. Temas de Biblioteconomía: Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia, descriptores y tesauros. 2009. URI: <http://hdl.handle.net/10760/14817>.
3. Méndez Rodríguez E M, Moreiro González J A. Lenguaje natural e Indización automatizada. Cienc Inf [internet].1999 [acceso el 1 de junio de 2016];30(3):11-24. Disponible en: <http://www.mdp.edu.ar/humanidades/documentacion/licad/archivos/modulos/proces/archivos/bibliografia/procesamiento/Eje3/P020.pdf>.
4. Muñoz-Martín B. Incrementa el impacto de tus artículos y blogs: de la invisibilidad a la visibilidad. Rev Soc Otorrinolaringol Castilla Leon Cantab La Rioja. 2015;6(Supl. 4):S6-S32. URI: <http://hdl.handle.net/10366/126907>.
5. Gil-Leiva I, Alonso-Arroyo A. La relación entre las palabras clave aportadas por autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. Rev Esp Doc Cient. 2005;28(1):62-79.
6. Rockelle Strader C. Author-Assigned Keywords versus Library of Congress Subject Headings. Implications for the Cataloging of Electronic Theses and Dissertations. Libr Resour Tech Serv. 2009;53(4):243-50.
7. Ansari M. Matching between assigned descriptors and title keywords in medical theses. Libr Rev. 2005;54 (7):410-4.
8. Mattar V S, González Tous M. Las claves de las palabras clave en los artículos científicos. Rev Mvz Cordoba [internet].2012 [acceso el 1 de junio de 2016]; 17 (2): 2955-6. Disponible en: <http://revistas.unicordoba.edu.co/revista/mvz/mvz-172/V17N2A1.pdf>.
9. Macías-Alegre A. ¿Tienen sentido las palabras clave de los artículos científicos?. COMeIN [internet]. 2016 mayo [acceso el 1 de junio de 2016]. Disponible en: <http://www.uoc.edu/divulgacio/comein/es/numero55/articles/Article-Adrian-Macias-Alegre.html>.
10. Ferreras-Fernández T. Repositorios de acceso abierto: un nuevo modelo de comunicación científica. Rev Soc Otorrinolaringol Castilla Leon Cantab La Rioja. 2015;6(Supl.4):S33- S65. URI: <http://hdl.handle.net/10366/126908>.
11. Pardal-Refoyo JL. La gestión del conocimiento. Rev Soc Otorrinolaringol Castilla Leon Cantab La Rioja. 2015;6(1):1-3. URI: <http://hdl.handle.net/10366/125163>.