

# ESTRATEGIAS PARA LA RECUPERACIÓN DE LA INFORMACIÓN

William Hersh

# 5

## OBJETIVO

**Ayudar a entender el funcionamiento de los sistemas de búsqueda de información, para que los profesionales sanitarios aprendan a usarlos con destreza para atender una mayor variedad de necesidades clínicas y de otros tipos.**

*«Las computadoras son buenas para seguir instrucciones, pero no para leer tu mente.»*

**DONALD KNUTH**

## INTRODUCCIÓN

La recuperación de la información (RI), a veces llamada «búsqueda», es el campo relacionado con la adquisición, la organización y la búsqueda de informaciones basada en el conocimiento (Hersh, 2009). Aunque en el ámbito biomédico esta labor se ha concentrado tradicionalmente en la recuperación de documentos de la bibliografía biomédica, la información sobre la que puede aplicarse de manera eficaz ha aumentado sustancialmente con la aparición de las publicaciones multimedia y los grandes repositorios de imágenes, video, estructuras químicas, secuencias de genes y proteínas, y una gran variedad de otros recursos digitales de interés para la formación en biomedicina, investigación y el cuidado de los pacientes. Con la proliferación de los sistemas de búsqueda de información y el contenido *online*, la noción de biblioteca ha cambiado sustancialmente, y han aparecido nuevas bibliotecas digitales.

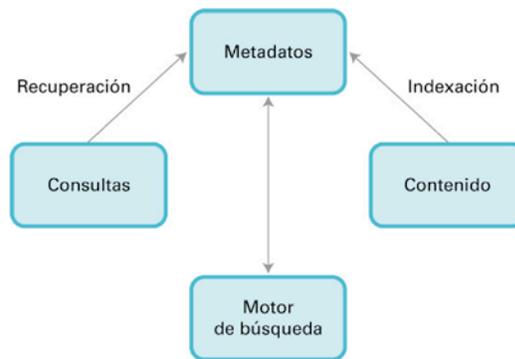
## CASO CLÍNICO

El paciente es un varón de 65 años con insuficiencia cardíaca congestiva de instauración reciente. Se desea encontrar un régimen médico que le permita mantener sus actividades normales y evitar el ingreso hospitalario. Se está valorando el uso de fármacos inhibidores de la enzima convertidora de la angiotensina (ECA), pero se desea revisar las evidencias sobre su empleo.

## DESARROLLO

Los sistemas de RI han almacenado y divulgado históricamente informaciones basadas en conocimientos, que se han obtenido y organizado a partir de investigaciones basadas en la observación o en los experimentos. En el caso de las investigaciones clínicas, esta información ofrece a los profesionales sanitarios, los administradores y los investigadores conocimientos derivados de los experimentos y las observaciones, que pueden aplicarse a pacientes individuales. Esta información se contiene sobre todo en libros y revistas, si bien puede adoptar una amplia variedad de formas, como guías de práctica clínica, bibliografía sobre salud de los consumidores, páginas web y otras fuentes.

La **Figura 5.1** ilustra una visión general del proceso de recuperación de la información. El objetivo principal del mismo es encontrar un *contenido* que cumpla con las necesidades de información de una persona, ya sea mediante artículos, páginas web, imágenes u otras fuentes. Esta labor se lleva a cabo mediante una *consulta* al sistema de recuperación de la información. Un *motor de búsqueda* compara la consulta con los contenidos mediante *metadatos*, que son «datos sobre los datos» en los que se describen los elementos de los contenidos. En la recuperación de la información entran en juego dos procesos intelectuales. Se llama «*indexación*» al proceso consistente en asignar metadatos a elementos del contenido, mientras que la *recuperación* alude al proceso por el cual el usuario introduce su consulta y recupera la información.



**FIGURA 5.1** Visión general del proceso de recuperación de información. La recuperación se hace posible mediante metadatos, que se producen a través de la indexación y de la realización de consultas por parte de los usuarios. Los metadatos se utilizan por el motor de búsqueda, que dirige al usuario hacia el contenido. (Copyright, William Hersh.)

El empleo de sistemas de recuperación de la información y bibliotecas digitales se ha extendido enormemente en la actualidad. Aunque las estimaciones manejadas son variables, entre las personas que utilizan Internet en Estados Unidos, más del 80% han hecho uso del mismo para buscar información relevante sobre su salud o la de alguien que conocen (Fox y Duggan, 2013). Prácticamente todos los médicos de Estados Unidos (y probablemente del resto del mundo) utilizan Internet (anónimo, 2012; Davies, 2010). Además, el acceso a los sistemas ha ido más allá del ordenador personal tradicional para extenderse a los nuevos dispositivos, como teléfonos inteligentes y tabletas.

