



Elementos sobre indización y búsqueda de la información por medio de vocabularios controlados en bases de datos biomédicas

Lic. Rubén Cañedo Andalia ^I

Lic. Maria Caridad Small Chapman ^{II}

^I Licenciado en Información Científico-Técnica y Bibliotecología. Departamento Fuentes y Servicios de Información. Centro Nacional de Información de Ciencias Médicas-Infomed.

^{II} Licenciada en Tecnología de la Salud en Gestión de la Información en Salud. Facultad de Tecnología de la Salud. Universidad de Ciencias Médicas de Holguín.



Fecha de actualización

14 de abril de 2011.

Correspondencia

Lic. *Rubén Cañedo Andalia*. Grupo de Alfabetización Informativa. Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas Holguín. Ave. V. I. Lenin No. 4 e/ Aguilera y Agramonte. CP. 80100. Holguín. Cuba. Correo electrónico: ruben@infomed.sld.cu

Derechos de copia



Copyright: © Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas Holguín. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 3.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original. Las imágenes del cabezal de la portada se tomaron de: <http://decs.bvs.br/E/homepagee.htm> y <http://preview.ncbi.nlm.nih.gov/mesh>.

Contenido

Desde las perspectivas de la alfabetización informacional y la búsqueda de la información, existe la imperiosa necesidad de conocer algunos elementos básicos sobre los procesos de indización y recuperación de la información antes de introducir a los alumnos en la búsqueda de la información en bases de datos que ofrecen la posibilidad de explorar sus colecciones mediante vocabularios controlados. Se define la indización y la búsqueda de información; se exponen sus elementos, nociones e importancia; y se estudian los tesauros MeSH y su versión en español, el DeCS. El conocimiento de algunos elementos teóricos y prácticos sobre los procesos de indización y recuperación de la información es un escalón importante en la correcta comprensión de la actividad de búsqueda en bases de datos bibliográficas que utilizan vocabularios controlados.

Introducción

Los lenguajes documentales –también llamados de búsqueda informativa- existen desde la creación de las primeras bibliotecas, porque ellos surgen precisamente cuando el volumen de documentos depositados en un lugar es tan alto que impide o hace impracticable la localización de uno o varios de ellos. Es entonces imprescindible su organización para permitir su localización en el momento oportuno.

Los primeros lenguajes, aparecidos en los orígenes de la biblioteca, fueron sistemas rudimentarios de clasificación.

Los lenguajes documentales son vocabularios que se utilizan, tanto para organizar los documentos que ingresan a las instituciones de información o sus solicitudes, así como para representar sus contenidos. Son, por tanto, herramientas de organización y de recuperación.

Durante el último cuarto del siglo XIX y a lo largo del siglo XX, los sistemas de clasificación bibliográfica y listas de encabezamientos de materia experimentaron una consolidación progresiva en la práctica de la recuperación de la información. A comienzos del siglo XXI, al parecer, ocurrirán transformaciones sustanciales en relación con dichos lenguajes y nuevas concepciones apuntan hacia los fundamentos teóricos de la indización, la recuperación y el empleo en general de los lenguajes para la búsqueda de la información. Tras algunos años de relativa calma, el debate entorno a las disyuntivas "*indización controlada versus indización libre*" y "*vocabulario controlado versus vocabulario libre*" ha vuelto a exacerbarse.

En un contexto más particular, relacionado con la alfabetización informacional en salud en Cuba, la experiencia indica que se requiere de un conocimiento previo general sobre los procesos de indización y recuperación de la información antes de introducir a los alumnos en la búsqueda de información en bases de datos que ofrecen la posibilidad de explorar sus colecciones mediante vocabularios controlados. Por ello a continuación, en apretada síntesis, se intentará bosquejar algunos de los elementos básicos necesarios para la comprensión de la búsqueda bibliográfica en bases de datos como *PubMed-Medline* y *Scopus*.

La indización

La indización^a busca la identificación y asignación de rubros, descriptores o encabezamientos de materia representativos, en esencia, de los contenidos^b principales y secundarios, tratados en un documento con el objetivo de recuperar información sobre un tema determinado.¹ Incrementa la relevancia de los resultados de las búsquedas de información en base de datos a partir de la representación precisa del contenido del documento y las solicitudes de información de los usuarios; se enfoca hacia una recuperación temática específica.

Precisamente, debido a su ostensible dualidad, se estudia de dos maneras:¹

- Como proceso. Consiste en representar el contenido -y las características- de un documento con la ayuda de los conceptos y aspectos considerados en un vocabulario controlado.
- Como finalidad. Permite la búsqueda eficaz de la información contenida en un fondo documental. La indización es un requisito importante para una adecuada recuperación de la información.

La indización es un proceso que implica dos fases:

- 1) La identificación de los contenidos tratados en el documento objeto de interés.
- 2) La conversión de las palabras y frases utilizadas en el lenguaje natural para representar una realidad en términos de un lenguaje controlado, que, como dijimos, pretende elevar la calidad de la recuperación de la información.

Entre los factores determinantes del éxito en la recuperación de información se encuentra el dominio del usuario del vocabulario empleado para representar los documentos de interés y las solicitudes de búsqueda.

Como producto, es una enumeración sucesiva de los diferentes encabezamientos (términos) que expresan el (los) tema(s) contenido(s) en un documento o solicitud. Requiere de la aplicación de criterios uniformes para su ejecución y del

^a Se refiere a la indización por medio de vocabularios controlados. Existe un tipo de indización llamada libre que se realiza a partir de la identificación de las palabras significativas en el texto de un documento.

