

Hacer resúmenes para la búsqueda en texto libre*

Raya Fidel

*Titular de la Escuela de Biblioteconomía,
Universidad de Washington*

Un estudio sobre las formas seguidas en la elaboración de resúmenes utilizadas por los productores de bases de datos bibliográficas revisó las instrucciones en la elaboración de resúmenes, que pretenden realzar la recuperación en texto libre. De las 123 prácticas de bases de datos examinadas, cincuenta y siete (un 46%) incluían dichas instrucciones. Los editores consideran el contenido de los resúmenes y su lenguaje como un factor principal para la mejora de la recuperación. La mayoría de ellos recomienda que una vez el resumidor decide qué conceptos incluir en un resumen y en qué forma los va a representar, estos términos deberían ser coordinados mediante términos de indización asignados por un vocabulario controlado. Las pautas acerca del tipo de resumen, por ejemplo informativo o indicativo, y su longitud no se ven afectas por la capacidad de la recuperación en texto libre.

A survey of abstracting policies used by producers of bibliographic databases examined abstracting guidelines which aim to enhance free-text retrieval. Of the 123 database policies examined, fifty-seven (46 per cent) included such instructions. Editors consider content of abstracts and their language as a primary factor in retrieval enhancement. Most recommend that once abstractors decide which concepts to include in abstracts and in which form to represent them, these terms should be coordinated with index terms assigned from a controlled vocabulary. Guidelines about the type of abstracts, i.e., informative or indicative, and about their length are not affected by the capability of free-text retrieval.

Introducción

Desde que los servicios de resúmenes e indización han ofrecido bases de datos bibliográficas en forma impresa, las funciones asignadas a la indización se han ido diferenciando casi completamente de aquéllas asignadas al proceso de hacer resúmenes. La indización facilita la recuperación eficiente de la información y el proceso de resumir facilita las valoraciones acerca de la relevancia de la información recuperada. Los resúmenes han sido siempre considerados como sustitutos de los documentos originales.

La disponibilidad de las bases de datos bibliográficas en línea alteró la función de los resúmenes. Los resúmenes de la mayoría de

bases de datos bibliográficas públicamente disponibles se pueden buscar en la modalidad de texto libre, que permite al usuario buscar en línea la aparición de cualquier término que crea apropiado. Entonces el sistema recupera las referencias de todos los documentos cuyos resúmenes contengan los términos requeridos o sus combinaciones. Este proceso se lleva a cabo sin el uso de un lenguaje de indización específico.

Sencillamente, los resúmenes se han convertido en una importante mejora para la recuperación de información.

A pesar de esta nueva función, las cuestiones vitales relativas a la elaboración de resúmenes para facilitar una recuperación eficiente aún

*© Journal of Documentation, 42 (1), 1986, pp. 11-21

Traducción: Pilar Mañas Lahoz

siguen sin respuesta. La investigación activa acerca de cómo elaborar resúmenes cesó a mitad de los años 60, después de que se estableciera mediante una serie de tests¹ la importancia de los resúmenes para evaluar la pertinencia de los documentos y para ser usados como sustitutos, y después de que un extenso estudio de 130 prácticas seguidas en la elaboración de resúmenes elaborara criterios que serían usados en la evaluación de resúmenes². Aunque estos estudios son hitos para la investigación de la elaboración de resúmenes y sus resultados son altamente valiosos, no examinan sin embargo la pertinencia de los resúmenes en la recuperación en texto libre.

Sin embargo, los investigadores de diseño y evaluación de los sistemas de información se han interesado en gran manera por la actuación de la búsqueda en texto libre. La indización es un proceso costoso, que requiere mucha mano de obra y los lenguajes de indización resultan caros de elaborar. Más aun, existen varias cuestiones teóricas relacionadas con dicha elaboración que aun no se han resuelto. La capacidad de recuperar documentos sin la ayuda de un lenguaje de indización artificialmente construido parece, por tanto, muy atractiva. Comenzando con los estudios de Cranfield³, numerosos ex-

perimentos se han intentado determinar si realmente la búsqueda en texto libre supera la búsqueda con ayuda de un lenguaje de indización controlado. Los resultados son contradictorios y continúa la polémica sobre si los futuros sistemas deberían o no disponer de vocabularios controlados y ser restringidos a la capacidad de búsqueda en texto libre^{4, 5, 6 y 7}.

