



INTRODUCCIÓN A LA RECUPERACIÓN DE LITERATURA DIGITAL EN CIENCIAS BIOLÓGICAS

AUTORES:

Layla Michán
Eduardo Alvarez
Jack Guillén

<http://www.ciib.unam.mx/libros/irldcb>



INTRODUCCIÓN A LA RECUPERACIÓN DE LITERATURA DIGITAL
EN CIENCIAS BIOLÓGICAS por Michán, L., Alvarez, E. y Guillén, J.
se encuentra bajo una Licencia Creative Commons Atribución-No
Comercial-Licenciamiento Recíproco 3.0 Unported.



Universidad Nacional Autónoma de México
Facultad de Ciencias
Primera edición
México, 2013

CÓMO CITAR ESTE LIBRO

Michán, L., Alvarez, E. y Guillén, J. (2013). *Introducción a la recuperación de literatura digital en ciencias biológicas*. Las Prensas de Ciencias, México. 131 p. ISBN 978-607-02-4093-5. Disponible en <http://www.ciib.unam.mx/libros/irldcb/index.html>

*A Judith y Ramón por alentarme y heredarme su
entusiasmo por el estudio y la investigación*

CONTENIDO

PRÓLOGO

La presente *Introducción a la recuperación de literatura digital en ciencias biológicas* es un texto didáctico cuyo objetivo es compenetrar al lector en contenidos que necesariamente pasarán a ser una asignatura indispensable en el currículo y formación de todo estudiante de licenciatura y de posgrado en Biología y en general, para todo aspirante a investigador y profesionalista en cualquier área del conocimiento.

Especializada en recursos de información en biología, esta obra sustenta dos principios metodológicos que, en realidad, son aplicables para cualquier práctica científica en la actualidad. Figura, en primer lugar, el desarrollo de habilidades para la consulta, selección, extracción, análisis y preservación de grandes cantidades de información digital, para lo cual se requiere de una actitud proactiva por parte del usuario de la misma, superando la práctica pasiva o meramente receptiva en la consulta de catálogos, páginas web y bases de datos especializadas. En este sentido, las herramientas provistas por la Web 2.0, compendiadas en el presente libro, permiten potenciar las capacidades de gestión de los flujos de información de los que actualmente se disponen y manejan en prácticamente cualquier proyecto científico.

No obstante, la actual acumulación de información, expresada anualmente en más de dos millones de artículos científicos publicados en miles de revistas indizadas en una gran variedad de bases de datos y sistemas de información, convierten la tarea de consulta y recuperación de información en una práctica inabarcable para el investigador aislado; de ahí, el segundo principio rector de este texto consiste en la socialización del conocimiento como supuesto y objetivo del proceso de investigación. Las redes sociales aplicadas a la administración, la compartición y la comunicación de la información especializada permiten al investigador aprovechar el

conocimiento de los demás (inteligencia colectiva) al mismo tiempo que convertirse en un agente cooperante en la actividad científica, la cual adquiere cada vez más un perfil colectivo y multidisciplinario.

Finalmente, debe hacerse referencia al hecho de que si bien es cierto que las telecomunicaciones y las tecnologías de la información han permitido superar las barreras geográficas y temporales en el acceso a la información, permitiendo al usuario la consulta directa a una infinidad de recursos de información, también lo es en contrapartida, que esta desintermediación obliga al consumidor a familiarizarse y adentrarse en el conocimiento de una gama creciente de fuentes de información. Al respecto, resulta de suma utilidad la clasificación y descripción de los sistemas de información glosados en este documento, así como la explicación sobre el significado y la utilidad del meta-análisis y las tendencias más recientes en el trepidante mundo de la información científica.

Sin lugar a dudas, los estudiantes, profesores e investigadores resultarán beneficiados por la relevancia y oportunidad de este material de consulta, el cual es uno más de los frutos del trabajo del dinámico, laborioso y propositivo equipo de investigación liderado por la Dra. Layla Michán Aguirre y conformado en torno al Laboratorio de Información, Informática y Cienciometría Biológica, perteneciente al Departamento de Biología Evolutiva (Facultad de Ciencias, UNAM). Sin lugar a dudas también, nuestra casa de estudios seguirá resultando beneficiada por el empuje y los productos aportados por este grupo de vanguardia.

Antonio Sánchez Pereyra
Jefe del Departamento de Bibliografía Latinoamericana
Dirección General de Bibliotecas, UNAM

PRESENTACIÓN

Haciendo mi doctorado en Historia de la Biología cerca del año 2000 mi director de tesis, Jorge Llorente, me dijo: “Tienes que hacer una base de datos bibliográfica, en la que sistematices la literatura que se ha publicado sobre taxonomía en México en el siglo XX” (este era el tema de mi investigación). Nunca imaginé que esa indicación trazaría en el sendero que definiría mi profesión biológica y mi pasión. Así, me involucré con las bases de datos, las colecciones bibliográficas, y poco después, con la bibliometría, además del estudio de los documentos y de sus características y propiedades, lo que ya era parte de mis objetivos primarios de atención.

Era precisamente la época en la que se transformaban las publicaciones periódicas impresas al formato electrónico como producto del impacto de la web en la difusión de la información; este proceso y sus repercusiones me fascinó de tal manera que lo seguí muy de cerca y lo experimenté en todas sus posibilidades. El cambio era vertiginoso, imposible de concebir en esa época, aunque ya me parezca normal. Me asombraba a diario con la aparición de colecciones como PubMed, nuevas revistas electrónicas (muchas de ellas con diseños geniales y formatos novedosos) y ni hablar de las aplicaciones disponibles como las alertas o la opción de enviar un correo electrónico a los autores.

De repente me encontré las redes sociales, los RSS, el acceso abierto y la aparición del *DOI* (*Digital Object Identifier*). Me volví una seguidora implacable de todas las novedades que aparecían, y conforme se me iban presentando, las probaba todas. Combinando mi interés en la historia reciente de la ciencia, mi pasión por las bases de datos bibliográficas, el gusto por el uso de la bibliometría y mi afinidad por la informática se constituyó la línea de investigación que actualmente desarrollo en la Facultad de Ciencias de la UNAM y que trata sobre “información, informática y cienciometría para ciencias biológicas”. Se trata de la investigación, enseñanza y difusión de estrategias para la resolución de problemas de información biológica respecto a sus características, flujo, acceso, ciclo, clasificación, búsqueda, sistematización, alfabetización, dominio, manejo, administración, procesamiento, gestión, colección, análisis, preparación, presentación, evaluación, visualización, contextualización, publicación, difusión, uso y aplicación para la obtención de nuevo conocimiento y toma de decisiones.

Resultado de ello surge este primer libro, en el que este se resume lo básico para un estudiante o académico que trabaja con temas biológicos.

La literatura especializada se aplica en varios ámbitos de la práctica biológica, que van desde la producción de nuevo conocimiento biológico hasta la administración de recursos, evaluación, gestión y política científica; su uso es indispensable para realizar este trabajo científico de manera eficiente. La búsqueda, el acceso, el análisis y la actualización de la literatura a través de la Web se ha vuelto una tarea cotidiana, la cual implica una complejidad cada vez más grande en el momento en que crecen en cantidad y diversidad los documentos científicos, y los diferentes recursos electrónicos (en especial las bases de datos bibliográficas) por medio de los cuales se puede acceder a dicha información se hacen más sofisticados. Es importante reiterar que estos recursos cambian y se actualizan constantemente, lo que hace prácticamente imposible estar al tanto de todos los recursos por lo que es difícil identificar cuáles y cuántos se pueden y deben utilizar.

Previo a 2009, en la Universidad Nacional Autónoma de México (**UNAM**) no existía, ni en licenciatura, ni en posgrado, algún curso teórico o práctico en el que se presentara el procedimiento, las aplicaciones y los conceptos básicos para el dominio del uso de la literatura científica, mucho menos para el caso de la biología. Al reconocer esta carencia diseñamos, propusimos e impartimos por primera vez el curso optativo "E-investigación para Ciencias Biomédicas en la Licenciatura en Ciencias Biomédicas de la UNAM". Para el año 2010 se inició el curso "Manejo de información digital para Biología" como asignatura optativa en la Facultad de Ciencias de la UNAM. A principios del año 2011 se impartieron también ambas materias en el posgrado de Ciencias Biomédicas y la segunda en el de Ciencias Biológicas de la UNAM. Además, hemos impartido en el transcurso de estos tres años más de 30 cursos, talleres y diplomados de introducción y actualización sobre el tema en otras universidades, centros de investigación y hospitales.

Estos cursos representan la aplicación de un enfoque interdisciplinario e integral para utilizar la literatura especializada en biomedicina y conocer las tácticas y principios para extraer, analizar y administrar literatura especializada de manera eficiente, inmediata, actualizada, exhaustiva y organizada. La estructura de los cursos en cuanto a sus contenidos se organiza de la siguiente manera: 1) se presentan, ordenan y clasifican los recursos electrónicos disponibles en la web

para la recuperación y el análisis de información especializada de manera eficaz y eficiente; 2) se revisan los tipos y las características de la literatura biológica digital; 3) se explican las definiciones básicas, se explora la importancia e implicaciones de las colecciones digitales, se sintetizan y explican las colecciones bibliográficas en línea más relevantes y prácticas a partir de las cuales se puede y debe extraer información; 4) se detalla el método para la recuperación de literatura en línea; 5) se explican el flujo y el ciclo de la literatura biológica; 6) se exponen los principales indicadores bibliométricos que se utilizan para la evaluación de la literatura; 7) se aplican técnicas novedosas de análisis de las referencias y los contenidos de los documentos científicos para analizar grandes cantidades de documentos simultáneamente por meta-análisis, los cuales permiten entre otras cosas representar la estructura, el desarrollo, las relaciones y las tendencias de la literatura; 8) se exponen los rudimentos sobre gestión de la bibliografía personal; y 9) se discuten los problemas actuales y las tendencias de la bioinformación acorde a las necesidades y retos de nuestro tiempo. Los cursos permiten a los estudiantes preparar un artículo académico e introducirse a las herramientas suficientes para realizar reportes con información sobre el impacto, colaboración y citación de la producción propia en revistas y documentos indizados, comúnmente solicitados por los comités de evaluación del desempeño académico como Consejo Nacional de Ciencia y Tecnología (**CONACYT**) y Dirección General Asuntos del Personal Académico (**DGAPA**).

En la impartición de estos cursos nos hemos enfrentado al problema de que hay muy pocos documentos que traten estos temas y ninguno que los sintetice, mucho menos para biología. Por esta razón nos hemos dado a la tarea de preparar un texto introductorio en el que se presenten los temas básicos para el manejo de literatura especializada en biología (temas relacionados y afines) disponible en la Web. Nuestra exposición está dividida en seis secciones: 1) una introducción al tema; 2) descripción, clasificación y comentario de las aplicaciones, utilidades, herramientas y servicios electrónicos disponibles en la Web 2.0 y la Web 3.0 más útiles para la búsqueda, manejo, análisis y publicación; 3) breve explicación del proceso de recuperación de información electrónica; 4) síntesis de las características y los tipos de colecciones bibliográficas más relevantes y útiles para obtención de literatura; 5) breve introducción sobre los métodos más utilizados para realizar meta-análisis de literatura y algunos ejemplos de las aplicaciones en línea para este propósito; y 6) presentación sobre lo más innovador dentro de la Web para el manejo de literatura. La estructura general del libro se presenta en la figura 1.

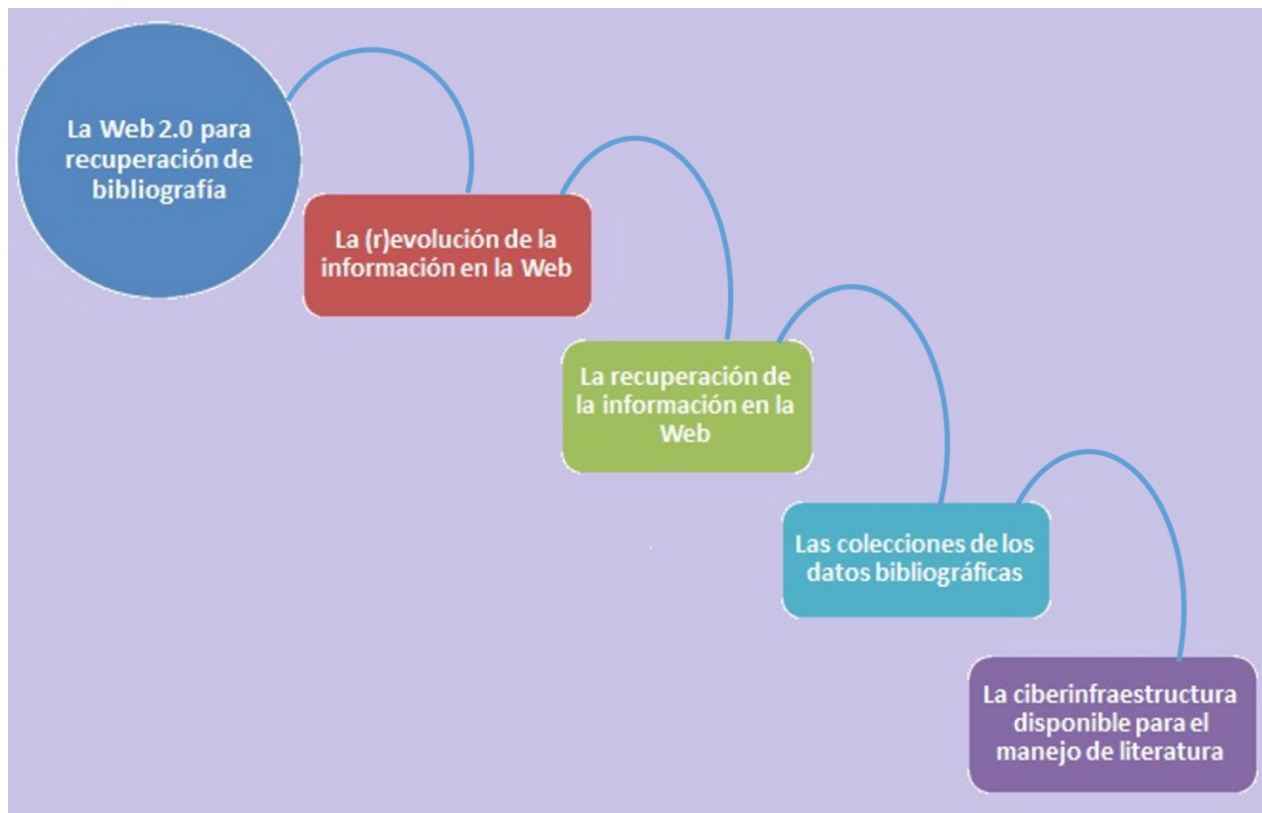


Figura 1. Estructura del libro.

Hemos agregado tanto las direcciones electrónicas como la liga correspondiente para cada referencia electrónica para que el libro pueda utilizarse tanto en formato impreso como electrónico. Al final hemos agregado una breve lista de bibliografía que se puede consultar y además hemos puesto un breve glosario (www.ciib.unam.mx/libros/irldcb/finales/glosario.html) en el que se hace referencia a las definiciones básicas.

Considerando la velocidad a la que evolucionan los contenidos y las aplicaciones electrónicas, hemos diseñado dos blogs que permitirán mantener actualizada toda la información relacionada con los temas de este libro: *Ciberliteratura* “www.ciberliteraturaunam.blogspot.com” (Fig. 2) en el que publicamos lo más novedoso sobre literatura académica y *Biiiogeek* “www.biiiogeek.blogspot.com” (Fig. 3) en el que publicamos las noticias más recientes sobre información e informática biológica para el manejo de literatura.

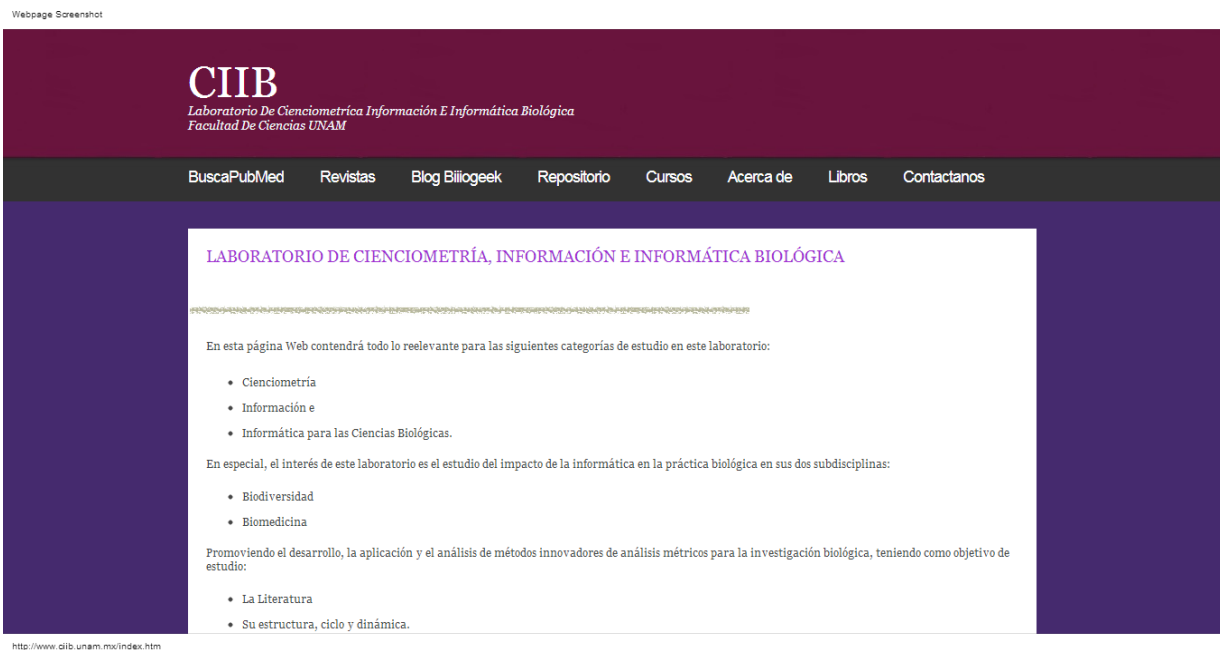


Figura 2. Página web del laboratorio CIIB, UNAM.

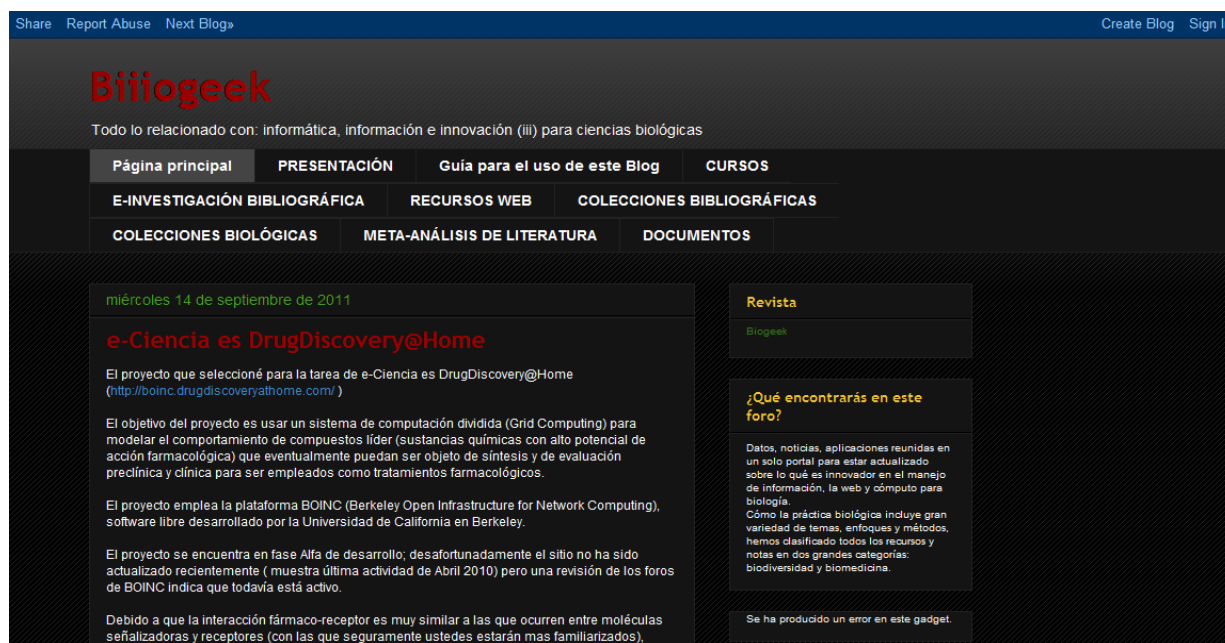


Figura 3. Blog Biiogeek.

Vale la pena mencionar que en la realización de este libro participaron, en mayor o menor medida todos los estudiantes que han trabajado en mi laboratorio, pero en especial Eloy Valtierra, Lyssania Macías, Roberto Calderón y Claudia Pedraza. Agradecemos la corrección de estilo realizada por Anel Jiménez Carrizosa y el cuidado de la edición estuvo a cargo de Patricia

Magaña y Mercedes Pereyó.

Su edición fue posible gracias a los fondos aportados por los proyectos DGAPA, PAPIME PE201509, 2009-2011 "Recursos didácticos para el manejo electrónico de literatura en biología", Ciencia básica proyecto 13276, "Análisis de las ciencias biológicas en la actualidad 1980-2010" CONACYT, y DGAPA, PAPIME PE212112 "Web 2.0 y 3.0 para dominio de la literatura biológica".

Esperamos que les sirva, pero sobretodo que lo disfruten y se asombren, igual que nosotros, con el potencial que todas estas herramientas y enfoques representan.

Layla Michán
7 de septiembre 2012

I. INTRODUCCIÓN

La (r)evolución de la información es causa y efecto del progreso científico y tecnológico del siglo XX. La cantidad de información que se produce ahora sobre temas científicos es diversa, además de colosal: puede ser electrónica o impresa; existe en texto, imágenes y sonidos; se encuentra sistematizada en bases de datos, catálogos o listas; su consulta puede ser libre o restringida; trata sobre los seres vivos o sus partes, fenómenos y explicaciones; versa sobre las publicaciones, los investigadores, los proyectos, los grupos y las líneas de investigación, los convenios, los subsidios, la producción científica, las instituciones de investigación y enseñanza o las colecciones biológicas, por mencionar algunas.

Referirse a información en el siglo XXI implica la mención de términos, métodos y teorías novedosas e innovadoras que están en boga como: sociedad del conocimiento, **sociedad de la información**, globalización, infodiversidad, acceso a la información, **e-ciencia**, e-investigación, **arquitectura grid**, **colaboratorios**, conocimiento basado en la literatura, minería de textos, red semántica, índice de impacto, cocitación, Web 1.0, 2.0 y 3.0, redes sociales, plagio, repositorios, acceso libre, entre otros. Esto ha repercutido de manera dramática en la visión del mundo contemporáneo, la práctica científica, las relaciones científicas, sociales, económicas, políticas y culturales. Inclusive la esencia misma de la ciencia está cambiando, especialmente el manejo de la información a través del empleo de las redes electrónicas y las computadoras de alta velocidad.

La producción y popularización de la computadora personal con la capacidad de procesar datos y el uso masivo del Internet como medio de comunicación electrónico han generado una gran cantidad de información sistematizada en formato digital a través de la web, lo que ha favorecido el acceso, la automatización, la inmediatez y democratización de la información y el conocimiento. En especial, la sociedad científica genera y recibe información y queda expuesta a ella como representación del pensamiento y del conocimiento; en todos los casos, se crea un interés consciente o inconsciente de transmitirla de manera individual o colectiva.

La información digital publicada sobre biología es tal en cantidad, diversidad y complejidad, son tantos, tan diferentes y complicados los recursos electrónicos por medio de los cuales se puede

acceder a ella, que se ha hecho necesario el estar informado y actualizado sobre la continua aparición y modificación de estas herramientas, tanto que se ha vuelto un problema a resolver: se publican continuamente en las revistas especializadas un gran número de artículos sobre estrategias de recuperación, colecciones y análisis de la información en general o en específico para las disciplinas biológicas. Se ha hecho necesario el estar informado y actualizado sobre la continua aparición y evolución de programas para el manejo de la literatura.

Los recursos electrónicos con literatura especializada en ciencias biológicas que pueden ser consultados electrónicamente vía Internet permiten el acceso inmediato a las colecciones digitales actualizadas con información generada por los especialistas. El poder de las nuevas tecnologías electrónicas ha aumentado de manera exponencial. Se han diseñado gran cantidad de aplicaciones que permiten agrupar, clasificar y visualizar los documentos, esto ha reducido la energía, el costo y el tiempo requeridos para el análisis de la literatura especializada. Más aún, en menos de 15 años se ha modificado la práctica científica. Ya no se explora la realidad sólo a través de experimentos y modelos *in vivo* y/o *in vitro* sino que ahora se hacen también *in silico* con herramientas y métodos informáticos. Este fenómeno ha repercutido a tal grado en la forma de producir conocimiento científico que se han desarrollado nuevos campos: la bioinformática, la e-ciencia, la e-biología, la informática médica (*medical informatics*), la informática biológica (*biological informatics*), la neuroinformática, la simbiomática (*symbiomatics*), la informática de la ecología (*ecological informatics*), la e-taxonomía, la informática de la biodiversidad (*biodiversity informatics*), entre otras.

El cambio ha sido importante a varios niveles: teórico, metodológico y práctico. Incluso en la forma en la que se recupera y se analiza la bibliografía, se han propuesto nuevas formas de acceder a la información, dejando de lado el enfoque reduccionista y adoptando uno sistémico que consta de visualizar grandes cantidades de documentos (miles) simultáneamente para identificar patrones y tendencias de la literatura obtenida e implica el uso de la ciberinfraestructura y los recursos de la Web 2.0 y la Web 3.0, tales como exploradores, buscadores, metabuscadores, blogs, redes sociales, folksonomías, marcadores, bases de datos bibliográficas, software para manejo de literatura, portales especializados, lectores RSS, wikis, entre otros acordes al progreso de la propia disciplina biológica.

La ambigüedad que existe en el actual entorno de la información científica es resultado, por un

lado, de un acceso más extenso e integrado a un amplio espectro de fuentes de información, por otro, las dificultades relacionadas con el acceso a los materiales especializados. Para esto, se empleaba el método tradicional, esto es la investigación documental (o bibliográfica) en la que se localizan los documentos y se revisan uno por uno para cada tema; pero ya no depende únicamente de los documentos impresos disponibles en las librerías o bibliotecas, sino de la información que se encuentra en catálogos electrónicos (de hecho, hay bibliotecas que ya no cuentan con catálogos impresos) lo que ha favorecido el acceso, la automatización, la inmediatez y democratización de la información. Este tipo de investigación bibliográfica a la que se ha denominado “e-investigación bibliográfica” consiste en el análisis sistémico, integral y simultáneo de toda la literatura electrónica disponible.

El manejo adecuado de la literatura electrónica especializada en ciencia se centra en la capacidad de las personas para colaborar y compartir información en línea, es decir, la Web 2.0 se ha convertido en un laboratorio en el cual se ha desarrollado una enorme cantidad de aplicaciones, herramientas y servicios que ofrecen la posibilidad de integrar los procesos de recuperación de bibliografía. Uno de los principales problemas a los que se enfrentan los usuarios es el desconocimiento de los recursos digitales disponibles en la Web, de tal manera que la mayoría de herramientas y aplicaciones más sofisticadas para el manejo de literatura científica se utiliza en una mínima parte de su potencial y de la oferta disponible en la red. Muchas de estas aplicaciones, herramientas, utilidades y *mashups* son programas que permiten marcar, agrupar, clasificar, compartir, visualizar, recuperar, analizar y administrar grandes cantidades de información de manera sistematizada, automatizada, eficiente, inmediata, actualizada y exhaustiva, lo que ha hecho más eficiente la catalogación, consulta y análisis de la literatura especializada.

En las últimas décadas el proceso de publicación científica ha sufrido cambios fundamentales; la sobreabundancia de información, la facilidad para su distribución en los medios electrónicos, la publicación de preimpresos (*pre-prints* y *post-prints*) y la amplia gama de fuentes de información sin certificación han creado mayor confusión para quienes se dan a la tarea de realizar búsquedas de bibliografía en Internet. Estas herramientas promueven la inter y multidisciplinariedad, es decir, la colaboración y la participación de un gran número de investigadores (miles) localizados en diversas regiones y con diferentes especialidades que conforman grupos de trabajo virtuales,

en algunos casos llamados laboratorios.

En la figura 4 se presenta una síntesis del proceso y las aplicaciones más útiles para el dominio de la literatura especializada en biología a través de la Web, la cuál se utilizará como referencia para estructurar este libro.

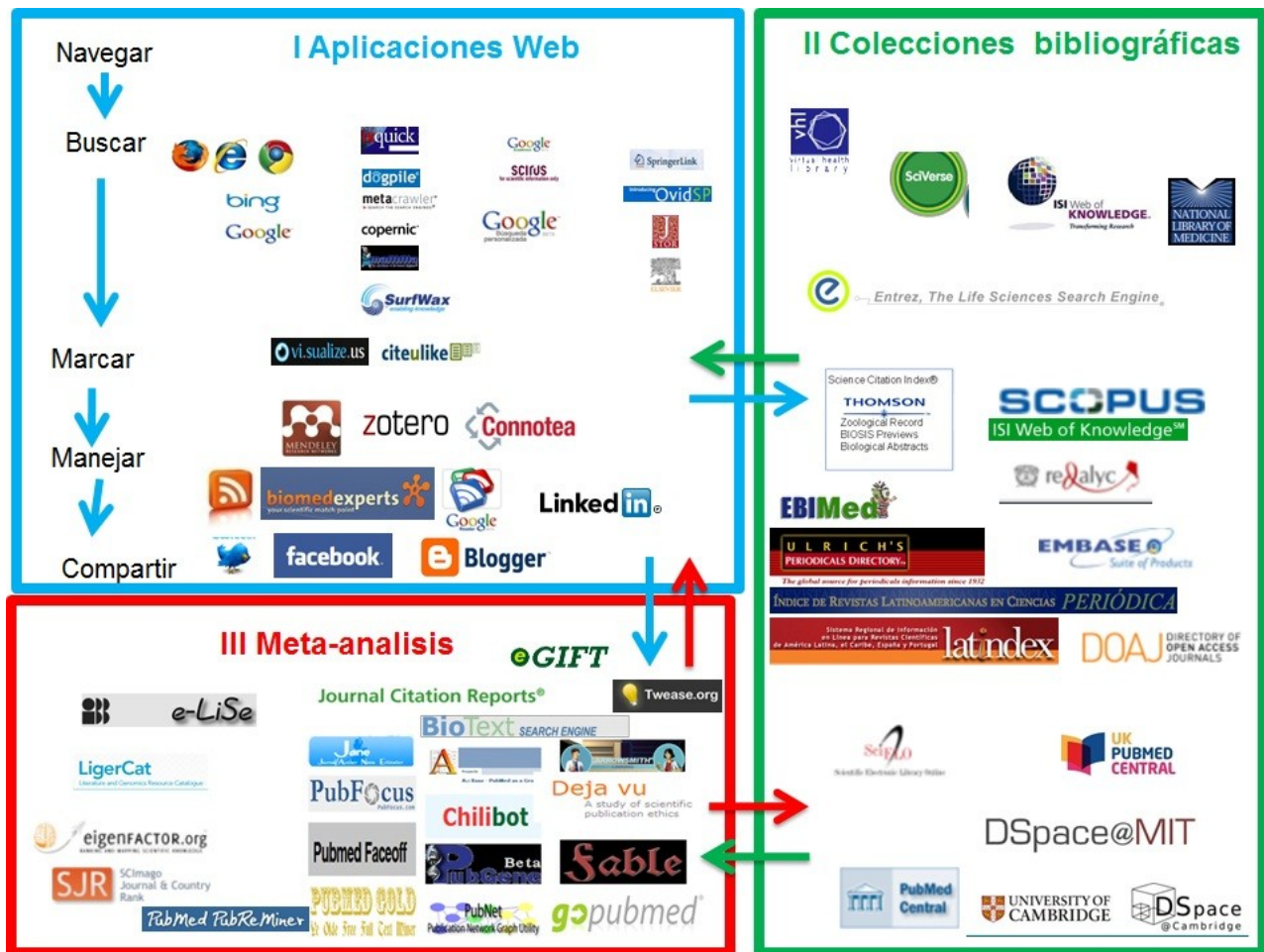


Figura 4. Principales aplicaciones presentadas en este libro (capítulos 2, 4 y 5).

II. LAS APLICACIONES WEB 2.0

El presente capítulo clasifica, describe y presenta los recursos Web 2.0 más relevantes accesibles y básicos para la recuperación y el manejo de literatura especializada en biología. Cada uno de los recursos electrónicos se expondrá utilizando el orden lógico en el que se realiza la recuperación de literatura en la red (Fig. 5).

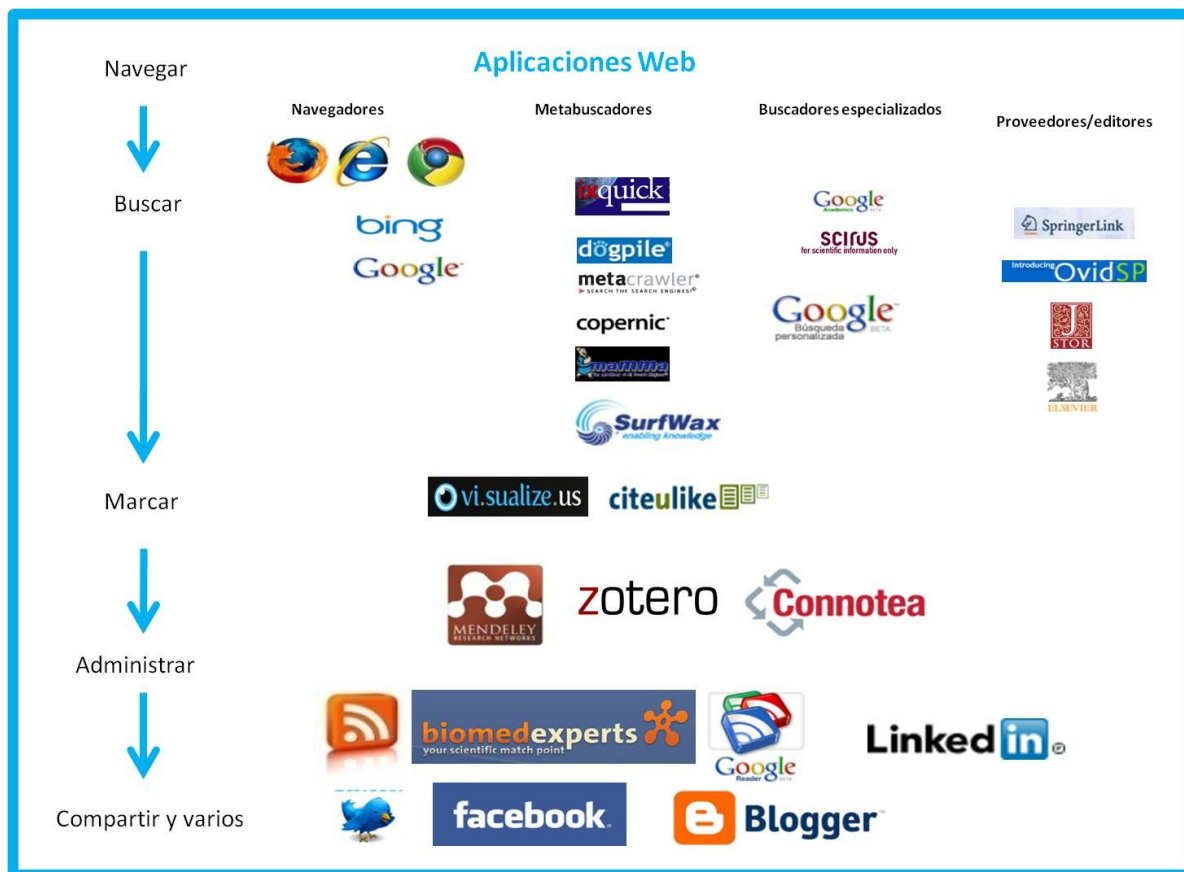


Figura 5. Recursos o aplicaciones Web 2.0 disponibles y más relevantes para la recuperación y el manejo de literatura.