^b Puede incluir aspectos relacionados con la forma del documento, el tipo de artículo del que se trata u otros aspectos.

establecimiento previo de una lista de términos autorizados en la cual se basa el proceso. Dicho proceso se identifica, de igual modo, con otras denominaciones como asignación de epígrafes y clasificación alfabética por materias.

La indización es una de las etapas del procesamiento analítico sintético de la información (organización), que se ubica por su complejidad intelectual entre la clasificación y la confección de resúmenes. El producto final de este proceso es generalmente un índice bibliográfico, una base de datos automatizada o simplemente un catálogo alfabético de materias manual, indispensables para asegurar el acceso y consulta de la información a los usuarios. Esta aseveración significa que se toman en cuenta los mismos criterios para la ejecución. Ellos son exhaustividad, multiplicidad, especificidad, coherencia, imparcialidad, fidelidad y buen juicio.

Atributos esenciales

En la indización, la especificidad y la exhaustividad constituyen los dos atributos cardinales. Estos son esenciales para determinar la clase de recuperación que se realizará.

La exhaustividad se define por el grado de cobertura que presentan los tópicos o asuntos tratados en los documentos o solicitudes en los resultados de la indización, es decir, en el producto. Cuanto mayor sea la representación de los tópicos expuestos en un documento o solicitud en su indización, mayor será la exhaustividad del proceso. Se expresa cuantitativamente por el número de términos utilizados para realizar la representación. Comprende tópicos centrales y colaterales (principales y secundarios).

Por su parte, la especificidad es el nivel de detalle y exactitud con que se representa cada tópico tratado en un documento o solicitud en la indización. La recuperación de información de manera eficiente requiere de la determinación de los términos más específicos que representan los conceptos presentes en un documento o solicitud. La representación de un concepto particular mediante un término, cuyo significado es más general que el concepto tratado en la entidad que se indiza, conduce a una pérdida de especificidad y por consiguiente de información.

Los tesauros

El tesoro es un tipo de vocabulario controlado, compuesto esencialmente por una lista de términos autorizados, que se utilizan para indizar (representar semánticamente) los registros que ingresan a una base de datos o sistema de información, así como las solicitudes de búsqueda que formulan sus usuarios. Es el medio más eficaz para realizar una búsqueda por materia. Sin embargo, los tesauros, como es obvio, son incapaces de reunir el total los conceptos tratados en la literatura especializada. Por esta razón, los índices y los estudios de la frecuencia de empleo de las palabras y frases relacionadas con un tema de interés son fundamentales. Tesoro e índice son herramientas complementarias. De hecho es frecuente el empleo en una búsqueda de vocablos controlados y de palabras y frases del lenguaje natural, o vocablos libres.

Las bases de datos disponen además de un índice alfabético ascendente del total de palabras y frases significativas que contienen sus registros. Para cada palabra, nombre, abreviatura, siglas y otros. El índice relaciona a menudo el número de veces en que éste aparece y el número de registros que le corresponden.

El índice es útil para:

- 1) Comprobar la ortografía de los términos, conocer sus sinónimos o verificar si la palabra tecleada aparece en la base de datos.
- 2) Elegir las entradas más apropiadas y eliminar aquellas por las cuales no se desea buscar.
- 3) Hallar las palabras que con mayor especificidad representan el contenido de una solicitud.
- 4) Delimitar la búsqueda a los registros de un autor determinado; en ocasiones, en una base de datos se consigna un mismo autor de diferentes maneras (Lowe, HJ o Lowe, Henry J.); al conocerlas todas es posible maximizar los resultados de la búsqueda, como es obvio esto es válido para la exploración por cualquiera de los campos de la base.
- 5) Combinar de manera automática dos términos relacionados con la búsqueda que se seleccione y crear con ellos una prescripción de búsqueda basada en operador lógico.
- 6) Usar varios términos de búsqueda con una misma raíz, pues al inspeccionarse los que presentan una raíz común, se pueden elegir todos los que sean relevantes

(y obviar los que no lo son, pero poseen la misma raíz), con lo cual se evita la elección de términos no deseados.

El MeSH y el DeCS

En el área de la salud, uno de los tesauros más empleados es el MeSH (*Medical Subject Headings* - <http://preview.ncbi.nlm.nih.gov/mesh>), un vocabulario controlado y jerárquico de encabezamientos de materias^c (*Headings*), elaborado por la Biblioteca Nacional de Medicina de los Estados Unidos. Este a su vez, presenta una versión en español, el DeCS (Descriptores en Ciencias de la Salud - <http://decs.bvs.br/E/homepagee.htm>), útil para aquellos usuarios que no dominan la terminología médica en inglés. El DeCS está muy difundido en Latinoamérica y el Caribe. Ambos son totalmente compatibles. Y posibilitan:

- La indización de los conceptos y los aspectos fundamentales expuestos en un documento o solicitud.
- El empleo de un lenguaje consistente para indizar las materias; por ello, no será necesario realizar búsquedas mediante distintos sinónimos cuando se pretenden recuperar materiales sobre un mismo tema (por ejemplo, *coriza aguda - resfriado común*).
- La revisión en el tesoro, si se desean localizar términos adicionales relacionados con la temática seleccionada, que pueden ayudar a tornar más específica la búsqueda o perfilarla.