Entre los profesionales, los diseñadores de sistemas de información ofrecen resúmenes para la búsqueda en texto libre, y los usuarios del sistema utilizan en gran medida esta capacidad. Diferentes servicios de resúmenes e indización han investigado el uso de la búsqueda en texto libre para determinar cómo diseñar sus lenguajes de indización y las pautas seguidas en la elaboración de resúmenes^{8 y 9}. Además, los usuarios en línea intentan seleccionar estrategias eficientes para la búsqueda en texto libre de los resúmenes¹⁰.

En vista de la amplia experiencia que se ha conseguido en los últimos años acerca de la búsqueda en texto libre, parece que ha llegado el momento de hacer una investigación sistemática acerca de las características de los resúmenes que son esenciales para una recuperación eficaz.

1. RESNICK, A.: "Relative effectiveness of documents titles and abstracts for determining relevance of documents". En: *Science*, 134 (3484), 1961, pp. 1004-1005.
2. BORKO, H. and CHATMAN, S.: "Criteria for acceptable abstracts: a survey of abstracters instructions". In: *American Documentation*, 14(2), 1963, pp. 149-160.
3. CLEVERDON, C. W.: *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Cranfield: College of Aeronautics, 1962.
4. CLEVERDON, C. W.: "Optimising convenient online access to bibliographics databases. In: *Information Service & Use*, 4(1/2), 1984, pp. 37-47.
5. LANCASTER, F.W.: "Trends in subject indexing from 1957 to 2000". In: *New trends in documentation and information*, London: Aslib, 1980.
6. HENZLER, R. G.: "Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information system". In: *International Classification*, 5(1), 1978, pp. 21-26.
7. DUBOIS, C. P. R.: "The use of thesauri in on-line retrieval". In: *Journal of information Science*, (2), 1984, pp. 63-66.
8. CARROW, D. and NUGENT, J.: "Comparison of free-text and index search abilities in an operating information system," In: *Proceedings of the American Society for information Science Annual Meeting*, 14(2), 1977, 2-e8.
9. MARKEY, K, and others: "An analysis of controlled vocabulary and free text search statements in on-line searches" In: *Online Review*, 4(3), 1980, pp. 225-236.
10. WAGERS, R.: "Effective searching in databases abstracts". In: *Online*, 7(5), 1983, pp. 60-77.

Con el fin de considerar los criterios vitales para la admisibilidad de los resúmenes en la búsqueda en texto libre, este estudio examinó las prácticas seguidas en la elaboración de resúmenes utilizadas por los productores de bases de datos bibliográficas con el fin de determinar el grado en que los servicios de resúmenes e indización incorporan ya tales criterios en sus resúmenes. De modo muy importante, el estudio sintetiza las características de los resúmenes que son importantes en la recuperación en texto libre, tal como las definen las prácticas para la elaboración de resúmenes, y alude a sus efectos sobre la calidad de los resúmenes.

El método de estudio

Para estudiar las prácticas en la elaboración de resúmenes, se complicó una lista de productores de bases de datos bibliográficos que elaboran resúmenes disponibles para la búsqueda en texto libre, por medio de varios directorios de bases de datos. La información acerca de bases de datos específicas, tomada primeramente de la bibliografía ofrecida por los distribuidores de sistemas de búsqueda, eliminó a aquellos productores que utilizan solamente resúmenes de autor.

A 159 productores de bases de datos se enviaron cartas posteriormente explicando el propósito del estudio y solicitando copias de las pautas seguidas en la elaboración de resúmenes y en la indización. Las cartas a los productores americanos y canadienses fueron dirigidas a los directores responsables de los resúmenes dentro de sus respectivas organizaciones. Si no se obtenía una respuesta rápida, se le volvía a avisar por carta, seguido de una llamada telefónica. Las cartas enviadas a productores de otros países iban dirigidas a "el director" y no fueron seguidas de ningún aviso posterior.

Setenta y cinco de los noventa y seis productores de bases de datos norteamericanos respondieron a la petición, resultando un porcentaje de respuesta del 78%. De los productores de otros países respondieron un 59%, treinta

y siete de los sesenta y tres del total respondieron a la petición. Aquellos que respondieron ofrecieron diversos tipos de respuestas: Aquellos que tenían las pautas impresas enviaron una copia, mientras que otros escribieron cartas o telefonearon describiendo en cada caso su política y práctica utilizadas en la elaboración de resúmenes.