1. Internet y Web

La mayoría de la gente utiliza las palabras **Internet** y **Web** (*World Wide Web*) como sinónimos, pero el hecho es que los dos términos no significan lo mismo, aunque están estrechamente relacionados. Internet comenzó siendo una red informática y tenía como misión la interconexión de computadoras entre varias universidades y laboratorios de investigación de los Estados Unidos. Es la red masiva de redes, una infraestructura que conecta millones de computadoras a nivel mundial formando una trama en la que cualquier computadora puede comunicarse con otra sin importar su cercanía geográfica. La información que viaja a través de Internet tiene una gran variedad de lenguajes para transmitir los datos, conocidos como protocolos, de los cuales los más usados son: el envío de correo electrónico (SMTP), la transmisión de archivos (FTP y P2P), las conversaciones en línea (IRC), la mensajería instantánea, la transmisión de contenidos y la comunicación multimedia (telefonía [VoIP] y televisión [IPTV]), los boletines electrónicos

(NNTP) y el acceso remoto a otras máquinas (SSH y Telnet).

La red o Web, como la conocemos ahora, nació en 1989 y fue diseñada por el informático inglés Timothy Berners-Lee para el Consejo Europeo de Investigación Nuclear. Se refiere a una forma de acceder a la información que viaja en Internet y se basa en un modelo de información compartida por medio del protocolo HTTP (HiperText Transfer Protocol). Se utilizan los programas llamados **exploradores** o **navegadores** para acceder a los documentos electrónicos llamados **páginas web** que están ligadas unas a otras a través de **hiperligas** o **enlaces de hipertextos** (*hyperlinks*) los documentos pueden contener texto, gráficas, sonidos y/o vídeo. La Web es sólo una de las formas de distribuir información a través de Internet, la cual ha sufrido grandes modificaciones estructurales, tecnológicas, filosóficas y sociales desde que fue creada. Con base en estas transformaciones, históricamente se ha clasificado su evolución en tres etapas denominadas Web 1.0, Web 2.0 y Web 3.0.

El modelo de la Web 1.0 con el que se popularizó esta tecnología en los años noventa se limitó a un espacio electrónico de publicación de las entradas corporativas y de servicios, sin participación abierta ni gratuidad en contenidos o servicios de alta relevancia. Las comunidades se formaban fundamentalmente a partir de la oferta de servicios, prescindiendo de espacios para que los miembros publicaran. Es decir, los usuarios fueron relevantes en tanto eran consumidores. Las aplicaciones características de esta generación son el **correo electrónico**, el **chat** y los **directorios de páginas web**.

Web 2.0 es el término dado para describir a la segunda generación iniciada en el año 2000, y se centra en la capacidad de las personas para colaborar y compartir información en línea. Básicamente se refiere a la transición de una web estática a una dinámica, más organizada, basada en el servicio de aplicaciones web, que consta de la transformación del usuario en productor de información, fomenta la comunicación abierta con énfasis en la conformación de comunidades de usuarios y en la posibilidad de intercambio de información. Aplicaciones distintivas de éstas son los blogs, las wikis y las redes sociales, así como la sistematización de datos digitales en bases de datos (Fig. 6). Estas transformaciones en el rol del usuario indudablemente han repercutido en la generación de nuevas formas de buscar información; ya no sólo se trata de ofrecer la posibilidad de encontrar información, sino de lograr objetivos específicos pues es factible, crear, etiquetar, jerarquizar y compartir datos, por mencionar las

funciones más comunes de la Web 2.0. Este enfoque es básicamente social y pretende ofrecer al usuario un conjunto de contenidos agrupados y con significado que ofrece la posibilidad de generar una siguiente generación, la **Web semántica** (Web 3.0) basada en la idea de añadir **metadatos semánticos** y **ontológicos**.

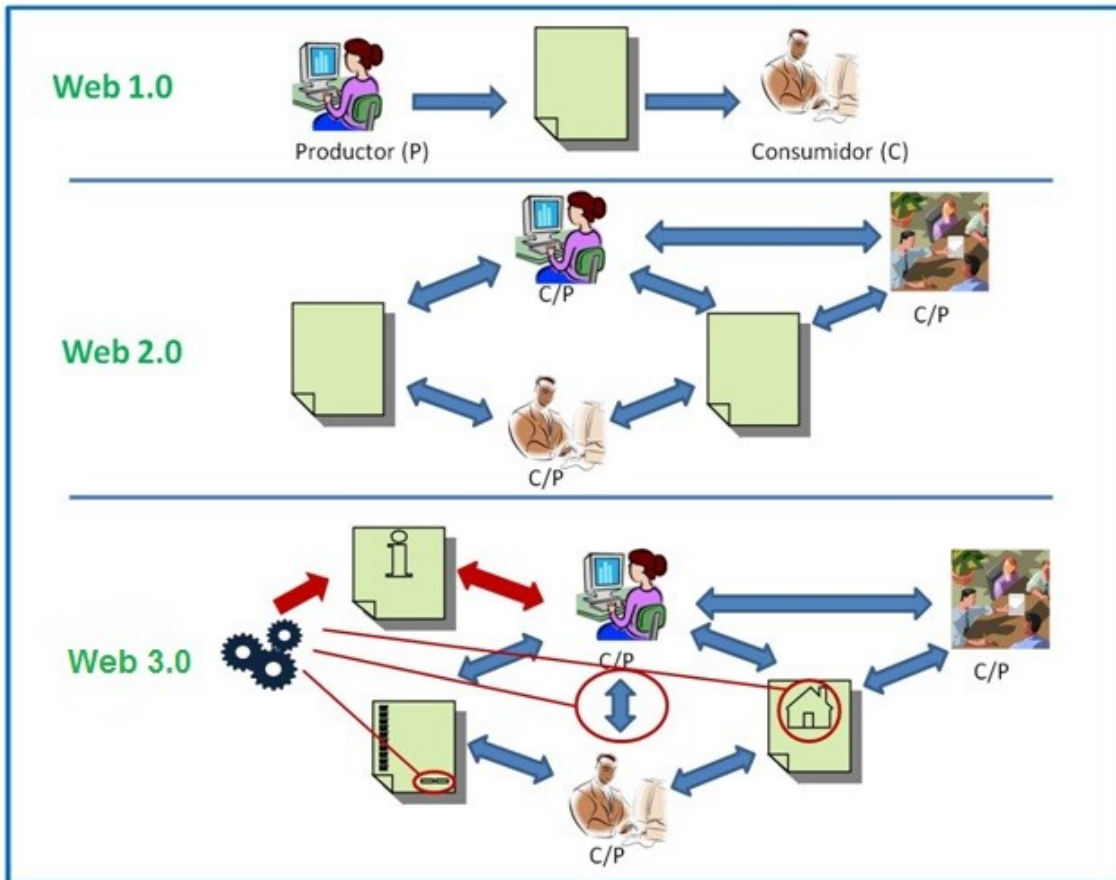


Figura 6. Evolución de la Web.(P) productor; (C) consumidor; (C/P) consumidor/proveedor. Los círculos rojos se refieren a la inclusión de ontologías.

Muchos de los próximos desarrollos contribuirán seguramente a consolidar la web semántica y, ésta a su vez, permitiría estructurar la información de la manera más similar posible a como los humanos almacenan datos en el cerebro a través de mapas cognitivos, pues se ha incluido el desarrollo de nuevos sistemas de interoperabilidad que permiten interpretar metadatos para adaptarse a las acciones de cada usuario. A continuación se describen con más detalle las aplicaciones de la Web 2.0 más relevantes para el dominio de la literatura biológica digital.

2. Clasificación de los recursos en la Web 2.0 para literatura

Hemos definido siete categorías para clasificar todos los programas disponibles para la recuperación de literatura. A continuación, se caracteriza cada uno de acuerdo con su función y su evolución en la Web. El proceso de dominio inicia con la búsqueda de información en literatura electrónica publicada en las páginas electrónicas, blogs y wikis a las cuales se accede a través de los exploradores.

Recurso (clasificación)	Ejemplo
-------------------------	---------

EXPLORAR	Mozilla Firefox, Opera, Internet Explorer, Safari, Google Chrome, Midori, Epiphany.
----------	---

Exploradores o navegadores



BUSCAR	Alexa, Ask, Bing, Exalead, Google, Wikia.
--------	---

Buscadores (*search engines*) y metabuscadore



Copernic, Dogpile, Ipselon, Ixquick, Kartoo, Mamma, Metacrawler, T10 (Turbo10), WebCrawler, Yippy.

Recurso (clasificación)	Ejemplo
-------------------------	---------



ACTUALIZAR

Feedly, Bloglines, Netvibes, Yahoo! .

RSS o lectores de feeds



GUARDAR

Diigo, Delicious, Google Bookmarks, VisualizeUs, StumbleUpon, Evernote, YouTube, Picasa.

Marcadores



COMPARTIR

Badoo, Facebook, Hi5, MySpace, Twitter, LinkedIn, Google +.

Redes Sociales



GESTIONAR

BibCiter, BibMe, Citation Machine, CiteULike, Connotea, EndNoteX, Mendeley, NoodleBib, Express, Ottobib, ProCite, RefWorks, Zotero.

CITAR

Manejadores de literatura

Recurso (clasificación)	Ejemplo
	
Varios	<p data-bbox="602 472 1398 579">PDF-XChange Viewer, PrimoPDF, 4Shared, Dropbox, <i>Diccionario de la Real Academia de la Lengua Española</i>, <i>Diccionario y Tesauro Merriam's Webster Online</i>, Logmein.</p> 

Tabla 1. Síntesis de la Web 2.0 y clasificación de los recursos.

Para acceder y manejar la literatura digital se pueden utilizar herramientas y **aplicaciones** (muchas de ellas gratuitas) para realizar una o varias de las funciones características de la Web 2.0 que en conjunto le añaden valor agregado a la información. Entre las funciones más comunes resaltan almacenar, etiquetar, jerarquizar, clasificar, editar, compartir y actualizar (Fig. 7).

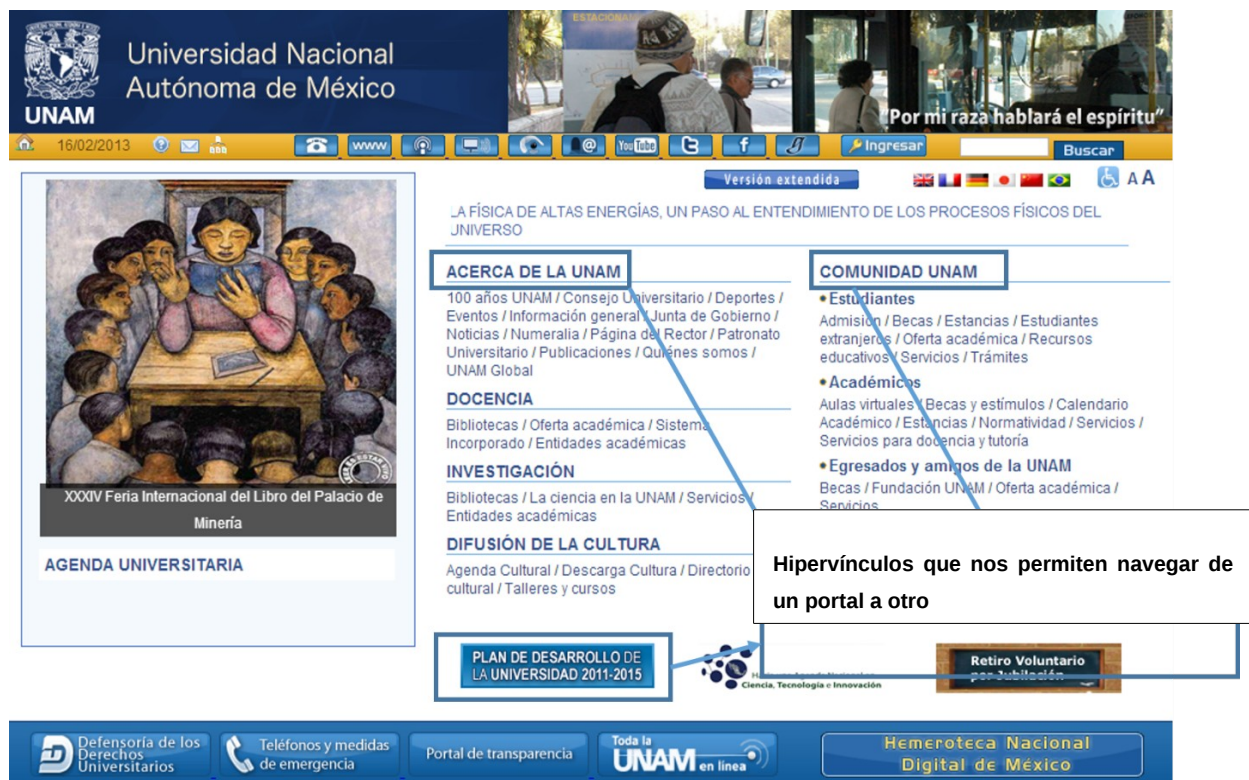


Figur
a 7. Herramientas de la Web 2.0 y sus funciones características. Las tablas corresponden a las numeradas en este capítulo.

A continuación se hace una descripción detallada de cada una de las categorías en las que se clasificaron las herramientas y se presentan los ejemplos más usados con su respectiva explicación.

3. Navegadores o exploradores / Navegar

Los nombres de navegador, navegador red, navegador web u hojeador (traducción literal aunque su uso es mínimo) provienen del inglés *web browser*, que se refiere al seguimiento en la computadora de los enlaces de una página a otra. Es un programa que permite visualizar la



información que contiene una página web (o electrónica) al interpretar el código **HTML** en el que está escrita. Puede estar alojada en la propia computadora, en un servidor local o en la red. El navegador presenta en pantalla el documento electrónico permitiendo al usuario interactuar con su contenido de dos formas: navegando (o explorando) hacia otra página mediante los enlaces o hipervínculos que contiene (Fig. 8) o buscando (Fig. 11).

Figura 8. Página web y la función de navegar.

Actualmente son muy frecuentes, además de las páginas web y los **portales** (conjunto de páginas web que comparten un dominio), versiones más dinámicas y sociales como son los blogs y las wikis, que también son visibles a través de los exploradores. **Blog** es el nombre corto de la

palabra *weblog* (en español puede traducirse como “bitácora”), una publicación en forma de diario o foro virtual interactivo donde se publica información general, personal o sobre algún tema específico lo que permite aclarar dudas, añadir comentarios, hacer preguntas, publicar artículos e información, entre muchas otras cosas (Fig. 9). Consta de entradas etiquetadas



Figura 9. Estructura de un blog.

Una **wiki** es un sitio web cuyas páginas pueden ser editadas por múltiples voluntarios a través del navegador, quienes pueden crear, modificar o borrar un mismo texto compartido. Los textos o páginas wiki tienen títulos únicos y conservan un historial de los cambios y es posible recuperar

fácilmente cualquier versión anterior e identificar quién hizo cada cambio, lo cual facilita enormemente el mantenimiento y el control de los usuarios destructivos. La primera wiki, la más grande, conocida y consultada es *Wikipedia* (<http://www.wikipedia.org>) (Fig. 10).

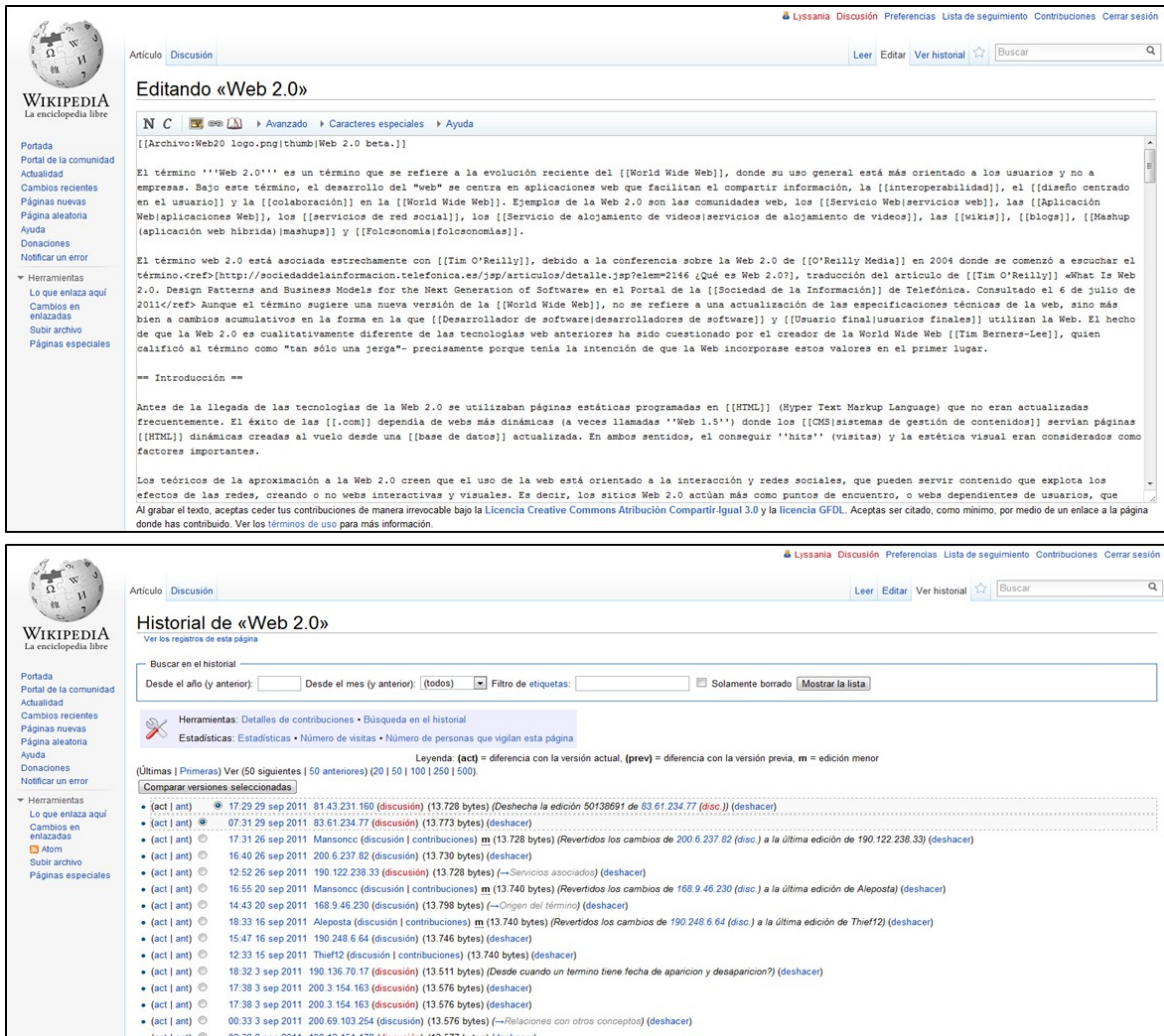


Figura 10. Edición e historial en Wikipedia.

Para localizar las páginas web, los blogs y las wikis publicadas en la web sobre algún tema determinado es necesario buscarlas entre las cientos de millones que se calcula existen hoy en día (una búsqueda con la letra "a" en Google actualmente da como resultado en más de 25 billones de páginas). Por eso, para acotar la búsqueda se utilizan programas llamados buscadores y metabuscadores.

4. Buscadores y metabuscadores / Buscar

Un buscador es un sistema de bases de datos que incorpora automáticamente páginas web mediante robots de búsqueda en la red, los cuales indizan las páginas web disponibles en los servidores gracias a su araña (*spider*). Las búsquedas se hacen a partir de palabras clave o por

Webpage Screenshot

The screenshot shows the Scientific WebPlus search interface. At the top, there is a navigation bar with the logo 'Scientific WebPlus BETA' and a search bar containing the text 'bioontology'. The search bar has a 'Search' button. Below the search bar, there are tabs for 'TOPIC', 'PERSON OR AUTHOR', 'ORGANISM', 'GENE', and 'MORE'. The main content area is titled 'All Results (Results 1-10 of 134)' and includes a 'Filter Results: Domain' section. The filter section shows a bar chart with the following data:

Domain	Count
.org	53
.com	42
.edu	20
.uk	3
.gov	2
Other	14

The main results list includes three entries:

- Pruning Bio-Ontologies - IEEE Xplore - Home**: A paper about pruning methods from ontology engineering. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076776&isnumber=4076362
- Project Prospect Features - Royal Society of Chemistry | Advancing...**: A paper about bio-ontology terms. URL: <http://www.rsc.org/Publishing/Journals/ProjectProspectFeatures.asp>
- Welcome | bioontology.org**: National Center for Biomedical Ontology. URL: <http://www.bioontology.org/>

Each result entry includes a '+ add tags' button, a 'bookmark' icon, an 'add comment' button, and a 'vote' section with up and down arrows.

<http://scientific.thomsonwebplus.com/SearchResults.aspx?query=bioontology&searchContext=TOPIC>

temas y pueden ser generales o especializadas (es decir, que establecen límites respecto a los campos de búsqueda). El resultado de la búsqueda es un listado de direcciones web en las que se mencionan las palabras clave buscadas o los temas.

Figura 11. Resultados de búsqueda en un buscador especializado.

Otro recurso es el **metabuscador**, que permite realizar búsquedas en distintos motores seleccionados respetando el formato original de cada buscador, de tal manera que exploran simultáneamente varios buscadores, analizan los resultados y presentan los resultados según un orden definido por el sistema estructural del metabuscador.

Los buscadores y metabuscadores (Tablas 2 y 3 respectivamente) se seleccionaron según el

criterio del sitio web Alexa, que muestra información jerarquizada (**ranking**) de los sitios más populares o más consultados.

Buscador	Editor	Características generales
Alexa http://www.alex.com	Amazon.com	Sitio web que provee información acerca de la cantidad de visitas que recibe un sitio web dado. Recolecta información de los usuarios que tienen instalado <i>Alexa Toolbar</i> , lo cual le permite generar estadísticas acerca de la cantidad de visitas y de los enlaces relacionados. También proporciona una gráfica donde se puede apreciar perfectamente el crecimiento o disminución de las visitas realizadas a cada página web.
Ask http://www.ask.com	InterActive Corporation	Primer buscador comercial de tipo pregunta-respuesta. Soporta una amplia variedad de consultas de usuarios realizadas en inglés, así como las tradicionales búsquedas con palabras claves; consigue que las búsquedas sean más intuitivas y amigables que otros buscadores; posee tecnología de búsqueda basada en temas de popularidad para calcular el grado de autoría y posee una licencia para calcular la popularidad de cada clic. El algoritmo de ExpertRank provee resultados ordenados según la autoría de los sitios web. Permite guardar las búsquedas y hacer anotaciones en ellas.
Bing http://www.bing.com	Microsoft Corporation	Bing es el sustituto de <i>Live Search</i> y <i>Yahoo!</i> y fue puesto en línea el 3 de junio de 2009 con una versión preliminar publicada el 1 de junio del mismo año. Incluye una lista de sugerencias de búsqueda en tiempo real y una lista de las búsquedas relacionadas; está basado en tecnología semántica de PowerSet que Microsoft compró en el 2008. Bing incluye también la posibilidad de guardar y compartir el historial de búsqueda a través de <i>Windows Live SkyDrive</i> , <i>Facebook</i> y correo electrónico. El 29 de julio de 2009 Microsoft y Yahoo! anunciaron que habían hecho un trato en el que Bing reemplazaría el motor de búsqueda de Yahoo!
Exalead http://www.exalead.com	Qualis, S. C. A.	Exhibe estadísticas e información categórica y geográfica sobre los resultados en el lado izquierdo de la página y permite la búsqueda de términos próximos, uso de expresiones jerarquizadas de la búsqueda, consejos de búsqueda específica (por localización geográfica, idioma y tipo del archivo) sugiere palabras adicionales. Permite búsqueda fonética (para utilizarse cuando se desconoce la ortografía), búsqueda dentro de los resultados obtenidos, y búsqueda con expresiones regulares.
Google http://www.google.com	Google Inc.	Google es el buscador en Internet más grande y más usado. Usa un algoritmo para ordenar la relevancia de los documentos conocido también como PageRank que es muy efectivo. Usa varias arañas (<i>spiders</i>) para recolectar y ordenar

Buscador	Editor	Características generales
Wikia http://answers.wikia.com/wiki/Wikianswers	Wikia	<p>la información. La araña que más tiempo lleva rastreando la red es el <i>Googlebot</i>, encargado de recoger las ligas que después se encontrarán en Google. Además cuenta con otros robots como <i>FreshBot</i>, que escanea los principales sitios con información actualizada frecuentemente, como los portales de noticias. Cuando una página ya no está disponible muestra la última versión guardada en el caché.</p> <p>Wikia es un buscador creado por la organización del mismo nombre; usa el buscador Grub, que usa computación distribuida y aprovecha los ciclos de inactividad de los clientes conectados para que su araña web tenga mayor potencia. Ofrece la opción de que los usuarios mejoren los resultados de manera colaborativa introduciendo enlaces de lo que consideran que se debería mostrar en una búsqueda determinada e incluye mini-artículos explicativos sobre algunos sitios web cuando estos son buscados.</p>

Tabla 2. Buscadores generales más utilizados.

Meta buscador	Editor	Características generales
Copernic http://www.copernic.com	Copernic Inc.	Además de ser un metabuscador en línea, se instala en el ordenador. Existe una versión gratuita y otra de pago. Entre sus ventajas más relevantes se encuentran el filtrado, la clasificación de los resultados, sus criterios propios de pertinencia y/o posibilidad de salvar y recuperar las búsquedas para actualizar los resultados.
Dogpile http://www.dogpile.com	InfoSpace Inc.	Hace búsquedas en About, Ask.com, FinWhat, Google, LookSmart, MSN Search, Teoma y otros buscadores populares; permite ver la búsqueda en tiempo real.
Ipselon http://ipselon.com/es/index.html?q=&cx=008968604428227682493:8qmmgtg-gfc&ie=UTF-8&cof=FORID:9#	Ipselon	Es un metabuscador que permite realizar búsquedas por rango de fechas, ubicación, tipo de archivo, sitios web y sugerencias. Rastrea resultados por Google, Yahoo! o Bing y ofrece un <i>widget</i> para descargar e instalarlo en el escritorio. También incorpora otras utilidades como el Suggest, función que sugiere de forma automática las palabras clave más buscadas.
Ixquick http://www.ixquick.com	Surfboard Holding BV	Da relevancia especial a los primeros diez resultados de los buscadores que utiliza: Teoma, Entireweb, Gigablast, Go, LookSmart, Netscape, ixdmoz, goto, Wisenut. Usa un sistema de estrellas (<i>Star System</i>) para calificar sus resultados, coloca una estrella por cada resultado devuelto dentro de los diez primeros

Meta buscador	Editor	Características generales
Kartoo http://www.kartoo.com	Kartoo	por cada buscador. En la página principal tiene la opción para elegir cuáles buscadores usar. Devuelve los resultados de la búsqueda en forma de mapas. Según la relevancia de los sitios encontrados aparecen esferas más o menos grandes. Además se pueden realizar búsquedas avanzadas por temas que se pueden ir ajustando dependiendo de los criterios de búsqueda.
Mamma http://www.mamma.com	Copernic, Inc.	Busca simultáneamente en AltaVista, Infoseek, Lycos, Excite, Webcrawler, About y AlltheWeb.
Metacrawler http://www.metacrawler.com	InfoSpace Inc.	Envía las consultas formuladas a Google, Yahoo! Search, Bing, Ask Jeeves, About, MIVA, LookSmart y muchos más.
Turbo10 http://www.turbo10.com	Turbo10 Limited	Es capaz de trabajar con más de 1, 000 buscadores diferentes, de entre los cuales cada usuario puede personalizar su propio paquete de 10, guardarlos y realizar las búsquedas. No permite el empleo de signos de puntuación o expresiones lógicas, algo que según sus desarrolladores no es necesario. Tiene una filosofía de búsqueda diferente que pretende conectar a miles de buscadores especializados, mejorando los resultados de búsqueda al construir un listado ponderado.
Webcrawler http://www.webcrawler.com/webcrawler_yaylf/ws/index/_iceUrlFlag=11?IceUrl=true	Infospace Inc.	Ofrece a los usuarios los resultados de búsqueda de Google, Bing y otros motores de búsqueda populares, también proporciona resultados multimedia, incluyendo imágenes, video, noticias e información local. Los resultados de búsqueda son una combinación de los mejores comerciales (anuncios patrocinados) y no comercial (algoritmos) de los motores de búsqueda más populares en la Web.
Yippy http://yippy.com	Yippy, Inc.	Rastrea e integra información de otros buscadores y fuentes muy variadas, como Ask, Gigablast, Bing, Open Directory, Wikipedia, NY Times, Yahoo News y Shopzilla, además de los enlaces patrocinados. Combina los resultados y genera una lista ordenada basada en un <i>ranking</i> comparativo y los agrupa por <i>clusters</i> (categorías).

Meta buscador	Editor	Características generales
---------------	--------	---------------------------

Tabla 3. Metabuscadores más comunes.

Los buscadores que se pueden clasificar en generales son aquellos que tienen como intención registrar las páginas web existentes, mientras que los especializados registran determinados tipos de páginas con base en criterios particulares, por lo que permiten hacer búsquedas más específicas sobre el tema que se desea. En este caso, seleccionamos los buscadores especializados que tuvieran mayor cantidad de páginas web para biología. Todos son de acceso libre.

Buscador especializado	Editor	Características generales
Bio-Netbook http://www1.pasteur.fr/recherche/BNB/bnb-en.html	Instituto Pasteur	Recopilado por el Instituto Pasteur, es un repertorio de páginas web de biología con un motor de búsqueda que permite hacer interrogaciones muy especializadas.
Google Scholar http://scholar.google.com	Google	Permite buscar bibliografía especializada desde un solo sitio y en un gran número de disciplinas y fuentes como estudios revisados por especialistas, tesis, libros, resúmenes y artículos de editoriales académicas, sociedades profesionales, depósitos de impresiones preliminares, universidades y otras organizaciones académicas. Ayuda a encontrar el material más relevante dentro del mundo de la investigación académica.
Intute http://www.intute.ac.uk	The Intute Consortium	Es un servicio gratuito que ayuda a encontrar recursos en Internet sobre agricultura, ciencias biológicas, arquitectura, comunicación, ingeniería, matemáticas, medicina, humanidades y ciencias sociales.
NextBio http://www.nextbio.com/b/nextbio.nb	NextBio	Provee una plataforma que permite a investigadores en ciencias biológicas buscar, descubrir y compartir conocimientos restringidos. Se conecta a grandes bases de datos como ArrayExpress, Stanford Microarray Database y Gene Expression Omnibus.
OmniMedical http://www.omnimedicalsearch.com	OmniMedicalSearch.com	Localiza información médica en seis categorías: web, noticias, imágenes, foros, profesionales médicos y compras.
Scirus http://www.scirus.com	Elsevier B.V.	Localiza información científica, académica, técnica y médica en la web. Encuentra los últimos informes, artículos revisados, las patentes, las impresiones y las revistas antes que otros motores de búsqueda. Ofrece funcionalidades únicas diseñadas para científicos e investigadores. Se

Buscador especializado	Editor	Características generales
Scientific WebPlus http://scientific.thomsonwebplus.com/BasicSearch.aspx?lbu=WebOfKnowledge&key=49jf934i20s0g4i290s0&bhcp=1	Thomson Reuters	centra sólo en las páginas web que contienen contenido científico. Buscador abierto creado por Thomson Reuters que aprovecha el poder de su experiencia editorial, vocabularios controlados y algoritmos de relevancia. Está diseñado para complementar los resultados de la búsqueda con los recursos web más relevantes y de vanguardia para el investigador profesional.

Tabla 4. Buscadores más útiles para localizar páginas con información biológica, relacionada o afín.

5. RSS o agregadores de feeds / Actualizar

Dentro de todos los recursos web actuales, son de los más innovadores y útiles los agregadores; en especial, los RSS (*Really Simple Syndication*) responsables de automatizar la inmediatez, que se refiere a enterarse en el instante mismo en el que se ha publicado algún cambio o noticia en una página web (usualmente blog) al que se está suscrito mediante la distribución de contenidos categorizados que alimentan automáticamente a otros sitios y programas lectores (*readers*). Estas herramientas ayudan a conectar a los usuarios con aquellas fuentes que son de su interés y mantenerse actualizados sin necesidad de saturar el correo electrónico de mensajes. Una de las principales cualidades de la sindicación es que posibilita un monitoreo inteligente de la información a través de **agregadores** o **fuentes web**, simplificando enormemente la tarea de encontrar información útil. Esta es una tecnología representativa de la Web 2.0, ya que el usuario puede enlazar o etiquetar una página web (Fig. 12), pero también su contenido al recibir notificaciones en un solo lugar cada vez que se produce una actualización, sin necesidad de consultar distintas páginas (blogs, revistas, etc.).