El MeSH y el DeCS sugieren relaciones entre los términos a partir de una estructura jerárquica de términos genéricos y específicos, que enlaza términos no autorizados y autorizados, así como los relacionados entre ellos. Los tesauros, al igual que el índice, ayudan a la localización de los términos más específicos. Sus referencias cruzadas remiten de los vocablos no autorizados –muchas veces sinónimos o subordinados- a los términos autorizados. Los tesauros referidos permiten recorrer una jerarquía de términos en dos direcciones, con vistas a delimitar o ampliar los resultados de la búsqueda.

A continuación, sobre la base del criterio que tanto el conocimiento de los términos del MeSH y el DeCS como de su estructura y organización es determinante para la ejecución eficiente de las búsquedas, se estudiarán una serie de detalles útiles para su empleo.

^c Un encabezamiento de materia es un término controlado similar a un descriptor. La diferencia entre ellos radica en su amplitud semántica. Un descriptor con frecuencia es un término más específico; y por ello, comprende un contenido menor que el que abarca un encabezamiento de materia. A pesar de esto, en muchas ocasiones, es imposible determinar si un término controlado pertenece a una u otra categoría.

Los tesauros mencionados disponen básicamente sus términos en 3 formas de organización diferentes: alfabética, jerárquica y permutada. La relación alfabética organiza sus elementos en forma ascendente; estos elementos son los descriptores principales y secundarios, subencabezamientos (o calificadores) y referencias cruzadas. Por su parte, en la estructura jerárquica se organizan los términos desde los más genéricos hasta los más específicos (de la A a la Z) y comprende términos y calificadores. La totalidad de los encabezamientos se distribuye en 20 categorías principales que responden a una compleja jerarquía de las especialidades o ramas del saber en el área de la salud (fig. 1). También se les denomina árboles estructurales.

Fig. 1. Estructura jerárquica principal.



Finalmente, la lista permutada contiene una relación que abarca, tanto los términos no autorizados como los autorizados y lleva de los primeros a los segundos.

Estos tesauros contienen referencias cruzadas que relacionan a los términos con sus sinónimos o cuasi-sinónimos y ofrecen una especie de definición de cada término denominadas notas de alcance.

Adicionalmente, los tesauros poseen índices permutados de casi todos los términos donde se excluyen los subencabezamientos, cuya función es ayudar a la

identificación del término deseado, cuando éste no se recuerda o conoce con exactitud. Estos índices contienen una relación alfabética de todas las palabras significativas que son similares o se relacionan con el término.

A continuación se explican algunos de los términos componentes de estos vocabularios. Ellos son:

1) Descriptor principal.

Es cualquier término del *MeSH* o el *DeCS* que se acepta directamente y sin conversión alguna a otro término y bajo el cual se almacenan las referencias bibliográficas. Aparece destacado tipográficamente en la lista alfabética. Entre los descriptores principales se encuentran algunos denominados especiales, los cuales confieren gran especificidad a la indización y recuperación. Estos tipos especiales de términos nunca designan un concepto desde el punto de vista temático, aunque sí se pueden utilizar para efectuar búsquedas. Ellos son:¹

- *Tipos de publicación.* Estos términos caracterizan el tipo de publicación, en lugar de reflejar el tema.
- *Encabezamiento precodificado.* Es un concepto que siempre debe considerarse al indizar cualquier documento o solicitud. Este grupo de términos designa atributos muy generales de su contenido. Por ejemplo: puede tratarse de un tipo de animal (perros), una condición (embarazo), el sexo (femenino), un grupo de edad (niños), entre otros. Estos términos resultan de gran utilidad, porque contribuyen a restringir el resultado de las búsquedas a los documentos más específicos. Por ejemplo: si a una búsqueda sobre el "tratamiento medicamentoso del asma", se adiciona el término precodificado *niño*, su resultado disminuye considerablemente.²
- *Encabezamiento geográfico.* Es un término que identifica regiones geográficas particulares, continentes, países, estados y ciudades seleccionadas. Por ejemplo: en Estados Unidos.
- *Término No-MeSH.* Es un descriptor no disponible para la indización de documentos pero que puede utilizarse en la prescripción de búsquedas.

2) Calificador o subencabezamiento.

Es una palabra o frase con la cual se califica un término principal. Comprende aspectos generales frecuentemente utilizados y describe con mayor precisión los temas principales que discute el autor. Como proporciona un nivel adicional de especificidad, permite limitar la recuperación a referencias bibliográficas que traten sobre un aspecto específico de un concepto biomédico, por ejemplo: diagnóstico, tratamiento, cirugía, estadísticas, clasificación, etcétera.

El calificador se utiliza combinado con el descriptor. Cuando sea posible, es preferible aplicar una combinación encabezamiento/subencabezamiento en lugar de una combinación encabezamiento/encabezamiento para recuperar referencias bibliográficas sobre un aspecto específico de un tema. Es válido subrayar que no se permiten todas las combinaciones descriptor/calificador, porque éstas se rigen por ciertas reglas relacionadas con la categoría permisible a cada término.

Se acostumbra a preceder a los calificadores por una diagonal. Entre ellos, es posible citar a modo de ilustración: */epidemiología* o */química*. El empleo adecuado de los subencabezamientos, aumenta el grado de especificidad de las búsquedas.

3) Estructuras jerárquicas.

Las estructuras jerárquicas constituyen otro de los recursos disponibles en los lenguajes controlados para lograr especificidad en la búsqueda. Este tipo de organización de los términos posibilita la formulación de solicitudes de búsqueda, amplias o específicas, porque permite la revisión ascendente o descendente una estructura jerárquica hasta localizar el término que nos resulta más representativo del concepto o noción de nuestro interés.