Basándose en la información ofrecida por los productores de bases de datos, se analizaron los datos a nivel de bases de datos específicas. Algunos directores informaron que sus resúmenes ya no estaban disponibles para la búsqueda, mientras que otros afirmaron que no utilizaban ninguna política en la elaboración de resúmenes. Estas bases de datos no se tomaron en consideración. Se identificó un total de 123 bases de datos que ofrecen resúmenes susceptibles de búsqueda, escritos de acuerdo con algún tipo de pautas. Estas pautas se analizaron para extraer la instrucciones dadas por los editores de directrices utilizadas en la elaboración de resúmenes con el fin de mejorar la recuperación en línea.

Este documento describe y analiza estas instrucciones. Aunque en el texto no se hace mención a las bases de datos específicas, se incluyen en el apéndice una lista de los productores de bases de datos cuyas políticas y prácticas para hacer resúmenes contribuyeron directamente a la complicación de estas instrucciones.

Del número total de bases de datos consideradas para este estudio, ochenta y nueve (72%) ofrecen al resumidor pautas escritas, mientras que las otras treinta y cuatro informan al resumidor de manera informal. Las estadísticas nos muestran que los directores que ofrecen las directrices por escrito es más probable que se preocupen por las instrucciones acerca de la recuperación en texto libre que sus contrarios: mientras que el 50% de las pautas (cuarenta y nueve) incluyen instrucciones en texto libre, solamente el 35% (doce) de las bases de datos que abastecen pautas informales mantienen este criterio.

Se debe tener en cuenta que los product-

res de bases de datos específicas suelen utilizar un grupo de pautas de resúmenes en algunas de sus bases de datos. Treinta y seis productos (incluidos en el apéndice) facilitaron las pautas de las cincuenta y siete bases de datos (46%) que incluyen instrucciones para el texto libre: estas bases de datos son la base de las

instrucciones para el texto libre que se describen en este documento.

Al examinar la relación entre indización por descriptor e instrucciones para el texto libre, encontramos que existen dos tendencias, como se puede apreciar en la Figura 1.

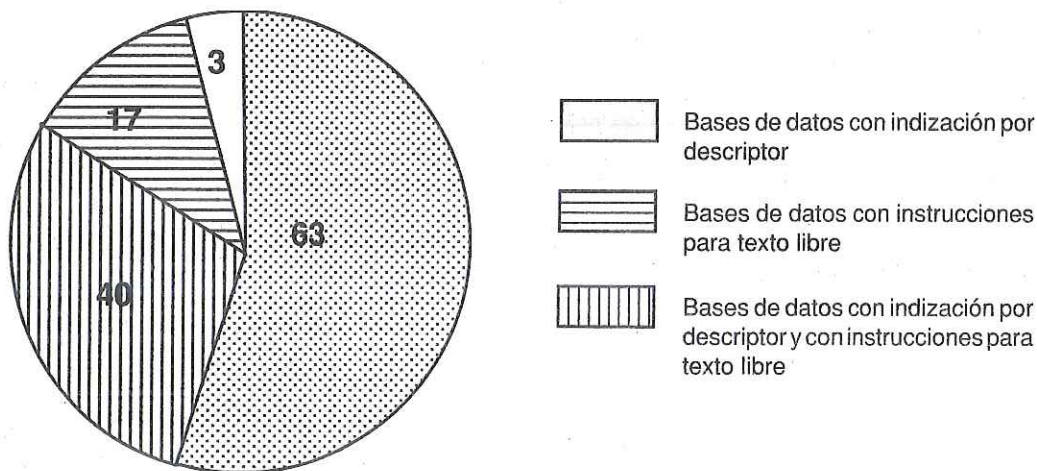


Figura 1. Análisis de las 123 bases de datos incluidas en el estudio

En primer lugar, un gran número de bases de datos que utilizan descriptor de indización aun eluden las instrucciones para la recuperación en texto libre. De entre las 103 bases de datos que utilizan descriptor de indización, sólo cuarenta (un 39%) ofrecen instrucciones para la recuperación en texto libre. De hecho, sesenta y tres bases de datos (el 51% de las 123 consideradas para este estudio) utilizan descriptor de indización pero *no* ofrecen instrucciones para la recuperación en texto libre, aunque incluso sus resúmenes pueden ser buscados en línea e incluyen algún tipo de pautas para resúmenes. En otras palabras, las instrucciones para texto libre no son muy comunes entre los directores que utilizan descriptores para la indización.