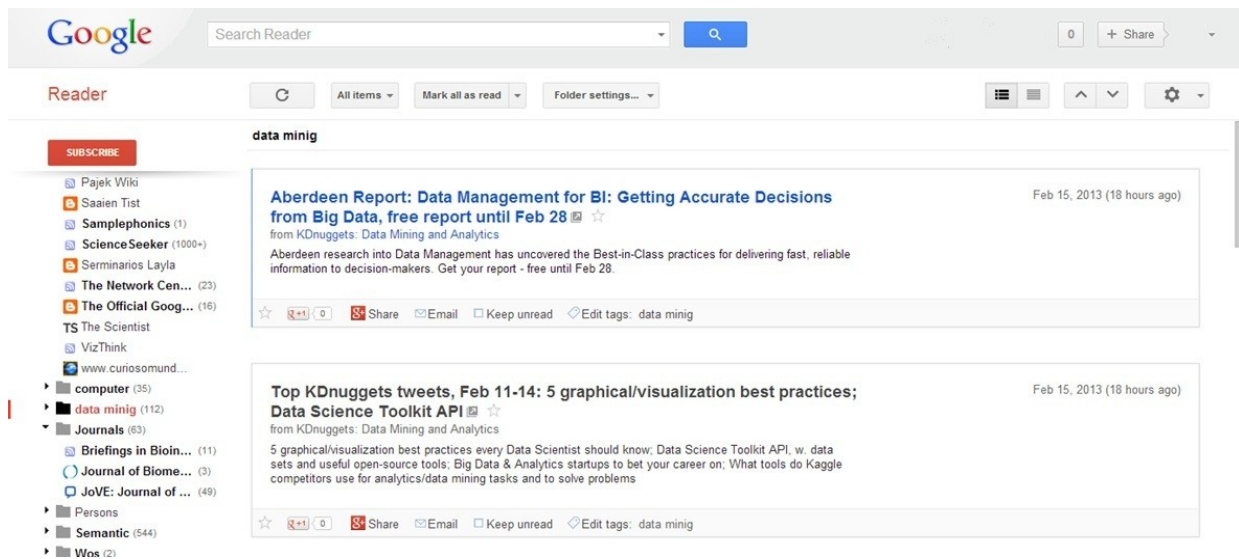


Figura 12. Ejemplo de RSS Feed de Google.

6. Marcadores / Guardar

Los llamados **marcadores** o **favoritos** son herramientas que permiten guardar la **dirección electrónica** (URL) de una página web seleccionada. En la actualidad, la mayoría de ellos ofrecen las funciones de etiquetar, organizar y compartir las ligas almacenadas y presentan la opción de una arquitectura diseñada bajo la idea de la colectividad del conocimiento, que permite recoger las opiniones de todos los que participan, para establecer los contenidos de mayor relevancia o los más populares (*ranking*) y crear, por ejemplo, listas de las páginas más vistas. Los marcadores se clasificarán en este libro de acuerdo con los siguientes criterios: 1) el lugar donde se almacenan (en la memoria de la computadora personal o en línea); 2) si se pueden compartir o no (sociales o personales); y 3) los objetos que guardan (páginas web, imágenes o bibliografía) (Tabla 5).

Existen muchos servicios de marcadores, como los incluidos en los propios exploradores y buscadores, que guardan las direcciones de las páginas marcadas en la memoria de computadora; los más comunes son los “Favoritos” de Explorer y Firefox. Si es posible compartir las páginas guardadas con otros usuarios, son sociales; si no, son sólo de uso personal (aunque la tendencia ha sido permitir la sociabilidad). Además, actualmente existen marcadores que guardan las

páginas web registradas en línea (en servidores) a los cuales se puede acceder desde cualquier computadora porque se actualizan y además son sociales. Lo ideal es utilizar los marcadores en línea y sociales como Google Bookmarks, Delicious y Diigo, que tienen gran variedad de funciones y son compatibles con otros.

Dentro de los marcadores, se han producido aplicaciones especializadas para guardar algún tipo de información específica, por ejemplo, las mencionadas anteriormente sirven muy bien para administrar páginas web, pero hay otras como VisualizeUs, que es un servicio conveniente para marcar, etiquetar y compartir imágenes específicamente, o CiteULike, que permite marcar bibliografía con todas las ventajas mencionadas anteriormente.

Una de las funciones posibles en la Web 2.0 más relevantes consiste en la asignación de etiquetas (**folksonomía** *otagging*) a las páginas electrónicas con base en las categorías que cada usuario establezca; este procedimiento permite generar datos producidos por la participación de miles de usuarios. El sistema colaborativo usado por muchas aplicaciones en red, como las redes sociales y los marcadores, materializa la arquitectura de la participación, así como las ideas de la inteligencia colectiva y la intercreatividad. Muchos sistemas utilizan esta organización cooperativa al sugerir al usuario las etiquetas que un contenido ya tiene registradas y que han sido empleadas por los usuarios anteriores. Esta manera de clasificar la información que circula por Internet se basa en una estructura social construida sobre el principio de “cómo ponerse de acuerdo”, al sumar la colaboración de todos bajo un mismo objetivo y aprovechar el conocimiento de los demás. Es decir, por una parte es conveniente apoyar los métodos y herramientas taxonómicas que hacen posible clasificar, etiquetar, jerarquizar y ordenar y por otra, incorporar nuevas metodologías orientadas a organizar de manera colectiva aquella información relevante. Esta es una de las actividades clave que han dado a la Web 2.0 características emergentes como intercreatividad, inteligencia colectiva, multitudes inteligentes, sabiduría de las multitudes y arquitectura de la participación. La Web 2.0 se ha convertido en un laboratorio en constante evolución bajo esta idea (un ejemplo característico son las opciones Beta). A mayor colaboración de los usuarios se logra un menor nivel de incertidumbre informacional.

De tal manera que los marcadores personales y sociales pueden ser utilizados como administradores de favoritos para almacenar, etiquetar, organizar y compartir colectivamente los hipervínculos más relevantes en la red de manera sencilla y gratuita. Facilitan la obtención de la

bibliografía digital y el reingreso a la dirección electrónica de donde se obtuvo la información, así como la obtención de los metadatos del documento. La mayoría de los marcadores sociales incorporan dentro de sus funciones la de etiquetar, como herramienta para facilitar la búsqueda de información dentro del sitio y para encontrar personas con los mismos intereses.

Los marcadores sociales son sistemas que permiten añadir enlaces a un sitio público de tal forma que siempre estén disponibles en cualquier computadora con acceso a Internet de tal forma que las fuentes de información quedarán almacenadas en una cuenta. Su funcionamiento tiene un carácter colectivo pues se pueden compartir información con todos los usuarios que se desee, aspecto muy importante en los grupos de investigación que generalmente son multidisciplinarios. De esta forma un grupo de investigación puede tener un sitio exclusivo en el que guardan todas las fuentes pertinentes e independientes del lugar donde se encuentre cualquiera de los integrantes del grupo.

Marcadores	Características generales
Favoritos (P)	Almacena direcciones electrónicas (URL) en la computadora personal; no se pueden compartir.
Herramienta de Internet Explorer	
Marcadores (P)	Almacena direcciones electrónicas (URL) en la computadora personal; no se pueden compartir.
Herramienta de Firefox	
Google Bookmarks (S) http://www.google.com/bookmarks Herramienta de Google	Servicio en línea que permite guardar los sitios web favoritos y agregar etiquetas y notas. Permite el acceso desde cualquier computadora que tenga instalada la barra de Google. También permite exportar los favoritos a un marcador social como Delicious o importar los favoritos almacenados a la PC a través de los navegadores mencionados.
Delicious (S) http://delicious.com	Servicio de marcadores sociales que permite a los usuarios etiquetar, almacenar, gestionar y compartir páginas web a partir de una fuente centralizada. En vez de tener diferentes marcadores, facilita tener un solo conjunto de marcadores y mantenerlos sincronizados entre todos los equipos en los que se instala. Incluso si se usa una computadora de uso común, se pueden usar directamente en la página web de Delicious; además de compartirlas con otros usuarios.
Diigo (S)	Diigo ofrece un complemento (<i>add-on</i>) en el navegador que permite marcar la página y resaltar las partes de las páginas web

Marcadores	Características generales
http://www.diigo.com	<p>que son de particular interés para el usuario. También se pueden adjuntar notas adhesivas permanentes a partes específicas de una página web, de tal forma que cada vez que se vuelve a la página web original se pueden ver los textos que se han resaltado y las notas adhesivas superpuestas; además de compartirlas con otros usuarios.</p>
<p>Evernote (S)</p> http://www.evernote.com/about/intl/es	<p>Es una aplicación informática destinada a la organización de información personal por medio de notas. Existen versiones que se instalan para diversos sistemas operativos y la versión web. La versión para windows tiene soporte para pantallas táctiles y reconocimiento de escritura. Permite la sincronización entre varias computadoras y hay una buena versión móvil, además de compartirlas con otros usuarios.</p>
<p>Picasa (S)</p> http://picasa.google.com	<p>El programa Picasa permite el inventariado de todos los archivos gráficos de la computadora, su clasificación, orden e incluye además herramientas de edición y retoque fotográfico. El programa interactúa con Picasa web permitiendo colocar las fotos directamente en los álbumes.</p>
<p>StumbleUpon (S)</p> http://www.stumbleupon.com	<p>Marcador social con clasificación por temas y formato de vídeo, imágenes y noticias; se puede compartir a través de redes sociales y agregar comentarios.</p>
<p>VisualzeUs (S)</p> http://vi.sualize.us	<p>Sitio web de marcadores sociales de contenidos visuales; permite almacenar, compartir, etiquetar y guardar las imágenes favoritas junto con la dirección del documento electrónico que las contiene.</p>
<p>YouTube (S)</p> http://www.youtube.com	<p>Permite descubrir, mirar, etiquetar, marcar, traducir y compartir videos. Ofrece un foro para que la gente se conecte y se informe; también actúa como una plataforma de distribución para creadores de contenido y publicistas originales, grandes y pequeños.</p>

(S)=marcadores sociales; (P)=marcadores personales.

Tabla 5. Ejemplos más comunes de marcadores.

7. Redes sociales / Compartir

Las **redes sociales** son servicios que ofrecen un espacio virtual para escribir y compartir contenidos multimedia con personas de intereses similares que contribuyen a fortalecer las relaciones sociales, en su mayoría son gratuitas y de fácil uso. La popularidad de estas tecnologías ha ido a la par de un aumento en los niveles de intercambio de contenidos a través de la red; esto ha hecho de la web un medio más social en el que se produce y se consume información, pero también sirve para comunicarse, entretenerse y compartir. Este fenómeno del uso colectivo de las tecnologías ha contribuido al valor agregado de la información y la reciprocidad de la red, es decir, mientras más personas la usan, se vuelve cada vez mejor.

A continuación mencionan las redes sociales científicas con mayor difusión y uso en la actualidad. Entre sus funciones destaca la de agrupar a especialistas en ciencia; en muchas de ellas existen grupos de discusión sobre biología, servicios de literatura y bibliotecas.

Existe un gran número de redes sociales científicas como BiomedExperts (<http://www.biomedexperts.com>), AuthorMapper (<http://authormapper.com>) WorldWideScience (<http://worldwidescience.org/index.html>), WebMD (<http://www.webmd.com>), Medscape (<http://www.medscape.com>), Physician's Online (<http://www.physiciansonline.com>), LinkedIn (<http://www.linkedin.com>) y ResearchGate (<http://www.researchgate.net>). Su uso es común y cada vez más frecuente; sin duda alguna, es una herramienta que, utilizada adecuadamente, permitirá aumentar y mejorar la colaboración científica.

Red social	Características generales
Badoo http://www.badoo.com	Badoo <cuyo lema "Yo estoy aquí"> es un sitio web de redes sociales creado en la ciudad de Londres. La idea era crear una red social de Internet que pudiera romper barreras a nivel mundial pero, al mismo tiempo, resguardar de manera local la identidad cultural de los usuarios. El objetivo era permitir a la gente compartir sus vidas y atraer la atención de otras personas a nivel global y local (<i>glocal</i>). Hoy Badoo se ha convertido en una de las principales 500 webs del mundo, según Alexa, y goza de una gran popularidad como red social en los mercados europeos y americanos.
Facebook	Facebook es un sitio web gratuito de redes sociales creado por Mark

Red social	Características generales
http://www.facebook.com	<p>Zuckerberg. Originalmente era un sitio para estudiantes de la Universidad de Harvard, pero actualmente está abierto a cualquier persona que tenga una cuenta de correo electrónico. Los usuarios pueden participar en una o más redes sociales, agrupados por su situación académica, su lugar de trabajo o región geográfica. Ocupa el sitio número cuatro dentro del Top 500 de Alexa.com.</p>
<p>Hi5</p> <p>http://www.hi5.com</p>	<p>Es una red social lanzada en 2003 y uno de los 500 sitios web más visitados del mundo según Alexa. Hi5 es famoso por su interactividad, pues hace de una simple cuenta de usuarios una especie de tarjeta de presentación virtual.</p>
<p>MySpace</p> <p>http://www.myspace.com</p>	<p>Es un sitio web de interacción social formado por perfiles personales de usuarios que incluye redes de amigos, grupos, blogs, fotos, vídeos, música, una red interna de mensajería que permite comunicarse y un buscador interno.</p>
<p>Twitter</p> <p>http://www.twitter.com</p>	<p>Twitter es un servicio gratuito de microblogging innovador y en boga que hace las veces de red social y permite a sus usuarios enviar micro-entradas basadas en texto, denominadas Tweets, de una longitud máxima de 140 caracteres. El envío de estos mensajes se puede realizar tanto por el sitio web de Twitter, como por vía SMS (<i>Short Message Service</i>) desde un teléfono móvil, desde programas de mensajería instantánea, o incluso desde cualquier aplicación de terceros, como puede ser Twitterrific, Tweetie, Facebook, Twinkle y TweetDeck.</p>
<p>LinkedIn</p> <p>http://www.linkedin.com</p>	<p>LinkedIn es un sitio web orientado a negocios. Fue fundado en diciembre de 2002 y lanzado en mayo de 2003, principalmente para red profesional. Es una red social orientada a un perfil de público profesional que cuenta actualmente con más de 65 millones de usuarios y empresas. Tiene la finalidad de favorecer el intercambio de información, ideas y oportunidades entre estos, fomentando el capital social, los emprendimientos y las alianzas estratégicas, entre otros tópicos para generar negocios beneficiosos para los usuarios de esta red social. LinkedIn es catalogada como una red social, sin embargo, se distingue de las conocidas redes sociales como Facebook, MySpace o Twitter porque no permite almacenar o compartir fotos, videos, modificar diseños de perfil y otras cosas, debido a que está orientada al currículum y este es el que se puede subir y modificar.</p>
<p>Google +</p> <p>https://plus.google.com</p>	<p>Google + permite relacionarse en Internet como en la vida real. Comparte lo que piensas, enlaces y fotos con los círculos que quieras. Utiliza un vídeo chat sencillo y espontáneo para hablar con hasta nueve personas simultáneamente.</p>

Tabla 6. Redes sociales.

8. Manejadores de bibliografía / Citar

La investigación biológica implica de forma indispensable la revisión bibliográfica porque la literatura es fuente y producto del nuevo conocimiento biológico. La investigación bibliográfica consiste en tomar fuentes escritas, bien sean libros, publicaciones periódicas, tesis u otros materiales impresos y digitales. En vista del avance tecnológico y la disponibilidad generalizada de los documentos electrónicos (*e-books*, *e-journals*, *e-libraries*) y su registro digital, ahora es posible guardar y organizar las referencias bibliográficas por medio de marcadores y manejadores o gestores de bibliografía disponibles en la red, los cuales permiten importar, organizar, exportar, editar y compartir referencias bibliográficas, así como crear bibliografías personales, dar formato e incluso algunos cuentan con la función “citar mientras escribo”, que permite insertar la cita correspondiente al documento que se está escribiendo y agregar la bibliografía correspondiente en el estilo predeterminado que elija el usuario de manera automática y personalizada.

En parte, la razón de su éxito es que no es necesario poseer conocimientos especiales para usarlos y rinden beneficio inmediato a cada usuario y a la comunidad por medio de una creación colaborativa, pues en muchos casos estos servicios también son redes sociales que han adquirido gran cantidad de usuarios en poco tiempo. El servicio mejora cuanto más gente participe, creando gran cantidad de información en un breve periodo de tiempo. Es tal la cantidad de tiempo y los recursos que ahorran estos servicios que muchas editoriales científicas, como es el caso de BMC (<http://www.biomedcentral.com>) hacen descuento en el costo de publicación si se utilizan.

La mayoría de los manejadores (Tabla 7) tienen en común que son sistemas abiertos, es decir, poseen una interfaz web que permite acceder a las referencias desde cualquier ordenador con conexión a Internet e incluyen las siguientes tareas básicas:

1. Crear una base de datos bibliográfica personal o grupal para almacenar referencias importadas o añadidas manualmente.
2. Buscar y recuperar referencias relacionadas.
3. Organizar las referencias con base en distintos criterios.
4. Dar formato según los estilos bibliográficos más usados (MLA, Chicago, etc.).
5. Opción de respaldar, exportar y compartir la bibliografía.

Por ejemplo, actualmente Mendeley es una de las mejores opciones de este tipo pues permite almacenar, compartir, ordenar y sistematizar las referencias consultadas. De hecho, se utilizó para citar y poner las referencias en este libro, debido a las siguientes ventajas:

1. Acceso libre
2. Sincronización entre la versión de escritorio y la de red.
3. Acceso remoto a la bibliografía desde cualquier equipo conectado a Internet.
4. Se puede compartir la bibliografía con otros usuarios.
5. Permite cargar el documento en formato **PDF** y la ficha bibliográfica.
6. Sincronización con el marcador de bibliografía CiteULike.
7. Extracción automática de la referencia bibliográfica de un documento por medio de **minería de textos**. Si uno importa todo un archivo que contenga documentos en PDF, genera la bibliografía y la adiciona junto con los archivos.

Nombre del manejador de bibliografía	Desarrollador URL	Sistema operativo	Formato de importación	Formato de exportación	Estilo de citas	Formato de archivo de lista de referencias	Plug-in para procesador de texto	Conexión a base de datos
BibCiter	BibCiter developers http://bibciter.sourceforge.net	Windows, Mac OS X, Linux, BSD, Unix	BibTeX	BibTeX	APA, Harvard, otros; Método de extensión: plantilla de exportación de BibCiter	HTML, RSS	Ninguno	Ninguna
BibMe (OL)	Bibme.org http://www.bibme.org	No aplica, basado en un servidor web	Solo crea la bibliografía manualmente o a partir del ISBN	Word (RTF)	APA, Chicago/Turabian, MLA	HTML	No, pero permite exportar la bibliografía en formato RTF a Word.	Amazon, FindArticles, Yahoo! News, CiteULike Academic Papers
Landmark Project	The Landmark Project	No aplica, basado en un servidor	Sólo crea bibliografía manualmente	No tiene	APA, MLA, Chicago/Turabian	HTML	No, solo permite copiar y	No

Administrador de bibliografía	Desarrollador URL	Sistema operativo	Formato de importación	Formato de exportación	Estilo de citas	Formato de archivo de lista de referencias	Plug-in para procesador de texto	Conexión base de datos
CiteSpace	http://citationmachine.net/index2.php	web	de y de un solo documento a la vez				pegar en Word	
CiteULike	Richard Cameron http://www.citeulike.org	No aplica, basado en un servidor web	BibTeX	BibTeX, RIS, Otros (COinS)	APA, Chicago/Turabian, Harvard, MLA, Otros	RSS	Ninguno	Amazon, ArXiv, CiteSeer, IEEE Xplore, PubMed, Otros
Connotea	Nature Publishing Group http://www.connotea.org	Windows, Mac OS X, Linux, BSD, Unix	BibTeX, Endnote/Refer/BibIX, ISI, MODS XML, RIS, otros (marcadores de Firefox)	BibTeX, Endnote/Refer/BibIX, MODS XML, RIS, otros (RDF)	No tiene	RSS	Ninguno	ArXiv, CiteSeer, PubMed, Otros
EndNoteW	Thomson Corporation http://www.myendnoteweb.com/EndNoteWeb.html	Windows, Mac OS X	CSA, Endnote/Refer/BibIX, ISI, Medline, Ovid, PubMed, RIS, SciFinder, Otros	BibTeX, RIS, otros (CSV)	APA, Chicago/Turabian, Harvard, MLA, Otros; Método de extensión: Formato EndNote a través de la interfaz gráfica de usuario	HTML, RTF, Texto plano, Otros (clipboard, XML)	Microsoft Word, OpenOffice.org/Writer	PubMed, Otros
Mendeley	Mendeley http://www.mendeley.com	Windows, Mac OS X, Linux	BibTeX, Endnote/Refer/BibIX, RIS, Otros (Marcadores o favoritos)	BibTeX, Endnote/Refer/BibIX, RIS, Otros (EndNote XML)	APA, Chicago/Turabian, Harvard, MLA, Otros; Método de extensión: CLS	RTF, Texto plano	Microsoft Word, OpenOffice.org/Writer	ArXiv, CiteSeer, IEEE Xplore, PubMed, Otros

Nombre de gestor de bibliografía	Desarrollador	Sistema operativo	Formato de importación	Formato de exportación	Estilo de citas	Formato de archivo de lista de referencias	Plug-in para procesador de texto	Conexión base de datos
NoodleTools, Inc. (OL)	http://www.noodletools.com/tools/index.php	No aplica, basado en un servidor web	Extrae la información de algunas bases de datos	Word	APA, MLA, Chicago/Turabian	HTML	No, pero permite exportar la bibliografía a un formato de Word	EBSCO
Jonathan Otto (L)	http://www.ottobib.com	No aplica, basado en un servidor web	Crea la bibliografía a partir del ISBN	Sólo permite imprimir	APA, Chicago/Turabian, MLA, Otros (formato Wikipedia para libros)	Texto plano	No	Amazon libros y Google Books
Thomson Corporation	http://www.procite.com	Windows	CSA, ISI, Medline, Ovid, PubMed, RIS, Otros	RIS	APA, Chicago/Turabian, Harvard, MLA, Otros	HTML	Microsoft Word, Otro (Word Perfect)	PubMed, Otros (varios)
RefWorks (P, L)	http://www.refworks.com	No aplica, basado en un servidor web	BibTeX, Copac, CSA, Endnote/Refer/BibIX, ISI, Ovid, RIS, SciFinder	BibTeX, RIS, Otros	APA, Chicago/Turabian, Harvard, MLA, Otros; Método de extensión: a través del formato de la interfaz gráfica del usuario	RTF, Texto plano, RSS, Otros (DOC, ODT)	Microsoft Word	PubMed, Otros
Center for History and New Media (OL)	http://www.zotero.org	Windows, Mac OS X, Linux, BSD, Unix	BibTeX, Endnote/Refer/BibIX, ISI (parcial), MODS XML, Ovid (parcial), PubMed (parcial: directamente	BibTeX, Endnote/Refer/BibIX, MODS XML, RIS, Otros (RDF, plantilla de citas de Wikipedia)	APA, Chicago/Turabian, Harvard, MLA, Otros, Método de extensión: CLS	HTML, RTF	Microsoft Word, OpenOffice.org/Writer, Otros (RTF Scan)	ArXiv, CiteSeer, IEEE Xplore, PubMed, Otros

Manejador de bibliografía	Desarrollador	Sistema operativo	Formato de importación	Formato de exportación	Estilo de citas	Formato de archivo de lista de referencias	Plug-in para procesador de texto	Conexión base de datos
---------------------------	---------------	-------------------	------------------------	------------------------	-----------------	--	----------------------------------	------------------------

de página web, no desde archivo guardado), RIS, Otros(COinS, MARC, RDF, unAPI)

(C)costo por licencia; (P) requiere password; (OL) versión en línea.

Tabla 7. Comparación de características básicas entre algunos manejadores de bibliografía.

9. Varios

Es tanta la cantidad de recursos disponibles -más los que se generan diariamente- y son tan diferentes que resulta imposible colocarlos todos en una sola categoría bajo una misma particularidad ya que comparten un gran número de aplicaciones disponibles en la web para manejo de literatura. En la categoría “Varios” se han incluido todos aquellos recursos básicos que resuelven problemas comunes de procedimiento, formato o accesibilidad. Todos ellos son libres y se seleccionaron de entre sus similares por mostrar eficiencia en su desempeño y por la sencillez de su manejo (Tabla 8).

Asimismo, se incluyen en esta clase a los *add-ons*, *gadgets*, *plug-ins* y *widgets*, pequeños programas que se instalan como complementos de otros programas principales, diseñados para mejorar una aplicación, servicio de una computadora, o bien cualquier tipo de interacción a través de la web (Tabla 9). Más adelante se muestran las aplicaciones y complementos que ofrece Google y se presenta una imagen con una pantalla de Google Chrome personalizada con algunos de los complementos mencionados (Figs. 12 y 13).

Complemento	Descripción
4Shared almacenamiento en línea http://www.4shared.com	4Shared tiene un solo propósito: almacenar y compartir archivos en línea; pueden tener cualquier formato aunque los más compartidos son las imágenes y los vídeos. Los usuarios no registrados pueden descargar los archivos gratis, pero si se registran tienen más beneficios. Los archivos se pueden proteger con una contraseña para mayor seguridad, una característica muy útil para estudiantes y para profesionales que desean hacer sus proyectos exclusivos para un grupo. Es útil en presentaciones y agendas de reuniones altamente confidenciales.
Diccionario de la Real Academia de la Lengua Española http://buscon.rae.es/draeI	El Diccionario académico es actualmente una base informática de datos con más de 200, 000 registros, además ha adquirido el compromiso de ir haciendo públicas con periodicidad semestral las adiciones, supresiones y enmiendas que la Real Academia Española y sus Academias asociadas vayan aprobando. Contiene el texto de la última edición en papel, en este caso, la vigésima segunda, de 2001 con el conjunto de modificaciones aprobadas. Los consultantes acceden inicialmente a la vigésima segunda edición y, en los casos en que se añada un nuevo registro o un artículo haya sido modificado, verán en la pantalla un aviso que

Complemento	Descripción
	<p>les permitirá contemplar la nueva versión. Ofrece instalar un botón en la barra de herramientas de Google que permite el acceso inmediato al diccionario.</p>
<p>Diccionario y Tesauro Merriam's Webster Online</p>	<p>El diccionario Merriam-Webster contiene más de 225, 000 definiciones, frases, expresiones idiomáticas y ejemplos de uso. Las definiciones incluyen pronunciaciones, etimologías, derivados e ilustraciones, más enlaces a palabras y definiciones relacionadas.</p>
<p>Diccionario y Tesauro Español-Inglés-Español</p>	<p>El tesauro Merriam-Webster contiene más de 340, 000 sinónimos, antónimos y expresiones idiomáticas, así como palabras contrastadas y relacionadas. Cada entrada del tesauro presenta una definición breve que describe los significados que comparten los sinónimos, estableciendo vínculos con otras entradas, frases o expresiones idiomáticas que constan en él. Se pueden instalar complementos para el navegador Firefox, botones en la barra de Google y en la barra de herramientas para Internet Explorer.</p>
<p>http://www.merriam-webster.com</p>	
<p>Dropbox (almacenamiento en línea)http://www.dropbox.com</p>	
<p>Dropbox</p>	<p>Dropbox es un servicio de almacenamiento en línea totalmente gratuito, con el que se puede sincronizar y compartir archivos entre varios equipos. Se crea una cuenta en la página web del editor y se usa este cliente para crear la carpeta Dropbox en la PC. Una vez creada la cuenta, se obtiene un espacio de 2GB de forma gratuita y se asocia uno o varios equipos a la cuenta; los datos se colocan arrastrando o copiándolos a la carpeta Dropbox creada por el programa. Estos se sincronizan automáticamente con la cuenta y, por ende, ya están al alcance de todos los equipos asociados a la misma. También es posible acceder al servicio desde cualquier navegador web y descargar directamente los ficheros alojados ahí. Una carpeta pública permite compartir ficheros con otras personas, estén o no registradas en el servicio. Por otro lado, todos los datos sincronizados gozan de una protección adicional ya que están codificados con un robusto algoritmo. Por último, ofrece un sistema de control de versiones, así como un diario de la actividad de la cuenta con todas las modificaciones realizadas.</p>
<p>LogMeIn https://secure.logmein.com (acceso remoto a la PC)</p>	<p>LogMeIn sirve para tener acceso remoto a la PC, o a la de algún conocido o cliente que necesite asistencia. Una vez conectado con el equipo remoto, se podrá controlar su ratón y teclado, además cuenta con prácticas herramientas como un láser para señalar posición; y una pizarra para escribir directamente sobre</p>

Complemento	Descripción
<p>NOTASUno (notas <i>Post-it</i> virtuales para la PC) http://notasuno.uptodown.com</p>	<p>el escritorio de la computadora remota. Su menú de control se maneja completamente desde el navegador (comprobado en Firefox e Internet Explorer), lo que permite incluso acceder desde un dispositivo móvil con conexión a Internet.</p> <p>NOTASUno es un generador y gestor de notas de escritorio, incluyendo avisos y transparencias para mayor comodidad. Como administrador de notas incluye todo lo necesario para crearlas y avisar, todo lo que se busca en una nota <i>Post-it</i>. Las alarmas se introducen individualmente en cada nota, de la misma manera que los niveles de transparencia. Podrán ser opacas y mantenerse siempre visibles o pasar a disimularse e incluso desaparecer en el escritorio de la computadora.</p>
<p>PDF-XChange Viewer (visor de documentos PDF) http://pdf-xchange-viewer.uptodown.com</p>	<p>PDF-XChange Viewer es un visor de documentos PDF capaz de acelerar la carga de documentos pesados, entre otras posibilidades. El programa incluye una serie de funciones adicionales que permiten añadir notas y comentarios, insertar texto directamente en los documentos, extraer los contenidos (tanto texto como imágenes) y exportar páginas enteras a imágenes BMP, JPEG, TIFF o PNG. Al igual que otros visores del estilo, PDF-XChange Viewer facilita la navegación por los documentos previsualizando las páginas en el lateral e incluye las típicas funciones para aumentar el área de trabajo y ejecutar búsquedas.</p>
<p>Publish or Perish http://www.harzing.com</p>	<p>Es un complemento que permite buscar, visualizar y ordenar los resultados de Google Académico; presenta un software que realiza análisis de citas.</p>
<p>PrimoPDF (convertidor de documentos [.txt, .doc, html, ect.] a formato PDF) http://www.primopdf.com</p>	<p>PrimoPDF es un programa que funciona instalando una impresora virtual en el sistema. Así, se puede convertir a PDF cualquier documento que se pueda abrir con una aplicación con soporte para impresión. Sólo se tiene que ir al menú <Imprimir> correspondiente, seleccionar la impresora instalada por PrimoPDF y hacer clic en <Aceptar>. Se optimizará el documento para pantalla o impresión y crear a un archivo PDF de forma totalmente gratuita: ni limitaciones, ni marcas de agua, ni frases publicitarias al final de los documentos.</p>
<p>Complementos (<i>add-ons</i>) para el navegador Firefox https://addons.mozilla.org/es-ES/firefox</p>	<p>Son más de 5, 000 complementos que permiten personalizar y extender Firefox del modo que se necesite, como barras de herramientas, temas y buscadores que ayudan a realizar tareas comunes. Son fáciles de instalar, y con envío de avisos cuando</p>

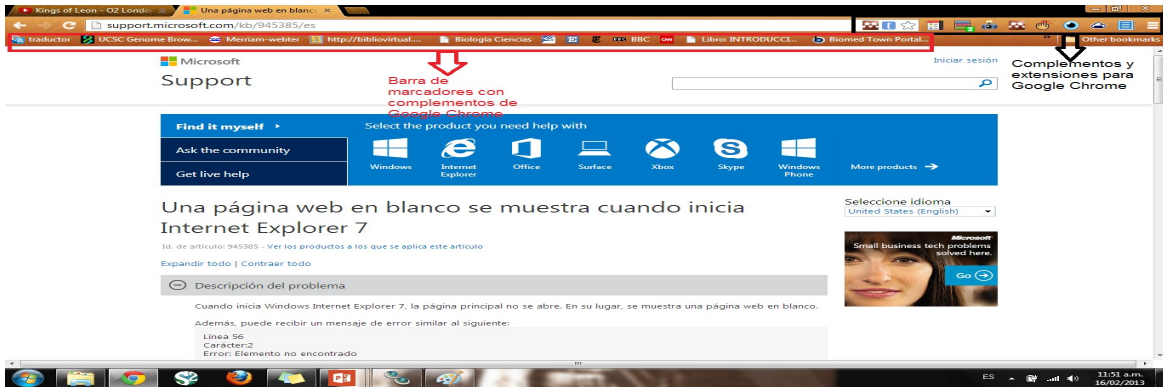
Complemento	Descripción
	hay actualizaciones disponibles. Dentro de estos complementos encontramos los botones de acceso rápido, que se agregan como barra de herramientas o en la barra de herramientas de navegación de Firefox; también se pueden agregar en la barra de herramientas de Favoritos o en el menú contextual que se despliega al presionar el botón derecho del ratón.
<p>Chrome Web Store</p> <p>https://chrome.google.com/webstore</p>	<p>La tienda virtual de Chrome es un mercado en línea donde hay miles de aplicaciones, extensiones y temas para Google Chrome. Una vez en la tienda, encuentra interesantes aplicaciones, extensiones o temas utilizando el cuadro de búsqueda o navegando a través de diferentes categorías. Cada elemento tiene su propia página, donde se puede contribuir con revisiones y calificaciones. Si se usan muchas computadoras, se pueden sincronizar las aplicaciones, extensiones y temas en todas las computadoras a través de la función <i>Browser sync</i> del menú <Opciones> del navegador. Muchas de ellas son gratuitas.</p>
<p>Barra de Google</p> <p>http://www.google.com/tools/firefox/toolbar/FT2/intl/es</p>	<p>Es una aplicación que se integra con el navegador Internet Explorer y permite realizar búsquedas desde cualquier página web, además de bloquear anuncios de ventana emergente (<i>pop-ups</i>). También traduce páginas web y corrige la ortografía de los mensajes, se pueden añadir botones para personalizar la barra y tener acceso más rápido a las páginas que más se visitan como el correo electrónico y redes sociales, o de herramientas como diccionarios o correctores de ortografía.</p>
<p>iGoogle</p> <p>http://www.google.com.mx/ig</p>	<p>Es una página de inicio personalizable. Sus características incluyen la capacidad de añadir extensiones de páginas web y <i>gadgets</i> de Google y especialmente desarrollados para mostrar el contenido a un usuario de la página. Estos pueden colocarse en distintas pestañas dentro de iGoogle.</p>

Tabla 8. Complementos útiles para leer, manejar y compartir literatura.