Con el empleo de la estructura jerárquica para hallar el término más específico, se logra mejorar la precisión de la búsqueda, disminuir el número de referencias recuperadas no relevantes. Igualmente, permite ampliar el alcance de la búsqueda y por ende mejorar el recobrado, porque gracias a ella es posible encontrar el término más general.

Es oportuno destacar que los árboles estructurales (estructuras jerárquicas) son valiosos auxiliares en la modificación de la especificidad de la búsqueda

y aportan un método para localizar determinado término a partir, únicamente del conocimiento de su área conceptual general.

Los elementos disponibles en el *MesH* o el *DeCS*, aseguran la especificidad en el proceso de recuperación de información.

4) Referencias cruzadas.

Son términos utilizados para señalar los encabezamientos autorizados establecidos previamente y sirven para remitir de un término a otro: *Término de entrada - Término autorizado*.

5) Notas de alcance.

Tanto los calificadores como la mayoría de los términos de estos vocabularios contienen una información para definirlos denominada *Notas*. En el caso de los calificadores sólo presentan notas de alcance, una especie de definición del término que ayuda a seleccionar el subencabezamiento adecuado.

En los encabezamientos además de las notas de alcance, pueden aparecer notas para los indizadores y personas que realizan búsquedas, notas históricas y notas para efectuar búsquedas automatizadas. Todo este conjunto de datos ofrece una información amplia y muy útil para propiciar tanto la indización como la recuperación.

Procedimiento para la indización con un vocabulario controlado

- Identificación de los conceptos fundamentales.
Se debe tratar de comprender desde un inicio el alcance, intención, contenido y significación de cada tema tratado en el documento. No se debe iniciar la asignación de los vocablos controlados sin entender previamente en qué consiste el estudio o la solicitud de información, cuáles son sus propósitos y su importancia.
- Identificación de los aspectos fundamentales que se tratan de los conceptos identificados.
- Conversión de los conceptos a vocablos controlados.

Es imprescindible identificar el término controlado que más se adecua al concepto tratado, el que con mayor precisión indica el contenido del documento o la solicitud. Se indizarán sólo los temas discutidos. Debe diferenciarse entre vocablos controlados principales y secundarios. Se utilizarán tantos términos autorizados como sean necesarios para representar el contenido del documento o solicitud. Para la indización se emplearán las herramientas complementarias disponibles para estos propósitos: permutados, notas de alcance para indizadores, tablas de especificaciones para calificadores, leyendas de abreviaturas y símbolos, etcétera. Debe procurarse identificar el descriptor pre-coordinado si existe, en lugar de utilizar una combinación descriptor/calificador. Por ejemplo, *Accidentes/prevenición*, existe *Prevención de accidentes*. En caso de que nos sea imposible comprender sobre qué trata el documento o cómo indizarlo, deben consultarse expertos en el tema. La indización en todo momento debe orientarse a la recuperación.

- Conversión de los aspectos a calificadores.

Deben asignarse sólo los tratados en el documento. Asimismo, deben diferenciarse en principales y secundarios, como sucede con los términos controlados. Debe evitarse el uso de más de 3 calificadores para un mismo descriptor; si es posible agruparlos, por ejemplo, *radiografía, ultrasonografía y cintigrafía*, puede emplearse sólo diagnóstico.

- Determinación de validez de los calificadores para los vocablos controlados utilizados.

Cada término controlado acepta una cantidad limitada de calificadores. Por ello, cuando se asigna un calificador, primero debemos asegurarnos de que es válido para ese término. Para determinar cuáles son los autorizados, deben utilizarse las herramientas complementarias disponibles para las labores de indización. Éstas, insistimos, son las indicadas para poder determinar con exactitud cuáles son los calificadores autorizados en cada caso.

- Registro de los descriptores u otros términos controlados, seleccionados para representar el contenido – en caso necesario también su forma- del material procesado.

Búsqueda de información y búsqueda bibliográfica

La búsqueda de información es una secuencia ordenada de pasos, ejecutados con la finalidad de localizar los documentos que contienen cierta información o de entregar datos/información concreta que responde a determinada pregunta. La búsqueda bibliográfica se circunscribe a la identificación, selección, ordenamiento y entrega de las referencias de los documentos y sus enlaces, si están disponibles, al texto completo de ellos.

La búsqueda bibliográfica es una etapa condicionante de la investigación como proceso central de la actividad científica. Puede dividirse en tres etapas principales: planificación, ejecución y evaluación.

- Planificación

En esta fase deben identificarse los conceptos, aspectos y límites que comprenden cada necesidad de información; así como expresar estos de manera que el sistema de recuperación los entienda. Para poder realizar esta conversión es esencial el conocimiento, tanto de los requisitos semánticos y formales de la solicitud de información como de las características y facilidades que presentan las herramientas lingüísticas, la interfaz de recuperación -en especial, la avanzada-, así como la base de datos que nos proponemos consultar.

La planificación de la búsqueda es una fase esencial, previa a su ejecución y de la que depende en gran parte, la evaluación de sus resultados. Sin embargo con frecuencia recibe una escasa atención entre los investigadores como consecuencia de la existencia de una inadecuada concepción sobre la búsqueda bibliográfica que le niega su carácter de investigación documental, con exigencias nada inferiores a los de la investigación histórica, social u otras. Esta insuficiencia se convierte entonces en una de las causas principales de sus pobres resultados de investigación.

El conocimiento de las características de las bases de datos como son su cobertura temática, documental, geográfica e idiomática; la retrospectividad de la colección y su período de actualización^d entre otros; así como de las facilidades que ofrece su sistema de recuperación: vocabularios controlados

^d Es el período que transcurre desde el momento de la publicación de la contribución hasta la aparición de su referencia en la base de datos.