Puede ser que algunos directores aún no hayan visto la conexión entre las pautas para resúmenes y la recuperación en texto libre; otros se dan cuenta de la necesidad de incorporarlas a sus directrices pero aún no lo han hecho. Otros directores han decidido ignorar el efecto de los

resúmenes sobre la recuperación, basándose en consideraciones de tipo económico. Algunos de estos editores creen que mientras los resúmenes y títulos, que ya han sido elaborados ofrecen puntos de acceso adicionales sin suponer un esfuerzo extra en el período de entrada de datos, en cambio las instrucciones especiales para elaborar dichos resúmenes hacen que la recuperación en texto libre sea menos económica. De modo alternativo, otros creen que la búsqueda por descriptor es el mejor método de recuperación para sus bases de datos; por tanto, la inversión de recursos para la mejora de la recuperación en texto libre no resulta económica para ellos.

La segunda tendencia que deducimos de los datos es que la mayoría de las bases de datos que ofrecen instrucciones para texto libre no consideran que la indización por descriptor sea obsoleta. Entre las cincuenta y siete bases de datos que ofrecen instrucciones para el texto libre, cuarenta (70%) utilizan indización por

descriptor. En otras palabras, un porcentaje considerable de los directores que estiman la búsqueda en texto libre como un método importante para la recuperación también creen que no es sustituible por la búsqueda por descriptor. Por tanto, el debate acerca de si la recuperación en texto libre debería o no ser sustituida por la indización por descriptor resulta un tanto académico; es muy probable que las bases de datos que mejoran la recuperación en texto libre ofrezcan ambos tipos de puntos de acceso.

Pautas para elaborar resúmenes

En su estudio acerca de las instrucciones generales para elaborar resúmenes, Borko y Chatman abordaron tres cuestiones: la función, el contenido y la forma de los resúmenes¹¹. De modo interesante, solamente las dos últimas son consideradas por los productores de bases de datos en sus instrucciones específicas para bases de datos automatizadas.

La primera cuestión, la función de los resúmenes, determina si los resúmenes deben ser informativos o indicativos: si deben condensar la información en documentos o informar acerca de lo que tratan los documentos. Aunque los directores de bases de datos automatizadas suelen indicar a los resumidores qué tipo de resumen deben elaborar, ninguno supone que esto se relaciona con la recuperación en texto libre. Dicho de otro modo, la naturaleza informativa o indicativa de los resúmenes, está aún determinada por el propósito y uso de los mismos, y no por ningún factor que tenga que ver con la recuperación de información.

Cuando las prácticas generales para elaborar resúmenes abordan la segunda cuestión, su contenido, indican qué *información* se debe incluir en el resumen, pongamos por caso un método de estudio o los resultados de una investigación. Los directores que están interesados en la recuperación con resúmenes señalan de modo adicional qué *conceptos* se deben incluir

en los resúmenes. Como veremos más tarde, aunque los editores suelen estar de acuerdo acerca de la naturaleza de los conceptos a incluir, presentan una variedad de opiniones, algunas contradictorias, acerca de los criterios específicos que se deben usar para seleccionar conceptos.

La tercera cuestión, la forma de los resúmenes, se relaciona principalmente con la longitud y el estilo en que son escritos. Sorprendentemente, en ninguno de los manuales de resúmenes revisados se menciona la posible relación entre la longitud de los resúmenes y la eficacia de la recuperación. El estilo en que son escritos, sin embargo, sí se relaciona con la recuperación: un buen número de editores suministra sistemas acerca del modo en que se deben representar los conceptos para mejorar la recuperación. Desgraciadamente, algunas instrucciones hacen que sea difícil el mantener un estilo de escritura aceptable, afectando en la mayoría de los casos en la brevedad y claridad en la escritura.

En la figura 2 se incluye un sumario de las directrices para elaborar resúmenes que este estudio siguió, y que tienen como fin la búsqueda en texto libre.

El contenido de los resúmenes

Existe una máxima sobre la cual todos están de acuerdo: los conceptos "importantes" para la recuperación deben ser incluidos en los resúmenes. Evidentemente, algunos productores de bases de datos consideran suficiente incluir un enunciado general de este tipo en sus pautas para elaborar resúmenes, sin ofrecer más detalles. Pero cuando se diseñan criterios específicos para seleccionar conceptos "importantes", los directores difieren, especialmente con respecto a la relación entre estos conceptos y otros términos de indización.

Enunciados generales

Ya que un enunciado de este tipo recuerda al resumidor que sus resúmenes serán utiliza-

11. Ver nota 2.

El contenido de los resúmenes

Menciones generales

Utilizar conceptos y términos "importantes" (por ejemplo, aquéllos que mejoren la recuperación en texto libre, aquéllos acerca de los que un documento da suficiente información, o palabras clave).