Complemento	Descripción
Diigo	Es una extensión que permite almacenar las direcciones web en el servidor de Diigo, también proporciona herramientas como marcador de textos y notas para resaltar lo más importante de la página que se está almacenando.
CiteULike	Es un botón que se agrega como extensión del navegador; se utiliza para ingresar la información de un documento electrónico de forma automática en CiteULike.

Complemento	Descripción
Impresionante captura de pantalla	A través de un clic permite capturar, recortar o seleccionar cualquier parte de una página web que se desee. Permite editar, ocultar información y compartir en la Web, anotar con rectángulos, círculos, flechas, líneas y texto.
Mendeley	Es un botón que se instala en la barra del navegador; se utiliza para agregar la información de la página que contiene el documento deseado directamente en Mendeley.
VisualizeUs	Esta extensión agrega la dirección y la información del documento que contiene la imagen que se desea almacenar directamente en VisualizeUs.
Xmarks	Es una herramienta que ayuda a guardar y sincronizar con un servidor y las contraseñas de todos los sitios y páginas web a las que se accede frecuentemente.
<i>Plugs</i> de Mendeley y Zotero	Estos dos complementos se agregan en los procesadores de texto como Word u Open Office y facilitan agregar citas y referencias en un escrito automáticamente y de forma personalizada.

Tabla 9. Complementos para Google Chrome y procesadores de texto recomendados.



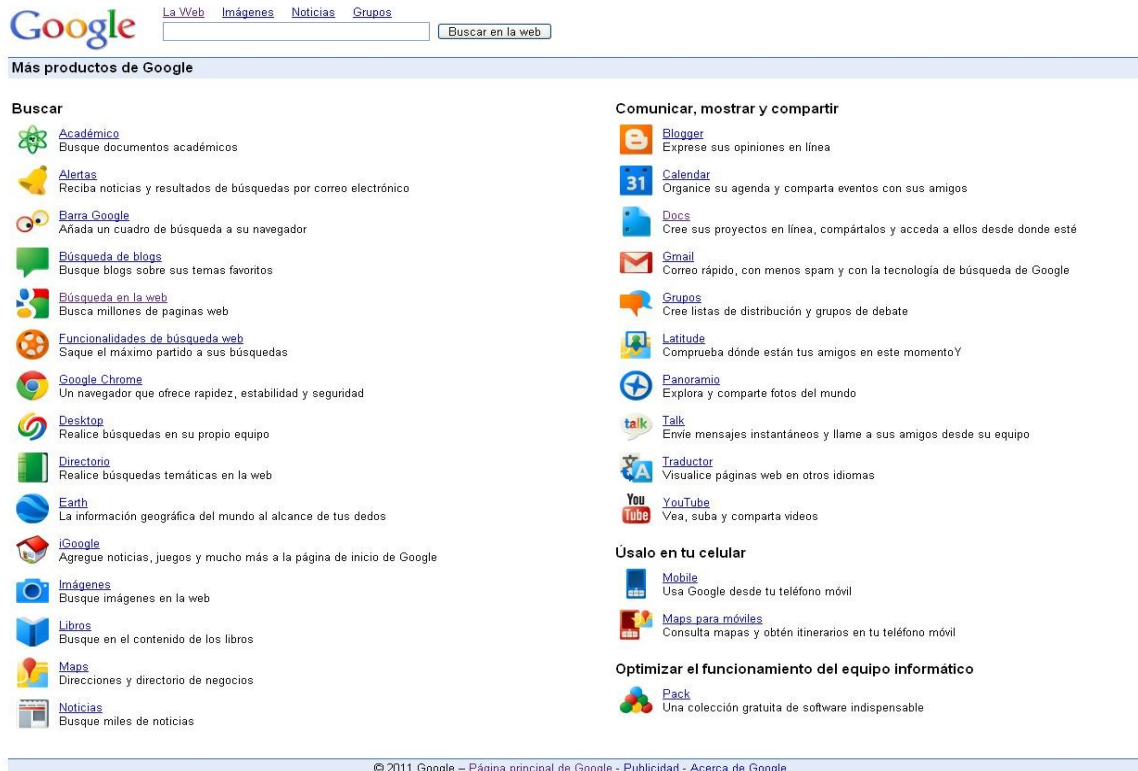


Figura 12. Algunas aplicaciones de Google.

Figura 13. Complementos para Google Chrome.

III. LA RECUPERACIÓN DE INFORMACIÓN EN LA WEB

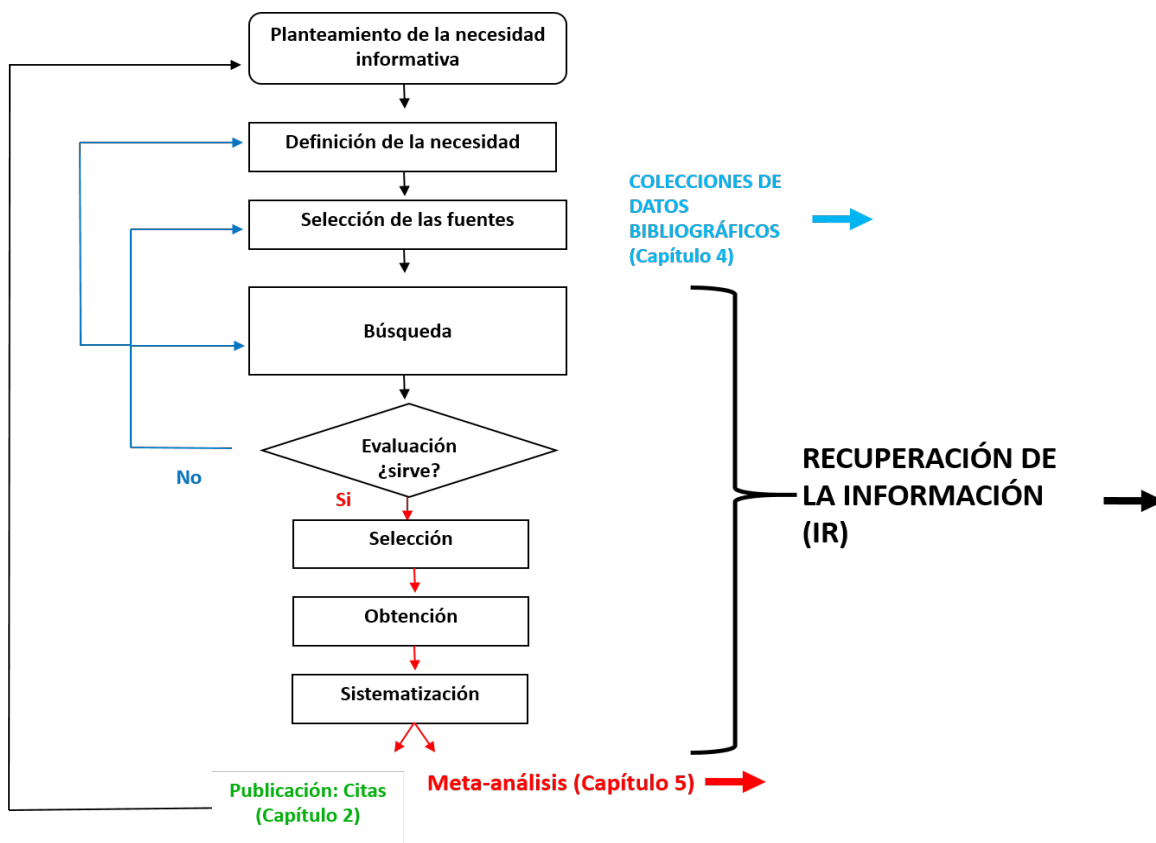
Las actividades científicas y sus productos crecen de forma exponencial: el número de investigadores y sus publicaciones se duplica aproximadamente cada veinte años, viven actualmente entre un 80 o 90% de los científicos del siglo XX y cada año se publican más de dos millones de artículos, se editan miles de revistas científicas y se conceden más de un millón de patentes. Estamos inmersos en la era de la información y en la sociedad del conocimiento, donde la información constituye un recurso de alto valor, pues es la materia prima del proceso de enseñanza-aprendizaje, la generación de nuevo conocimiento y la toma de decisiones.

La profunda transformación que han impulsado el formato digital y la existencia de la Web y ha trascendido la forma en la que se almacenan, sistematizan, recuperan y difunden los datos. Esto implica que conocer sobre un tema ya no dependa únicamente de los

documentos impresos disponibles en las librerías o bibliotecas, sino también de aquellos que existen en catálogos electrónicos, formato en el que hoy se encuentra la mayor cantidad de información en línea, y ha favorecido el acceso, la inmediatez y democratización de la información.

1 La recuperación de información

Para iniciar el proceso de **recuperación de información** es importante plantear la necesidad informativa y definirla claramente; se requiere reflexionar y analizar a profundidad el objeto de estudio y definir claramente una pregunta documental. Es necesario establecer el universo documental (conjunto de documentos sobre el objeto de estudio), definir el tema general, el dominio académico y las áreas de interés involucradas para utilizar el enfoque correcto y eliminar la información no deseada (Fig. 14). La pregunta documental se puede definir con base en los siguientes cuestionamientos: ¿qué necesitamos saber sobre el tema?, ¿cuáles son los conocimientos previos?, ¿qué se espera resolver?, ¿qué trabajos se han realizado para resolver el problema?, ¿qué investigadores trabajan con el tema?



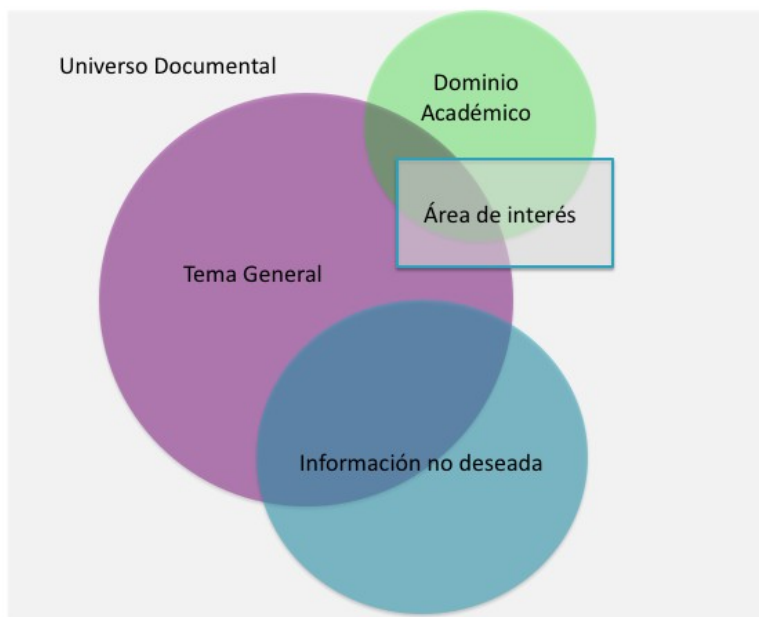


Figura 14. Planteamiento y definición de la necesidad informativa.

Figura 15. Proceso de recuperación de información en la web.

Analizar el objeto de estudio a detalle mejora el proceso de recuperación de información, pues este proceso es cíclico; consta de seis etapas: 1) la selección de las fuentes, 2) la búsqueda, 3) la evaluación, 4) la selección, 5) la obtención y 6) la sistematización (Fig. 15) y termina con el uso y la aplicación de la información recuperada, pero sólo se acaba cuando la necesidad informativa ha sido satisfecha.

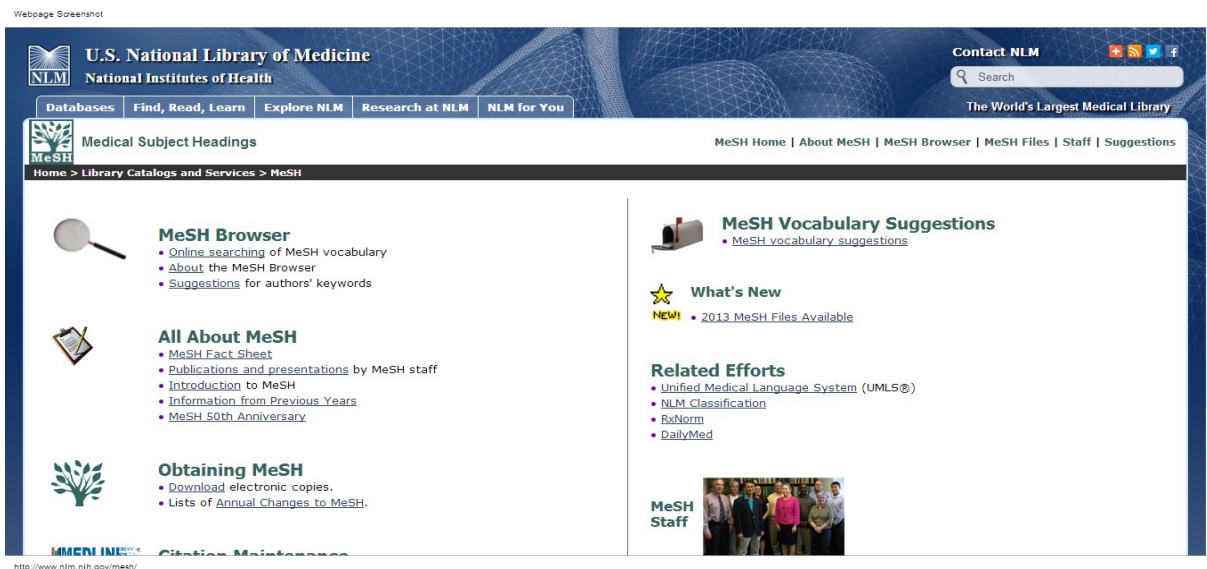
	Búsqueda general	Búsqueda avanzada
Registros recuperados	Muchos resultados	Menos resultados, pero más específicos
Ruido documental (documentos no útiles)	Mucho	Poco
Campos de búsqueda	Uno solo, busca en toda la base de datos	Año Idioma Tipo de documento Titulo

	Búsqueda general	Búsqueda avanzada
		Autor
		Revista
Uso de filtros para refinar la búsqueda	No	Sí
Velocidad de introducción de los datos para la consulta	Rápida	Lenta
Registros recuperados	Generales	Específicos, se eliminan muchas páginas que contienen información que no es de nuestro interés
Consulta	Se utilizan pocos términos	Se incluyen sinónimos, homónimos y palabras relacionadas Se incluyen y excluyen términos

Tabla 10. Comparación entre búsqueda general y avanzada.

La búsqueda general es rápida y básica; se usa cuando se desea obtener un resultado general y consiste en introducir uno o varios términos elegidos. La búsqueda avanzada explora toda la información, pero hace pruebas por medio de distintas consultas (*queries*) utilizando tres elementos básicos: los términos, los operadores indicados y los límites correspondientes, lo cual permite aplicar criterios más específicos que arrojan resultados más exactos.

El primer elemento de la consulta son los términos, esto es, palabras que responden específicamente a la pregunta documental planteada en un principio. También es importante identificar todas aquellas palabras relacionadas o afines y se deberán incluir o excluir los homónimos y sinónimos con el fin de acotar detalladamente el conjunto de documentos que



se obtendrán de la búsqueda. Los términos adecuados para la definición también se pueden buscar en los tesauros o catálogos de palabras clave de las colecciones (por ejemplo MeSH <http://www.nlm.nih.gov/mesh/MBrowser.html>, Fig. 16) que muestran la relación jerárquica existente entre los diversos descriptores y, en la mayoría de los casos, incluyen una definición.

Figura 16. Tesauro MeSH.

El segundo elemento, los **operadores lógicos**, permiten establecer relaciones entre los términos de búsqueda. Existen varios tipos de operadores, los más comunes son los booleanos, los de truncamiento y los de proximidad (Tabla 11); estos funcionan generalmente de la misma manera en todas las bases de datos, sin embargo, es recomendable leer los consejos de búsqueda, la ayuda o los tutoriales disponibles en cada caso para obtener mejores resultados.

Tipo de operador	Operador	Descripción	Ejemplo
Booleanos	AND (en inglés), Y (en español)	Recupera registros que tengan todos los términos	estudiantes AND universitarios
	OR (en inglés), O (en español)	Recupera registros que tengan cualquiera de los términos	estudiantes OR universitarios
	NOT (en inglés), NO (en español)	Excluye el término indicado para sesgar los resultados	estudiantes NOT universitarios
	?	Sustituye una letra o número en cualquier posición de una palabra o serie de números	mexican?- mexicana, mexicano
Truncamiento	*	Sustituye una o más letras o números en cualquier posición de la palabra o serie de números	mexican*.- mexicana, mexicano, mexicanas, mexicanos
Proximidad	NEAR	Recupera documentos en los que los términos de búsqueda estén cerca (máximo mil palabras)	becas NEAR extranjero
	ADJ	Recupera documentos en los que los términos estén uno junto al otro; pueden obtenerse los mismos resultados entrecomillando los términos	servicio ADJ social

Tabla 11. Tipos de operadores para realizar las búsquedas más comunes.

Ya que se han definido los términos y los operadores (relaciones) que se utilizarán para la búsqueda, la estrategia se completa con la elección del o los campos de búsqueda específicos; los más usados son: título, autor, año, revista y palabra clave. Contar con una estrategia de búsqueda definida conduce a la recuperación de resultados pertinentes de

forma más eficiente, rápida, completa y exacta, reduce la incertidumbre y permite satisfacer la necesidad informativa planteada. Si uno se inscribe a la **base de datos** (colección) que se está usando, es posible guardar las consultas hechas con estrategias anteriores en el historial de búsqueda (Fig. 17).

WEB OF KNOWLEDGESM | DISCOVERY STARTS HERE THOMSON REUTERS

Signed In | My EndNote Web | My Citation Alerts | My Journal List | My Saved Searches | Log Out | Help

Open / Manage Saved Searches

<< Back

Open from the Web of Knowledge Server

Use this box to open histories that were saved to your private account on our server.

Display histories from: All Products

History Name	Product	Description	RSS Feed	Alerting	Modify Settings	Delete Select All Delete	Open/Run History
e-Taxonomy	Biological Abstracts	Para Colombia	<input checked="" type="checkbox"/> XML	Status: On Expires: 2012-01-19 <input type="button" value="Renew"/>	<input type="button" value="Settings"/>	<input type="checkbox"/>	<input type="button" value="Open"/>
FACZOOLOGICAL	Zoological Record	Primera busqueda en Zoological	<input checked="" type="checkbox"/> XML	Status: Off Expires: --	<input type="button" value="Settings"/>	<input type="checkbox"/>	<input type="button" value="Open"/>
PrimeraBusBIOSIS	Biological Abstracts	Primera busqueda en BIOSIS	<input checked="" type="checkbox"/> XML	Status: Off Expires: --	<input type="button" value="Settings"/>	<input type="checkbox"/>	<input type="button" value="Open"/>

WEB OF KNOWLEDGESM | DISCOVERY STARTS HERE THOMSON REUTERS

Signed In | Marked List (0) | My EndNote Web | My ResearcherID | My Citation Alerts | My Journal List | My Saved Searches | Log Out | Help

All Databases | Select a Database | Web of Science | Additional Resources

Search | Author Finder | Cited Reference Search | Advanced Search | Search History

Web of ScienceSM

Advanced Search

Use 2-character tags, Boolean operators, parentheses, and set references to create your query. Results appear in the Search History at the bottom of the page.
 Example: TS=(nanotub* SAME carbon) NOT AU=Smalley RE #1 NOT #2 [more examples](#) | [view the tutorial](#)

Searches must be in English

Restrict results by any or all of the options below:

All languages: English, Afrikaans, Arabic
 All document types: Article, Abstract of Published Item, Art Exhibit, Review

Current Limits:

Timespan: All Years (updated 2011-10-17) From 1899 to 2011 (default is all years)

Booleans: AND, OR, NOT, SAME, NEAR

Field Tags:
 TS= Topic, TI= Title, AU= Author, RID= ResearcherID, GP= Group Author, ED= Editor, SO= Publication Name, DO= DOI, PY= Year Published, AD= Address, OG= Organization, SC= Suborganization, SA= Street Address, CI= City, PS= Province/State, CU= Country, ZP= Zip/Postal Code, FA= Funding Agency, FG= Grant Number, FT= Funding Text

Figura17. Búsqueda avanzada e historial de búsquedas.

10. Las etapas siguientes

La recuperación no termina cuando se realiza la primera búsqueda; es un proceso que se debe repetir varias veces hasta encontrar información relevante y oportuna, haciendo varias

pruebas y modificaciones. Una vez que se tiene una buena consulta que arroja los resultados esperados, se procede a las siguientes etapas: la evaluación, la selección, la extracción y la sistematización de la información encontrada, las cuales no necesariamente deben seguir un orden determinado y pueden realizarse de manera simultánea.

La evaluación consiste en la revisión de los documentos con detenimiento y detalle de manera sistémica esto es, un conjunto de artículos simultáneamente. Se revisa el título, el resumen, las palabras clave, las figuras y tablas, los pies de figuras, todo el contenido o las referencias de cada documento encontrado para seleccionar aquellos que satisfagan nuestra necesidad. Al observar los resultados, se deduce si la **selección** fue óptima; si es así, se procede a elegir aquellos registros que se consideran útiles o interesantes; para proceder al análisis de sus contenidos, como la aplicación de **indicadores bibliométricos** o **minería de textos** (*text mining*).

Por ende, es en este proceso donde se determina si nuestra necesidad informativa ha sido satisfecha; en caso contrario, es preciso sesgar la lista refinando la búsqueda o generando una nueva consulta, también se pueden elegir otras fuentes y/o redefinir la necesidad informativa hasta que se obtengan resultados favorables.

La siguiente etapa es la **extracción**. La mayoría de bases de datos permiten extraer (exportar) los documentos a una lista en distintos formatos y pueden ser transferidos a archivos personales como es posible: documentos, bases de datos o incluso agregar directamente los registros bibliográficos a un software de administración de referencias (capítulo 2) para consultarlos con detenimiento una vez que se hayan obtenido suficientes documentos; a partir de estos se puede leer, analizar o citar uno o varios textos.

La última etapa del proceso de recuperación de información es la **sistematización** de los documentos. Esta puede ser una tarea individual o grupal; consiste en mantener, ordenar y utilizar las referencias, citas y documentos coleccionados. Siempre es necesario conservar la bibliografía para el uso continuo y mantenerla actualizada; esta tarea se vuelve viable y veloz cuando se utiliza un software de administración de citas.

Con base en lo expuesto en este capítulo, la recuperación de información no es un proceso

lineal, sino cíclico y continuo porque la información recuperada propicia nuevas necesidades de información que deberán cubrirse. Una vez que se obtiene la habilidad para recuperar la información pertinente e identificar las fuentes idóneas, sólo hace falta un poco de práctica para recuperar exitosamente la información en diversas colecciones. Explotar estos recursos permite tener información prácticamente imposible de obtener por otro medio; de igual manera, implementar estrategias de recuperación de información digital en la actualidad constituye una práctica indispensable para investigadores, profesores y estudiantes de ciencias biológicas, ya que esto facilita la investigación bibliográfica, permite explotar los recursos informativos con los que cuentan las universidades e instituciones de investigación, desarrolla la capacidad de análisis y síntesis, promueve el uso de fuentes certificadas y facilita el uso de las tecnologías de la información actualmente disponibles.

El procedimiento para la recuperación de información aquí planteado deberá ir acompañado de las herramientas de personalización, actualización y difusión presentadas en el capítulo 3, de tal manera que el proceso, una vez realizado, genere posteriormente información actualizada sin que tengamos que visitar los servicios de información constantemente. La nueva literatura publicada que cumpla con la consulta generada se programará para que llegue a nuestro correo electrónico o mejor aún al lector de *feeds* (RSS) inmediatamente después de que sea producida.

IV. LAS COLECCIONES DE DATOS BIBLIOGRÁFICAS

Durante los últimos años, el crecimiento exponencial de la información digital ha sido simultáneo al aumento acelerado de las colecciones que la albergan. Esto se debe principalmente al diseño de manejadores sofisticados y a la reducción en los gastos de almacenamiento que han hecho factible, e incluso rentable, crear y mantener grandes bases de datos cada vez más complejas. De tal manera que los usuarios tienen acceso en línea, de manera inmediata, a los datos provenientes de una gran variedad de fuentes que se pueden consultar, extraer, compartir y analizar en línea, lo que permite con una relativa facilidad el intercambio de los datos digitales (en comparación a los registrados en el papel) y fomenta la colaboración entre usuarios de diferentes disciplinas, instituciones y ubicaciones geográficas. La aparición, difusión y consulta de la gran cantidad de bases de datos que hay

en la actualidad, particularmente las que se pueden consultar en la Web, fue sin duda el resultado de un fenómeno representativo surgido en la industria de la información a finales del siglo XX.

1 Las bases de datos

La herramienta por excelencia para sistematizar información digital son las **bases de datos**, un sistema informático de registros almacenados en tablas con un orden establecido que permite guardar, ordenar, mantener, procesar, recuperar, presentar y generar información. Estos programas, manejadores o gestores permiten la administración total de los datos, realizar búsquedas a partir de diferentes criterios, procesar información de forma cuantitativa y cualitativa, interrelacionar los resultados, utilizar distintas variables, actualizar datos fácil y rápidamente, efectuar diversos cálculos por medio de consultas (*queries*), añadir módulos y se pueden diseñar, almacenar, manejar y analizar en una computadora personal.

En 1951, la Oficina de Censo estadounidense diseñó la primera base de datos formal. En la década de los 70, se dio acceso en línea a las bases de datos existentes, aunque estaba limitado a las instituciones gubernamentales como los laboratorios de la NASA, la Fuerza Aérea y la biblioteca médica del Estado de Nueva York. En esa década, el desarrollo del sistema de almacenaje sobre cinta magnética impedía consultas simultáneas, es decir, sólo se podía realizar una consulta a la vez. Posteriormente, con la posibilidad de hacer consultas de modo *batch* (por lotes), la aparición del sistema de almacenamiento en discos, la popularización de la Internet y la amplia **indexación** de documentos, hubo un crecimiento acelerado de las bases de datos, lo que generó ahorro de tiempo y dinero a quienes las utilizaban.

A finales de los 80, las bases de datos eran producidas generalmente por particulares, estaban principalmente destinadas a los sectores privados, financieros, de la información, de la mercadotecnia y de la planeación económica. El crecimiento de esta industria en el mercado comercial se dirigió al usuario o cliente. Los impulsores de las bases de datos como DIALOG y BRS iniciaron la investigación satelital y perfeccionaron los servicios de consulta. Los sistemas de cómputo personal continuaron su proceso hacia un ambiente

amigable, se volvieron más sencillos e invitaron al usuario a su aplicación.

La cantidad de bases de datos conocidas en 1980 era de 300 y, para 1984, alcanzó la cantidad de 2, 400 (ocho veces más). A partir de esta proliferación, se integró un comité de regulación de los derechos, propiedad y estatus legal de las bases de datos. Durante esta época, los editores de periódicos y revistas habían adoptado la computadora para la fotocomposición en sus procesos de impresión y algunos empezaron a utilizarlas para el manejo de información.

Las bases de datos se pueden clasificar de acuerdo con varios criterios, pero los más comunes son:

1. **Estáticas o dinámicas, por la posibilidad de modificar los datos por los usuarios.** Las primeras son bases de datos de solo lectura, se usan primordialmente para almacenar datos históricos que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo. En las segundas, la información almacenada sufre modificaciones con el tiempo, permitiendo operaciones de actualización por los usuarios, como borrado y adición de datos, además de las operaciones de consulta. Un ejemplo de esto puede ser la base de datos utilizada en un sistema de información de inventario de cualquier almacén o la de un repositorio institucional (Fig. 18).
2. **Por la estructura de los datos.** 1) Jerárquicas: se basan en una estructura escalonada en forma de árbol en donde hay un nodo raíz que puede tener varios nodos derivados; tienen la desventaja de no representar eficientemente la redundancia de datos. 2) En red: los datos son representados por colecciones de datos relacionados por medio de ligas, su organización es abstracta y a la vez gráfica, permite que cada camponodo tenga varios ascendientes; este tipo es muy poco utilizado. 3) Relacionales: están compuestas por una colección de tablas, cada una con campos en común que las relacionan entre sí, formando un conjunto de ellas y no hay relevancia en la manera en que se almacenan (Fig. 18). Este modelo es el más utilizado en la actualidad para administrar bases de datos de una manera dinámica y eficiente porque pueden ser diseñadas e interpretadas fácilmente y los

datos pueden ser recuperados y manipulados mediante consultas a través de SQL (*Structured Query Language*), creado para este propósito. Existen otros modelos como las multidimensionales que se emplean para realizar análisis métricos, las orientadas a objetos que se utilizan en los lenguajes de programación y las distribuidas que se organizan estratégicamente en distintos puntos de una red para ser consultadas y después congregar los datos.

3. **Respecto al tipo del propietario, editor o proveedor de la base de datos.** Se pueden clasificar, por ejemplo, en gubernamental, privado, o académico.
4. **Por la forma en la que se almacenan y distribuyen las bases de datos.** Puede ser en CD, en una intranet o por Internet.
5. **Con base en su contenido.** Por ejemplo, las biológicas pueden registrar textos, imágenes, sonidos, software, especies, secuencias de genes, estructuras proteicas, organismos, sustancias químicas, etc.

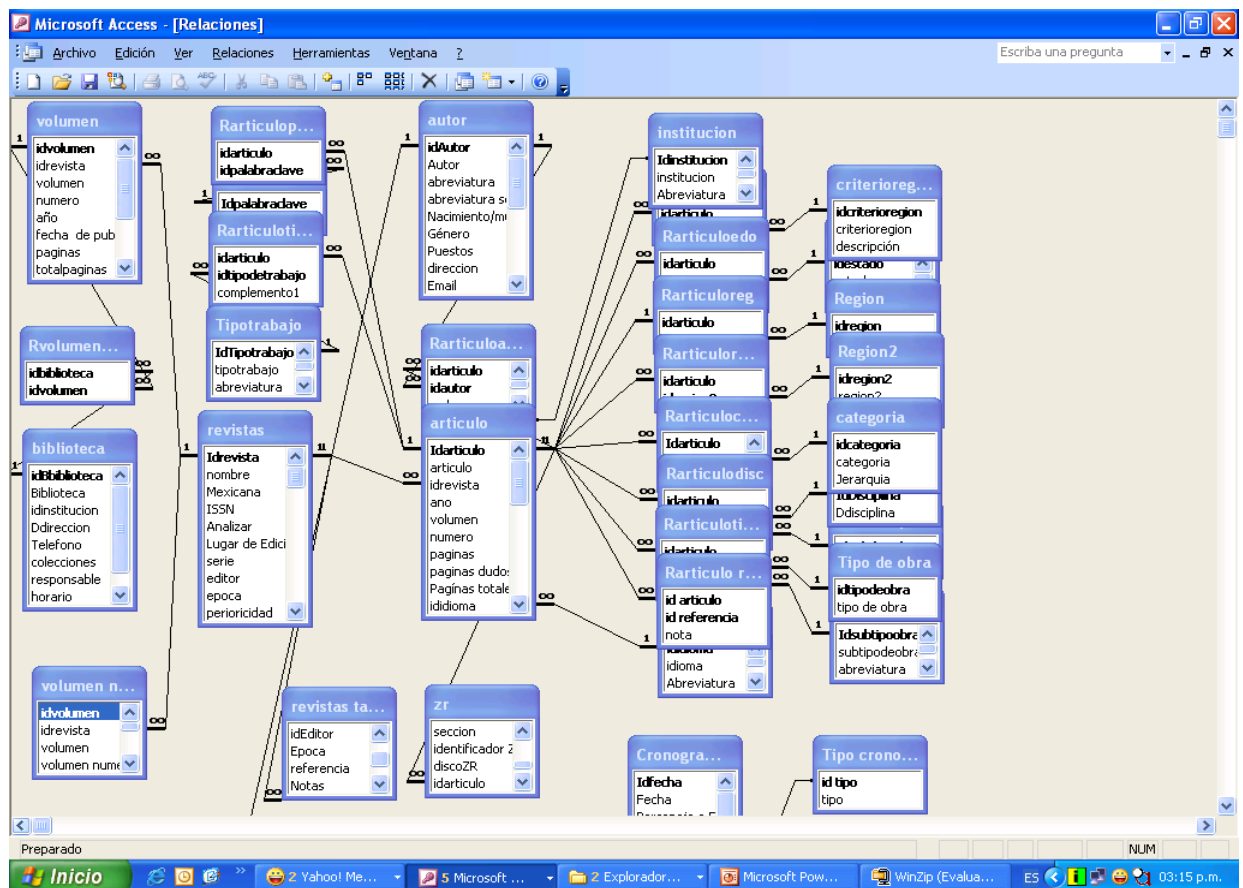


Figura 18. Estructura relacional de la base de datos bibliográfica TaXMeXX. Contiene artículos taxonómicos publicados en México durante el siglo XX, disponible en: <http://repositorio.fcencias.unam.mx:8080/xmlui/handle/123456789/56181>.

11. Las colecciones bibliográficas

Las bases de datos que registran documentos son llamadas comúnmente bibliográficas, documentales o de literatura. En un principio contenían resúmenes e índices de las bibliotecas y se almacenaban en cintas magnéticas; posteriormente, incluyeron fotocomposiciones y citas. Entre las más destacadas históricamente se encuentran: *Medical Literature Analysis and Retrieval System* (MEDLARS), producida en 1960 por la National Library of Medicine, ahora llamada *PubMed*; *Chemical Registry System* y *Chemical and Biological Activities* (CBAS), publicadas en 1965 por el *Chemical Abstracts Service* en forma impresa y cinta magnética para computadora; *MachineReadable Catalog* (MARC) emitida por la Biblioteca del Congreso de los Estados Unidos de América en 1969; *Engineering Information*, creada en 1967; *BioSciences Information Service* (BIOSIS), en 1969. Se calcula que en 1965 existían entre 12 y 20 bases de datos bibliográficas; en 1968, había 25; para 1970, en promedio eran 75; en 1986, eran 3, 337; y, en el año 2011, se podrían computar varios miles.

Llamaremos **colección** a toda aquella información sistematizada en una base de datos digital con acceso inmediato, diseñadas con la intención de hacerlas disponibles para quienes se interesen por su consulta, realizada con fines académicos y disponibles a través de la Web. Denominaremos **colección bibliográfica** o de **literatura científica** a toda aquella que registre documentos científicos, en especial, información sobre artículos y libros producto de la investigación científica. Desecharemos de esta categoría a los directorios de bibliografía que consisten en simples listados de referencias que no permiten búsquedas y los cuales no están guardados en una base de datos, de tal forma que el usuario debe inspeccionar cada recurso para discriminar entre los que le son útiles y los que no.