La información basada en el conocimiento puede subdividirse en dos categorías. Se llama «*información basada en conocimientos primarios*» (o bibliografía primaria) a la investigación original que aparece en revistas, libros, informes y otras fuentes. Este tipo de información alude al descubrimiento inicial de los conocimientos sanitarios, en general con otros datos originales o con el análisis de los mismos (p. ej., revisiones sistemáticas y metaanálisis). La *información basada en conocimientos secundarios* consiste, por su parte, en escritos que revisan, condensan y/o sintetizan la bibliografía primaria. Los ejemplos más comunes de este tipo de bibliografía son libros, monografías y artículos de revistas. La bibliografía secundaria incluye la información sanitaria de creciente calidad orientada a los pacientes/consumidores que abunda cada vez más en la Web. Comprende asimismo escritos basados en opiniones (como editoriales o artículos sobre políticas o posturas personales), guías de práctica clínica, revisiones narrativas e información sanitaria en páginas web. Además, incluye la multitud de manuales de bolsillo que antes los profesionales de numerosos ámbitos manejaban como volúmenes encuadernados y que hoy existen en formato electrónico. Como se verá más adelante, la bibliografía secundaria es el tipo utilizado con más frecuencia por los médicos.

Internet y la Web han tenido un impacto profundo en la publicación no solo de informaciones basadas en conocimientos, sino en todos los tipos de datos e información que existen en el mundo. Prácticamente la totalidad de las revistas científicas se publican hoy en formato electrónico. Para acceder a la mayor parte de su contenido, basta con una conexión moderna a Internet.

## CONTENIDOS

Para comprender mejor la estructura y la función de las informaciones basadas en conocimientos, resulta útil establecer una clasificación de las mismas. En este apartado se ordenan los contenidos en las siguientes categorías: bibliográficos, de texto completo, anotados y agregados; no obstante, algunos contenidos no encajan exactamente dentro de los límites de estos grupos.

La primera categoría está formada por los *contenidos bibliográficos*, e incluye los elementos que durante décadas constituyeron el centro principal de actuación de los sistemas de recuperación de informaciones: las bases de datos bibliográficas. Estos contenidos constan de citas o referencias directas a la bibliografía médica, es decir, artículos de revistas. La base de datos bibliográfica de tipo biomédico más conocida y utilizada es Medline, producida por la National Library of Medicine (NLM) de Estados Unidos y que contiene referencias bibliográficas de todos los artículos, editoriales y cartas a los editores de aproximadamente 5 000 revistas científicas. Las revistas son seleccionadas por un comité asesor de expertos acordado por los National Institutes of Health (NIH). En la actualidad, cada año se añaden a Medline unas 900000 referencias, lo que elevó este número a más de 27 millones de referencias a mediados de 2017.

Es posible acceder a Medline por distintos medios y sin costo a través del sistema PubMed (pubmed.gov). Algunos distribuidores de información (p. ej., Ovid Technologies, [www.ovid.com](http://www.ovid.com)) emiten licencias para el acceso al contenido de Medline y otras bases de datos, y ofrecen servicios de valor añadido a personas individuales e instituciones a cambio de una tarifa.

Medline es solo una de las numerosas bases de datos producidas por la NLM. Existen otras bases, desde libros a secuencias génicas y estructuras de proteínas, a las que es posible acceder por medio de su interfaz global GQuery (<https://www.ncbi.nlm.nih.gov/sites/gquery>). Se dispone también de un gran número de bases de datos bibliográficas ajenas a la NLM, en general más centradas en áreas temáticas o tipos de recursos. La principal de estas bases no elaboradas por la NLM para el ámbito de la enfermería es el Cumulative Index to Nursing and Allied Health Literature (CINAHL, CINAHL Information Systems, <http://www.ebscohost.com/cinahl/>), que abarca bibliografía relacionada con la enfermería y otras disciplinas afines, como fisioterapia, terapia ocupacional, tecnología de laboratorio, educación sobre salud, auxiliares sanitarios y registros médicos.

Un segundo tipo más moderno de contenidos bibliográficos es el *catálogo web*. Estos catálogos, cuyo número aumenta a gran velocidad, consisten en páginas web que contienen principalmente enlaces a otras páginas y sitios web. Entre los catálogos web más conocidos se incluyen los siguientes:

- HealthFinder (healthfinder.gov): información sanitaria orientada a los consumidores y mantenida por la Oficina de Prevención de Enfermedades y Promoción de la Salud del Departamento de Sanidad y Servicios Humanos de Estados Unidos.
- MedHunt de HON Select (<http://www.hon.ch/HONsearch/Patients/medhunt.html>): catálogo de contenidos web orientados al paciente con filtros de calidad de la fundación HON.
- Translating Research into Practice (TRIP, [www.tripdatabase.com](http://www.tripdatabase.com)): base de datos de contenidos para los que se considera que cumplen las exigentes normas de calidad de la medicina basada en evidencias.