Términos de indización

- Coordinar los conceptos, utilizados en los resúmenes con descriptores asignados.
 - a) Asignar conceptos que sean en los resúmenes idénticos a los descriptores.
 - b) Asignar conceptos que complementen a los descriptores (por ejemplo términos que se hayan omitido en la indización por descriptor y los títulos, términos más específicos que los descriptores, o algún tipo concreto de término que sea importante para el tema, como nombres geográficos)
 - c) Asignar términos que complementen y sean idénticos a los descriptores. Mejorar la indización, independientemente del lenguaje de indización utilizado.

Listas de comprobación

Seguir una lista de elementos pertinentes a la indización que deberían ser incluidos en los resúmenes.
Formas de las listas de comprobación:

- a) Categorías que se deberían incluir en los resúmenes (por ejemplo materiales, propiedades y procesos) y condiciones bajo las cuales se deben incluir (por ejemplo sólo cuando son discutidas con detalle, o siempre que sean mencionadas).
- b) Pautas concretas y específicas (por ejemplo, siempre que se trate de un producto nuevo, mencionar el nombre de la compañía).

El lenguaje de los resúmenes

Uso del lenguaje del autor.

- Utilizar el lenguaje del autor.
- No utilizar el lenguaje del autor.

a) Usar términos concretos y nominalizados, específicos para un área temática. Usar tanto el lenguaje del autor como sinónimos.

Relación con el lenguaje de indización utilizado

- Coordinar los términos usados en el resumen con los descriptores.
- Complementar los descriptores con términos incluidos en el resumen (por ejemplo, utilizar sinónimos o términos más específicos).
- Utilizar términos específicos y bien aceptados para categorías concretas (como materiales, procesos y productos).

Prácticas que se deben evitar:

- No utilizar la negativa (por ejemplo utilizar *enfermo* en lugar de *no sano*).
- No incluir términos que tengan la primera palabra en común como una serie (como "clase alta, media y trabajadora").

Formas de las palabras:

- Seguir el uso del lenguaje local (por ejemplo, cambiar la ortografía americana en las bases de datos británicas). Deletrear siempre los términos para ciertas categorías (como procesos, materiales y productos).
- Cuando un término y un descriptor son iguales, mencionar el término en la forma utilizada por el descriptor.
- Mencionar los términos tanto en su forma abreviada como en su forma completa.

Figura 2. Una presentación resumida de las directrices para elaborar resúmenes en la búsqueda en texto libre.

dos para la recuperación, debe mencionar los conceptos centrales de los documentos resumidos. El enunciado facilitado por la Organización Internacional de Normalización (International Organisation for Standardization), por ejemplo, pide a los resumidores el "utilizar palabras significativas del texto, que ayudarán a la búsqueda de texto automatizada"¹². Aunque este enunciado puede ser aceptado como una norma internacional, los directores individualmente son los que pueden definir más concretamente qué materias son "significativas para la búsqueda de texto automatizada".

Los directores pueden elegir el ser más o menos rigurosos en sus definiciones, pero las pautas disponibles resultan inadecuadas cuando no ofrecen ayuda adicional a los resumidores. Sin ser demasiado concreto, un director puede pedir a los resumidores que incluyan conceptos acerca de los cuales un documento ofrece suficiente información. Se puede incluso especificar la forma en que se desea la información y requerir la inclusión de tan sólo aquellos temas para los cuales se ofrecen, digamos, procedimientos de comprobación y resultados. Otro enfoque es el de sugerir que las palabras clave sean incorporadas a los resúmenes, especialmente si se repiten con frecuencia en el texto.

Términos de indización

Al requerir que los conceptos importantes sean explícitamente representados en los resúmenes para la búsqueda automatizada, lo que realmente se pretende es que los documentos sean indizados utilizando un lenguaje natural. Aunque los enunciados generales acerca de los contenidos de los resúmenes no hacen referencia al proceso de indización, numerosos manuales de resúmenes indican claramente que los resúmenes deberían mejorar la indización. De hecho, algunas organizaciones requieren que los

resumidores revisen a fondo la indización de los documentos antes de editar o escribir un resumen. Los editores de manuales de resúmenes, sin embargo, no coinciden en el modo en que los resúmenes deberían mejorar la indización.