La información mínima que puede registrarse en una colección bibliográfica es el registro bibliográfico básico, también llamado **asiento** o **ficha bibliográfica** (en el área de la documentación), que corresponde a la información mínima necesaria para localizar un

documento: autor, año de publicación, editorial y páginas. La cantidad, calidad y organización de la información de cada colección es heterogénea, porque responde a distintos propósitos. Las más complejas generalmente están organizadas de la siguiente manera: 1) Los documentos, artículos fundamentalmente, contienen elementos como título, tipo de documento, idioma, palabra clave, descriptores, volumen, número y páginas (Fig. 19); 2) La fuente de donde provienen (principalmente revistas), contiene título, año de publicación y tema (Fig. 20); y 3) La autoría, esto es, el o los nombres de los autores, institución de adscripción y país. Para cada documento (o fuente en algunos casos) se asignan descriptores con la finalidad de clasificarlos y utilizarlos posteriormente como referencia para recuperarlos por medio de las búsquedas avanzadas. Las bases de datos más completas también contienen resúmenes, referencias, citas, ligas al documento en texto completo (libre o restringido a un pago), documentos relacionados, análisis bibliométricos sencillos (cuentas de autores, temas, países, tipos de documento, idioma y descriptores) o complejos (citación, vida media e índice h), catálogos y tesauros, herramientas o aplicaciones electrónicas para salvar, **etiquetar** (*tagging*), almacenar, sistematizar, analizar y manejar las referencias recuperadas.



Figura 19. Tabla de artículos de la base de datos TaXMeXX.



Figura 20. Tabla de revistas de la base de datos TaXMeXX.

12. Características de las colecciones bibliográficas

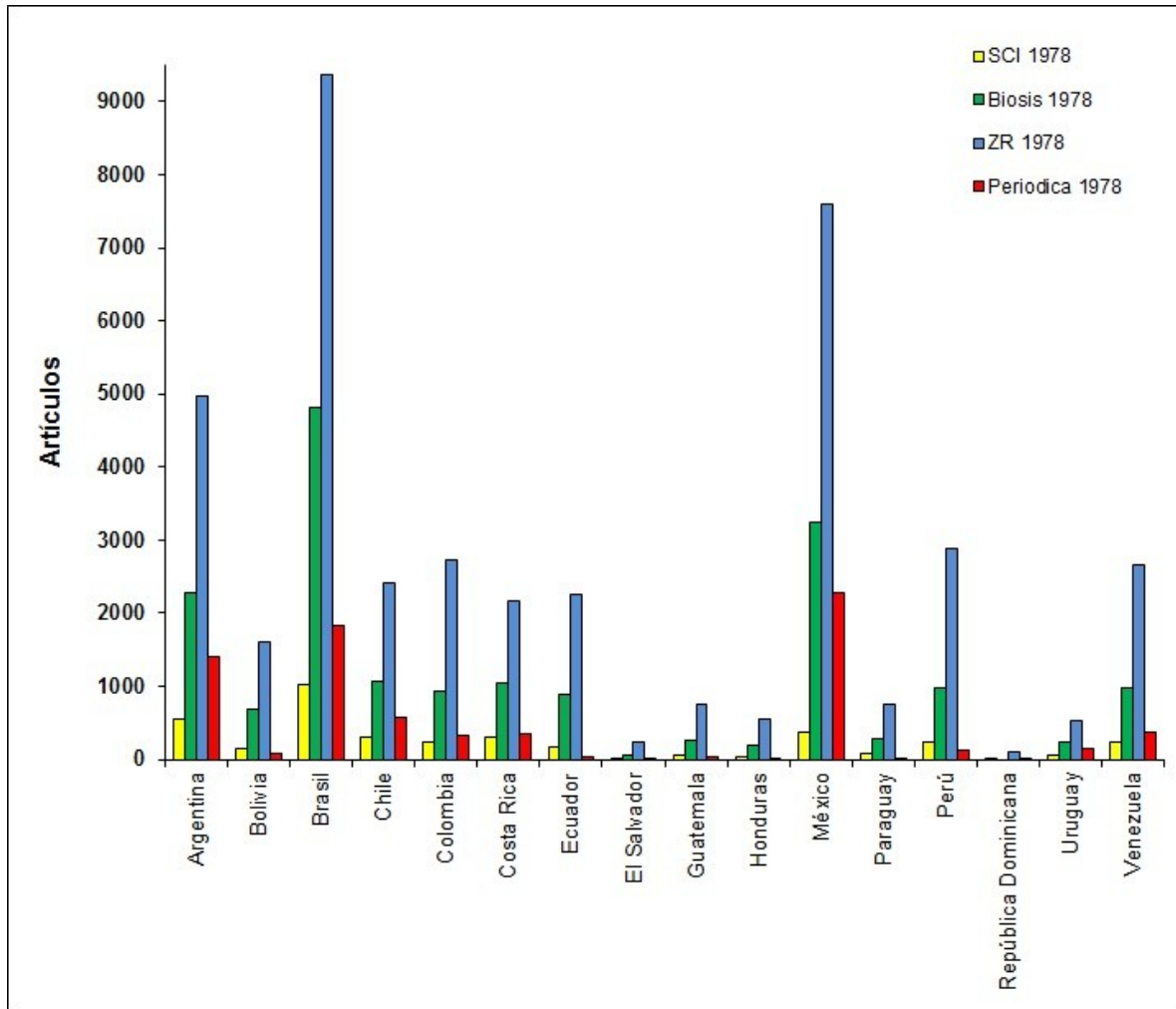
Entre las características más determinantes para elegir cuál o cuáles bases de datos documentales utilizar para recuperar literatura están:

- **El tipo de instancia** (compañía o institución) que produce y mantiene los datos y la que provee el acceso. Ésta define en gran medida las características de los registros, que están de acuerdo con el propósito de cada base de datos.
- **El tamaño digital que tiene la colección.** Por ejemplo, terabytes, gigas o megas.
- **La cantidad de registros que contiene.** Los más usados son los documentos y las revistas indizadas; en algunos casos también se usa como indicador la velocidad con que adicionan nuevos documentos (Fig. 21).

Figura 21. Representación gráfica del tamaño de las colecciones bibliográficas más utilizadas en biología; se esquematiza también las revistas que comparten.

- **Cobertura tipológica.** Se refiere al tipo de documentos que privilegia el catálogo. En general, los más frecuentes son los artículos de revistas, series, memorias y libros; estos últimos son los que constituyen la literatura primaria.
- **Los criterios de inclusión de documentos.** Están dados en cada base de datos de acuerdo a su propósito, el cual determina los lineamientos para elegir los documentos que están o serán incluidos en la colección.

- **La cobertura temática y de disciplinas.** Se refiere a las áreas de conocimiento y



tópicos que abordan los documentos registrados en la colección. Las primeras son, por ejemplo, multidisciplinarias, disciplinarias o especializadas. Los segundos serían zoología, botánica, taxonomía o genética, por ejemplo (Fig. 22).

Figura 22. Países con más documentos sobre biología en cuatro de las colecciones bibliográficas. Se representan búsquedas sobre taxonomía de América Latina (consulta realizada en el año 2009).

- **El tipo de campos.** Son las entidades que se capturan para cada registro (título, autor, revista, idioma, palabras clave, resumen, referencias y citas).
- **El tipo de acceso a la base de datos.** Puede ser libre o restringido; el segundo se refiere a que sólo se puede acceder previa contratación y pago, que generalmente se

tramitan de forma institucional pues el precio es demasiado alto (miles de dólares) para hacerlo de manera individual. Por lo tanto, es más común que estos servicios se puedan consultar, en las bibliotecas de instancias académicas o gubernamentales.

- **Cobertura idiomática.** Se refiere a los idiomas de los documentos registrados, la mayoría utilizan el inglés por ser el idioma recurrente en las publicaciones científicas (Fig. 23).

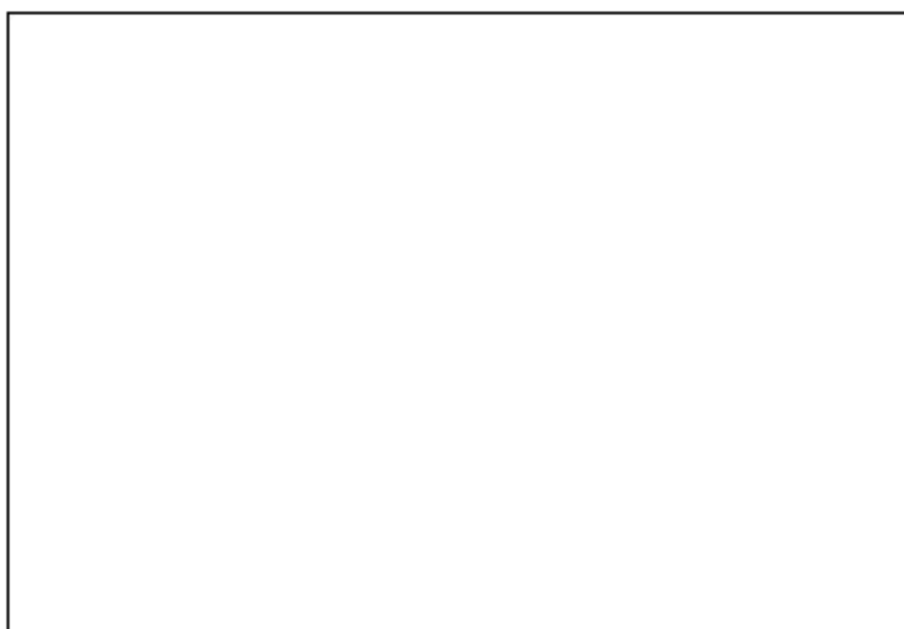


Figura 23. Porcentaje de documentos según su idioma en cinco de las colecciones bibliográficas con más documentos sobre biología; búsquedas sobre taxonomía de América Latina (consulta realizada en el 2009).

- **Cobertura geográfica.** Corresponde al lugar de edición de los documentos (generalmente las revistas) que se indizan. Puede ser mundial, si se registran documentos producidos en todos los países; regional, para un área determinada como América Latina; y local, por ejemplo, las que contienen registros publicados solamente en México.
- **Cobertura temporal.** Cita los años que abarca la literatura registrada en el catálogo (Fig. 24).

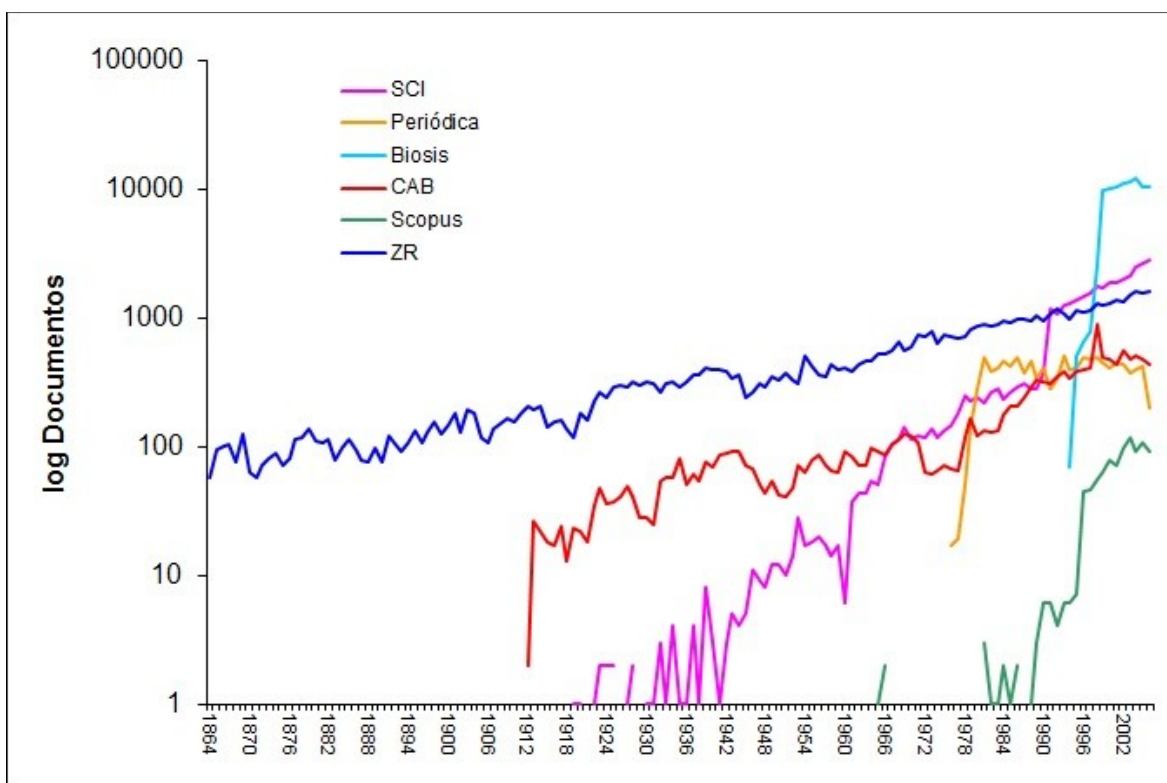


Figura 24. Años que abarcan los documentos en siete de las colecciones bibliográficas con más documentos sobre biología. Se representan documentos sobre taxonomía de América Latina (consulta realizada en el año 2009).

- **Herramientas de la Web 2.0.** Las características de la aplicación para acceder a la información y las herramientas que contienen para búsqueda, almacenamiento, sistematización, análisis y gestión de la literatura.
- **Tesauros.** El uso de un catálogo controlado para la clasificación de los documentos con base en descriptores constituye una herramienta documental importante que permite buscar y analizar la información de manera completa, consistente y eficaz.

La consulta de bases de datos bibliográficas se ha vuelto una tarea indispensable para los académicos. Es tal la cantidad, diversidad y complejidad de las bases de datos disponibles en la red que es necesario consultar varias de ellas para tener una representación completa de la literatura sobre el tema de interés. No sólo eso, sino que constantemente se originan nuevas colecciones, cambian, se renuevan y se actualizan, lo que hace difícil estar actualizado sobre las posibilidades disponibles. Por todas estas razones, es necesario

permanecer informado y actualizado sobre las características y evolución de las colecciones bibliográficas. Por ejemplo, ha sido tan grande el éxito de estas colecciones que en la actualidad existen varias maneras de localizarlas en la Web, pues utilizan indistintamente varios nombres (Fig. 25). El término más reciente es repositorio, que se refiere a colecciones bibliográficas que contienen para cada registro el documento en texto completo accesible, generalmente de manera abierta; por otro lado, si contiene la producción de una instancia académica como la Facultad de Ciencias de la UNAM; se le denomina institucional (Fig. 26).



Figura 25. Distintos nombres utilizados frecuentemente en la web para referirse a las colecciones bibliográficas.

Ciencias, UNAM

DSpace Principal → FACULTAD DE CIENCIAS, UNAM

FACULTAD DE CIENCIAS, UNAM

Listar por

- Por fecha de publicación
- Autores
- Títulos
- Temas
- Por fecha de envío

Búsqueda en el texto completo: Ir

Búsqueda avanzada

Subcomunidades en esta comunidad

- Departamento de Biología Celular [193]
- Departamento de Biología Comparada [193]
- Departamento de Biología Evolutiva [22308]
- Departamento de Ecología y Recursos Naturales [335]
- Departamento de Física [25]
- Departamento de Matemáticas [66]
- Unidad Multidisciplinaria de Docencia e Investigación (UMDI-Sisal) [0]

Colecciones en esta comunidad

- Facultad de Ciencias, UNAM, Artículos de difusión [3]
- Facultad de Ciencias, UNAM, Artículos de investigación [3734]
- Facultad de Ciencias, UNAM, Capítulos de libros [4]
- Facultad de Ciencias, UNAM, Las Prensas de Ciencias [61]

La Facultad de Ciencias produce libros, manuales y, en general, textos que apoyan el proceso de aprendizaje en las distintas carreras que se imparten en la Facultad, y que pueden ser también de utilidad para las escuelas ...

- Facultad de Ciencias, UNAM, Libros [9]
- Facultad de Ciencias, UNAM, Memorias de reuniones científicas [6]
- Publicaciones [7]
- Tesis [90]

Tesis de la Facultad de Ciencias

Envíos recientes

- Título: Archiver: Archivo del mar
 Autor: Ifremer
 Fecha: 2005
- Título: BIODIVERSITY INTERNATIONAL LIBRARY CATALOG
 Autor: Biodiversity International
 Fecha: 1913
- Título: Biblioteca del Instituto de Biología de la UNAM
 Autor: Universidad Nacional Autónoma de México
 Fecha: 2007
- Título: Biblioteca del Instituto de Ecología de la UNAM
 Autor: Universidad Nacional Autónoma de México
 Fecha: 2012-11-05
- Título: Google Scholar
 Autor: Google
 Fecha: 2012-11-05

DSpace software copyright © 2002-2010 Duraspace
 Contacto | Sugerencias

Theme by @MIRE

<http://repositorio.ciencias.unam.mx:8080/xmlui/handle/11154/5>

Figura 26. Colecciones del repositorio institucional de Ciencias, mantenido por la Facultad de Ciencias, UNAM; contiene índices o catálogos, la producción de los investigadores e imágenes, disponible en: <http://repositorio.ciencias.unam.mx:8080/xmlui>

A continuación, se presenta una síntesis de las colecciones de literatura sobre biología más relevantes disponibles en línea. Se explican sus tipos, particularidades, usos y aplicaciones web diseñadas para fines académicos, con la finalidad de compararlas (Tabla 12).

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
AGRICOLA	National Agriculture Library	Libros, artículos y monografías, simposios, conferencias, reportes	1970	Agricultura: ciencia animal, acuicultura, biotecnología, botánica, conservación, citología, forestal, horticultura, ecología humana, hidrología, hidroponía, microbiología, pesticidas, fisiología, plantas, animales, ciencia de las plantas, contaminación, calidad del agua	Mundial	Libre	http://agricola.nal.usda.gov/cgi-bin/Pview/recon.cgi
FAO International	Food and Agriculture Organization (FAO)	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes	1975	Producción animal, ciencias acuáticas, forestal, fuentes naturales, contaminación, producción de plantas, protección de plantas, transgénicos	Mundial	Restringido UNAM	http://www.nti.gov/products/nti_b.aspx
Ecology and Microzoology Abstracts	Cambridge Scientific Abstracts	Artículos de revistas (<i>T</i>)	1982	Ficología: reproducción, crecimiento, ciclos de vida, bioquímica, genética, infección e inmunidad en el hombre, otros animales y plantas.	Mundial	Restringido UNAM	http://www.lib.jmu.edu/resources/more.asp?id=1427
Waterline	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos publicados	1960	Calidad del agua, legislación sobre el agua, análisis de los desechos del agua y su uso, afluentes industriales	Mundial	Restringido	http://www.environmentaldinsight.com/insight/ac_browse.do?hdAction=link&rowse_title&contentChar=A
Bacteriology Abstracts	Cambridge Scientific Abstracts	Artículos de revistas (<i>T</i>)	1982	Bacteriología, bioquímica, genética, inmunología, vacunas y enfermedades del hombre y los animales en sus aspectos básicos y clínicos	Mundial	Restringido UNAM	http://www.lib.csu.edu/search/lection/databases/more_info.pl?database=663
BIODATABASE	Elsevier	Conferencias, simposios,	1994	Microbiología aplicada y biotecnología, investigación	Mundial	Restringido	http://www.biobase-international.com

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		reuniones, artículos de revistas		sobre el cáncer, biología celular, ciencias ambientales y ecológicas, endocrinología y metabolismo, genética y biología molecular, inmunología, neurociencias, ciencias de las plantas, bioquímica de proteínas, toxicología			com/index.php?id=214
Business	Thomson Reuters	Registro bibliográficos y artículos públicos en diferentes medios	1995	Producción animal, biotecnología, medio ambiente, forestal, toxicología, tratamiento de residuos	Mundial	Restringido	http://portal.ac...org/citation.cf?id=15357
COMMERCE ia's	BioCommerce Data Ltd.	Directorios y registros bibliográficos	1981	Bioquímica, control biológico, conversión de biomasa, biopesticidas, hibridización celular, genómica, horticultura, inmunología, microbiología, biología molecular, recombinación de RNA, transgénicos, tratamiento de residuos	Mundial	Restringido	http://www.str...nternational.de...n_news_messa..._04.html?&cHash=bf79...e011&tx_ttnev...pointer]=1&tx...news[tt_news]...19&tx_ttnews...ckPid]=5152
Engineeri Abstracts	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes	1991	Aplicaciones marinas, aplicaciones medioambientales, bioinformática, acuicultura y piscaderías	Mundial	Restringido UNAM	http://www.lib...su.edu/search...ection/databas...more_info.php...database=289
Biological Abstracts	Thomson Reuters	Revistas, libros, memorias y reportes (<i>T</i>)	1926	Botánica, microbiología y farmacología	Mundial	Restringido UNAM	http://www.ov...com/site/catal...DataBase/24.j
Biological igest	Cambridge Scientific Abstracts	Compilación de resúmenes e índices de	1989	Ciencias biológicas	Mundial	Restringido UNAM	http://portal.ac...org/citation.cf

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		literatura nacional e internacional					id=607994
One	Cambridge Scientific Abstracts	Colección de resúmenes e índices de literatura de texto completo (T)	1915	Ciencias biológicas y de la salud	Mundial	Restringido UNAM	http://www.bioone.org
OSIS reviews	Thomson Reuters	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes (T)	1926	Biología aeroespacial, agricultura, bacteriología, bioquímica, bioingeniería, biofísica, biotecnología, botánica, biología celular, biología ambiental, genética, inmunología, microbiología, parasitología, patología, sistemática, toxicología, ciencia veterinaria, virología	Mundial	Restringido	http://www.ov.com/site/catalog/DataBase/26.j
OSIS toxicology	Thomson Reuters	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes	1969	Toxicología, agricultura, producción animal, biotecnología, medio ambiente, forestal, ingeniería genética, microbiología, pesticidas, toxicología, ciencia veterinaria, tratamiento de residuos	Mundial	Restringido	http://www.ov.com/site/catalog/DataBase/26.j
eme	Secretaría de la Salud del Estado de São Paulo	Revisiones sistemáticas, ensayos clínicos, sumarios de la evidencia, evaluaciones de tecnologías en salud	1967	Ciencias de la salud	Latinoamericano y el Caribe	Restringido UNAM	http://bases.bvud.org/publications/php/page_ow_main.php?home=true&language=en&form=simple
B	CABI	Revistas, libros, actas de congresos(T)	1910	Agricultura, ciencias ambientales(en particular, la ecología y el cambio	Mundial	Libre	http://www.cabi.org/datapage.a

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
				climático), microbiología y parasitología (entre ellos la micología, bacteriología y virología), ciencias de la plantas(incluida la biotecnología, protección de las plantas, los cultivos y la ciencia)			?iDocID=165
Chemical Abstracts	American Chemical Society	Colección de información sobre sustancias químicas	1800	Ciencias biomédicas, química, ingeniería, ciencia de materiales, ciencias agrícolas	Mundial	Restringido UNAM	http://www.cas.org/expertise/content/index.html
Directory of Open Access Journals	Universidad de Lund, Copenhague	Artículos de revistas científicas y académicas de acceso abierto	2002	Ciencia	Mundial	Libre	http://www.doaj.org
Enviroline	Congressional Information Service, Inc	Registros bibliográficos, informes, conferencias, simposios, reuniones, artículos de revista, artículos de periódicos	1975	Contaminación del aire, diseño ambiental y ecología urbana, educación ambiental, uso de la tierra, contaminación, océanos y estuarios, recursos renovables terrestres, agua, toxicología ambiental y seguridad, gestión de residuos, contaminación del agua, fauna	Mundial	Restringido	http://goliath.exlibris.com/coms/0199-76707/Enviroline-Ft-Lauderdale-FL.html
Environment Complete	EBSCO	Reportes científicos, revistas, monografías y libros	1940	Agricultura, ecología, energía, derecho ambiental, tecnología medio ambiental, ciencia marina y de agua dulce, recursos naturales	Mundial	Restringido UNAM	http://search.ebscohost.com
Environment Science Abstracts	Cambridge Scientific Abstracts	Revistas científicas, actas de congresos, informes,	1967	Biotecnología agrícola y ambiental, contaminación acuática y calidad ambiental, bacteriología, microbiología,	Mundial	Restringido UNAM	http://www.oxfordjournals.org/site/catalog/DataBase/50.js

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		monografías, libros y publicaciones gubernamentales		ecología , declaraciones de impacto ambiental, ingeniería ambiental, manejo de recursos naturales, contaminación, toxicología, recursos hídricos			
General Science Abstracts	The H. W. Wilson Company	Bibliografía, libros, monografías, artículos de revista, reseñas de libros	1984	Astronomía, ciencias de la atmósfera, biología, botánica, química, conservación, ciencias de la tierra, medio ambiente, genética, salud, microbiología, oceanografía, física, fisiología, zoología	Mundial	Restringido	http://www.oclc.org/Support/documentation/financial/research/databases/dbdetails/details/GenSciAbs.htm
Google Scholar (académico)	Google Inc	Documentos académicos	2007	Multidisciplinario	Mundial	Libre	http://scholar.google.com.mx/scholar_preferencias?hl=es
Industrial and Applied Microbiology Abstracts	Cambridge Scientific Abstracts	Artículos de revistas (<i>T</i>)	1982	Prácticas en la agricultura, alimentos y bebidas, químicos industrias farmacéuticas	Mundial	Restringido UNAM	http://www.lib.olyu.edu.hk/node/711
Pharmacology Abstracts	Dialog LLC	Artículos de revistas	1970	Cuestiones, como la toxicidad, relacionadas con la industria farmacéutica en general	Mundial	Restringido	http://library.dialog.com/blueshirts/html/bl0153.html
Life Sciences Abstracts	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes	1982	Ciencias biológicas, amino-ácidos, bacteriología, bioingeniería, biología de membranas, biotecnología, ecología, entomología, genética, manejo de recursos naturales, inmunología, microbiología, biología molecular, micología, neurociencias, ácidos nucleicos, oncogénesis protozoología, toxicología, virología, zoología	Mundial	Restringido	http://lib.stanford.edu/falconer/biology-library/databases-life-sciences

Nombre	Editor-Productores	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
Oceanic Abstracts	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos de revista	1981	Oceanografía biológica, ecología, oceanografía física y química, geología marina, geofísica, geoquímica, contaminación marina, recursos marinos, desalinización, biología marina	Mundial	Restringido UNAM	https://login.hcon.vuw.ac.nz/gin?url=http%3a%2f%2fwww.csa.m%2fhtml%2fdbrng.cgi%3fusername%3dvicwell%26access%3dvicwell12%26cat%3doceanic%26adv%3d1
Periódica	Dirección General de Bibliotecas UNAM	Registros bibliográficos de artículos originales, informes técnicos, estudios de caso, estadísticas y documentos publicados	1978	Agrociencias, arquitectura, astronomía, biología, ciencias de la atmósfera, computación, física, geofísica, geología, geografía, ingeniería, matemáticas, medicina, química, oceanografía, veterinaria	Latinoamérica y el Caribe*	Libre	http://132.248.1:8991/F/-/?func=find-b-0ocal_base=PE1
Pollution Abstracts	Cambridge Scientific Abstracts	Libros, documentos de conferencias, procedimientos, revistas, documentos de investigación, e informes técnicos	1981	Contaminación del aire, contaminación de agua dulce, contaminación terrestre, contaminación marina y tratamiento de aguas residuales	Mundial	Restringido	http://www.nestle.edu.au/section/library/data/se/pollution.htm
PubMed	U.S. National Library of Medicine	Artículos y revistas en ciencias de la vida con una concentración en biomedicina (T)	1950	Ciencias en enfermería, odontología, medicina veterinaria, farmacia, pre-clínica y ciencias afines	Mundial	Libre	http://www.ncbi.nlm.nih.gov/pmc/submitter/submitter.html

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
DALyC	Universidad Autónoma del Estado de México	Registros bibliográficos, artículos de revistas	2002	Agrociencias, arquitectura, astronomía, biología, ciencias de la atmósfera, computación, física, geofísica, geografía, geología, ingeniería, matemáticas, medicina, oceanografía química, veterinaria	Iberoamé-rica*	Libre	http://redalyc.emex.mx
ELO	FAPESP, UNAM	Revistas, números, artículos, citas	1997	Ciencias	Iberoamé-rica*	Restringido UNAM	http://www.scio.org/php/index.php?lang=es
ence ation lex*	Thomson Reuters	Artículos, bibliografías, reseñas de libros, correcciones y agregados, debates, editoriales, artículos de revisión, reseñas de software y hardware de computadora y bases de datos	1899	Ciencia y tecnología	Mundial	Restringido UNAM	http://wok.wiley.com
Search	Thomson Reuters	Artículos de revista, reseñas de libros	1990	Bioquímica, biología, ciencias biomédicas, ciencias de la tierra, ciencias ambientales, genética, microbiología, zoología	Mundial	Restringido	http://www.scisearch.international.com/scisearch.htm
opus	Elsevier	Registro bibliográfico, resumen, conferencias, citas e indicadores bibliométricos	1869	Multidisciplinaria	Mundial	Restringido UNAM	http://www.scopus.com/home
ter sources	Cambridge Scientific	Informes, libros monografías,	1967	Agua subterráneas, lagos, estuarios, erosión y	Mundial	UNAM	http://www.naturalhistory.oxford.ac.uk/science

Nombre	Editor-Pro ductor	Cobertura tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
Abstracts	Abstracts	conferencias, simposios, reuniones, artículos de revistas, patentes, casos de corte		sedimentación, suministro de agua y conservación, desalinización			ce/library/data/se/water.html
Biology & Ecology Worldwide	EBSCO	Artículos de revistas	1935	Ciencias de la vida	Mundial	UNAM	http://www.nis.com/Frame/N/C_products-fodified.htm
Biological & Agricultural Index	The H.W. Wilson Company	Artículos de revistas	1983	Agricultura, bioquímica, biología, biotecnología, botánica, citología, ecología, entomología, ciencias ambientales, ciencias de la pesca, ciencia de los alimentos, forestal, genética, horticultura, limnología, microbiología, neurociencias, nutrición, fisiología, patología de plantas, ciencia del suelo, medicina veterinaria, zoología	Mundial	Restringido	http://www.ov.com/site/catalog/DataBase/923.p
Biological Record Line	Thomson Reuters	Revistas, boletines, monografías, libros, revistas y actas de congresos (T)	1864	Comportamiento, biodiversidad, conservación y ciencias ambientales, biología marina y agua dulce, paleontología, parasitología, sistemática y taxonomía, técnicas, zoogeografía	Mundial	UNAM	http://www.ov.com/site/catalog/DataBase/200.p

(T) La base de datos tiene un tesauro UNAM, en diciembre del 2012 con suscripción a este servicio
 *La colección privilegia documentos en español

Tabla 12. Colecciones bibliográficas más relevantes con documentos de biología.

Cada base de datos tiene características diferentes, para elegir cuáles y cuántas es pertinente consultar, es necesario reconocer las particularidades de cada una.

Entre las colecciones bibliográficas más grandes están:

- *Google Scholar*.
- *Scopus* (de Elsevier)
- *Web of Science* (de Thomson Reuters)
- *Chemical Abstracts* (de la American Chemical Society)
- *Biosis Previews* (de Thomson Reuters)
- *PubMed* (de la National Library of Medicine)

Además del volumen de información, la cobertura de temas y disciplinas son características básicas a considerar para hacer una buena elección de la base de datos que se debe consultar. Del conjunto de las colecciones más grandes, resaltan tres multidisciplinarias que son las más utilizadas por la comunidad científica, son mantenidas por empresas de información, tienen cobertura mundial y ofrecen análisis de citas porque capturan las referencias de los documentos registrados:

- *Google Scholar* (académico)
- *Scopus*
- *Science Citation Index*

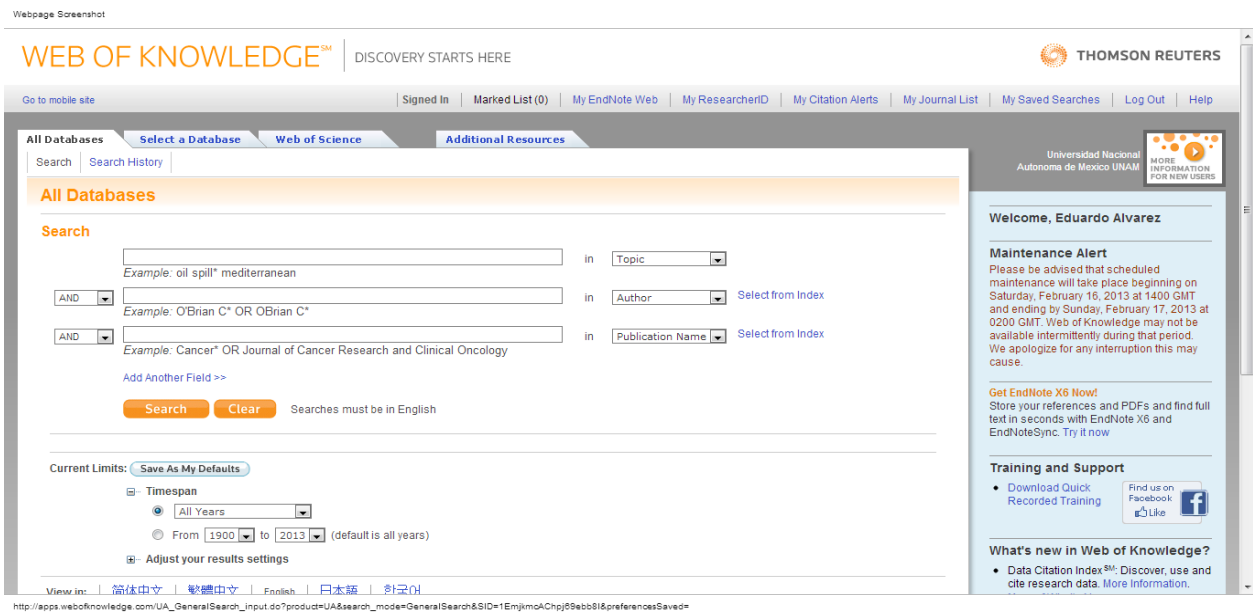
La primera es de libre acceso y las dos últimas son restringidas. Vale la pena resaltar en este rubro a SciELO, una base de datos también multidisciplinaria que cuenta con análisis de citas, pero con cobertura limitada a América Latina.

Dentro del conjunto de las especializadas en alguna área biológica sobresalen por su tamaño:

- *Biosis*
- *Zoological Record*
- *PubMed*

En la actualidad hay un debate entre la comunidad científica sobre si Google Académico es una fuente confiable de información o no, a causa de sus deficiencias por ejemplo que muchos de sus registros son de origen desconocido y que no se conocen los criterios de inclusión de documentos ni las coberturas que tiene. Éstos son datos indispensables para saber a qué tipo de información se está teniendo acceso.