Otro recurso bibliográfico es el ofrecido por la National Guidelines Clearinghouse (NGC, [www.guideline.gov](http://www.guideline.gov)). Producido por la Agency for Healthcare Research and Quality (AHRQ) de Estados Unidos, contiene información exhaustiva sobre guías de práctica clínica. El objetivo general de la NGC es poner a disposición de los profesionales sanitarios y de otras disciplinas las guías de práctica clínica pertinentes y otros materiales sobre resúmenes, compendios y comparaciones.

Una última forma de contenido de tipo bibliográfico consiste en *aportaciones RSS* (originalmente RDF Site Summary [resumen de sitios RDF], también denominada «Really Simple Syndication» [sindicación verdaderamente sencilla]), que son breves resúmenes de contenido web, en general noticias, artículos de revistas, apuntes de blogs y otros. Los usuarios configuran un agregador de RSS, que puede considerarse un navegador web, un cliente de correo electrónico o un programa de *software* autónomo, configurado para la aportación RSS deseada, con la opción de añadir un filtro para contenidos específicos.

La segunda categoría es la llamada *contenido de texto completo*. Uno de sus principales componentes es el constituido por versiones *online* de libros y publicaciones periódicas. Como ya se ha indicado, la mayor parte de la bibliografía médica basada tradicionalmente en papel, desde libros de texto a revistas, está hoy disponible en formato electrónico. Las versiones electrónicas pueden verse potenciadas por medidas que comprenden desde el suministro de datos suplementarios de un artículo de una revista a enlaces y contenidos multimedia de un libro de texto. El componente final de esta categoría es el sitio web. Ha de admitirse que la diversidad de información en sitios web es enorme, y que tales sitios podrían incluir contenidos diferentes al descrito en este capítulo. Sin embargo, dentro del contexto de esta categoría, «sitio web» alude al inmenso número de páginas web estáticas y dinámicas ubicadas en un lugar web individual.

Los libros de texto electrónicos ofrecen características adicionales más allá del texto propio de la versión impresa. Si bien numerosos libros impresos contienen imágenes de alta calidad, las versiones electrónicas abren la posibilidad de contar con un mayor número de imágenes e ilustraciones. Además, permiten el empleo de sonido y video, aunque por ahora no se recurre demasiado a ello. Al igual que las revistas en texto completo, los libros electrónicos pueden contener enlaces a otros recursos, como referencias de revistas y los artículos enteros. Muchos sitios de libros de texto basados en la Web ofrecen asimismo acceso a preguntas de autoevaluación y a noticias médicas para la formación continua. Por último, los libros de texto electrónicos permiten a los autores y editores presentar actualizaciones más recientes de la información de lo que se facilita mediante el ciclo habitual de ediciones impresas, en el cual las nuevas versiones suelen aparecer en períodos de 2-5 años.

Como se indicó anteriormente, los sitios web definen otra modalidad de información de texto completo. Probablemente el proveedor más eficaz de información sobre salud basada en la Web es el

Gobierno de Estados Unidos. No solamente produce bases de datos bibliográficas, sino que, a través de instituciones como la NLM, la AHRQ, el National Cancer Institute (NCI) y los Centers for Disease Control and Prevention (CDC), entre otras, han impulsado iniciativas innovadoras de suministrar una extensa información en texto completo para los proveedores y consumidores de la atención sanitaria. Un ejemplo es el popular sitio CDC Travel (<http://www.cdc.gov/travel/>). Algunas de estas fuentes se describirán más adelante como agregaciones, debido a su capacidad de proporcionar muchos tipos de recursos diferentes.

En los últimos años ha aparecido un gran número de sitios web comerciales biomédicos y sanitarios. Desde el punto de vista del consumidor, comprenden más que simples colecciones de texto, con interacción con expertos, almacenes de datos *online* y catálogos de enlaces con otros sitios. Uno de los más conocidos es NetWellness ([www.netwellness.com](http://www.netwellness.com)). Existen también sitios web, de sociedades o empresas médicas, que aportan información para los proveedores de atención sanitaria, en general visiones generales de enfermedades, diagnósticos y tratamientos; a menudo se añaden también noticias médicas y otros recursos.

La tercera categoría es la formada por los *contenidos anotados*, que pueden subdividirse en distintos tipos de información:

- Bases de datos de imágenes, con colecciones de imágenes de radiología, anatomía patológica y otros ámbitos.
- Bases de datos genómicas, con información de secuenciación de genes, caracterización de proteínas y otros datos genómicos.
- Bases de datos de citas, con enlaces a la bibliografía científica.
- Bases de datos basadas en evidencias, consistentes en colecciones altamente estructuradas de evidencias clínicas.
- Otras bases de datos, con colecciones diversas de información.