Un enfoque afirma que los resúmenes se deben componer teniendo en cuenta la indización, pero independientemente de cualquier lenguaje de indización utilizado. Este enfoque hace que sea difícil dar instrucciones específicas, y normalmente se da un enunciado general con el fin de que los resúmenes incluyan términos bajo los cuales se pueda buscar el tema de un documento. Los directores puede que pretendan recordar a los resumidores que esta función es particularmente importante en aquellos documentos cuyo título es corto y falto de información. Teniendo en cuenta la idea de que los términos de los resúmenes son puntos de acceso, se pueden ofrecer ejemplos a los resumidores con el fin de llamar su atención. Se puede hacer que excluyan conceptos generales o inclusivos, como "diversas localizaciones", y que usen conceptos específicos, como los nombres de las localizaciones.

El otro enfoque coordina los conceptos utilizados en los resúmenes con términos de indización que son asignados por un tesauro, por ejemplo, con descriptores. Aquí hacemos frente a la siguiente cuestión: los conceptos en los resúmenes, ¿deben ser idénticos a los descriptores o deben complementarlos? Aunque la mayoría de los servicios eligen para sus resúmenes conceptos que complementen a los descriptores, algunos directores requieren específicamente de sus resumidores que escriban resúmenes con descriptores incluidos. Otros no consideran contradictorias las dos opciones. Aunque los resúmenes se hacen con el fin de complementar la indización, piensan que es realmente deseable el que exista una repetición en los términos de indización. De hecho sostienen que un resumen es deficiente si no incluye

12. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Information transfer. Geneva: ISO, 1982, p. 23.

los términos bajo los cuales se recuperó usando descriptores.

Los directores que deciden que los conceptos de los resúmenes deberían complementar a los descriptores, pueden dar instrucciones diversas para conseguir una mejora en la indización con estos conceptos. Ejemplos de dichas instrucciones incluyen: conceptos que no aparezcan en el título o en el campo del descriptor y que el autor quiera que sean recuperados; cualquier término relevante siempre que mejore los términos del tesoro, ya sea un sinónimo apropiado o alguno que no esté adecuadamente presentado por los descriptores; o equivalentes lingüísticos respecto a los códigos temáticos para la base de datos donde los descriptores tomen la forma de códigos temáticos.

Estas instrucciones son aún muy generales y confían en los juicios subjetivos de los resumidores. Se puede ofrecer ayuda adicional indicando la naturaleza de los conceptos que se han de seleccionar para la mejora de la indización. La idea predominante es que los conceptos en los resúmenes deberían ser más concretos, o más detallados, que los conceptos que pretenden realzar. Para bases de datos específicas, sin embargo, los directores pueden tener requerimientos más específicos, como la inclusión de nombres geográficos, términos que pertenecen a un campo concreto, reclamos, jergas y citas pertenecientes a un documento, o nombres propios que aparezcan junto a los términos.

Lista de control.

El enfoque más riguroso respecto a la indización con términos que provienen de resúmenes incluye la confección de una lista de elementos pertinentes a la recuperación que deben ser incluidos en los resúmenes, sean o no tratados por la indización por descriptor. Dichas listas de control pueden adoptar varias formas.

La lista de control más estructurada incluye categorías que deben ser incluidas en los resúmenes. Las categorías reales suelen variar de

una base de datos a otra. En un área tecnológica, por ejemplo, una base de datos siempre incluye el tipo de proceso de soldadura y tipos de acero, y otros sugieren la inclusión sugieren la inclusión de materiales, formas, condiciones, propiedades y procesos. En otra área se consideran especialmente significativos los nombres de compañías, nombres de productos y nombres concretos. Un servicio incluye en la lista once categorías y requiere que cualquier categoría relevante sea incluida en los resúmenes.

Otro uso de las listas de control es el de escribir directrices muy concretas, como: siempre que se trate de un producto, proceso o tecnología nuevos, mencionar el nombre de la compañía; incluir el nombre normalizado completo de todos los tests, cuestionarios...; si el documento trata de un estudio, hacer constar quién dirigió el estudio; o, si una compañía es subsidiaria, mencionar el nombre de la organización fuente si no está incluido en el nombre de la compañía.

Para conseguir su meta de mejorar la indización con conceptos provenientes de resúmenes, los directores pueden pasar por alto otras consideraciones. Por ejemplo, insistir sobre la inclusión en los resúmenes de listas detalladas de categorías puede dificultar mucho la brevedad. Los directores deben, por tanto, informar a los resumidores bajo qué condiciones han de incluir cada categoría. Dichas condiciones se pueden especificar de muy diversas maneras.