(tesauros, encabezamientos de materias); operadores (anexo 1); opciones para restringir la recuperación de resultados (límites); posibilidades para el agrupamiento de los términos (paréntesis o corchetes) y para el truncado; estructura de los campos y otras; son imprescindibles para formular una estrategia de búsqueda moderadamente adecuada. En dependencia de las posibilidades que ofrece el sistema, la estrategia tomará una forma u otra y será necesario trabajar más o menos para conseguir los resultados deseados.

En la fase de planificación se formulan las necesidades referidas en los términos y formas que comprende el sistema de recuperación de la información; a esta formulación se denomina indistintamente: prescripción, condición o estrategia de búsqueda. La estrategia propiamente dicha abarca el contenido o temática de la necesidad, así como algunos de sus aspectos formales como son: el idioma, las fechas de publicación, los tipos de documentos, etcétera.

- Ejecución

Durante la fase de ejecución el usuario introduce su estrategia de búsqueda, observa los resultados y realiza los ajustes necesarios a su estrategia inicial con vistas a obtener una mejor correspondencia entre los resultados que entrega el sistema y sus necesidades de información.

- Evaluación

Para la evaluación de los resultados de la utilización de una estrategia de búsqueda, generalmente se emplean diversos puntos de vista y conceptos como son: relevancia, pertinencia, recobrado y precisión. La primera se mide por el número de documentos recuperados cuyo contenido responde a la estrategia formulada; la pertinencia depende más de la capacidad de quien interactúa con el sistema para desarrollar una prescripción de búsqueda eficaz y se evalúa según el grado de correspondencia que existe entre los documentos recuperados y la necesidad de información del usuario.

El recobrado (también llamado sensibilidad), por su parte se refiere a la capacidad de la estrategia de búsqueda para recuperar la mayor parte posible de los documentos existentes en la base de datos, relacionados con el tema objeto de búsqueda. Finalmente la precisión (también denominada especificidad) se refiere a la capacidad de la estrategia para discernir entre

los documentos existentes en la base de datos objeto de exploración, aquellos que mejor responden a la prescripción de búsqueda.

La evaluación de los resultados de la búsqueda depende en gran medida de la intención de quien la realiza; si el interés es recuperar el mayor número de referencias posible entonces un alto recobrado será un indicador apropiado para medir la calidad de la exploración; si lo que se requiere es que los resultados presenten la mayor correspondencia semántica posible con la prescripción de búsqueda entonces será el nivel de precisión de los resultados el que medirá su calidad como proceso de búsqueda y así sucesivamente. En general estas nociones son difíciles de determinar debido a los volúmenes de información que es necesario manejar pero sobre todo a causa del grado de subjetividad que encierran.

Sin embargo con frecuencia la necesidad de información como tal, no se satisface sólo como resultado de la utilización de una estrategia de búsqueda eficaz en la recuperación de la información en uno o varios recursos y requiere a menudo del diseño de un servicio personalizado que considere además de los aspectos relativos a la semántica de la necesidad, los relacionados con la estructura y requerimientos de la actividad o actividades profesionales que desempeña el usuario y en el seno de las cuales se originan las necesidades de información que el servicio pretende satisfacer; las condiciones en que se realizan éstas, así como las características socio-psicológicas y culturales de quienes se propone servir.³

Consejos útiles para aumentar la precisión de los resultados de una búsqueda

Existen diversas alternativas para aumentar la precisión o especificidad^e de los resultados de una búsqueda de información, una vez seleccionada la base de datos más apropiada^f para realizar la búsqueda, entre ellas:

- Mediante un mejor manejo de la terminología.

^e Los aspectos tratados en el acápite titulado *Procedimiento para la indización con un vocabulario controlado* son útiles para esta sección del documento.

^f En esta selección debe atenderse especialmente al tipo de base de dato (bibliográfica, factográfica, textual u otra); su accesibilidad (libre o por suscripción); su cobertura (temática, documental, geográfica, idiomática, etc.); así como la consistencia, adecuación y facilidades que ofrece su interfaz de búsqueda.

En todo momento deben utilizarse las palabras y frases claves o términos controlados, que con mayor exactitud representen el contenido del tema de la búsqueda. Para esto se debe poseer una clara comprensión del alcance, intención, contenido y significación del tema objeto de interés. Por ejemplo: si se desea recuperar trabajos que traten sobre tumores malignos, deberán emplearse términos como *cáncer* que indica malignidad en lugar de *tumor* y *neoplasia* que no necesariamente señalan esta condición.

De ser posible se deben utilizar vocablos controlados en lugar de palabras y frases claves del lenguaje natural. Cuando los términos del vocabulario controlado no presentan la especificidad requerida para la búsqueda, la combinación de términos libres y controlados es muy útil para aumentar la especificidad de los resultados recuperados.

De manera general la búsqueda por términos controlados supera a la búsqueda por palabras y frases claves del lenguaje natural.

- Mediante una mejor explotación de la búsqueda por campos.