En la Web existe un gran número de bases de datos de imágenes biomédicas, entre las que pueden citarse las siguientes:

- Visible Human: [http://www.nlm.nih.gov/research/visible/visible\\_human.html](http://www.nlm.nih.gov/research/visible/visible_human.html).
- eRadiology de Lieberman: <http://eradiology.bidmc.harvard.edu>.
- WebPath: <http://library.med.utah.edu/WebPath/webpath.html>.
- Pathology Education Instructional Resource (PEIR): [www.peir.net](http://www.peir.net).
- DermIS: [www.dermis.net](http://www.dermis.net).
- Health Education Assets Library (HEAL): <http://library.med.utah.edu/heal>.
- Open-I (Open Access Biomedical Image Searching Motor): <https://openi.nlm.nih.gov/>.

En la Web existen muchas bases de datos genómicas. El primer número anual de la revista *Nucleic Acids Research* (*NAR*, <https://academic.oup.com/nar>) cataloga y describe estas bases de datos, y en la actualidad está disponible de forma abierta. Asimismo, *NAR* mantiene una base actualizada de dichas bases de datos, la Molecular Biology Database Collection (<http://www.oxfordjournals.org/nar/database/a/>). Entre las más importantes de estas bases de datos figuran las accesibles a través del National Center for Biotechnology Information (NCBI). Todas sus bases de datos están vinculadas entre sí, y también con PubMed, y es posible indagar en sus contenidos mediante el sistema GQuery descrito anteriormente.

Las bases de datos de citas proporcionan enlaces a artículos que citan a otros dentro de la bibliografía científica. Las primeras bases de datos de citas fueron el Science Citation Index (SCI, Clarivate Analytics) y el Social Science Citation Index (SSCI, Clarivate Analytics), en la actualidad integrados en la Web of Science. Dos bases de datos bibliográficas bien conocidas para asuntos relacionados con la salud y la biomedicina con enlaces a citas son SCOPUS ([www.scopus.com](http://www.scopus.com)) y Google Scholar ([scholar.google.com](http://scholar.google.com)). Recientemente se llevó a cabo una comparación entre estos tres recursos para analizar sus características y su cobertura (Kulkarni et al., 2009). Finalmente, debe mencionarse la base de datos de citas CiteSeer (<http://citeseerx.ist.psu.edu/>), centrada en ordenadores y ciencias de la información, incluida la Informática Biomédica.

Las bases de datos de medicina basada en evidencias tienen como finalidad ofrecer informaciones anotadas en este ámbito. Algunos ejemplos (todos disponibles por suscripción) son:

- Cochrane Database of Systematic Reviews, una de las colecciones originales de revisiones sistemáticas ([www.cochrane.org](http://www.cochrane.org)).
- Clinical Evidence, un «formulario de evidencias» ([www.clinicalevidence.com](http://www.clinicalevidence.com)).
- UpToDate, un contenido centrado en cuestiones clínicas ([www.uptodate.com](http://www.uptodate.com)).
- Essential Evidence Plus ([www.essentialevidenceplus.com](http://www.essentialevidenceplus.com)).

Se ha desarrollado un mercado creciente para una clase relacionada de contenidos basados en la evidencia en forma de conjunto de órdenes de apoyo a las decisiones clínicas, reglas y plantillas de gestión de salud/enfermedad. Entre los editores se incluyen los proveedores de registros sanitarios electrónicos cuyos sistemas emplean este contenido y el de otras fuentes, como Zynx ([www.zynxhealth.com](http://www.zynxhealth.com)).

Se dispone de una variedad extensa de otros contenidos anotados. La base de datos ClinicalTrials.gov nació como una base de ensayos clínicos patrocinada por los NIH. En los últimos años ha extendido su alcance hasta convertirse en un registro de todos los ensayos clínicos (DeAngelis et al., 2005; Zarin et al., 2017) y contener los resultados reales de los ensayos (Zarin et al., 2016). Otra base de datos importante para los investigadores es la llamada NIH RePORTER (<http://projectreporter.nih.gov/reporter.cfm>), que se extiende a todas las investigaciones financiadas por los NIH. Con la creciente cantidad de conjuntos de datos procedentes de las investigaciones publicadas se ha creado otro recurso anotado, con el nombre de «biomedical and healthCAre Data Discovery Index Ecosystem» (bioCADDIE; <https://datamed.org/>).

La última categoría de contenidos es la formada por las *agregaciones* de los procedentes de las tres primeras clases. Es preciso admitir que no existe una distinción clara entre esta categoría y algunos de los tipos descritos anteriormente, salvo porque las agregaciones suelen incluir una variedad extensa de tipos diferentes de información que atienden las diversas necesidades de los usuarios. Se han desarrollado contenidos agregados para todos los tipos de usuarios, desde consumidores a profesionales clínicos y científicos.