Para la selección de categorías, por ejemplo, tests o cuestionarios, se puede requerir que siempre que se mencionen en un documento se informe de ello en el resumen, incluso si sólo se les menciona de pasada o si el autor no los hace aparecer como significativos. Para otras categorías, los directores pueden requerir que sean relevantes o significativas en el documento, o que se las trate con detalle. Utilizando un enfoque cuantitativo, se puede diseñar una medida indirecta de la significatividad y estipular, por ejemplo, que en los resúmenes se especifiquen los nombres de las compañías, a no ser que en el documento se mencionen más de diez compañías.

El lenguaje de los resúmenes

Una vez que se ha determinado los conceptos que se van a incluir en un resumen, se deben dar instrucciones acerca del uso del lenguaje. Más específicamente, los directores deben indicar qué términos se deberían seleccionar para representar estos conceptos y en qué forma se han de representar. Aquí de nuevo, los directores coinciden en la naturaleza general de los términos que no se deben utilizar; las prácticas para elaborar resúmenes incluyen enunciados que recomiendan que se evite el uso de lenguaje enrevesado o de frases que dieran como resultado referencias erróneas cuando se busque el documento. Parece, sin embargo, que semejantes enunciados tan vagamente definidos no son de gran ayuda a los resumidores, y los directores deben ser más concretos acerca de la selección de términos y la forma de las palabras usadas. Llegado este momento, las instrucciones parecen contradictorias, cada una apoyada por argumentos plausibles pero no probados. Lo que es más, algunas directrices violan un gran número de reglas en cuanto a la claridad del estilo.

Un ejemplo de dichas instrucciones es la cuestión sobre si los resúmenes deberían o no utilizar el lenguaje del autor. Aunque ésta es cuestión que concierne a las prácticas para elaborar resúmenes en publicaciones impresas, parece ser especialmente significativo en la recuperación en texto libre. La razón fundamental para el uso del lenguaje del autor estriba en que los usuarios suelen recordar frases-clave utilizadas por los autores, y las incorporan a sus estrategias de búsqueda. El argumento contrario se basa en que el uso de citas directas no es positivo, ya que no suelen ofrecer suficiente información. Como postura intermedia, un director puede requerir que los resúmenes incluyan tanto el lenguaje del autor como sinónimos. Aunque los sinónimos representan una protección para no perder un concepto en la recuperación, se debe tener en cuenta que su uso afecta a la calidad del estilo. Los lectores pueden encontrar confuso que un concepto, que suele ser significativo, se vea representado por un

número de términos diferentes dentro del mismo párrafo.

Cuando a un resumidor no se le pide que se atenga al lenguaje del autor, la recomendación más frecuente es el uso de términos concretos, normalizados, y de un lenguaje que sea específico de la materia. No se encuentran argumentos en contra de este enfoque en las prácticas para elaborar resúmenes, pero los directores evidentemente no se ponen de acuerdo acerca de qué instrucciones específicas pueden conseguir una uniformidad en la asignación de términos en los resúmenes.

Una vez más, es importante decidir si se deben o no coordinar los términos de los resúmenes con los descriptores. En las prácticas para elaborar resúmenes existen ambos enfoques, pero aun no se ha decidido en qué manera se deberían coordinar éstos términos.

Los directores que desean conseguir una consistencia en el uso de los términos de los resúmenes exigen que se utilicen términos del tesoro, siempre que sean aplicables. Por el contrario, aquellos que creen que la recuperación se puede mejorar si se ofrece una variedad de puntos de acceso, recomiendan que se evite el uso de los mismos términos en la indización y el resumen. A los resumidores se les ordena, por tanto, que complementen verdaderamente los términos de indización y usen sinónimos o términos más concretos. El uso de sinónimos, sin embargo, puede empobrecer la calidad del estilo del estilo de escritura, como se ha explicado anteriormente.

Como término medio entre estos dos enfoques, los directores pueden elaborar una lista de categorías seleccionadas para las cuales los resumidores deben utilizar términos concretos, normalizados o bien aceptados. Ejemplos de dichas listas de categorías son: (a) materiales, formas, procesos, condiciones, propiedades y productos; (b) nombres de los tests, nombres de los programas, instituciones y procesos intelectuales; o (c) nombres de compañías subsidiarias. Aunque no se encontró en las prácticas para elaborar resúmenes ninguna lista de autoridades especialmente elaborada para términos utiliza-

dos en resúmenes, algunos productores de bases de datos prefieren elaborar sistemas detallados para el uso de los términos que han de designar las categorías relevantes. Un típico ejemplo de esta costumbre es el requerimiento de que los nombres latinos además de los nombres comunes o vernáculos sean explícitamente mencionados. Otro ejemplo lo constituyen una serie de reglas que indican cómo registrar las localizaciones geográficas.