Scopus es una base de datos reciente que está creciendo con gran rapidez; abarca una amplia gama de revistas, pero la captura de las referencias está actualmente limitada solo a los artículos recientes. *Science Citation Index* es la base de datos más utilizada porque es la más antigua; presenta herramientas sofisticadas de recuperación y análisis de información, cubre revistas de



corriente principal (alto impacto), tiene criterios estrictos de selección de los títulos, presenta una cobertura desde 1899, ha afinado el modo de recuperación de información y tiene registradas las referencias de todos los documentos que indiza (Fig. 27).

Figura 27. *Web of Science*.

PubMed es producida por la Biblioteca Nacional de Medicina (NLM) de los Estados Unidos que introdujo la primera búsqueda interactiva de base de datos (*Medline*) en 1971 y posteriormente en 1996, añadió el *Old Medline* con las publicaciones entre 1950 y 1965. En 1997, fue liberada PubMed (una combinación de ambos *Medline*) y desde ese momento se convirtió en el recurso bibliográfico en biomedicina más popular y uno de los más fiables. Entre sus ventajas están: acceso libre, gran volumen de información, vínculos entre muchos de los documentos y una gran cantidad de colecciones e información biológica a través de la plataforma llamada Entrez (Fig. 28), la actualización constante de los recursos y el uso de descriptores para cada documento con base en el tesoro del *MeSH* que permite hacer búsquedas más detalladas (Figs. 29 y 30).

The screenshot displays the Entrez search engine interface. At the top, there is a navigation bar with the NCBI logo on the left and the Entrez logo with the text "Entrez, The Life Sciences Search Engine" on the right. Below the navigation bar, there is a search bar with the text "Search across databases" and a "GO" button. The main content area is titled "Welcome to the Entrez cross-database search page" and is divided into several sections, each containing a list of databases with their respective icons and descriptions. The databases are organized into three main groups: a top group with PubMed, PubMed Central, and Site Search; a middle group with Nucleotide, EST, GSS, Protein, Genome, Structure, Taxonomy, SNP, Gene, SRA, BioSystems, HomoloGene, GENSAT, Probes, and Genome Project; and a bottom group with Journals, NLM Catalog, Books, OMIM, OMIA, dbGaP, UniGene, CDD, 3D Domains, UniSTS, PopSet, GEO Profiles, GEO DataSets, Cancer Chromosomes, PubChem BioAssay, PubChem Compound, PubChem Substance, Protein Clusters, and Peptidome. Each database entry includes a small icon and a brief description of the database's content.

Figura 28.Entrez (<http://www.ncbi.nlm.nih.gov/sites/gquery>).

MeSH Browser (2013 MeSH):
 The files are updated every week on Sunday.
[Go to 2012 MeSH](#)

Enter term or the beginning of any root fragments: or

Search for these record types:

Search in these fields of chemicals:

- Main Headings
- Qualifiers
- Supplementary Concepts
- All of the Above
- Search as MeSH Unique ID
- Search as text words in Annotation & Scope Note
- Heading Mapped To (HM) (Supplementary List)
- Indexing Information (II) (Supplementary List)
- Pharmacological Action (PA)
- CAS Registry/EC Number/UNII Code (RN)
- Related Registry Number (RR)

National Library of Medicine - Medical Subject Headings

2013 MeSH

MeSH Descriptor Data

[Return to Entry Page](#)

Standard View. [Go to Concept View](#). [Go to Expanded Concept View](#)

MeSH Heading	Epigenomics
Tree Number	H01.158.273.180.350.074
Tree Number	H01.158.273.343.350.042
Scope Note	The systematic study of the global gene expression changes due to EPIGENETIC PROCESSES and not due to DNA base sequence changes.
Entry Term	Epigenetics
Allowable Qualifiers	CL EC ED ES HI IS LJ MA MT OG SN ST TD
Previous Indexing	Epigenesis, Genetic (2002-2010)
History Note	2011
Date of Entry	20100625
Unique ID	D057890

[Natural Science Disciplines \[H01\]](#)
[Biological Science Disciplines \[H01.158\]](#)
http://www.nlm.nih.gov/cgi/mesh/2013/58B_spl

National Library of Medicine - Medical Subject Headings
 2013 MeSH
MeSH Descriptor Data
[Return to Entry Page](#)
 Standard View. [Go to Concept View](#). [Go to Expanded Concept View](#)

MeSH Heading	Epigenomics
Tree Number	H01.158.273.180.350.074
Tree Number	H01.158.273.343.350.042
Scope Note	The systematic study of the global gene expression changes due to EPIGENETIC PROCESSES and not due to DNA base sequence changes.
Entry Term	Epigenetics
Allowable Qualifiers	CL EC ED ES HI IS LJ MA MT OG SN ST TD
Previous Indexing	Epigenesis, Genetic (2002-2010)
History Note	2011
Date of Entry	20100625
Unique ID	D057890

MeSH Tree Structures

[Natural Science Disciplines \[H01\]](#)
[Biological Science Disciplines \[H01.158\]](#)
http://www.nlm.nih.gov/cgi/mesh/2013/58B_spl

Figura 29. Sitio del tesoro de *PubMed MeSH* (<http://www.nlm.nih.gov/mesh>).

Figura 30. Ejemplo de un descriptor de *MeSH*.

Otro indicador interesante de las bases de datos tiene que ver con la velocidad de captura de los documentos, la cual es cada vez más rápida, como resultado de la evolución de las tecnologías y de alianzas estratégicas entre las compañías o instancias encargadas de las colecciones.

Respecto a la cobertura tipológica, los documentos más comunes en las colecciones son, por

mucho, los artículos de revistas (de investigación y revisiones), que abarcan cerca del 90% de las bases de datos, seguidos por memorias, libros, folletos, bibliografías, monografías, artículos de prensa, simposios, patentes, directorios, resúmenes, citas, reseñas de libros, artículos de periódicos y conferencias. Este resultado era de esperarse, pues es prioritario catalogar los registros bibliográficos de la literatura primaria, que constituye la forma más común en que los nuevos conocimientos científicos son puestos a consideración de los pares. La excepción es Google que contiene más libros que revistas y por eso, junto con Amazon, es de las colecciones digitales más grandes de libros (Fig. 31).

El acceso a los documentos en texto completo de los registros bibliográficos constituye un factor fundamental para la elección de una base de datos. El proceso hacia el acceso abierto (*open access*) del software, las colecciones, las revistas y los documentos, especialmente en las ciencias biológicas, está en boga y sigue avanzando. El ahorro en los altos costos de impresión e insumos y la disminución del tiempo en los procesos de edición de los documentos electrónicos han facilitado la distribución gratuita y abierta a literatura académica. Con respecto a las bases de datos de acceso limitado, la UNAM paga suscripciones para que la comunidad tenga acceso a muchas de las bases de datos de acceso restringido más relevantes, por lo que es necesario que estas colecciones sean conocidas y aprovechadas en todo su potencial.

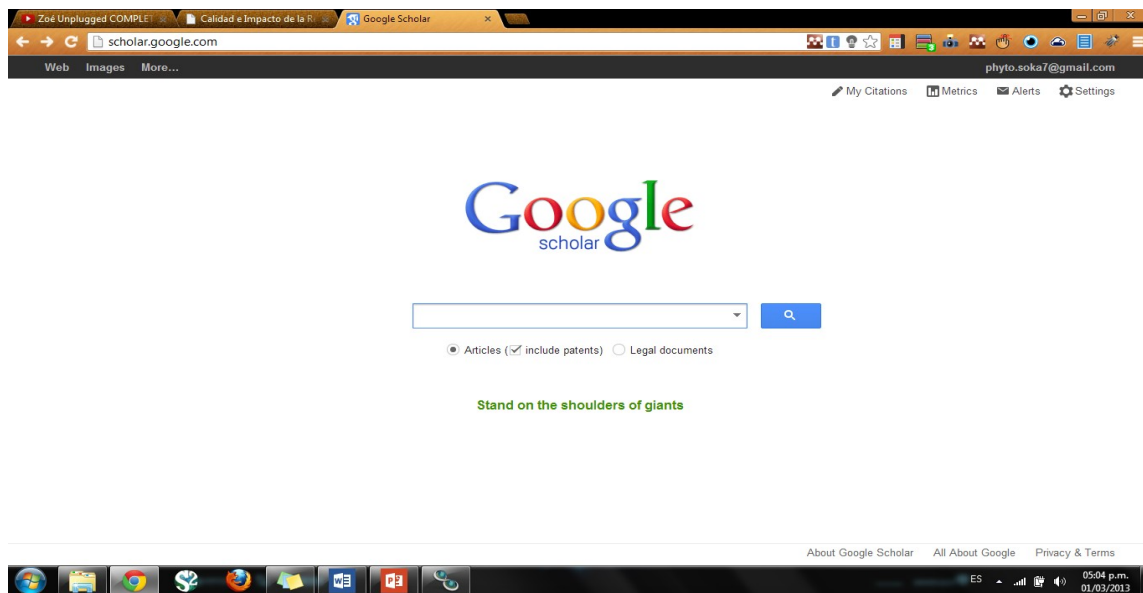



Figura 31. Google Académico.

El alcance de la información que se encuentra contenida en una base de datos está determinado fundamentalmente por la cobertura geográfica de los documentos que contiene. Tres bases de datos contienen únicamente documentos publicados en América Latina, en dos de ellas participa la UNAM y una más contiene información sobre Iberoamérica. Estas colecciones son realmente importantes para los científicos de la región latinoamericana y de países como el nuestro en los que la ciencia es publicada en su mayor parte en revistas nacionales o regionales, razones por las cuales estos documentos están pobremente representados en las bases de datos mundiales o de idioma distinto al inglés. Por lo tanto, cualquier científico mexicano para estar bien documentado sobre lo que se hace en su región deberá consultar necesariamente alguna base de datos local o regional como:

- Clase
- Periódica
- SciELO
- Redalyc
- BVS
- LILACS

La cobertura temporal permite identificar el periodo de tiempo en el que fueron publicados los documentos registrados en una base de datos determinada. Los documentos recién indizados van ampliando las bases de datos en el momento en que se capturan, pero también se da el crecimiento gracias a la recuperación de registros antiguos, es decir, los editores de muchas bases de datos están realizando el rescate de los archivos publicados en papel para someterlos al proceso de digitalización. Es importante recalcar que la cobertura temporal está en constante transformación por la adición de registros antiguos. Una colección que crece rápidamente con registros históricos es *Biodiversity Heritage Library* que además proporciona el texto completo en todos los casos (Fig. 32).


Now Online [Details](#)
47,594 titles
94,129 volumes
35,216,953 pages
[Recent additions](#) 

Connect with BHL



Featured Collection



BHL Updates  [Details](#)
[New Article Download](#)
[Features](#)
[BHL and Our Users: Megan Pahr and the History of Biology](#)

Search Results ([New Search](#))

Results for "mexico"

[Books/Journals](#) displayed : 100 [view more books/journals...](#)
[Authors](#) displayed : 8
[Subjects](#) displayed : 4
[Scientific Names](#) displayed : 16

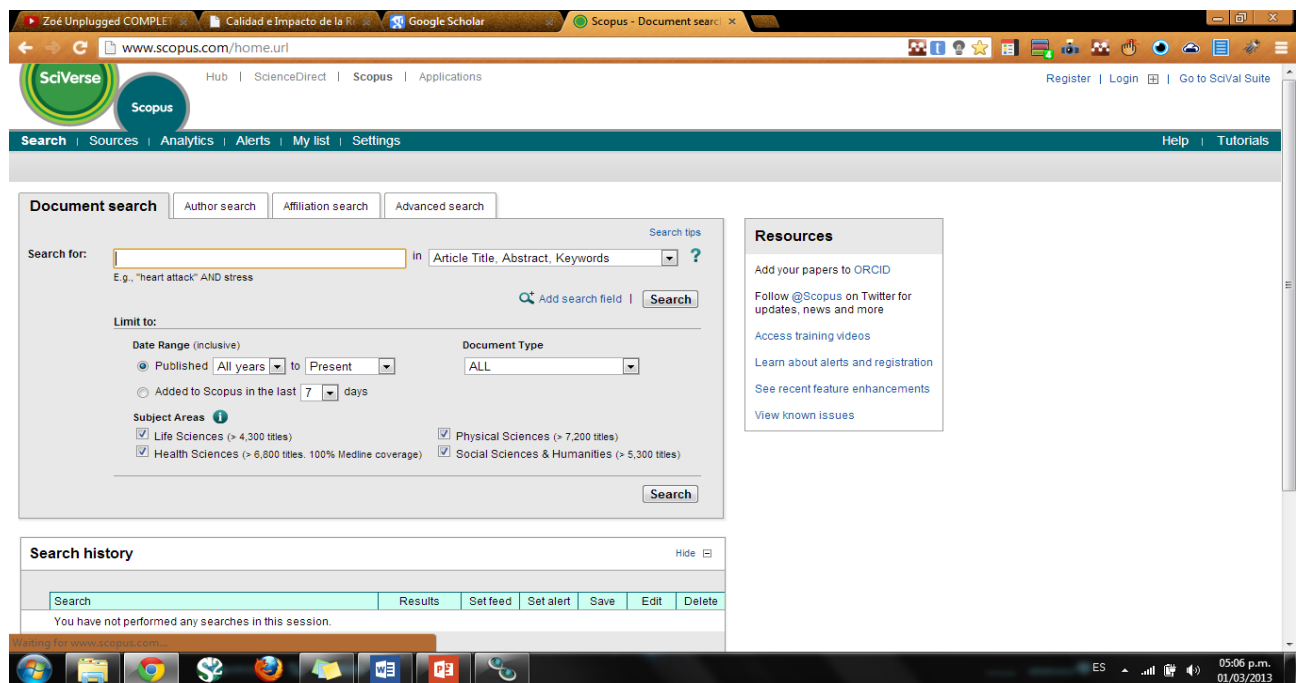
Books/Journals [Download results](#)

Sort By: [Title](#) [Author](#) [Year](#)

1. **An address delivered before the Oswego County agricultural society at Mexico, September 11,**
By: Tucker, Luther Henry,
Publication info: Albany,Printed by C. Van Benthuysen,1861.
Contributed by: Library of Congress (archive.org)
[View record](#) [View book](#)
2. **Agrarian reform and agricultural development in Mexico / by Juan Antonio Figueras.**
By: Figueras, Juan Antonio,
Publication info: 1972.
Contributed by: University of Florida, George A. Smathers Libraries (archive.org)
[View record](#) [View book](#)
3. **El algodón en México. Trabajo escrito de orden de la Secretaría de fomento, por el ingeniero agrónomo, Alberto Ruiz y Sandoval.**
By: Ruiz y Sandoval, Alberto. - Mexico.
Publication info: México :Oficina tip. de la Secretaría de Fomento,1884.
Contributed by: University of California Libraries (archive.org)
[View record](#) [View book](#)

Figura 32. Portal de la biblioteca *Biodiversity Heritage*, una de las iniciativas más importantes para la digitalización de documentos sobre biodiversidad.

Web of Knowledge (o *ISI Web of Knowledge*, la aplicación de Thomson Reuters para acceder a sus bases de datos, entre ellas *Web of Science*) es la que presenta la plataforma más completa (Fig.27) y le sigue de cerca *Scopus* (Fig. 33). Ambas contienen utilidades y herramientas informáticas que permiten buscar, recuperar, guardar, etiquetar e incluso gestionar la bibliografía de manera eficiente (*Endnote* para la primera y *Refworks* para la segunda). Además, presentan herramientas de meta-análisis de la literatura, como los siguientes indicadores bibliométricos: cantidad de citas e índice H, factor de impacto, vida media, índice de inmediatez y *eigenfactor* (<http://www.eigenfactor.org>) por mencionar las más relevantes. Todos estos recursos también



están en *Web of Science* y *Journal Citation Reports* esta última se publica en junio de cada año con acceso restringido. Para *Scopus*, algunos análisis se acceden desde la plataforma principal o a través del portal se *Scimago* (<http://www.scimagojr.com>) que es libre.

Figura 33. *Scopus*, disponible desde la UNAM en <http://www.scopus.com/home.url>.

A estos lujosos y atractivos sistemas de información les siguen algunos más sencillos subsidiados con fondos federales como es el exitoso caso de *PubMed* (<http://www.ncbi.nlm.nih.gov/pubmed/advanced>) (Fig. 34). Con una plataforma menos atractiva pero muy eficiente, registra gran cantidad de meta-información asociada a la literatura (genes, sustancias químicas, proteínas,

especies, etc); además tiene la ventaja de que distribuye toda su información de manera libre, por lo que existen cientos de aplicaciones gratuitas diseñadas por diversos grupos de investigación para extraer, manejar y analizar la información de manera automatizada, inmediata y sistematizada. Basadas en la Web 2.0, aplican métodos innovadores de análisis como la bibliometría, las redes y el descubrimiento basado en literatura (minería de textos). Una lista preliminar se puede consultar en <http://www.diigo.com/user/lmichan/pubmed>.

The image shows a screenshot of the PubMed advanced search interface. The top section is titled "Search by Author, Journal, Publication Date, and more". It contains a search form with the following fields:

- Author: [dropdown] [text input] [Index]
- Journal: [dropdown] [text input] [Index]
- Publication Date: [dropdown] [text input] to [text input] [Index]

Below the search fields, there are radio buttons for "All of these (AND)" (selected) and "Any of these (OR)". A link "Click here! Add More Citation Search Fields" is also present. At the bottom of the search section are "Clear All" and "Search" buttons.

The bottom section is titled "Limit by Topics, Languages, and Journal Groups". It contains several filter panels, each with a "Clear" button:

- Full Text, Free Full Text, and Abstracts:** Includes checkboxes for "Links to full text", "Links to free full text", and "Abstracts".
- Humans or Animals:** Includes checkboxes for "Humans" and "Animals".
- Gender:** Includes checkboxes for "Male" and "Female".
- Type of Article:** Includes checkboxes for "Clinical Trial", "Editorial", "Letter", "Meta-Analysis", and "Practice Guideline".
- Languages:** Includes checkboxes for "English", "French", "German", "Italian", and "Japanese".
- Subsets:** Includes checkboxes for "Core clinical journals", "Dental journals", and "Nursing journals".
- Ages:** Includes checkboxes for "All Infant: birth-23 months", "All Child: 0-18 years", "All Adult: 19+ years", and "Newborn: birth-1 month".

Figura 34. Búsqueda avanzada en *PubMed*.

Los portales y recursos que ofrecen las bases de datos de América Latina como SciELO (Fig. 35), BIREME (Biblioteca Virtual en Salud) (Fig. 36), REDALyC (Fig. 37) y Periódica (Fig. 38) se han modernizado de manera acelerada durante los últimos años ofrecen, junto con la información bibliográfica, algunos indicadores bibliométricos y varias herramientas para manejar la bibliografía, aunque falta todavía mucho por hacer.

Webpage Screenshot

SciELO

revistas | articulos

alfab materia búsqueda autor materia búsqueda

Colección de la biblioteca

Base de datos : title Formulario libre

Buscar por : [Formulario básico](#)

Ingrese una o más palabras

Todas las palabras (AND) Cualquier palabra (OR)

Notas :

- No entre operadores lógicos (AND, OR o AND NOT) entre las palabras. Separe las palabras con espacio.
- Use el símbolo de truncar \$ para buscar palabras con la misma raíz. Ejemplo: educ\$ recupera educación, education, educação, etc. No es necesario usar operadores booleanos (AND, OR o AND NOT) entre las palabras. Seleccione la opción Todas las palabras (AND) para relacionar las palabras (restringe la búsqueda), o la opción Cualquier palabra (OR) para sumar las palabras (amplia la búsqueda).
- Para buscar por otros campos o para especificar el campo de busca, use el Formulario Básico.

Search engine: [iAH](#) powered by [WWWISIS](#)

BIREME/OPS/OMS - Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud

<http://www.scielo.org.mx/cgi-bin/wxis.exe/iah/?luisScript=iah/iah.xis&base=title&fmt=iso.pt&lang=es>

Figura 35. Portal de *SciELO*.



Figura 36. Sitio de la Biblioteca Virtual en Salud (BVS).

SISTEMA DE INFORMACIÓN CIENTÍFICA REDALYC
RED DE REVISTAS CIENTÍFICAS DE AMÉRICA LATINA Y EL CARIBE, ESPAÑA Y PORTUGAL

Reconocimiento
al Sistema de Información Científica Redalyc por la
diseminación de la comunicación científica y democratización
del conocimiento en estudios territoriales Iberoamericanos

XII SEMINARIO
INTERNACIONAL RII
Y TALLER DE EDITORES RIER

Una ventana de acceso abierto al conocimiento iberoamericano

Acceso a editores

808
REVISTAS
CIENTÍFICAS

20,645
NÚMEROS

260,310
ARTÍCULOS A
TEXTO COMPLETO

Búsqueda

Artículos Revistas

English
Français
Português

Avanzada

ARTÍCULOS

REVISTAS

Atlas
Cienciométricos

Iberoamérica
Venezuela
Austral
Chile
Psicología
Inter ciencia

SAIGE
SAIGE Acceso

Portales

Redalyc
Comités
Postulación de revistas
Política editorial
Acceso abierto

Consulta por área

Ciencias Sociales y Humanidades

http://redalyc.uaemex.mx/

Figura 37. Servicio de REDALyC.

Contáct@ | Ayuda

ÍNDICE DE REVISTAS LATINOAMERICANAS EN CIENCIAS PERIÓDICA

ACERCA DE PERIÓDICA BÚSQUEDAS MIS REGISTROS HISTORIAL OTROS CATALOGOS FINALIZAR SESIÓN

Búsqueda avanzada

Campos de búsqueda

Texto a buscar

¿Palabras adyacentes?

No de registros

Autor

Autor

Autor

No Si

No Si

No Si

Toda la base

Solo registros con enlace a texto completo

Presione sobre el número para ver los registros

Total:

Buscar Limpiar

Sugerencias generales para realizar búsquedas

Uso de operadores booleanos

© 2009 DGB - UNAM

Figura 38. Periódica.

Muchas de las herramientas presentadas en el Capítulo II, como los marcadores sociales, las folkosonomías o las que permiten comentar y compartir documentos, en especial en formato PDF, están disponibles en muchas de las plataformas de las bases de datos, de tal manera que uno se puede inscribir a alguno de estos servicios y acceder a ellos desde la aplicación de la base de datos elegida.

Otra característica a considerar para consultar una base de datos bibliográfica es la existencia de un tesoro que se usa para clasificar cada uno de los documentos por medio de descriptores. Esta herramienta tiene una ventaja importante, pues implica que hay especialistas que están clasificando el contenido de los documentos, lo que produce un alto grado de certeza en las búsquedas. La mayoría de los tesauros se consultan en la aplicación en línea de la base de datos en cuestión; si es restringida, generalmente el tesoro también lo es. En otros casos están disponibles para todo público, como es el caso de *MeSH* de *PubMed* (<http://www.nlm.nih.gov/mesh>) con 177, 000 entradas y *CAB Abstracts*

Webpage Screenshot

CAB Thesaurus 2012

Home | All terms A-Z | Browse by subject / category | Cookie Information

Search Thesaurus :

Language: English

Search type: terms begin with text

Results format: simple term list

Terms per page: 15

Match terms *: Equal or Greater

Term type *: Any

search

* Available only to search type "term begins with text"

[Help with search >>](#)

About CAB Thesaurus

The CAB Thesaurus is the essential search tool for all users of the CAB ABSTRACTS™ and Global Health databases and related products. The CAB Thesaurus is not only an invaluable aid for database users but it has many potential uses by individuals and organizations indexing their own information resources for both internal use and on the Internet

Its strengths include:

- Controlled vocabulary that has been in constant use since 1983
- Regularly updated (current version released September 2012)
- Broad coverage of pure and applied life sciences, technology and social sciences
- Approximately 250,200 terms, including 139,300 preferred terms and 110,900 non-preferred terms
- Specific terminology for all subjects covered
- Includes about 191,200 plant, animal and microorganism names
- Broad, narrow and related terms to help users find relevant terminology
- Cross-references from non-preferred synonyms to preferred terms
- Multi-lingual, with Dutch, Portuguese and Spanish equivalents for most English terms, plus organism names in Danish, Finnish, French, German, Italian, Norwegian and Swedish
- American and British spelling variants
- Relevant CAS registry numbers for chemicals

Search string

Click the **CS** button beside a term to add it to the below search string box

clear search

Send to CAB Direct

> Send to Google

> Send to Yahoo!

<http://www.cabi.org/cabthesaurus/>

(<http://www.cabi.org/cabthesaurus>) con 136, 900 (Fig. 39).

Figura 39. Tesoro de CAB.

Ninguna de las bases de datos bibliográficas existente cumple con todas las características suficientes para ser utilizada como única fuente de información. Por lo tanto, lo conveniente es utilizar distintas combinaciones que satisfagan las necesidades informativas, y de ellas, elegir cuáles son las más adecuadas (Fig. 40). Por ejemplo, para recuperar artículos sobre la investigación sobre biodiversidad de América Latina, sería idóneo utilizar al menos tres bases de datos: una general sobre temas biológicos como BIOSIS, que es la más extensa y que tiene bien representados los últimos 30 años; una internacional especializada en el taxón de interés como CAB (*Centre for Agricultural Bioscience* para botánica) o ZR (*Zoological Records* para zoología) y una latinoamericana como *Periódica* que tiene la mayor cobertura documental producida en la región. SCI (*Science Citation Index*) o *Scopus* servirían para identificar las tendencias actuales e ideas innovadoras en el tema y para identificar los artículos de mayor impacto a partir de la cantidad de citas que reciben. Si se pretende hacer una búsqueda sobre la sistemática de las angiospermas latinoamericanas en los últimos 20 años sería preciso usar al menos dos bases de datos, BIOSIS y *Periódica*; pero si se quisiera estudiar un periodo anterior, entonces sería más conveniente usar CAB y *Periódica*. Una lista con las características de cada una se pueden consultar en Line@ (Fig. 41).



Figura 32. Las colecciones bibliográficas más relevantes para biología clasificadas por categoría para facilitar su elección, una lista con las características de cada una se pueden consultar en Line@ (Fig 33).

Figura 40. Las colecciones bibliográficas más relevantes para biología clasificadas por categoría para facilitar su elección.

The screenshot displays the Line@ website interface. At the top, there is a dark blue header with the UNAM logo and the text "Ciencias, UNAM". A navigation bar below the header shows the path: "DSpace Home → FACULTAD DE CIENCIAS, UNAM → Departamento de Biología Evolutiva → Laboratorio Cienciometría, Información e Informática Biológica → Line@: Colecciones Bibliográficas para investigación biológica relacionadas y afines".

The main content area is titled "Line@: Colecciones Bibliográficas para investigación biológica relacionadas y afines". It features a "Browse by" section with links for "By Issue Date", "Authors", "Titles", "Subjects", and "By Submit Date". Below this is a search box with a "Go" button and a link to "Advanced Search".

On the right side, there is a "Search DSpace" section with a search input field, a "Go" button, and radio buttons for "Search DSpace" (selected) and "This Collection". Below this is a "Browse" section with links for "All of DSpace", "Communities & Collections", "By Issue Date", "Authors", "Titles", "Subjects", "By Submit Date", and "This Collection", followed by similar links for "This Collection".

At the bottom right, there is a "My Account" section with links for "Login" and "Register".

The "Recent Submissions" section lists two items:

- Item 1:** Title: [Digital CSIC](#), Author: CSIC- Agencia Estatal Consejo Superior de Investigaciones Científicas, Date: 2011-06-03.
- Item 2:** Title: [Wilson Biological & Agricultural Index](#), Author: The H.W. Wilson Company, Date: 2011-06-03.

Figura 41. Line@, base de datos con detalles de las colecciones bibliográficas para investigación biológica, relacionadas y afines. Disponible en: <http://repositorio.fciencias.unam.mx:8080/xmlui/handle/123456789/43331>.

Los editores que producen y soportan la mayoría de las bases de datos privadas más utilizadas son compañías como Cambridge Scientific Abstracts, Thomson Reuters, The H. W. Wilson Company, Elsevier y EBSCO, exitosas empresas de la información que están en un momento de expansión comercial, abarcan muchos tipos de información, todas las áreas del conocimiento y diversas aplicaciones como la academia, la ciencia, las noticias, las finanzas, las fotografías, las gráficas y los videos. Constituyen un ejemplo innegable del valor de este producto en la sociedad del conocimiento en la que vivimos actualmente. El manejo de la información es indispensable para las instituciones gubernamentales y académicas, las cuales realizan un gran esfuerzo por

mantener repositorios propios, en la mayoría de los casos de acceso abierto, como es el caso de la UNAM, que mantiene las bases de datos más grandes de literatura publicada en el país e incluso en la región. Sin embargo, la mayoría de bases de datos en la actualidad se lleva a cabo en el sector privado y su uso es estrictamente comercial (pago por servicio); en tanto, los esfuerzos de instituciones académicas y gubernamentales avanzan a paso lento, frenadas por la burocracia y la falta de recursos.

Los registros de literatura en las bases de datos aumentan aceleradamente debido a procesos de recuperación mejores y más completos, digitalización y sistematización de las publicaciones. Cabe destacar que el crecimiento no es casual, lo impulsan grandes casas editoriales para su beneficio. De manera que acaparan el mercado, crecen con rapidez, recuperan con más celeridad la mayor cantidad de información documental de sus archivos antiguos logran conseguir nuevos documentos en contraste con la ingenua visión y labor de algunas instituciones académicas, lo que nos muestra el interés en acopiar datos, libros, publicaciones, memorias, tesis y patentes. Y como su restricción indudablemente finca fronteras a las sociedades del conocimiento.

V. META-ANÁLISIS DE LITERATURA PARA BIOLOGÍA

El análisis de la literatura digital, en especial de la información contenida en los títulos, resúmenes y descriptores, se ha consolidado como una práctica común en el área científica para contrastar hipótesis. Se han producido en los años recientes aplicaciones informáticas que facilitan la extracción y procesamiento de los artículos científicos, las cuales permiten de manera automática realizar meta-análisis de miles de registros bibliográficos o incluso el texto completo de los documentos. Actualmente existen distintas maneras de realizar meta-análisis dependiendo del objeto de estudio, la infraestructura y tecnología con la que se cuente. Se realiza para cualquier área del conocimiento con el fin de sostener y fortalecer hipótesis y conclusiones de trabajo.

El uso del meta-análisis se ha incrementado desde su implementación metodológica en una convención anual de la American Education Research Association en 1976; el término “meta-análisis” fue acuñado por Gene V. Glass en 1976, quién lo aplicó a las ciencias sociales y a la psicología. A partir de la década de los 80, su uso se incrementó en medicina y para los años 90 fue muy frecuente encontrar artículos biomédicos o afines que aplicaban el meta-análisis. En la actualidad, gracias al avance tecnológico y a la conjunción de áreas de estudio como las ciencias de la computación, las ciencias de la información, el desarrollo de las bases de datos y el uso del Internet, podemos realizar meta-análisis de literatura de manera automática, personalizada, eficaz, rápida y selectiva.

El **meta-análisis** es una técnica cuantitativa para indicar las relaciones entre las variables de cada uno de los estudios sujetos a análisis. La técnica hace hincapié en los resultados de múltiples estudios en comparación con los resultados a partir de una sola investigación. La obtención de dichos resultados se ve estrechamente relacionada con al menos cuatro alternativas distintas de realizar el método cuantitativo, dependiendo del objeto que se quiera estudiar y/o la hipótesis que se planteé.

En este capítulo se describirán los principales métodos de meta-análisis de literatura más comunes y se proporcionarán ejemplos de aplicaciones web para realizarlos. Existen en línea varias aplicaciones automáticas que realizan el meta-análisis; algunas proporcionadas por las propias colecciones bibliográficas como *Web of Science*, *Scopus* o *SciELO* y otras diseñadas por

terceros (empresas o instituciones académicas) que utilizan colecciones para procesar información, como *Scimago* o *Eigenfactor*. Los análisis pueden ser incluso en tiempo real, como los que hacen *PubMed PubReMiner* y *Ligercat* a partir de la base de datos de *PubMed*. Para esto se hace una búsqueda en la aplicación y los resultados del meta-análisis se visualizan en la pantalla como gráficas, patrones y relaciones de la información contenida en los documentos, por ejemplo, texto, sustancias químicas, genes, proteínas, enfermedades, etc. Están disponibles para el uso de la comunidad académica en general, no especializada en las áreas de ciencias computacionales o versadas en la programación. La mayoría de estas herramientas son iniciativas académicas diseñadas y mantenidas por los propios grupos de investigación, en su mayoría dedicados a la bioinformática, que liberan dichos servicios para el uso del público en general, muchos de ellos son amigables y de uso libre.

El propósito del meta-análisis para literatura es múltiple: sirve para mejorar el proceso de recuperación de información, para la obtención de nuevo conocimiento, funciona para evaluar las características de un conjunto de literatura determinado e incluso se usa para realizar el análisis, desarrollo, estructura, relaciones y dinámica de un área científica.

Para exponer los distintos tipos de meta-análisis los hemos clasificado en cuatro tipos que se pueden combinar: 1) bibliometría, 2) análisis de redes sociales, 3) minería de textos y 4) semántica.

1 Bibliometría

La bibliometría fue conceptualizada por Derek de Solla Price en 1963 en su gran obra titulada *Little Science, Big Science*, donde de manera gráfica explica el alcance de la ciencia en cuanto a la producción por años como una curva con crecimiento exponencial (Fig. 42). Se define como el método cuantitativo que permite observar la estructura y desarrollo de la literatura; existen variables de este método, cada una denominada de la siguiente manera: cienciometría, infometría, cibernetría y alt-metría (Fig. 43). Dependiendo del grupo de trabajo o los autores que representan esta área de estudio, estos términos pueden utilizarse indistintamente o como sinónimos.

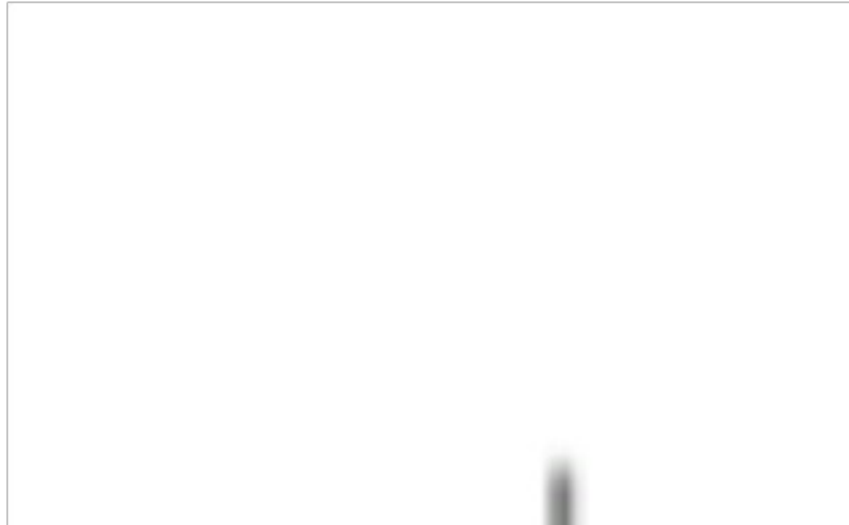


Figura 42. Representación gráfica del crecimiento exponencial de la ciencia según De Solla Price (1963).