Probablemente, el recurso de mayores dimensiones de información agregada para consumidores es *MedlinePlus* ([medlineplus.gov](http://medlineplus.gov)) de la NLM. MedlinePlus incluye todos los tipos de contenidos referidos anteriormente y reunidos para facilitar los accesos para una materia dada. MedlinePlus contiene materiales sobre salud, información de fármacos, diccionarios de medicina, directorios y otros recursos. En cada materia se incluyen enlaces a información sanitaria de los NIH y otras fuentes que los responsables han considerado creíbles. Existen asimismo enlaces con noticias de actualidad sobre salud (con actualizaciones diarias), una enciclopedia médica, referencias a fármacos y directorios, junto con una búsqueda en PubMed con formularios relacionada con el tema.

También se han desarrollado agregaciones de contenidos para los profesionales clínicos. La mayor parte de los principales editores de medicina agregan hoy todo su contenido en paquetes para los profesionales clínicos.

## INDEXACIÓN

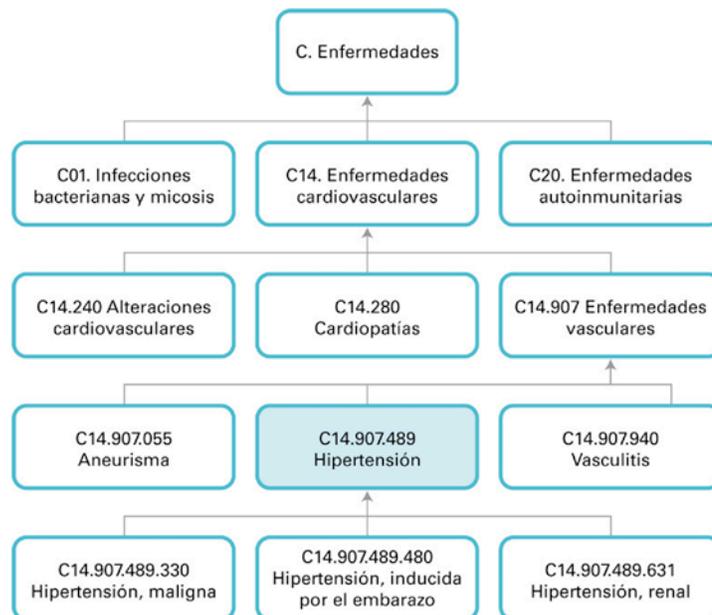
La indexación es el proceso consistente en asignar metadatos a los contenidos con el fin de facilitar su recuperación. En su mayor parte, los contenidos se indexan de dos maneras:

1. Indexación manual, cuando los indexadores son personas que, en general a partir de una terminología controlada, asignan términos de indexación y atributos a los documentos, a menudo de acuerdo con un protocolo determinado.
2. Indexación automatizada, en la cual los ordenadores se encargan de las asignaciones de indexación, normalmente destacando cada palabra del documento (o de una parte de él) como término de indexación.

Para indexar manualmente la mayor parte de las bases de datos producidas por la NLM se utiliza la terminología MeSH (Coletti y Bleich, 2001). En su versión más reciente contiene más de 27 000 epígrafes de materias (el término utilizado por MeSH para la representación canónica de sus conceptos). Contiene igualmente más de 87 000 sinónimos para estos términos, en la jerga de MeSH conocidos como «*términos de entrada*». Además, MeSH contiene los tres tipos de relaciones descritos en los párrafos precedentes:

1. **Jerárquicos:** MeSH está organizada jerárquicamente en 16 árboles, con conceptos como enfermedades, organismos, fármacos y sustancias químicas.
2. **Sinónimos:** MeSH contiene un gran número de términos de entrada, que son sinónimos de los epígrafes.
3. **Relacionados:** en muchos epígrafes se sugieren términos que pueden ser de utilidad para que los investigadores los incluyan en sus búsquedas cuando resulte pertinente.

Los archivos terminológicos de MeSH, sus datos asociados y la documentación de apoyo pueden consultarse en el sitio web MeSH de la NLM (<http://www.nlm.nih.gov/mesh/>). Existe asimismo un navegador que facilita la exploración de la terminología (<http://www.nlm.nih.gov/mesh/MBrowser.html>). La **Figura 5.2** muestra un fragmento de la jerarquía MeSH sobre «Hipertensión» y enfermedades cardiovasculares relacionadas del árbol C «Enfermedades».



**FIGURA 5.2** Fragmento de la jerarquía de epígrafes Medical Subject Headings (MeSH) para la «hipertensión» y otros términos relacionados, que muestra la posición del árbol C «Enfermedades». Las flechas muestran los enlaces con términos más amplios de la jerarquía, mientras que los códigos reflejan la dirección del árbol utilizada internamente por el sistema MeSH.

MeSH posee algunas características ideadas para ayudar a los indexadores a facilitar la recuperación de los documentos. Una de ellas es el tipo de publicación, que describe la clase de publicación o el tipo de estudio. La persona que desee obtener la revisión sobre un tema puede elegir el tipo de publicación, *review* (revisión) o *review literature* (revisión de la literatura), o bien, para buscar estudios que ofrezcan las mejores evidencias de un tratamiento, optar por el tipo *meta-analysis* (metaanálisis), *randomized controlled trial* (ensayo controlado aleatorizado) o *controlled clinical trial* (ensayo clínico controlado).