Circunscribir la búsqueda de las palabras^g y frases clave o términos controlados a campos específicos del registro bibliográfico incrementa la precisión de los resultados de la búsqueda. La exploración mediante palabras y frases claves requiere, como se advirtió antes, de la identificación cuidadosa del total de las equivalencias o al menos de las principales, utilizadas para representar una entidad o condición. Éstas pueden incluir no sólo sinónimos sino también siglas^h y hasta préstamos lingüísticos. Restringir la búsqueda de los términos controlados al campo de descriptores principales, cuando las bases de datos lo posibilitanⁱ, eleva sustancialmente la precisión de los resultados. La búsqueda libre si se utiliza con habilidad la estructura del registro, también puede ofrecer resultados precisos. La especificidad de los resultados según campos, crece desde el texto completo de la referencia al campo del resumen; de este al campo de palabras claves si existe; y alcanza su mayor valor en el campo título. Una búsqueda por medio de palabras y frases claves en el campo de título, es el mejor sustituto de una búsqueda por vocablos controlados cuando el sistema no permite su realización.

^g El campo de palabras claves de algunas bases de datos, como *Scopus*, abarca también los términos de indización del registro, es decir, los términos controlados. En este caso, *palabras* puede hacer referencia al nombre de un autor, de una revista, de una sustancia u otros. Buscar por ejemplo, el nombre de un autor o de una sustancia en el texto completo de un registro puede llevarnos a resultados irrelevantes como la biografía o el obituario de una persona llamada igual al autor que buscamos o a que el documento recuperado sólo mencione la sustancia pero no trate sobre su empleo.

^h Recuérdese el ejemplo sobre las ataxias.

ⁱ En *PubMed-Medline*, [MJ], principales, y [MN], secundarios. En *Scopus*, no es posible diferenciarlos.

- Mediante un mejor uso de los operadores y otras facilidades para la elaboración de las estrategias de búsqueda.

Los operadores lógicos por ejemplo: facilitan la combinación y exclusión de palabras y frases claves o términos. Los llamados operadores de vecindad, útiles para la búsqueda libre en el texto de la referencia, el documento o el recurso, aumentan o disminuyen la especificidad; generalmente en la medida en que los términos buscados se encuentren más cercanos. Los paréntesis por su parte, ayudan a agrupar y a establecer prioridades entre los distintos subconjuntos que forman una búsqueda compleja. Así el total de las equivalencias de una misma palabra clave pueden reunirse en una secuencia aparte en la condición de búsqueda que, a su vez, puede relacionarse con otros conceptos o aspectos de interés. Con el uso de los paréntesis se debe ser cuidadoso. Los subconjuntos de una prescripción encerrados entre paréntesis se ejecutarán antes que las demás operaciones explicitadas en la estrategia. Su presencia y su ausencia pueden llevar a que condiciones aparentemente similares de búsqueda produzcan resultados diferentes^j. El truncado bien utilizado, permite aumentar el recobrado sin reducir prácticamente la precisión de los resultados.

- Mediante un mejor empleo de los límites y otras facilidades para la búsqueda.

Los límites posibilitan establecer ciertas condiciones a la recuperación temática principal objeto de interés. Establecen una especie de entorno en el cual se debe realizar la exploración. Para la búsqueda libre puede emplearse la opción *Índice* y la facilidad de auto-completamiento de existir.

^j Obsérvese, $4 \times 5 + 3 = 23$ y $4 \times (5 + 3) = 32$.

Referencias bibliográficas

1. Jiménez Miranda J. La indización en el Sistema Nacional de Información de Ciencias Médicas. Parte I. El Medical Subject Headings (MESH). Acimed 2001;9(3). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352001000300002&lng=es&nrm=iso [Consultado: 15 de junio de 2010].
2. Alfonso Sánchez I, Armenteros Vera I. Indización biomédica. La Habana: Editorial Ciencias Médicas; 2008.
3. Fernández Valdés MM, Núñez Paula IA. Metodología para el estudio de las necesidades de información, conocimiento y aprendizaje en las bibliotecas y centros de documentación de salud. Acimed. 2007;15(4). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352007000400004&lng=es&nrm=iso [Consultado: 28 de junio de 2010].

Anexos

Anexo. Operadores y otras facilidades para la elaboración de las estrategias de búsqueda^k

Operadores lógicos

Los operadores^l lógicos posibilitan relacionar dos o más términos^m en una prescripción de búsqueda.

Operador	Acción	Ejemplos y observaciones
AND Espacio +	Permite recuperar las referencias ⁿ de los documentos o los recursos de información que contienen ambos términos en su registro bibliográfico ^o o en el texto completo. En algunos casos, sobre todo en los buscadores, basta con colocar los términos separados sólo por un espacio entre ellos. En ocasiones, se sustituye por el símbolo "+", seguido	Para buscar, por ejemplo, <i>asma en niños</i> pudiera utilizarse: asma AND niños asma niños diabetes +obesity. Precaución: En la medida en que se combine una mayor cantidad de palabras con el operador

^k A causa de la existencia de un amplio solapamiento entre las clases y los tipos de operadores que comprenden, se decidió tratar en este acápite, los operadores más utilizados tanto en la búsqueda de información en bases de datos como en los buscadores, generales y especializados, utilizados para explorar la Web.

^l En el caso de las bases de datos bibliográficas, el alcance de la acción de los operadores se circunscribe al texto de la referencia. En el caso de las bases de datos textuales y buscadores, los operadores pueden ir más allá de la referencia y comprender el texto completo del documento. En caso necesario se indicará el alcance.

^m Consideráramos *términos* aquellas palabras, datos, siglas, frases, o los subconjuntos de estos, agrupados mediante paréntesis, en una prescripción de búsqueda.

ⁿ Se entiende por referencia, en el contexto de las bases de datos bibliográficas y los buscadores, el conjunto de datos que describen y representan un documento o recursos de información. Puede acompañarse de un enlace para su localización en una base de datos particular, un sistema de información o en Internet; así como de otros datos.