Figura 43. Variables de meta-análisis según el objeto a estudiar.

La **cienciometría** se emplea comúnmente para la evaluación de la ciencia (tanto las naturales como las sociales). Los indicadores obtenidos miden de manera estandarizada y comparable la producción y la contribución de las publicaciones científicas al conocimiento sobre un tema determinado y examinan el desarrollo de las políticas científicas de países y organizaciones. Las revistas más importantes sobre bibliometría son *Scientometrics*, *Journal of the American Society*

for *Information Science and Technology*, *Information Processing Management* y *Journal of Information Science* donde se publican artículos sobre:

- El crecimiento cuantitativo de la ciencia.
- El desarrollo de las disciplinas y subdisciplinas.
- La relación entre ciencia y tecnología.
- La **obsolescencia** de los paradigmas científicos.
- La estructura de comunicación entre los científicos.
- La productividad y creatividad de los investigadores.
- Las relaciones entre el desarrollo científico y el crecimiento económico.
- La colaboración, autoría, citación y colaboración.

Un programa en línea que aplica la cienciometría es *Eigenfactor* (<http://eigenfactor.org>) (Fig 44); es una fuente de datos métricos sobre las revistas científicas internacionales basada en información de citas del *Journal Citation Reports* (<http://apps.webofknowledge.com> [en *Additional Resources*]) y en la aplicación de algoritmos de relevancia tipo *Page Rank* de Google. Proporciona datos anuales desde 1995 a 2010 acerca de las 7,000 revistas incluidas en JCR y de otras muchas fuentes citadas por ellas. Presenta tablas clasificatorias temáticas con dos indicadores: *Eigenfactor* y *article influence*. Además, ofrece análisis de calidad, el precio de las revistas y un módulo de representación gráfica de las relaciones entre las diferentes áreas de las ciencias.

LigerCat (<http://ligercat.ubio.org>), (Fig. 45) y es una propuesta de la Biblioteca de Biología como parte del proyecto de envejecimiento Woods Hole Marine Biological Laboratory. Es una herramienta de búsqueda y análisis de resultados de *PubMed* con base en tres criterios: artículos, revistas y genes. Utiliza nubes con etiquetas que proporcionan una visión general de los conceptos importantes y gráficas con las tendencias de publicación de los documentos recuperados en la búsqueda realizada. Como la búsqueda es en tiempo real y se utiliza directamente *PubMed*, simplemente hay que dar clic en un descriptor o una barra y se accede a los documentos correspondientes al criterio seleccionado dentro de *PubMed* (Fig. 46).



Figura 44. Interfaz de *Eigenfactor* (<http://eigenfactor.org>) con un ejemplo de análisis del área de biología molecular y las disciplinas relacionadas representadas en forma de red.

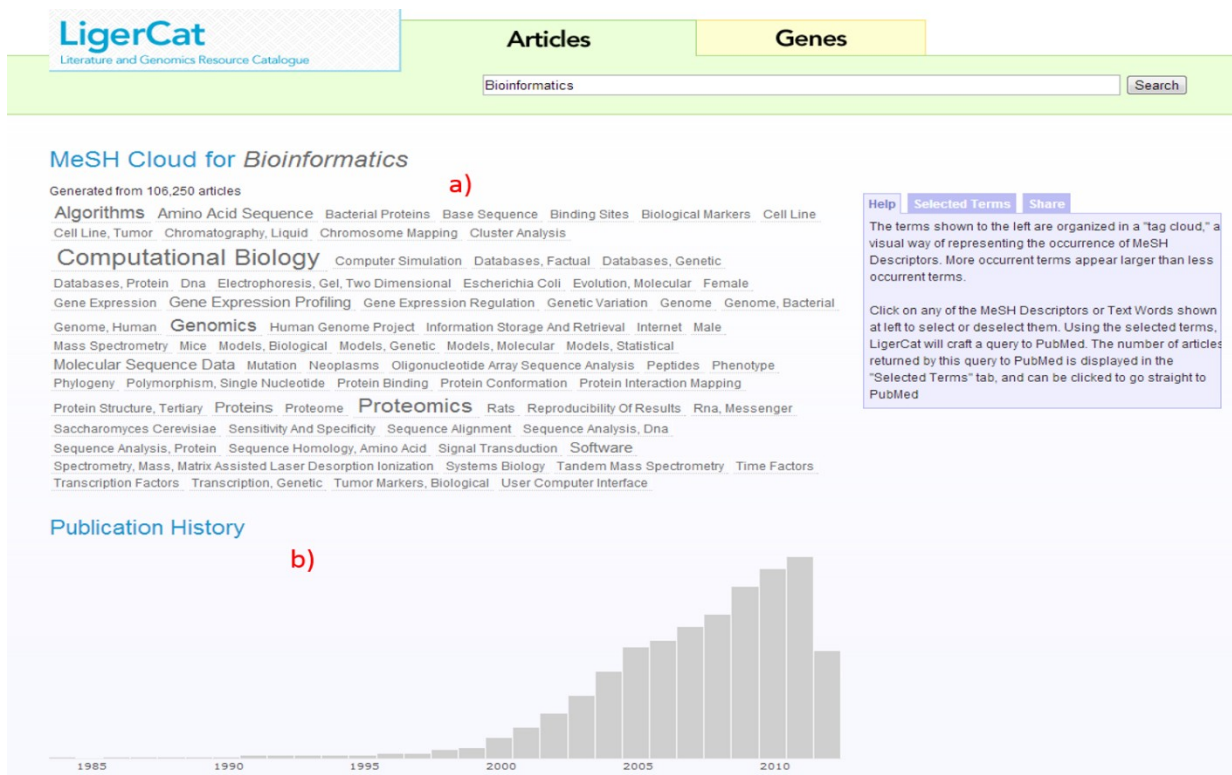


Figura 45. Interfaz de *LigerCat* con la búsqueda apoptosis. a) nube de términos relevantes; b) histograma con el número de artículos publicado por año.

PubMed PubReMiner es una aplicación creada por Jan Koster en 2007; está dirigida a aquellos que necesitan localizar y analizar la literatura biomédica sobre un tema en particular (Fig. 46). Permite encontrar información relevante mediante el análisis de los resúmenes de las publicaciones registradas en *PubMed* con tres propósitos:

Selección de una revista sobre su trabajo actual (mediante el escaneo de las revistas más utilizadas en investigaciones similares).

- Búsqueda de expertos en un área de investigación (mediante la visualización de los autores asociados a su consulta).
- Determinar el interés de la investigación de un autor (al ver las palabras clave asociadas con ese autor).

Los resultados aparecen en una tabla de frecuencias que resulta muy útil para refinar la búsqueda y seleccionar la bibliografía de interés, dando clic a cualquier opción permite desplegar automáticamente los registros en *PubMed*.

Webpage Screenshot



Figura 46. *PubMed PubReMiner*, plataforma para análisis bibliométricos en el área biomédica]: 1) Consulta; 2) indicadores bibliométricos; 3) ir a la consulta en *PubMed*; y 4) análisis realizado para la consulta “Gerontechnology”.

El laboratorio de cibermetría del Centro de Ciencias Humanas y Sociales (CCHS) dentro del

mayor centro nacional de investigación de España, el CSIC, se dedica al análisis cuantitativo de Internet y de los contenidos de la red, especialmente de aquellos relacionados con el proceso de generación y comunicación académica del conocimiento científico. Esta es una disciplina nueva y emergente que ha sido denominada **cibermetría**, también conocida como *Webometrics*. Existen revistas especializadas en el tema principalmente la revista *Cibermetrics*, editada por el mismo laboratorio. Un ejemplo es el Ranking Web de la Universidades del Mundo, creado y mantenido por el Laboratorio de Cibermetría Español (Fig. 47).

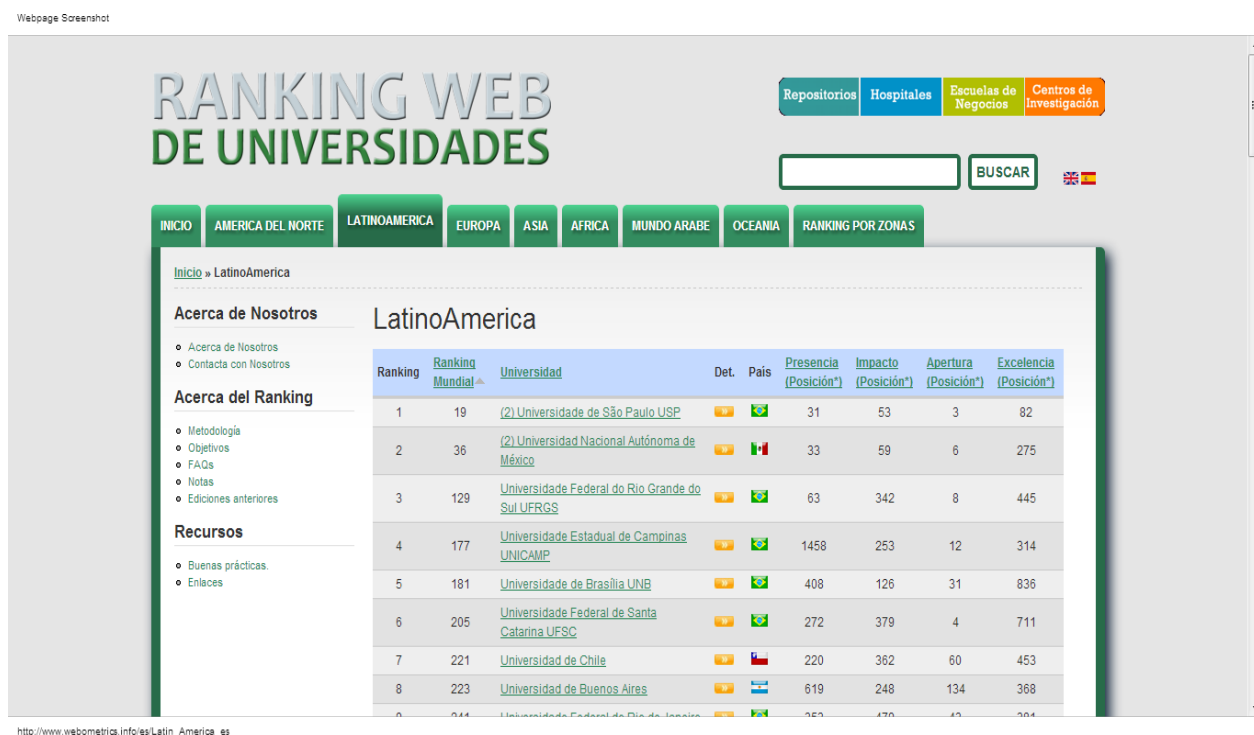


Figura 47. Ranking Web de Universidades del Mundo en el área de Latinoamérica.

Recientemente se ha desarrollado la **Alt-metría**, un nuevo método para estudiar las medidas sociales del impacto académico a través de las publicaciones con base en nuevos indicadores sociales, es decir, midiendo el uso de blogs, plataformas de **microblogging**, sistemas de producción entre pares y sistemas de colaboración y anotación, como son los servicios de marcadores sociales Mendeley (<http://www.mendeley.com>), CiteULike (<http://www.citeulike.org>) o Zotero (<http://www.zotero.org>). También se pueden usar para este fin las anotaciones electrónicas (marcar, calificar o comentar) que se realizan en revistas electrónicas como *PlosONE* (<http://www.plosone.org/home.action>), *BMC Research Notes*

(<http://www.biomedcentral.com/bmcresnotes>) o *BMJ Open* (<http://bmjopen.bmj.com>), etc. Este procedimiento se publicó por primera vez el 26 de octubre del 2010, fue redactado por Jason Priem de la Universidad de Carolina del Norte-Chapel Hill, Dario Taraborelli de Wikimedia Foundation, Paul Groth de la Universidad de Amsterdam y Cameron Neylon de Science and Technology Facilities Council y promete ser muy utilizado en los años futuros (Fig. 48).



Figura 48. Esquema de clasificación de los indicadores para cuantificar el impacto de la bibliografía en una red social según *Alt-metrics* (<http://altmetrics.org/manifesto>) (modificado de Priem *et al.*, 2010).

13. Análisis de redes

La teoría de redes es una aplicación matemática de la teoría de grafos. Tuvo su inicio con el matemático suizo Leonhard Euler, quien planteó en 1735 el curioso problema de los siete puentes sobre el río Pregel de la ciudad prusiana de Königsberg. Años después, Moreno en 1953 realizó la primera interpretación gráfica de una matriz de datos sociográfica que hasta la fecha se sigue utilizando para análisis, representación y visualización de las relaciones socio-académicas. Existen cuatro categorías de redes según su estructura y propiedades (Tabla 13) y tres de acuerdo al número de nodos que las componen (Tabla 14).

Tipos de redes	Descripción
Redes sociales	Las redes sociales están compuestas por individuos o grupos de individuos con patrones de contacto o interacciones entre ellos. Ejemplos de este tipo de redes son la colaboración entre investigadores, las relaciones de amistad o de negocios entre directivos de empresas y los lazos familiares a partir de sus matrimonios y descendencia (genealogías).
Redes de información	También denominadas redes de conocimiento. El ejemplo clásico de redes reales de esta categoría son las de citas y co-citas de trabajos científicos. Otro ejemplo ampliamente estudiado de redes de información es la estructura de la información dentro de la <i>World Wide Web</i> .
Redes tecnológicas	Son las redes diseñadas para la distribución de electricidad (energía), agua, gas, las redes de transportes (carreteras, ferrocarril, rutas aéreas), las redes telefónicas (sólo las redes físicas de cables y postes, puesto que las redes de llamadas formarían parte de las denominadas redes sociales) o Internet, como red de interconexión de las computadoras.
Redes biológicas	Son diversos los sistemas biológicos susceptibles de representarse en forma de redes. Las redes de reacciones metabólicas, las redes genéticas, las redes ecológicas y cadenas tróficas, las redes neuronales o las vasculares son algunos de los ejemplos de redes biológicas analizadas desde la perspectiva de la teoría de redes.

Tabla 13. Clasificación de redes según sus características y propiedades.

El **análisis de redes** implica un avance cuantitativo y cualitativo relevante en la representación e investigación de la estructura y las relaciones de los dominios científicos, ya sean geográficos, temáticos, institucionales o de individuos. Las redes son estructuras abiertas que se pueden expandir sin límite, por medio de la integración de nuevos nodos (Fig. 49).

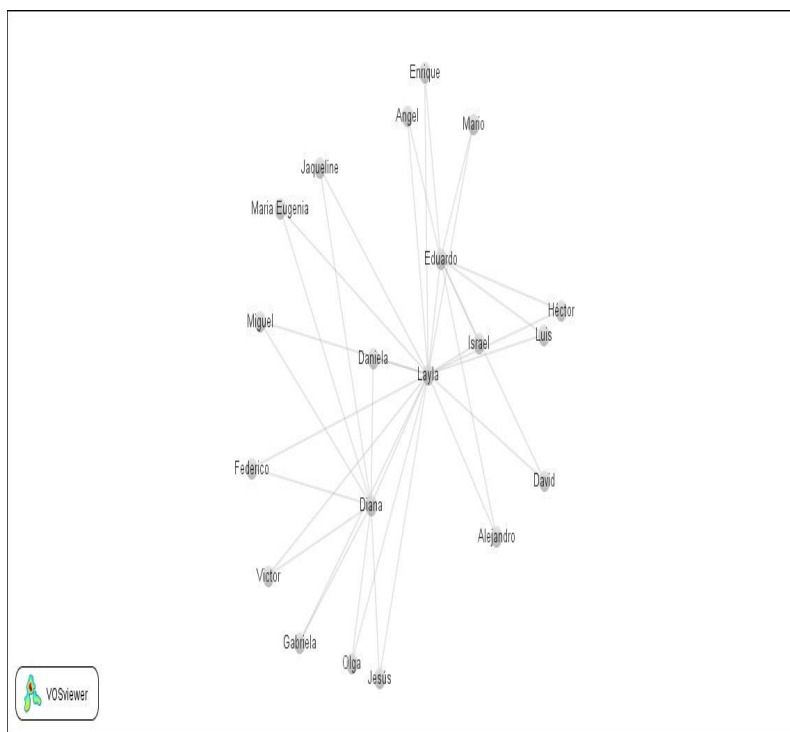


Figura 49. Ejemplo de un análisis de redes con los integrantes del laboratorio. A la izquierda el grupo que trabaja biodiversidad, a la derecha inferior biomedicina y derecha superior bioinformática.

Tamaño	Descripción
Redes pequeñas	Contienen un máximo de 100 nodos. Ejemplos son algunas redes sociales, de ecosistemas biológicos o de exportación-importación de productos entre países. En ellas es posible mostrar la totalidad de nodos, sus atributos y los enlaces que las unen. El tamaño de los nodos suele representar atributos como la importancia, el poder o el nivel de trabajo.
Redes medianas	Incluyen más de 100 y hasta 1,000 nodos. Ejemplos destacados son las redes genéticas, las metabólicas o las económicas. En ellas también es posible representar todos sus nodos, pero no todos sus atributos o etiquetas. En ocasiones es difícil mostrar la totalidad de sus enlaces, por lo que su número debe reducirse para realizar ajustes.
Redes grandes	Presentan más de 1,000 nodos, como Internet, las redes telefónicas, las redes de transportes o de carreteras. El principal reto que presenta la visualización de este tipo de redes consiste en identificar entre tantos nodos y relaciones los más relevantes, sin

Tamaño	Descripción
--------	-------------

perder las características del grafo principal.

Tabla 14. Clasificación de redes dependiendo del número de nodos (actores presentes).

Existen innumerables aplicaciones Web para realizar análisis de redes con base en información biomédica, están las que procesan literatura, genes, proteínas, células, enfermedades, enzimas, compuestos, entre otras. Algunas de las aplicaciones más novedosas en análisis de redes de literatura en línea sistematizada en la base de datos *PubMed* (Fig. 50) son:

- **Chilibot:**(<http://www.chilibot.net>) Encuentra relaciones entre proteínas, genes y palabras clave a partir de la búsqueda de dos términos en uno o dos resúmenes de *PubMed*. El resultado gráfico muestra los nodos con colores y dependiendo de los valores obtenidos, existe la posibilidad de re-graficar los resultados ya sea en subredes o eliminando los nodos y las relaciones elegidas.
- **PubGene:**(<http://www.pubgene.org>) Encuentra en los documentos las proteínas y los genes relacionados. Muestra la cantidad de artículos correspondientes a cada nodo y sus relaciones.
- **PubNet:**(<http://www.pubnet.org>) Realiza búsquedas en *PubMed* crea la red correspondiente. Los nodos representan autores, artículos, genes de *Gene Bank* o proteínas de *Swiss Prot*.



Figura 50. Cuatro aplicaciones para análisis de redes en el área biomédica.

14. Minería de textos

La **minería de textos** es un proceso que se utiliza para la generación de conocimiento nuevo interesante, plausible, e inteligible. Consiste en vincular dos o más conceptos de la literatura que hasta ahora no han sido relacionados (es decir, disjuntos) por medio del uso de programas y algoritmos diseñados para tal fin. Es, en muchos casos, utilizado como sinónimo de **descubrimiento basado en literatura** (*literature based discovery*) o **descubrimiento basado en bases de datos** (*Knowledge Discovery Database*). La minería requiere de dos procesos previos (pre-procesamiento): recuperar la información y automatizar la extracción de esa información. Dependiendo de los métodos utilizados en el pre-procesamiento será el tipo de representación de los contenidos de los textos construidos, de acuerdo con esta representación, será la clase de patrones por descubrir. Posteriormente el método consiste en relacionar dos o más conceptos al interior de los textos para descubrir patrones interesantes que ayuden a generar nuevos conocimientos.

Existen miles de aplicaciones para usar esta técnica en diversos conjuntos de información, de tal forma que se pueden categorizar según el tipo de datos que se analicen: la minería de datos que se refiere a números; la minería asociada a palabras que comprende el análisis del contenido completo, el resumen, el material, el método, las conclusiones y/o las referencias de la literatura basada sólo en términos; la minería de textos semántica se realiza utilizando en lugar de términos

el significado; y por último la minería de imágenes (Fig. 51). Además, se pueden combinar todas estas técnicas, aunque al hacerlo el tiempo de procesamiento aumenta, los algoritmos necesarios se hacen más complejos y los resultados llegan a ser más extensos. El National Center for Text Mining (NaCTeM) es un centro de investigación y desarrollo mantenido por la Universidad de Manchester en colaboración con la universidad de Tokio especializado en minería.



Figura 42. Distintos tipos de minería desarrollados en NaCTeM, institución de mayor prestigio en este tipo de análisis.

Figura 51. Distintos tipos de minería desarrollados en NaCTeM (<http://www.nactem.ac.uk>).

15. Semántica

Es un método que se basa en añadir metadatos semánticos y ontológicos a la información de la *World Wide Web*; en este proceso se describe el contenido, el significado y la relación de los datos procesados. Para lograr este propósito, los metadatos se deben estructurar de una manera específica, así que se usan los formatos RDF y OWL factibles de ser leídos automáticamente por máquinas inteligentes. El objetivo es mejorar el Internet ampliando la interoperabilidad entre los sistemas informáticos y usando programas de agentes inteligentes (algoritmos), cuya funcionalidad es buscar información sin operadores humanos. Este tipo de análisis puede incluir los tres métodos descritos anteriormente: bibliometría, análisis de redes y minería de textos, datos o imágenes.

La **semántica** se basa en las **ontologías** para formular un esquema conceptual que consta de un

mapa de los conceptos y sus relaciones y representa un proceso determinado en el que se refiere de manera explícita el significado de cada uno de los datos, propiedades, objetos y relaciones que lo componen; todo esto a través de la aplicación de una serie de reglas formales preestablecidas. Se basa en los enfoques y aplicaciones de los recursos semánticos y las herramientas para la investigación, el razonamiento, la predicción y los descubrimientos en el área a la que se refiere. Este procedimiento depende del desarrollo de la estructuración de datos en la Web a través de la creación de ontologías y lenguajes que comprendan significados; el área biomédica es una de las primeras que ha incorporado dicha estructura a sus documentos, en especial los que contengan genes y proteínas.

El primer gran avance en la estructuración de estas plataformas biomédicas para semántica consistió en la creación de una ontología semi-semántica, iniciada por El National Human Genome Research Institute (NHGRI), llamada *GeneOntology* (Fig. 52). El proyecto de ontología de genes es una importante iniciativa bioinformática con el objetivo de estandarizar la representación de los genes y los atributos del producto génico entre las distintas especies y bases de datos. El proyecto proporciona un vocabulario controlado de términos para describir las características del producto génico y los datos del producto génico, así como herramientas adecuadas para obtener y procesar estos datos.

GoPubMed (<http://www.gopubmed.org/web/gopubmed>) es un buscador semántico que realiza consultas con base en el significado de los términos (Fig. 53). Presenta al usuario un breve análisis bibliométrico por año, por autor o por país y dando a conocer dicho resultado a través de gráficas de barras, redes y un mapa mundial que representa la producción científica por países. *GeneOntology* (<http://www.geneontology.org>) y *PubMed* a través del uso de las ontologías procesan de manera lógica y semántica las consultas realizadas.

Figura 52. *Gene Ontology* y la relación entre su base de datos, su vocabulario de términos controlados y su interfaz de análisis.

La institucionalización de la semántica como disciplina se ha corroborado por los numerosos talleres impartidos por el consorcio de la Web semántica (W3C), la formación de grupos de trabajo de talla internacional y la edición de revistas altamente especializadas en el área biomédica, tal es el caso del *Journal of Biomedical Semantics* (<http://www.jbiomedsem.com>).

Webpage Screenshot



Figura 53. Interfaz de GOPubMed, presentando un histograma con los artículos por año procesados por minería de textos.

VI. DISCUSIONES ACTUALES

En este capítulo se presentarán algunos de los fenómenos más representativos de la web respecto a la bioinformación e informática biológica reciente, muchos de los cuales tienen repercusiones importantes para el manejo de literatura especializada en formato digital. Varios de estos términos marcan tendencias y son innovadores: acceso abierto, Creative Commons, duplicación y plagio, cómputo en nube, preservación digital y tipos de formatos de los documentos electrónicos.

1 Acceso abierto

Es el nombre de un movimiento internacional cuyo objetivo es que cualquier persona en el mundo, con una conexión a Internet, pueda acceder libremente, sin ninguna restricción de tipo económico, técnico o legal, a la información científica, académica y cultural disponible. El acceso abierto (AA, por sus siglas en español) u *open access* (OA, por sus siglas en inglés) se basa en la definición de la denominada Triple B, las declaraciones producto de tres conferencias realizadas en las ciudades de Budapest, Berlín y Bethesda.

En el 2002, la iniciativa de Budapest (BOAI, 2002) define como:

Acceso abierto[a la literatura científica revisada por pares], a su disponibilidad gratuita en la Internet pública, que permite a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o añadir un enlace al texto completo de esos artículos, rastrearlos para su indización, incorporarlos como datos en un software, o utilizarlos para cualquier otro propósito que sea legal, sin barreras financieras, legales o técnicas, aparte de las que son inseparables del acceso mismo a la Internet. La única limitación en cuanto a reproducción y distribución, y el único papel del copyright (los derechos patrimoniales) en este ámbito, debería ser la de dar a los autores el control sobre la integridad de sus trabajos y el derecho a ser adecuadamente reconocidos y citados (<http://www.opensocietyfoundations.org/openaccess/boai-10-translations/spanish>).

La Declaración de Berlín (2003) sobre el acceso abierto al conocimiento en ciencias y humanidades establece dos condiciones:

1. El (los) autor(es) y depositario(s) de la propiedad intelectual de tales contribuciones deben garantizar a todos los usuarios por igual, el derecho gratuito, irrevocable y mundial de acceder a un trabajo erudito, lo mismo que licencia para copiarlo, usarlo, distribuirlo, transmitirlo y exhibirlo públicamente, y para hacer y distribuir trabajos derivados, en cualquier medio digital para cualquier propósito responsable, todo sujeto al reconocimiento apropiado de autoría (los estándares de la comunidad continuarán proveyendo los mecanismos para hacer cumplir el reconocimiento apropiado y uso responsable de las obras publicadas, como ahora se hace), lo mismo que el derecho de efectuar copias impresas en pequeño número para su uso personal.
2. Una versión completa del trabajo y todos sus materiales complementarios, que incluya una copia del permiso del que se habla arriba, en un conveniente formato electrónico estándar, se deposita (y así es publicado) en por lo menos un repositorio online, que utilice estándares técnicos aceptables (tales como las definiciones del Acceso Abierto), que sea apoyado y mantenido por una institución académica, sociedad erudita, agencia gubernamental, o una bien establecida organización que busque implementar el acceso abierto, distribución irrestricta, interoperabilidad y capacidad archivística a largo plazo (<http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung>).

La Declaración de Bethesda (2003) define la investigación científica y sus objetivos de la siguiente forma:

La investigación científica es un proceso interdependiente donde cada experimento se nutre de los resultados de otros. Los científicos que investigan y las sociedades profesionales que los representan tienen un gran interés en garantizar que los resultados de investigación se difunden tan inmediatamente, ampliamente y efectivamente como sea posible. La publicación electrónica de resultados de investigación ofrece la oportunidad y la obligación de compartir los resultados de investigación, ideas y descubrimientos libremente con la comunidad científica y el público (http://ictlogy.net/articles/bethesda_es.html).

Acorde a estas declaraciones, se han instaurado dos mecanismos de publicación de acceso abierto para los artículos de investigación: aparecer en las revistas de acceso abierto y/o el autoarchivo del documento en texto completo, imágenes, videos, sonidos, etc. en repositorios, ya sean temáticos como *PubMed* o institucionales como el de MIT, <http://libraries.mit.edu/dspace-mit>, o el de la Facultad de Ciencias de la UNAM <http://repositorio.fciencias.unam.mx:8080/xmlui>, (Fig. 54).

Peter Suber es uno de los personajes más reconocidos dentro del movimiento de acceso abierto. Es investigador, escritor y consultor sobre el tema está adscrito a la Universidad de Harvard, en SPARC (Scholarly Publishing & Academic Resources Coalition) es investigador en filosofía en el Earlham College, director del Open Access Project de Public Knowledge y autor de la [SPARC Open Access Newsletter](#).

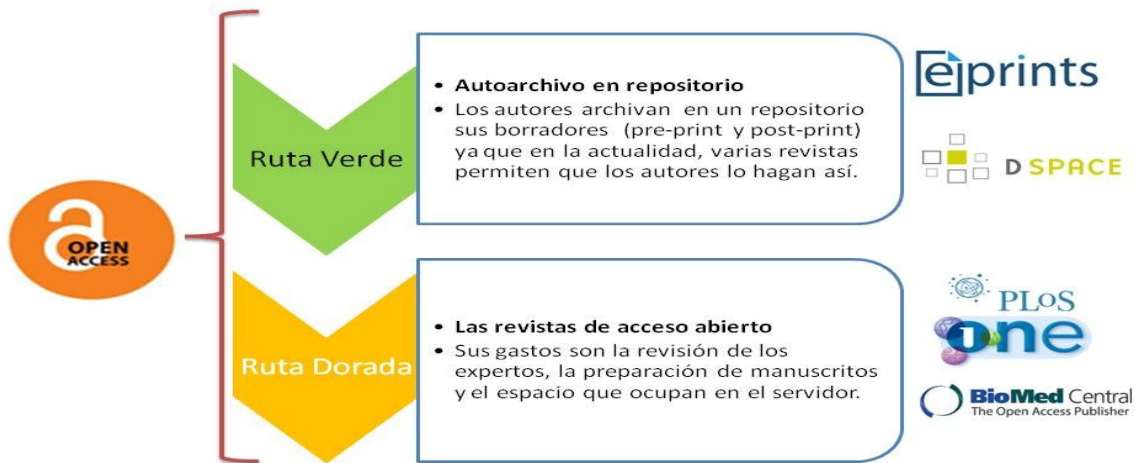


Figura 45. Esquema que representa las dos vías al acceso abierto a la información científica.

Figura 54. Esquema que representa las dos vías de acceso abierto a la información científica.

16. Creative Commons

Se refiere al espectro de permisos que define la protección absoluta de los derechos de autor (Fig. 55). Las distintas licencias ofrecen diferentes niveles para conservar los derechos de autor bajo el esquema de “Algunos derechos reservados”.



Figura 55. Esquema de las licencias Creative Commons©: (C) Todos los derechos reservados, (cc) Dominio público de Creative Commons, (dp) Ningún derecho reservado. Tomado y modificado de <http://creativecommons.org.mx/que>.

Creative Commons (CC) (<http://creativecommons.org.mx>) nació como iniciativa del profesor de derecho Lawrence Lessig de la Universidad de Stanford, especialista en fenómenos sociales y culturales del ciberespacio. Es una organización sin fines de lucro que persigue como principal objetivo ofrecer modelos de licencias que faciliten la distribución y el uso de contenidos disponibles a través del Internet. El portal de CC cuenta con herramientas para la búsqueda automática de obras con este tipo de licencia. El proceso de asignación de los diferentes tipos de licencia CC es flexible y fácil de usar, pues está pensado para que el autor decida de manera libre

bajo qué parámetros permite el uso de sus obras, con base en sus necesidades específicas.

17. Plagio

El **plagio** y la **duplicación de artículos** son de los principales problemas de la publicación científica en la era digital. El *Diccionario de la Real Academia Española* (<http://www.rae.es/rae.html>) define “plagiar” (del lat. *plagiāre*) como “copiar en lo sustancial obras ajenas, dándolas como propias”.

La mayoría de los científicos se enfrenta a una intensa competencia por el reconocimiento de sus pares, la visibilidad y la aceptación internacional; en la mayoría de los casos, la distinción académica se obtiene en función del número de publicaciones que tiene un investigador en revistas de vanguardia. Debido a la gran cantidad de tiempo y a los recursos que son necesarios para completar los estudios científicos, a menudo hay un deseo natural de buscar el máximo beneficio para el aprovechamiento de los gastos incurridos. Desafortunadamente, esto da lugar algunas veces a que los autores publiquen el mismo trabajo varias veces (duplicados) o, peor aún, que otra persona lo publique como propio (plagio).

Algunos documentos que suelen tener contenidos duplicados son las actas de congresos, actualizaciones importantes de estudios, la confirmación de los resultados impugnados en los estudios polémicos y las traducciones (que contienen duplicados en un idioma distinto). No se puede dudar de que en este contexto los duplicados tienen una función específica para la comunidad científica, pero en general esta práctica es vista como poco ética.

Dado el gran número de artículos publicados cada año sería razonable suponer que la duplicación de artículos científicos es una práctica común. Por esta razón en la actualidad existen herramientas informáticas para corroborar casos de duplicación y plagio, sobre todo en el área de la biomedicina que es una de las que más publica. Generalmente se utiliza la base de datos libres más grande de esta área que es *PubMed*. El ejemplo más interesante de estas herramientas es *Dejavu database* (<http://spore.vbi.vt.edu/dejavu>), propuesta por Mounir *et al*, en el año 2010 (Fig. 56); distingue a todos aquellos artículos que son similares y, a partir de ellos, permite identificar a los autores que publican artículos duplicados (igual contenido con distinto título) o plagio (cuando algún autor copia sin dar el crédito correspondiente del documento de otro). Además es posible consultar estadísticas sobre algunas de las variables del análisis de las publicaciones duplicadas respecto a revistas, curadores, idiomas, países e instituciones.



Figura 56. Portal web de *Deja vu* para detectar plagio en las publicaciones sobre el área biomédica.