En el procedimiento automatizado, la indexación es realizada por un ordenador. Normalmente, de ello se encarga un programa informático que extrae todas las palabras de un documento. Algunos diseños de la indexación automatizada asignan ponderaciones a las palabras de acuerdo con la medida calculada de su importancia, como las que son menos frecuentes y, por tanto, su uso resulta más discriminador.

Otro sistema de indexación automatizada ampliamente utilizado se basa en los métodos basados en enlaces, inspirados por el éxito del motor de búsqueda de Google ([www.google.com](http://www.google.com)). Este enfoque otorga una ponderación a las páginas de acuerdo con la frecuencia de cita de las mismas por parte de otras páginas. El algoritmo PageRank (PR) es matemáticamente complejo y aplica una ponderación mayor a las páginas web con un número superior de enlaces a otras páginas (Brin y Page, 1998). De este modo, la página principal de una revista o una organización médica muy conocida y de alta reputación probablemente tendrá un valor PR muy alto, a diferencia de una página más oculta (con menos enlaces).

Los motores de búsqueda genéricos, como Google y Microsoft Bing, hacen uso de estrategias basadas en palabras y en variantes del algoritmo PageRank para la indexación. Reúnen el contenido de sus sistemas de búsqueda mediante una indagación minuciosa de la Web, y reúnen e indexan todos y cada uno de los objetos que encuentran en la misma, lo que incluye no solo páginas HTML, sino también otros archivos, como Microsoft Word, Portable Document Format (PDF) e imágenes.

## RECUPERACIÓN

Para la recuperación se aplican dos grandes principios: la búsqueda con correspondencia exacta y la búsqueda con correspondencia parcial. Muchos sistemas modernos de recuperación de la información (p. ej., PubMed y Google) utilizan ambos. El primero se basa en una búsqueda con correspondencia exacta, que facilita al usuario un control preciso de los elementos recuperados. Por su parte, la búsqueda con correspondencia parcial reconoce el carácter inexacto de la indexación y la recuperación, e intenta

como alternativa mostrar al usuario el contenido ordenado por la cercanía con el criterio de búsqueda que ha planteado. Después de ofrecer unas explicaciones generales sobre estos enfoques, se describirán los sistemas reales que acceden a los distintos tipos de contenido biomédico.

En la búsqueda con correspondencia exacta, el sistema de recuperación de información muestra al usuario todos los resultados que coinciden exactamente con los criterios especificados en los enunciados de búsqueda. Dado que para crear conjuntos manejables de documentos se utilizan los operadores booleanos AND, OR y NOT, este tipo de búsqueda recibe a menudo el nombre de «booleana». Además, como el usuario construye normalmente conjuntos de documentos que se manipulan mediante operadores booleanos, este enfoque se conoce también como «búsqueda basada en conjuntos». La mayor parte de los primeros sistemas operativos de recuperación de la información de los años cincuenta a los setenta hacían uso de la correspondencia exacta, aun cuando en esa época Salton ya estaba desarrollando el sistema de correspondencias parciales (Salton y McGill, 1983). En la actualidad, la búsqueda con correspondencia exacta suele asociarse con la recuperación a partir de bases de datos bibliográficas y anotadas, mientras que la correspondencia parcial se relaciona más bien con búsquedas de texto completo, entre otros, en sitios web.

Por lo general, el primer paso en la recuperación mediante correspondencia exacta consiste en seleccionar términos para formar conjuntos. Para esta labor pueden incluirse además otros atributos, como el nombre del autor, el tipo de publicación o el identificador del gen (del campo de identificador de fuente secundaria de Medline). Una vez elegidos los términos y los atributos de búsqueda, se combinan con los operadores booleanos. El operador AND permite, en general, estrechar el conjunto de recuperación para contener únicamente documentos con dos o más conceptos. El operador OR suele utilizarse cuando existe más de una forma de expresar un concepto.

Aunque la búsqueda con correspondencia parcial se conceptualizó muy pronto, no fue objeto de un uso extendido de los sistemas de recuperación de información hasta la aparición de los motores de búsqueda web en los años noventa. Ello se debe probablemente a que la búsqueda con correspondencia exacta suele ser la preferida por los «usuarios expertos», mientras que la correspondencia parcial es una opción más propia de los noveles. Mientras que la búsqueda con correspondencia exacta exige un conocimiento de los operadores booleanos y (a menudo) de la estructura subyacente de las bases de datos (p. ej., los numerosos campos presentes en Medline), la de correspondencia parcial permite al usuario simplemente introducir algunos términos y empezar a recuperar documentos.