^o Es decir, en el total de datos dispuestos en la referencia del documento, que permiten su identificación, descripción, representación y en algunos casos, su ubicación en la red. La búsqueda de los términos puede realizarse en el texto completo del registro bibliográfico o en campos seleccionados, por ejemplo, campo de autor, de título, de resumen, u otro.

	inmediatamente ^a del término de interés y sin dejar espacio alguno entre el signo y la palabra.	lógico AND, se recuperará una menor cantidad de registros.
OR	Es útil para recuperar las referencias de los documentos o recursos que contengan al menos uno de los términos indicados en la condición de búsqueda. A diferencia del AND, el OR no exige que las referencias contengan ambos términos. Pero sucede que los que contengan ambos, también se recuperarán, porque cumplen con la condición mínima de contener uno de los dos.	Si se requiere buscar por medio de una sola prescripción, por ejemplo, documentos que traten por separado o en forma conjunta sobre bronquitis o neumonía, puede emplearse una estrategia como la siguiente: bronquitis OR neumonía
XOR	En la práctica, presenta un uso pobre, restringido sólo a algunos buscadores. Devuelve los documentos que contienen uno de los dos términos, pero excluye aquellos que contengan los dos a la vez.	Por ejemplo: bronquitis XOR neumonía
NOT AND NOT ANDNOT	Al insertar este operador se excluye el término que lo sucede. Este operador puede sustituirse en algunos casos por el símbolo "-", seguido inmediatamente del término de interés y sin dejar espacio alguno entre el signo y la palabra.	Si se busca <i>asma</i> , como en el primer ejemplo, pero no se desea que aparezca tratada en <i>niños</i> , puede escribirse según el caso: asma NOT niños asma AND NOT niños hypertension ANDNOT diabetes diabetes -obesity

		Con esto se recuperarán páginas que contengan sólo el término <i>asma</i> . Se excluyen, por tanto, los que presenten la palabra <i>niños</i> .
--	--	---

Fuente: Elaborado a partir de Oller Gómez J. Elementos teórico-prácticos útiles para comprender el uso de los motores de búsqueda en Internet. *Acimed* 2003;11(6).

Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352003000600007&lng=es

Operadores de adyacencia, proximidad o vecindad

Los operadores de proximidad permiten indicar al sistema de recuperación, la distancia que debe existir entre los términos utilizados para la búsqueda. Se utilizan preferentemente en la búsqueda por medio del lenguaje natural o libre. Generalmente en la medida que dos o más términos se encuentran más próximos entre ellos es mayor la probabilidad de que traten el tema objeto de exploración.

Operador	Acción	Ejemplo
NEAR	Indica cercanía. Se utiliza entre dos términos y establece la distancia máxima a la que pueden encontrarse estos en la referencia o en el texto completo del documento o recurso. Dicha cantidad puede variar. En los casos que conocemos, la cantidad máxima es de 25 palabras, la unidad en que se mide la distancia entre los términos. Pero ello no permite descartar la posibilidad de que se consideren otras	Si se buscan documentos que contengan las palabras <i>asma</i> y <i>niños</i> , y se desea que las dos se encuentren cerca una de la otra en una misma referencia o documento, puede introducirse la siguiente estrategia: asma NEAR niños

	distancias mayores como mínimas.	
NEAR/n	La <i>n</i> indica el número máximo de palabras permitidas entre los términos que se buscan.	<p>Por ejemplo:</p> <p>asma NEAR/5 niños</p> <p>El operador señala que la distancia máxima solicitada entre ambos términos es de hasta 5 palabras.</p>
ADJ	Indica adyacencia. Se usa entre palabras y recupera aquellos documentos o recursos en los que éstas se encuentran juntas sin importar el orden en que éstas se hallen.	<p>Si buscamos documentos que contengan la palabra <i>asma</i> y la palabra <i>niños</i>, pero que las dos palabras se encuentren juntas una al lado de la otra en ese mismo documento sin importar el orden, puede introducirse la siguiente estrategia:</p> <p>asma ADJ niños</p>
FAR	FAR señala lejanía. En los casos que conocemos, la distancia mínima admitida es de 25 palabras entre los términos de búsqueda que combina el operador. Pero ello no permite descartar la posibilidad de que se consideren otras distancias mínimas.	<p>Por ejemplo:</p> <p>asma FAR niños</p>
FAR/n	La <i>n</i> indica el número mínimo de palabras que pueden separar a los	Si se buscan documentos que contengan las palabras <i>asma</i> y <i>niños</i> ,

	términos de búsqueda que combina el operador.	pero que las dos se encuentren alejadas una de la otra en un mismo documento en 33 o más palabras de distancia, se introduciría: asma FAR/33 niños
BEFORE	Funciona como el conocido AND. Su única diferencia es que los términos deben aparecer en el orden especificado en la condición de búsqueda, pero pueden encontrarse a cualquier distancia en el registro bibliográfico o en el texto completo del documento o recurso.	Si se buscan documentos que contengan las palabras <i>asma</i> y <i>niños</i> , pero que la palabra "asma" aparezca antes que niños se introduciría: asma BEFORE niños.
FOLLOWED BY	Significa "Seguido de". Especifica claramente cuál debe ser el orden de los términos buscados. Presenta un uso escaso.	asma FOLLOWED BY niños
O NEAR FAR ADJ	De los operadores expuestos, sólo BEFORE y FOLLOWED BY tienen presente el orden en que deben aparecer los términos. No obstante, se puede añadir esta capacidad de ordenamiento a NEAR, FAR y ADJ, si se les antepone la letra O.	Una estrategia como: cardiopatía OADJ congénita Recuperará los recursos que contengan <i>cardiopatía congénita</i> , pero no <i>congénita cardiopatía</i> . ONEAR y OFAR generarán resultados similares,