Existen dos prácticas típicas del uso del lenguaje para la recuperación en texto libre que son consideradas negativas, pero que muchos directores pasan por alto. La primera es: evitar el uso de la forma negativa con el fin de minimizar o eliminar una recuperación falsa. Supongamos, por ejemplo que, un autor nos describe un test para evaluar a los distribuidores de sistemas de búsqueda en línea. En un resumen, una frase como: "evaluar todos los sistemas de búsqueda, excepto DIALOG, por el tiempo de respuesta", recuperará ese resumen cuando se desee recuperar documentos acerca de DIALOG. Es más el resumen será rechazado en una búsqueda acerca de tiempo de respuesta en BRS o en ORBIT. La segunda práctica es: evitar incluir en la lista términos que tienen una palabra común como una serie, con el término común escrito sólo una vez. Por ejemplo: escribir "primary care, secondary care and tertiary care" es mejor que "primary, secondary and tertiary care". Se debe, sin embargo, tener cuidado: mientras que el primer requerimiento puede afectar a la precisión de los resúmenes, el segundo se debería sopesar con la brevedad y claridad.

Por último, los directores pueden decidir la forma de las palabras que se han de utilizar. Mientras que se suele estar de acuerdo sobre la consistencia de la forma de las palabras, los modos de garantizar dicha consistencia son diversos. Un requisito normal es el de seguir una práctica de lenguaje local, como el cambiar la ortografía británica por la americana en las bases de datos utilizadas principalmente por americanos, o viceversa. Otra preocupación normal es el uso de abreviaturas y siglas. Los directores pueden exigir que ciertas categorías, como nombres geográficos, metales, elementos o

procesos, sean deletreadas. O pueden exigir que las iniciales de los nombres propios sean presentadas de modo diferente a las iniciales que pertenezcan a organizaciones. Se pueden utilizar descriptores que procuren esa consistencia en las formas de las palabras cuando los términos utilizados también como descriptores aparezcan en un resumen en la misma forma en que lo hacen en un vocabulario controlado.

Sin embargo, los directores no siempre consideran que la consistencia de las formas de la palabras sea positiva. Algunos mantienen que es preferible expresar un nombre tanto en su forma completa como en su forma abreviada, y que se deben incluir también los apodos. De este modo un documento puede ser recuperado bajo cualquier forma por el usuario.

Sumario

Existen dos cuestiones tratadas por los directores de prácticas para elaborar resúmenes respecto a la búsqueda en texto libre: la primera, qué conceptos se deben incluir en un resumen, y segunda qué lenguaje se debe usar para representar dichos conceptos.

Los conceptos que se han de incluir en los resúmenes par mejorar la recuperación, se pueden seleccionar en coordinación con descriptores asignados o se pueden añadir, independientemente de cualquier otro tipo de indización. Cuando los directores se deciden por coordinar los conceptos de los resúmenes con descriptores asignados, pueden dar a los resumidores instrucciones claras a cerca de si estos conceptos deberían complementar a los descriptores o duplicarlos. Si se desean ambas cosas, se debe especificar cuándo se ha de aplicar un enfoque u otro. Sin hacer caso de si la selección de conceptos para resúmenes está coordinada con la indización, puede ser útil para los directores el ofrecer a los resumidores una lista de categorías o elementos concretos que se han de incluir en un resumen. También puede ser ventajoso nombrar las condiciones bajo las cuales se debe incluir cada categoría.

El lenguaje de los resúmenes puede o no ser fiel al lenguaje de los autores. Si no está sujeto al documento original, los directores deben decidir si coordinar o no los términos de los resúmenes con un lenguaje de indización. Cuando se elige la coordinación, los editores deben decidir si los términos elegidos para un resumen deben duplicar o complementar los términos de indización. En algunas bases de datos, los directores pueden seleccionar categorías para las cuales se desea una forma concreta. Aunque se pueden excluir ciertos usos del lenguaje, sólo la investigación futura indicará si se debe requerir la consistencia o la variedad en las formas de las palabras.

Aunque la investigación de los resúmenes no se ha pronunciado acerca de la aceptabilidad de los resúmenes en la recuperación en texto libre, los directores de normas para elaborar resúmenes ofrecen instrucciones especiales que de modo intuitivo están basadas en la experiencia en la búsqueda en línea obtenida por los profesionales de la información en los últimos años. El desacuerdo entre los directores sobre un número de cuestiones, nos lleva a la necesidad de una investigación que evalúe el efecto de las pautas para elaborar resúmenes sobre el ejercicio de la recuperación.