18. Cómputo en la nube

El **cómputo en nube** (*cloud computing*) se refiere, tanto a las aplicaciones de la entrega de servicios a través de Internet como al hardware y software de los sistemas que proporcionan estos servicios. Ambos, el **hardware** del centro de datos y el **software**, son lo que llamamos una **nube**. Cuando una nube se pone a disposición de los usuarios, se denomina una nube pública, aunque generalmente esto implica un pago por estos servicios. Una nube privada se refiere a los centros de datos internos de una empresa u otra organización que no se pondrán a disposición del público en general. Algunos ejemplos de aplicaciones en la nube son: eyeOS (<http://www.eyeos.com/es>), un programa central que se ejecuta en el servidor de la empresa que gestiona los recursos de los usuarios y proporciona a los usuarios una interfaz web con la que operan sus archivos, correos electrónicos, calendario de eventos, aplicaciones, etc. desde la web; y Dropbox (<https://www.dropbox.com>), una aplicación que permite al usuario tener un archivo de almacenamiento de información disponible en línea y mantiene sincronizadas todas las cuentas asociadas a las versiones de escritorio y móvil descargadas. Trabajar en la nube implica tener respaldo de los datos en el ciberespacio, obtener actualizaciones automáticas y ayuda técnica en línea; estos servicios pueden conllevar un costo.

19. Preservación digital

Se basa en la búsqueda de soluciones para conservar aquellos documentos digitales almacenados, sea cual sea su formato, el software, hardware o sistema que se utilizó para su creación, manteniendo así la información resguardada pese a los rápidos cambios tecnológicos.

Generalmente se archivan los documentos y sus metadatos, que sirven para registrar la información que soporta y documenta el proceso de **preservación digital**. Existen distintos modelos de metadatos, cada uno de ellos con distintos esquemas de descripción. En los distintos modelos, cada objeto se describe por medio de una serie de atributos y el valor de estos atributos es el que puede servir para recuperar la información posteriormente (Tabla 15). De forma general podemos encontrar metadatos referidos a:

- El contenido (concepto).
- Aspectos formales (tipo, tamaño, fecha, idioma, etc.).
- Información del derecho de autor.
- Información de la autenticación del documento o recurso.
- Información sobre el contexto (calidad, condiciones o características de acceso, uso, etc.).

Los metadatos pueden estar almacenados dentro de una base de datos con una referencia al documento completo o ser incluidos en un encabezado dentro del propio texto. En el contexto de la Web, los metadatos se forman y almacenan para que puedan ser leídos por los motores de búsqueda. Las grandes ventajas del uso de metadatos radican en que se usa el mismo contenido del documento como un recurso de datos, y que los metadatos valen también para recursos que no tienen únicamente la morfología de texto, sino para cualquier tipo de formato tal como vídeo, audio o imágenes.

Las aplicaciones del uso de metadatos son muy amplias y van desde la recuperación de información, pasando por la descripción y catalogación de los documentos, su uso por parte de robots y agentes de software, comercio electrónico, firmas digitales, derechos de propiedad intelectual, valoración, evaluación y clasificación de contenidos, hasta trabajos bibliométricos e informétricos de todo tipo, entre otras. No obstante, aún hay muy pocos metadatos de preservación con sus políticas o estrategias asociadas (estabilidad, integridad y autenticidad).

Iniciativas

Liga electrónica

Metadatos para la descripción

Iniciativas	Liga electrónica
DC: Dublin Core Metadata Initiative.	http://dublincore.org
METS: Metadata Encoding and Transmission Standard.	http://www.loc.gov/standards/mets
MODS: Metadata Object Description Schema.	http://www.loc.gov/standards/mods
EAD: Encoded Archival Description.	http://www.loc.gov/ead
TEI: Text Encoding Initiative.	http://www.tei-c.org/index.xml
IFLA: Metadata Resources for Digital Libraries.	http://archive.ifla.org/II/metadata.htm
Metadatos para presentaciones	
MCF: Meta Content Framework.	http://www.textuality.com/mcf/NOTE-MCF-XML.html y http://www.w3.org/TR/NOTE-MCF-XML
Metadatos para la industria y el comercio electrónico	
UDEF: Universal Data Element Framework.	http://nd.undef.org
MPEG-21: Multimedia Framework.	http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
MPEG-7: Multimedia Content Description Interface.	http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
IMS: Instructional Management Systems.	http://www.imsproject.org
LOM: Learning Object Metadata.	http://ltsc.ieee.org/wg12
Metadatos para el gobierno y la administración	
EdNA Metadata Standard: Education Network Australia.	http://www.edna.edu.au/edna/go/pid/385
AGLS: Australian Government Locator Service.	http://www.agls.gov.au
Metadatos geospaciales	
Content Standard for Digital Geospatial	http://www.fgdc.gov/metadata/constan.html

Iniciativas	Liga electrónica
Metadata.	
Metadatos generals	
W3C Metadata activity	http://www.w3.org/Metadata/Activity.html
W3C Semantic Web activity	http://www.w3.org/2001/sw

Tabla 15. Iniciativas de metadatos más conocidas. Tomado y modificado de <http://www.hipertexto.info/documentos/metadatos.htm>

20. Tipo de formatos de documentos electrónicos

El proceso de preservación digital todavía no está completamente entendido ni valorado porque depende directamente del uso de la información sistematizada. En especial, para el caso de la biología ya no es necesario sólo tener la información archivada, sino su potencial en la creación de nuevo conocimiento con base en la interoperabilidad, integración y meta-análisis. Mucho de este potencial radica en las características de los meta-datos, que a su vez, dependerán de la estructura, las particularidades de los repositorios o páginas web donde están contenidos, de las políticas de preservación nacional, de la liberación y análisis de la información, y sobre todo del uso de estándares que sirvan como guía para establecer consensos respecto a la estructura y los formatos que deberán tener los objetos electrónicos y sus metadatos asociados.

Existen diferentes estándares para los documentos electrónicos aprobados por la ANSI (American National Standards Institute). El Consorcio de la Web Semántica (W3C) recomienda estrictamente seguir el estándar **XML** (eXtensible Markup Language), cuyo propósito es categorizar la información almacenada, constituyendo clases específicas para cada documento que sirvan para analizarlo posteriormente. XML fue diseñado especialmente para exportar y almacenar los datos; mientras que **HTML** (HyperText Markup Language) es, únicamente, para mostrar los datos en una página web (Fig. 57).

Un documento XML está formado por el prólogo y el cuerpo del documento con etiquetas y su texto correspondiente, de tal forma que se puede elegir específicamente información se va a exportar. Otro estándar ampliamente utilizado es **RDF** (Resource Description Framework), un

modelo estándar para el intercambio de datos en la Web; tiene características que facilitan la fusión de datos, incluso si los esquemas subyacentes difieren, y hace hincapié en la estructura de las relaciones entre los contenidos a través del uso de ligas entre **URI** (Uniform Resource Identifier). El uso de este modelo es simple, permite que los datos estructurados y semi-estructurados se puedan combinar, exponer y compartir a través de diferentes aplicaciones web.

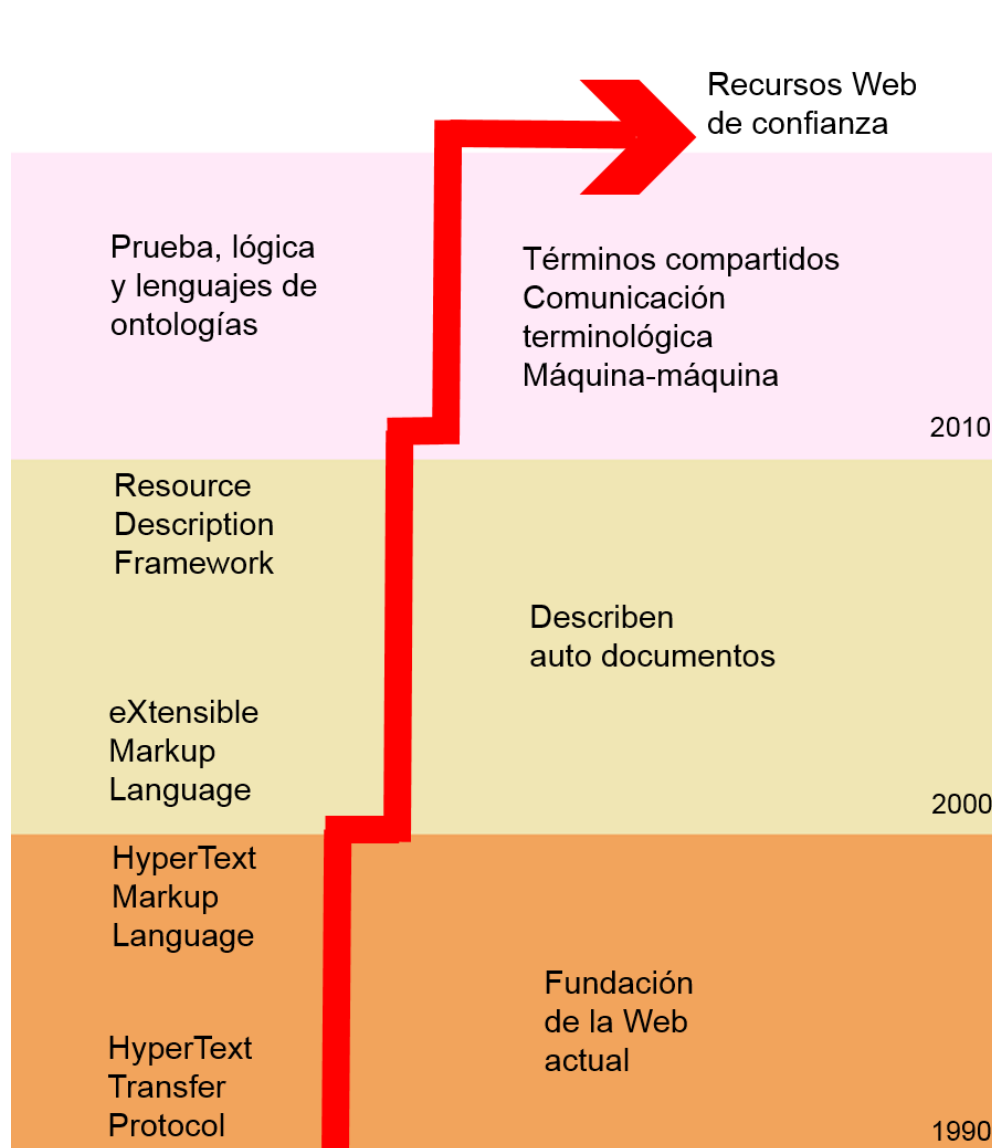


Figura 57. Evolución de los estándares web. Modificado de *Berners-Lee, T.* 2001.

Otro estándar usado para web semántica es el **OWL** (Web Ontology Language) o lenguaje de ontologías para la web, el cual se ha convertido en la recomendación del W3C desde 2004. Está

diseñado para usarse cuando la información contenida en los documentos necesita ser procesada por programas o aplicaciones semánticas. OWL puede usarse para representar explícitamente el significado de los términos sistematizados en vocabularios y las relaciones entre ellos; esta representación de los términos y sus relaciones se denomina una **ontología** (ver capítulo 5, Fig 52). En realidad, OWL es una extensión del lenguaje RDF, aunque es un lenguaje con más poder expresivo que éste.

VII. LITERATURA RECOMENDADA

- Berners-Lee, T. (2000). Tejiendo la red. El inventor del World Wide Web nos descubre su origen. *Siglo XXI*. Madrid. URL <http://news.bbc.co.uk/1/hi/technology/4132752.stm>
- Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics*, 39 (1), 17-23. URL <http://dx.doi.org/10.1038/ng1934>
- Bjaalie, J. G. (2008). Understanding the brain through neuroinformatics. *Frontiers in neuroscience*, 2 (1), 19-21. URL <http://dx.doi.org/10.3389/neuro.01.022.2008>
- Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. *Science*, 331 (6018), 649. URL <http://dx.doi.org/10.1126/science.1203354>
- Hull D, P. S., Kell DB (2008). Defrosting the Digital Library: Bibliographic Tools for the Nex Generation Web. *PLoS Comput Biol* 4(10).
- Larson, R. R. (2010). Introduction to information retrieval. *Journal of the American Society for Information Science and Technology* , 61 (4), 852-853. URL <http://dx.doi.org/10.1002/asi.21234>
- Michán-Aguirre, L., Calderón-Rojas, R., Nitxin-Castañeda-Sortibrán, A., & Rodríguez-Arnáiz, R. (2010). Aplicaciones web para recuperación y análisis de bibliografía de PubMed. *El Profesional de la Información*, 19 (3), 285-291. URL <http://dx.doi.org/10.3145/epi.2010.may.09>
- Michán, L. (2008). El acceso público a la información científica de corriente principal una utopía hecha realidad. *Rev. Cien. Med. Est. Antí.*, 2(1), 31-32+.
- Michán, Layla, Macías, L., Alvarez-López, E., Muñoz-Velazco, I., Medina-Hernández, A. E., Montoya, L., et al. (2010). *Propuesta de creación y mantenimiento de un repositorio de literatura institucional en la Facultad de Ciencias, UNAM*. Retrieved from http://sistemas.fciencias.unam.mx/~layla/2010/Repositorio_FC_mejorado.pdf.
- Michán, Layla, Muñoz-Velazco, I., Alvarez-López E., Macías, L. (2011). *Biomedical Web, Collections and Meta-Analysis Literature Applications*. InTech, 2011: In Biomedical Engineering - From Theory to Applications (August 2011), pp. 1-22.

- Mindell, D. P., Fisher, B. L., Roopnarine, P., Eisen, J., Mace, G. M., Page, R. D. M., & Pyle, R. L. (2011). Aggregating, tagging and integrating biodiversity research. *PLoS ONE*, 6 (8), e19491. URL <http://dx.doi.org/10.1371/journal.pone.0019491>
- NSF. (2007). *Cyberinfrastructure vision for 21st century discovery*. Arlington, VA: National Science Foundation. 57pp.
- Page, R. D. M. (2010). Enhanced display of scientific articles using extended metadata. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8 (2-3), 190-195. URL <http://dx.doi.org/10.1016/j.websem.2010.03.004>
- Parr, C. S., Guralnick, R., Cellinese, N., & Page, R. D. M. (2012). Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology & Evolution*, 27 (2), 94-103. URL <http://dx.doi.org/10.1016/j.tree.2011.11.001>
- Pirkola, A. (2009). The effectiveness of Web search engines to index new sites from different countries. *Information Research*, 14(2), paper 396. URL <http://informationr.net/ir/14-2/paper396.html>
- Russell, Jane M. (2001). *Scientific communication at the beginning of the 21st century*. *International Social Science Journal*, 168: 271-282.
- Samwald, M., Jentzsch, A., Bouton, C., Kallesoe, C., Willighagen, E., Hajagos, J., Marshall, M., Prud'hommeaux, E., Hassanzadeh, O., Pichler, E., & Stephens, S. (2011). Linked open drug data for pharmaceutical research and development. *Journal of Cheminformatics*, 3 (1), 19. URL <http://dx.doi.org/10.1186/1758-2946-3-19>
- Shapira, B., & Zabar, B. (2011). Personalized search: Integrating collaboration and social networks. *J. Am. Soc. Inf. Sci.*, 62 (1), 146-160. URL <http://dx.doi.org/10.1002/asi.21446>
- Sobral, B. W., Mangalam, H., Siepel, A., Mendes, P., Pecherer, R., & McLaren, G. (2001). Bioinformatics for rice resources. *Novartis Foundation symposium*, 236. URL <http://view.ncbi.nlm.nih.gov/pubmed/11387987>
- Specia, L. y Motta, E. (2007). *Integrating Folksonomies with the Semantic Web*. In *The Semantic Web: Research and Applications*, 639, 624.

VII. GLOSARIO

Add-on: extensión o tema que realiza y personaliza las aplicaciones básicas del explorador.

Alt-metría (Alt-metrics): análisis cuantitativo de la ciencia a través de enlaces sociales, marcadores sociales, redes sociales, etc.

API (Application Programming Interface): es un conjunto particular de códigos con especificaciones que los programas pueden seguir para comunicarse entre sí. Sirve como interfaz entre diferentes programas y facilita su interacción.

Aplicación: bajo esta denominación se llama a todo programa o conjunto de programas que nos proporciona una solución completa o parcial en algún campo concreto. No es sinónimo de programa; toda aplicación es un programa, pero no todo programa es una aplicación. Por ejemplo, un programa de desfragmentación de disco no se dice que sea una aplicación, sino se dice que es una utilidad.

App: una aplicación de software para dispositivos móviles.

Araña (spider): es un programa que inspecciona las páginas del *World Wide Web* de forma metódica y automatizada.

Arquitectura grid: disciplina que surge del conjunto de métodos y herramientas orientadas a estructurar, clasificar, buscar, manejar, organizar, navegar, etiquetar, recuperar, construir, capturar, orientar, conectar y ayudar en cuestiones relacionadas con el uso de la información; en este caso, en entornos.

Análisis de redes (Network analysis): se realiza a través del estudio de las teorías del comportamiento estructural, la dinámica y su influencia en los temas académicos y para establecer una probable explicación para el crecimiento y la evolución de las redes reales en cualquier tema. Su aplicación implica un salto cuantitativo y cualitativo en la representación y el análisis de la estructura de todo tipo de dominios científicos, sean geográficos, temáticos, institucionales e incluso individuales.

Bases de datos: sistema informático de registros con el propósito de almacenar, mantener y

generar información, sistematizada en tablas con un orden establecido. Programas que permiten guardar, ordenar, procesar y presentar los datos.

Bibliometría (*Bibliometrics*): evaluación cuantitativa de la literatura, que describe el impacto de la investigación y la institucionalización de la misma.

Blog (contracción del término *weblog*): bitácora, en un tipo de sitio web, donde las entradas se muestran en orden cronológico inverso. Puede utilizarse también como un verbo “mantener o añadir contenido a un blog”. La capacidad de los lectores de dejar comentarios en un formato interactivo es una parte importante de muchos blogs. La mayoría de blogs es principalmente textual, aunque algunos se centran en arte (artlog), fotografías (fotoblog), dibujos (sketchblog), vídeos (vlog), música (MP3 blog) o audio (podcasts), y todos forman parte de una red más amplia de medios de comunicación social. *Micro-blogging* es otro tipo de blog que consiste en blogs con pequeñas entradas (por ejemplo el caso de *Twitter*) (Orihuela 2006).

blogger: en inglés, autor de un weblog (Orihuela 2006).

Blogger: con mayúscula, designa un popular sistema de edición de weblogs, propiedad de *Google* (Orihuela 2006).

Buscadores o motores de búsqueda (*search engines*): permiten a los usuarios realizar búsquedas, en segundos, en sus catálogos, que contienen millones de registros con el texto de un gran número de páginas web, las cuales fueron previamente compiladas por máquinas o robots que mantienen actualizada la información continuamente (Alierta 2008).

Caché: memoria que almacena copias de los documentos que pasan por él, de forma que peticiones subsiguientes pueden ser respondidas por el propio caché, si se cumplen ciertas condiciones.

Chat (*messenger*): aplicación que permite la comunicación, fundamentalmente escrita, en tiempo real entre dos o más usuarios por vía electrónica.

Ciberinfraestructura: sistemas computacionales, datos, recursos de información, redes, sensores digitales, instrumentos, organizaciones virtuales y observatorios, en conjunto con

una serie interoperable de servicios de software y herramientas. Esta tecnología se complementa con los equipos interdisciplinarios de profesionales que son responsables de su desarrollo, divulgación y uso en propuestas transformadoras para el descubrimiento científico y la enseñanza.

Cibernetría (*Webometrics*): se dedica al análisis cuantitativo de Internet y los contenidos de la red, especialmente de aquellos relacionados con el proceso de generación y comunicación académica del conocimiento científico.

Cienciometría (*Scientometrics*): estudia los aspectos cuantitativos de la ciencia, especialmente artículos y patentes; encuentra aplicación en la evaluación y obtención de indicadores, también en el establecimiento de las políticas científicas.

Colaboratorios: punto de encuentro abierto a académicos, investigadores, estudiantes y público en general interesado en la conformación de espacios de aprendizaje en red, flexibles y participativos. Permite a varias personas trabajar juntas en un mismo proyecto, aunque se hallen lejos unos de otros.

Complemento (*plug-in*): es una aplicación que se relaciona con otra para aportarle una función nueva y generalmente muy específica. Esta aplicación adicional es ejecutada por la aplicación principal e interactúan por medio de la **API** (Application Programming Interface) propia del programa. Es un modo de expandir programas de forma modular, de manera que se puedan añadir sin afectar a las funciones ya existentes ni complicar el desarrollo del programa principal.

Cómputo en nube (*cloud computing*): se refiere tanto a las aplicaciones como a los servicios a través de Internet, cuya información se almacena de manera permanente en servidores electrónicos y se envía a memorias temporales de los clientes; algunos servicios pueden tener un costo.

Correo electrónico (*electronic mail* o *e-mail*): servicio de mensajería basado en Internet mediante el cual un usuario (ordenador) puede intercambiar mensajes con otros usuarios (ordenadores) a través de la red. Los mensajes son fundamentalmente de texto, pero se pueden adjuntar archivos de datos, imágenes, vídeo o música. El correo electrónico es uno

de los servicios más populares en Internet(Alierta 2008).

Dirección electrónica: ver URL.

Directorio web: tipo de sitio web que contiene un directorio organizado de enlaces a otros sitios web, con una estructura de categorías y subcategorías. Habitualmente los directorios web permiten a los creadores de sitios que informen el suyo para que sea incluido y después los editores autorizados revisan esas solicitudes antes de incluir sus enlaces para comprobar que se adecuan a los requisitos de aceptación determinados por el directorio. Los directorios web regionales integran en un mismo sitio a comercios o participantes de determinado sector, creando de esta manera una comunidad cerrada que facilita la navegación, localización y mercadeo. Estos directorios promueven el crecimiento económico del sector al que están enfocados ya que ponen al alcance del usuario la posibilidad de descubrir proveedores que desconocía y con esto resolver su necesidad de compra.

E-ciencia: es el proceso de investigación realizado en el espacio virtual con las siguientes características: colaboración internacional entre los investigadores (colaboratorios), la interconexión de computadoras, la aplicación de la arquitectura *grid* y el uso de repositorios de datos. Se refiere a la posibilidad de visualizar los datos comunes, desarrollar herramientas y procedimientos basados en Internet y construir estructuras de organización virtual en donde se lleve a cabo la investigación, distribución y publicación electrónica de los resultados en línea(Russell 2001).

Enlace de hipertexto (*hipertext, link, hyperlink*): el enlace entre dos documentos de texto o de una marca de texto a una posición denotada(Anon 2002).

Etiqueta (*tags*): los caracteres de identificación que son añadidos a un conjunto de datos. Son etiquetas pegadas por los usuarios a fragmentos de información virtual como marcadores (*bookmarks*) de páginas web, fotos o artículos de periódicos. Algunos ejemplos de aplicaciones o servicios de Internet que utilizan esta idea son los portales Flickr para la gestión de imágenes o Delicious para la gestión de marcadores(Alierta 2008).

Etiquetar: asignar categorías por medio de términos; las etiquetas son un tipo de metadato,

pues proporcionan información que describe al dato y facilita su recuperación.

Folksonomías (*folksonomy*): sistemas de clasificación grupal. Taxonomía social u organización de la información de manera colectiva, basada en la colaboración de las personas que ordenan contenidos mediante etiquetas. Las categorías usadas no obedecen a una lógica jerárquica, sino a las decisiones de etiquetado de los usuarios(Cobo and Pardo 2007).

Gadgets: dispositivos que tienen un propósito y una función específica, generalmente de pequeñas proporciones, prácticos y a la vez novedosos(Alierta 2008).

Hardware: componentes electrónicos, tarjetas, periféricos y equipo que componen un sistema de computación.

Herramienta: conjunto de utilidades que acompañan a un programa y que sirven para aumentar su capacidad. Un ejemplo puede ser la revisión ortográfica en un procesador de texto la cuál, sin ser necesaria para el funcionamiento del procesador, le añade sus prestaciones.

Hipertexto (ligas, vínculos, hipervínculos): sistema de presentación de la información en el cual el texto, las imágenes, los sonidos y las acciones están enlazados mediante una red compleja y no secuencial de asociaciones que posibilitan al usuario examinar los distintos temas que se están tratando, con independencia del orden de presentación de los mismos. El término “hipertexto” lo creó Ted Nelson en el año 1965 con el objeto de hacer una descripción de los documentos que se presentan en un ordenador expresando la estructura no lineal de las ideas.

HTML (Hyper Text Markup Language): acrónimo inglés de lenguaje de marcación de hipertexto; es un lenguaje de marcas diseñado para estructurar textos y presentarlos en forma de hipertexto, que es el formato estándar de las páginas web. Gracias a Internet y a los navegadores del tipo Explorer o Mozilla Firefox, el HTML se ha convertido en uno de los formatos más populares que existen para la construcción de documentos(Alierta 2008).

Indexación: 1) asignación de etiquetas (signos o términos) a los documentos de acuerdo

con su tema, concepto, clase, etc. 2) asignación de un término de entrada para localizar un documento por autor o por título (indicación de autor o de título) o por materia (indicación temática). Este concepto es utilizado por LISA (Library and Information Science Abstracts) en su afán por unificar la terminología bibliotecológica, documentaria y de la ciencia de la información.

Infometría (*Infometrics*): análisis métrico de la información.

Internet: sistema de redes de computadoras enlazadas, con alcance mundial y de continuo crecimiento, que facilita servicios de transmisión de datos.

Marcadores (*bookmarks*): herramienta de algunas aplicaciones que almacena direcciones de páginas web que el usuario encontró útiles y a las que quiere acceder fácilmente; conocidos también como “Favoritos”.

Mensajería instantánea: conjunto de programas que utilizan el protocolo TCP/IP, sirven para enviar y recibir mensajes instantáneos con otros usuarios conectados a Internet u otras redes; además, con ellas es posible saber cuando están disponibles para hablar. Los más utilizados son ICQ, Yahoo! Messenger, Windows Live Messenger, AIM (AOL Instant Messenger) y Google Talk (Alierta 2008).

Meta-análisis (*meta-analysis*): técnica cuantitativa que utiliza medidas específicas para indicar la fuerza de las relaciones de variables de los estudios incluidos en el análisis. Hace hincapié en los resultados de múltiples estudios en comparación con los resultados a partir de una sola investigación.

Metabuscaador: sitio web especializado en consultar varios motores de búsqueda simultáneamente y que permite interrogar varias bases de datos de páginas web desde una única interfaz; entre los más comunes están Copérnico (<http://www.copernic.com>) Dogpile (<http://www.dogpile.com>) Ixquick (<http://www.ixquick.com/esp>) Mamma (<http://www.mamma.com>) Metacrawler (<http://www.metacrawler.com/index.html/search/home>).

Metadatos (*metadata*): son datos que describen otros datos.

Microblogging: este concepto, relacionado con el de blog, consiste en una aplicación web

que permite insertar mensajes reducidos, a modo de posts de un blog, con una longitud máxima de 140 caracteres y con una elevada frecuencia de actualización. Twitter (<http://twitter.com>) y Jaiku (<http://www.jaiku.com>) son ejemplos de servicios de *microblogging*(Alierta 2008).

Minería de textos (*text-mining*): es la más reciente área de investigación del procesamiento de textos. Se define como el proceso de descubrimiento de patrones interesantes y nuevos conocimientos en una colección de textos, es decir, es el proceso encargado del descubrimiento de conocimientos que no existían explícitamente en ningún texto de la colección, pero que surgen de relacionar el contenido de varios de ellos.

Navegador (*browser*): aplicación software utilizada para localizar y mostrar páginas web. Los tres más conocidos son Microsoft Internet Explorer, Firefox y Google Chrome, que son navegadores gráficos, lo cual quiere decir que pueden mostrar gráficos y texto. La mayoría de navegadores actuales puede presentar información multimedia, incluyendo sonido y vídeo(Alierta 2008).

Obsolescencia (*obsolescence*): se refiere a los documentos que están cayendo en desuso.

Ontología (*ontology*): base de datos que describe los conceptos generales o sobre un dominio, algunas de sus propiedades y cómo se relacionan unos con otros. Hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados con la finalidad de facilitar la comunicación y el intercambio de información entre diferentes sistemas y entidades.

Páginas web: documentos situados en una red informática a la que se accede mediante enlaces de hipertexto.(Anon 2002)

PDF (Portable Document Format): acrónimo del inglés que significa “formato de documento portátil”, es un formato de almacenamiento de documentos desarrollado por la empresa Adobe Systems, es de tipo compuesto (imagen vectorial, mapa de bits y texto).

Podcast: tipo de archivo de audio y video distribuido a través de Internet (sitio web, blog, wiki, etc.). Una vez capturado, puede reproducirse en una computadora u otro tipo de dispositivo multimedia.

Portal: sitio web cuyo objetivo es ofrecer al usuario, de forma ordenada e integrada, el

acceso a gran variedad de recursos y de servicios, entre los que suelen encontrarse buscadores, foros, compra electrónica, etc(Alierta 2008).

Portales especializados: ofrecen recursos en la red sobre temas específicos. Alrededor de ellos se forman comunidades virtuales de intercambio de información.

Preservación digital (*digital preservation*):se basa en la búsqueda de soluciones para conservar documentos digitales almacenados sea cual sea su formato, el software, hardware o sistema que se utilizó para su creación, manteniendo así la información pese a los rápidos cambios tecnológicos.

Plagio (*plagiarism*): “copiar en lo sustancial obras ajenas, dándolas como propias” (DRAE).

Plug-in: ver **complemento**.

Ranking: ver **sistema de votación**.

Recuperación de la información: métodos, técnicas y procedimientos que permiten la búsqueda y el acceso a la información almacenada en estructuras voluminosas o complejas, con múltiples criterios de búsqueda y puntos de acceso (bases y bancos de datos).

Redes sociales (*social networking*): describe todas aquellas herramientas diseñadas para la creación de espacios que promuevan o faciliten la conformación de comunidades e instancias de intercambio social(Cobo and Pardo 2007).

Semántica (*semantic*): expresa el significado de los datos, las propiedades de los objetos y el complejo de relaciones entre ellos mediante una serie de reglas formales.

Sistema de votación (*ranking*): clasificación. En el entorno de Internet se dice del orden de presentación de los resultados en una búsqueda con un programa WAIS (Wide Area Information Server). Este orden se establece, según el grado de probabilidad, de mayor a menor. Es la relación entre un conjunto de elementos tales que, para uno o varios criterios, el primero de ellos presenta un valor superior al segundo, que a su vez es mayor que el tercero y así sucesivamente, permitiéndose que dos o más elementos diferentes puedan tener la misma posición. El orden se refleja asignando a cada elemento un ordinal, generalmente números enteros positivos.

Sitio web (*website*): conjunto de páginas web que comparten un mismo tema e intención y que generalmente se encuentran en un solo servidor(Anon 2002). Colección de páginas web a las que se accede a través de una dirección URL única.

Sociedad de la información: estado de desarrollo social caracterizado por la capacidad de sus miembros (ciudadanos, empresas y administración pública) para obtener y compartir cualquier información instantáneamente, desde cualquier lugar y en la forma que se prefiera(Alierta 2008).

Software: programa o programas de computadora, contraparte del equipo físico en que se ejecutan.

Software libre: programa informático que surge gracias a la colaboración de diversas personas y que permite a los usuarios copiar, modificar o distribuir su contenido sin tener que pagar permisos de propiedad intelectual, bajo ciertas normas de colaboración y uso(Cobo and Pardo 2007).

Tesauro: vocabulario controlado y dinámico de términos genérica y sistemáticamente relacionados que cubren un área específica del conocimiento.

URL: acrónimo del inglés “Uniform Resource Locator”, dirección de una página web(Orihuela 2006). Dirección global de documentos y de otros recursos en la Web. La primera parte de la dirección indica el protocolo a usar y la segunda parte especifica la dirección IP o el nombre de dominio donde el recurso está localizado.

VoIP (*Voice over IP*): sistema de comunicación de voz e imagen de gran calidad por vía del protocolo de Internet. Se realiza desde la computadora (Skype es el servicio VoIP más popular, aunque existen muchos otros) y puede ser gratuito o costar muchísimo menos que el servicio de telefonía fija tradicional.

Web 2.0: término acuñado para referirse a la segunda generación de web basada en comunidades de usuarios y una gama especial de servicios como redes sociales, blogs o wikis. Se puede definir como la transición que se ha dado de aplicaciones tradicionales hacia aplicaciones que funcionan a través de la web enfocada al usuario final(Alierta 2008).

Web semántica (*semantic web*): es la web de los datos; se basa en la idea de añadir metadatos semánticos y ontológicos a la *World Wide Web*. Esta información adicional —

que describe el contenido, el significado y la relación de los datos— se debe proporcionar de manera formal, para que así sea posible evaluarla automáticamente por máquinas de procesamiento. El objetivo es mejorar Internet al ampliar la interoperabilidad entre los sistemas informáticos y reducir la necesaria mediación de operadores humanos(Orihuela 2006).

Widget: es una pequeña aplicación o programa, usualmente presentado en archivos o ficheros pequeños que son ejecutados por un motor de widgets o *widget engine*. Entre sus objetivos están dar fácil acceso a funciones frecuentemente usadas y proveer de información visual.

Wiki: es una página o colección de páginas web diseñadas para permitir a cualquiera el acceso y contribuir o modificar su contenido, utilizando un lenguaje sencillo. Se usan con frecuencia para crear sitios web de colaboración, uno de los mejores ejemplos es la enciclopedia *Wikipedia*(<http://www.wikipedia.org>). Las wikis se utilizan en negocios para proveer intranet y sistemas de manejo de conocimiento. Ward Cunningham, el desarrollador del primer software wiki, WikiWikiWeb, originalmente lo describió como “la más sencilla base de datos en línea que posiblemente funcionaría”. Es un término tomado de la lengua hawaiana que significa “rápido”(Cobo and Pardo 2007).

WWW (*World Wide Web*, literalmente «malla que cubre el mundo»): servidor de información distribuido, basado en hipertexto, creado a principios de la década de 1990 por Tim Berners Lee, investigador en el CERN, Suiza. La información puede ser de cualquier formato (texto, gráfico, audio, imagen fija o en movimiento) y es fácilmente accesible a los usuarios mediante los programas navegadores(Alierta 2008).

XML (eXtensible Markup Lenguaje): estándar para transportar y almacenar datos.

**INTRODUCCIÓN A LA RECUPERACIÓN DE LITERATURA
DIGITAL EN CIENCIAS BIOLÓGICAS**

AUTORES

Layla Michán
Eduardo Alvarez
Jack Guillén

[<http://www.ciib.unam.mx/libros/irldcb/>]

Universidad Nacional Autónoma de México

Facultad de Ciencias

Primera edición

México, 2013



INTRODUCCIÓN A LA RECUPERACIÓN DE LITERATURA DIGITAL EN CIENCIAS BIOLÓGICAS,
por Michán, L., Alvarez, E. y Guillén, J. se encuentra bajo una LicenciaCreative Commons Atribución-No
Comercial-Licenciamiento Recíproco 3.0 Unported. Créditos