En el sistema de correspondencia parcial es posible clasificar además los documentos de acuerdo con un orden de relevancia en relación con la consulta. Una solución común consiste en otorgar una clasificación más alta a los documentos que contienen más de uno de los términos de la consulta, dado que esta múltiple concordancia suele asociarse a una mayor probabilidad de que el resultado sea relevante para el usuario.

Como se indicaba antes, PubMed es el sistema de la NLM que busca en Medline y en otras bases de datos bibliográficas. Aunque presenta al usuario una casilla de texto simple, PubMed realiza un proceso profundo a partir de la entrada del usuario para identificar términos MeSH, nombres de autor, frases comunes y nombres de revistas (descritos en el sistema de ayuda *online* de PubMed). En esta cartografía automática de términos, el sistema intenta establecer correspondencias entre la entrada del usuario, sucesivamente, con términos MeSH, nombres de revistas, frases comunes y autores. El texto restante para el que PubMed no encuentra correspondencias es objeto de búsqueda en palabras individuales (es decir, las que aparecen en cualquiera de los campos de Medline).

En la **Figura 5.3** se muestra un ejemplo de una búsqueda en Medline sobre insuficiencia cardíaca congestiva (*congestive heart failure*) y fármacos inhibidores de la ECA (*ACE inhibitors*). El sistema introduce automáticamente un AND entre ambos términos de búsqueda.

PubMed permite especificar otros atributos de indexación por medio de «filtros». Entre ellos se incluyen tipos de publicación, subconjuntos, intervalos de edades e intervalos de fechas de publicación. Se accede a ellos en la parte izquierda de la pantalla de resultados, donde se muestran los utilizados de forma más habitual y los otros accesibles mediante clics de ratón adicionales.

PubMed maneja otro enfoque para encontrar las mejores evidencias a través de su función de consultas clínicas (*Clinical Queries*), en las que los términos objeto están limitados por enunciados de búsqueda diseñados para recuperar los mejores resultados de acuerdo con los principios de la medicina basada en evidencias. Existen dos planteamientos diferentes. El primero aplica estrategias para recuperar las mejores evidencias de los cuatro tipos principales de preguntas clínicas. Estas estrategias proceden de investigaciones que valoran la capacidad de los enunciados de búsqueda de Medline de identificar los mejores estudios sobre tratamiento, diagnóstico, daño y pronóstico (Haynes et al., 1994). El segundo

The screenshot displays the PubMed search results page for the query "congestive heart failure ace inhibitors". The search results are listed in a numbered format, with the first four items visible. Each item includes the title, authors, journal information, and PMID. The interface also features a sidebar with navigation options, a search bar at the top, and various filters and tools on the right side.

**Search results**  
 Items: 1 to 20 of 10773

1. [\[Compliance with recommendations in secondary prevention of stroke in primary care\].](#)  
 Tamayo-Ojeda C, Parellada-Esquius N, Salvador-González B, Oriol-Torón PÁ, Rodríguez-Garrido MD, Muñoz-Segura D; en representación del grupo de investigación del proyecto «Adherencia a las recomendaciones de las guías de práctica clínica en la prevención secundaria de la enfermedad cerebrovascular».. Aten Primaria. 2017 Apr 7. pii: S0212-6567(16)30562-5. doi: 10.1016/j.aprim.2016.07.004. [Epub ahead of print] Spanish. PMID: 28395917 [Similar articles](#)

2. [Prognostic impact of heart rate in elderly with systolic heart failure and concomitant atrial fibrillation.](#)  
 Barywani S, Petzold M. Scand Cardiovasc J. 2017 Apr 10;1-7. doi: 10.1080/14017431.2017.1314549. [Epub ahead of print] PMID: 28393567 [Similar articles](#)

3. [Inverse Correlation of Venous Brain Natriuretic Peptide Levels with Body Mass Index Is due to Decreased Production.](#)  
 Shah Z, Wiley M, Sridhar AM, Masoomi R, Biria M, Lakkireddy D, Dawn B, Gupta K. Cardiology. 2017 Apr 8;137(3):159-166. doi: 10.1159/000464111. [Epub ahead of print] PMID: 28391273 [Similar articles](#)

4. [Achieved blood pressure and cardiovascular outcomes in high-risk patients: results from ONTARGET and TRANSCEND trials.](#)  
 Böhm M, Schumacher H, Teo KK, Lonn EM, Mahfoud F, Mann JF, Mancia G, Redon J, Schmieder RE, Sliwa K, Weber MA, Williams B, Yusuf S. Lancet. 2017 Apr 5. pii: S0140-6736(17)30754-7. doi: 10.1016/S0140-6736(17)30754-7. [Epub ahead of print] PMID: 28390695 [Similar articles](#)

**Filter your results:**  
 All (10773)  
[Free Full Text \(2552\)](#)  
[Review \(3777\)](#)  
[Manage Filters](#)

**Results by year**  
  
[Download CSV](#)