		siempre que los términos de la búsqueda aparezcan en el mismo orden en que se escribieron.
PRE/n	PRE/n (precedido por). El primer término precede al segundo en un número de palabras que va desde 0 hasta el número indicado como máximo. El número máximo de palabras que conocemos es de 255.	<p>Por ejemplo, en:</p> <p>behavioural PRE/3 disturbances</p> <p>El número de palabras máximo es 3. Es decir, el sistema devolverá referencias de documentos o recursos donde el término <i>behavioural</i> preceda a <i>disturbances</i> en una distancia no mayor de 3 palabras.</p>
W/	W/ (dentro). Los términos de búsqueda no deben estar a una distancia mayor que la indicada en el número de palabras.	<p>Por ejemplo,</p> <p>pain W/15 morphine.</p> <p>El sistema entregará referencias de documentos o recursos en los que <i>pain</i> y <i>morphine</i> no se encuentren a una distancia mayor de 15 palabras.</p>

Fuente: Elaborado a partir de Oller Gómez J. Elementos teórico-prácticos útiles para comprender el uso de los motores de búsqueda en Internet. *Acimed* 2003; 11(6).

Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352003000600007&lng=es

Operadores de exactitud y truncado

El operador de exactitud por excelencia son las comillas. El truncado puede utilizarse para sustituir un prefijo o un sufijo o caracteres particulares de un término.

Operador	Acción	Ejemplo
" "	<p>Frases literales entre comillas.</p> <p>El uso de frases entre comillas indican al sistema la intersección (AND) de las palabras en la búsqueda, pero, exige además, adyacencia. Es muy parecido al AND pero exige que la secuencia de palabras se encuentre en el mismo orden y sin términos intermedios.</p>	<p>La búsqueda de palabras aisladas como <i>hipertensión, arterial y esencial</i> puede arrojar resultados excesivos e imprecisos. Pero si se encierran entre comillas:</p> <p>"hipertensión arterial esencial"</p> <p>El resultado puede disminuir ostensiblemente, porque se buscará sólo esta secuencia en particular.</p>
*	<p>Truncado de partes de términos. El uso más conocido es a la derecha del término. Pero también es posible al inicio o en el medio del término. Generalmente se utiliza para truncar el asterisco (*), pero algunos buscadores admiten otros símbolos.</p>	<p>Por ejemplo:</p> <p>hipertens*.</p> <p>El sistema devolverá referencias documentos o recursos sobre hipertensión, hipertenso, etcétera.</p> <p>La búsqueda por la raíz común de un grupo de palabras aumenta considerablemente el recobrado.</p>

?	Truncado de caracteres específicos. Es útil para la búsqueda de términos en los que no estamos seguros sobre su ortografía. Hemos encontrado otros signos como \$, % y #, tanto para el truncado de partes como de caracteres.	Por ejemplo: anal?s?s No estamos seguros si se escribe primero con y después con <i>i</i> o viceversa.
---	--	--

Fuente: Elaborado a partir de Oller Gómez J. Elementos teórico-prácticos útiles para comprender el uso de los motores de búsqueda en Internet. Acimed 2003;11(6). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352003000600007&lng=es

Agrupamiento

Es propio de las búsquedas complejas donde se emplean varios operadores lógicos y otras facilidades.

Facilidad	Acción	Ejemplo
()	Los paréntesis se utilizan con mayor frecuencia para agrupar partes de secuencias lógicas y combinaciones de éstas en consultas complejas. Delimitan los subconjuntos de una búsqueda y establecen el orden para ella.	Si se buscan documentos que contengan las palabras "asma y niños", pero no "adultos y ancianos", se introduce: Asma AND niños NOT (adultos OR ancianos)
#	Numeral. Se utiliza sobre todo en las bases de datos bibliográficas para referirse a una búsqueda realizada con anterioridad	Por ejemplo, si se desea combinar las estrategias utilizadas en dos prescripciones previas, pudiera introducirse:

	en el sistema.	# 5 AND #7
--	----------------	------------

Precaución: Los requerimientos para la escritura tanto de los términos como de los operadores suele variar entre bases de datos y buscadores. Con frecuencia estos últimos aceptan la escritura de ambos en letra mayúscula y minúscula. Sin embargo, las bases de datos no se comportan de igual manera. Se recomienda en todo caso la consulta de las ayudas que ofrece cada sistema para verificar la forma correcta. Otro tanto sucede con las comillas, a veces no sólo son necesarias para las frases literales sino también para las palabras o los descriptores compuestos por más de una palabra e incluso para palabras simples.

¿Cómo citar este documento?

Cita (Vancouver): Cañedo Andalia R, Small Chapman MC. Elementos sobre indización y búsqueda de la información por medio de vocabularios controlados en bases de datos biomédicas. En: Cañedo Andalia R, Rodríguez Labrada R, Fernández Valdés MM, Zayas Mujica R, Nodarse Rodríguez M, Sánchez Tarragó N, *et al.* Lecturas avanzadas para la alfabetización informacional en salud. Holguín: Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas de Holguín; 2011. [citado día mes año]. Disponible en: http://www.hlg.sld.cu/sitios/CPICM/index.php?option=com_jdownloads&Itemid=87&view=viewcategory&catid=5