**PMC Images search for congestive heart failure ace inhibitors**  
  
[See more \(13\)...](#)

**Titles with your search terms**  
 Effects of combination of AT1-antagonist candesartan cilexetil ar [Srp Arh Celok Lek. 2013]

**FIGURA 5.3** Interfaz de búsqueda para PubMed que muestra una pregunta introducida, los resultados enumerados, los «límites» en la parte izquierda de la ventana y otras muchas funciones.

planteamiento para recuperar las mejores evidencias pretende obtener los recursos basados en la evidencia que son síntesis o sinopsis, en particular metaanálisis, revisiones sistemáticas y guías de prácticas. La estrategia se basa en parte en las investigaciones de [Boynton et al. \(1998\)](#). Cuando se utiliza la interfaz de consultas clínicas, el enunciado de búsqueda es procesado por la correspondencia automática de términos y el resultado se limita (con ayuda del operador AND) con el enunciado apropiado.

La información médica, en especial en lo que atañe al ámbito clínico, que todavía no está publicada en medios electrónicos es muy escasa (aun cuando existan todavía las versiones en papel). Existen numerosos sistemas de recuperación para facilitar su acceso. Al entender cómo funcionan estos sistemas, los profesionales sanitarios, los pacientes y otras personas aprenden a desarrollar técnicas para utilizar estos medios de la forma más eficaz.

## CONCLUSIONES

Prácticamente toda la bibliografía médica del mundo está disponible en Internet, con información adicional correspondiente para pacientes, consumidores y otras personas. No toda esta información está al alcance de cualquiera (podría ser accesible solo para usuarios registrados a cambio de una tarifa); tampoco necesariamente está bien organizada para su fácil recuperación. Sin embargo, con los conocimientos y las técnicas de búsqueda necesarios para el contenido en cuestión, los profesionales clínicos pueden mejorar su eficacia en el empleo de los sistemas de información. Esta capacidad constituye una competencia fundamental para los profesionales del siglo XXI, simplemente por el hecho de que los pacientes también pueden acceder a ella.

## ACTIVIDADES INDIVIDUALES Y GRUPALES

Existe una forma especial de formular preguntas clínicas conocida por el acrónimo PICO, que significa problema/paciente, intervención, comparación y resultado (*outcome*). Con la técnica PICO es posible diseñar un planteamiento de búsqueda sobre una pregunta clínica que puede plantearse, tal vez para

diagnosticar o tratar una enfermedad. Formule una pregunta clínica y elabore una búsqueda para los siguientes tipos de información:

- Las mejores evidencias de la bibliografía científica.
- Una guía de práctica clínica.
- Una visión general de un libro de texto clínico.
- Un sitio web con información sobre la materia.
- (Si fuera pertinente) Imágenes relacionadas con la cuestión.
- Información de una fuente de contenido sintético basado en la evidencia.
- Información sobre un paciente en relación con la pregunta.

## Referencias

- Anónimo (2012). From Screen to Script: The Doctor's Digital Path to Treatment. Recuperado de <https://www.thinkwithgoogle.com/research-studies/the-doctors-digital-path-to-treatment.html>.
- Boynton, J., Glanville, J., McDaid, D., y Lefebvre, C. (1998). Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. *Journal of Information Science*, 24, 137-157.
- Brin, S., y Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.
- Coletti, M., y Bleich, H. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association*, 8, 317-323.
- Davies, K. (2010). Physicians and their use of information: a survey comparison between the United States, Canada, and the United Kingdom. *Journal of the Medical Library Association*, 99, 88-91.
- DeAngelis, C., Drazen, J., Frizelle, F., Haug, C., Hoey, J., Horton, R., et al. (2005). Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Journal of the American Medical Association*, 293, 2927-2929.
- Fox, S., y Duggan, M. (2013). Health Online 2013. Recuperado de <http://www.pewinternet.org/Reports/2013/Health-online.aspx>.
- Haynes, R., Wilczynski, N., McKibbon, K., Walker, C., y Sinclair, J. (1994). Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *Journal of the American Medical Informatics Association*, 1, 447-458.
- Hersh, W. (2009). *Information Retrieval: A Health and Biomedical Perspective* (3a ed.). Nueva York: Springer.
- Kulkarni, A., Aziz, B., Shams, I., y Busse, J. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Journal of the American Medical Association*, 302, 1092-1096.
- Salton, G., y McGill, M. (1983). *Introduction to Modern Information Retrieval*. Nueva York: McGraw-Hill.
- Zarin, D., Tse, T., Williams, R., y Carr, S. (2016). Trial reporting in ClinicalTrials.gov — the final rule. *New England Journal of Medicine*, 375, 1998-2004.
- Zarin, D., Tse, T., Williams, R., y Rajakannan, T. (2017). Update on trial registration 11 years after the ICMJE policy was established. *New England Journal of Medicine*, 376, 383-391.