



INTRODUCCIÓN A LA RECUPERACIÓN DE LITERATURA DIGITAL EN CIENCIAS BIOLÓGICAS

AUTORES:

Layla Michán

Jack Guillén

Eduardo Alvarez

Lyssania Macías

Itzel Pedraza

[<http://www.fciencias.unam.mx/xxxx/>]

Universidad Nacional Autónoma de México
Facultad de Ciencias
Primera edición
México, 2012



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

CÓMO CITAR ESTE LIBRO


Michán, L., Guillén, J., Alvarez, E., Macías, L., y Pedraza, I. (2012). *Introducción a la Recuperación de Literatura Digital en Ciencias Biológicas*. Prensa Ciencias. México, D.F. 124 p.
Disponible en [www.....](#)

PREFACIO

Haciendo mi doctorado cerca del año 2000, en Historia de la Biología, mi director de tesis Jorge Llorente, me dijo: “tienes que hacer una base de datos bibliográfica”, en la que sistematice la literatura que se ha publicado sobre taxonomía en México en el siglo XX (este era el tema de mi investigación), nunca imagine que esa indicación se convertiría 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, sus características y propiedades que ya eran parte de mis objetivos primarios de atención.

Era precisamente la época en la que se transformaban las publicaciones periódicas impresas al formato electrónico, 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, ahora ya me es normal), me asombraba a diario descubriendo 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 invadieron las redes sociales, los rss, el acceso abierto y la aparición del doi, me volví una seguidora implacable de todas las novedades que aparecían, las probaba todas, y así combinando mi interés en la historia reciente, mi pasión por las bases de datos bibliográficas, el gusto por el uso de la bibliometría y mi afinidad por la informática, que se constituyó mi línea de investigación centrada en la literatura y que trata sobre “ información, informática y cienciometría para ciencias biológicas” que desarrollo en la Facultad de Ciencias de la UNAM, en el que investigamos, enseñamos y difundimos 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.

Este es nuestro primer libro, en el que resumimos lo que consideramos básico para un estudiante o académico que trabaja con temas biológicos. 

Esperamos que les sirva, pero sobre todo 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

CONTENIDO

CÓMO CITAR ESTE LIBRO	1
PREFACIO.....	2
CONTENIDO.....	4
PRÓLOGO	6
PRESENTACIÓN	8
INTRODUCCIÓN	12
I. LAS APLICACIONES WEB 2.0	16
1. Internet y Web.....	16
2. Clasificación de los recursos en la Web 2.0 para literatura.....	19
3. Navegadores o Exploradores / Navegar	22
4. Buscadores y Metabuscar / Buscar	24
5. RSS o agregadores de <i>feeds</i> / Actualizar	30
6. Marcadores / Guardar	32
7. Redes Sociales / Compartir	35
8. Manejadores de Bibliografía / Citar	37
9. Varios	42
II. LA RECUPERACIÓN DE INFORMACIÓN EN LA WEB	48
1. La recuperación de información.....	48
2. Las etapas siguientes	53
III. LAS COLECCIONES DE DATOS BIBLIOGRÁFICAS	56
1. Las bases de datos	56
2. Las colecciones bibliográficas.....	59
3. Características de las colecciones bibliográficas.....	62
Agradecimiento	92
IV. META-ANÁLISIS DE LITERATURA PARA BIOLOGÍA.....	93
1. Bibliometría.....	94
2. Análisis de redes.....	100
3. Minería de textos	103
4. Semántica	104

V. DISCUSIONES ACTUALES	107
1. Acceso abierto	107
2. Creative Commons	109
3. Plagio.....	110
4. Computo en la nube.....	111
5. Preservación digital	111
6. Tipo de formatos de documentos electrónicos	114
VI. GLOSARIO	117
VII. LITERATURA RECOMENDADA	127

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 la currícula y formación de todo estudiante de licenciatura de biología y posgraduante en el área, y en general para todo aspirante a investigador y profesionista en cualquier área del conocimiento.

Específicamente 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 masa 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, consistente 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 es cierto, en contrapartida, que esta desintermediación obliga al consumidor de información 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 de información 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 Ciencimetría 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 de información aportados por este grupo de vanguardia.

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

PRESENTACIÓN

La literatura especializada se aplica en varios ámbitos de la práctica biológica, que va 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, pero es tal la cantidad y diversidad de los documentos científicos, emergentes día a día que complementario a la complejidad y diferencia de los diferentes recursos electrónicos (en especial las bases de datos bibliográficas) por medio de los cuales se puede acceder a dicha información, y no solo eso, sino es importante reiterar que estos recursos cambian y se actualizan constantemente, esto hace prácticamente imposible estar al tanto de todos los recursos y es difícil identificar cuáles y cuántos se pueden y deben utilizar.

Previo a 2009, en la UNAM no existía en la licenciatura o el posgrado de la UNAM, algún curso teórico o práctico en el que se presentará 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 en el año 2009 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 optativa en la Facultad de Ciencias 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 integrativo para utilizar la literatura especializada en biomedicina, explicar las tácticas y principios para extraer, analizar y administrar literatura especializada de manera eficiente, inmediata, actualizada, exhaustiva y organizada. La estructura de los cursos y sus contenidos son: 1) presentar, ordenar y clasificar 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, 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 explica el flujo y el ciclo de la literatura biológica; 6) se exponen los principales indicadores bibliométricos que se utilizan frecuentemente 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, que 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. Además los cursos permiten a los estudiantes preparar un artículo académico e introducirse a las herramientas suficientes para realizar reportes con información sobre revistas y documentos indizados, impacto, colaboración y citación de la producción propia, comúnmente solicitados por los comités de evaluación del desempeño académico como CONACYT y 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 intención es presentar en seis secciones: 1) una introducción al tema, 2) describir, clasificar y comentar las aplicaciones, utilidades, herramientas y servicios electrónicos disponibles en la Web 2.0 y Web 3.0 más útiles para la búsqueda, manejo, análisis y publicación, 3) explicar brevemente el proceso de recuperación de información electrónica, 4) sintetizar las características y los tipos de colecciones bibliográficas más relevantes y útiles para obtención de literatura, 5) realizar una breve introducción sobre los métodos más utilizados para realizar meta-análisis de literatura y dar algunos ejemplos de las aplicaciones en línea para este propósito, y 6) terminamos con una presentación sobre lo más innovador dentro de la Web para el manejo de literatura. La estructura general del libro se presenta en la Fig. 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 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 toda información actualizada relacionada con los temas de este libro: **Ciberliteratura** (<http://ciberliteraturaunam.blogspot.com/>) (Fig. 2), en el que publicamos lo más novedoso sobre literatura académica y **Biiiogeek** (<http://biiiogeek.blogspot.com/>) (Fig. 3), en la que publicamos las noticias más recientes sobre información e informática biológica para el manejo de literatura.

Introducción a la Recuperación de Literatura Digital en Ciencias Biológicas
Layla Michán, Jack Guillén, Eduardo Alvarez, Lyssania Macías e Itzel Pedraza



Figura 2. Blog de ciberliteratura UNAM.

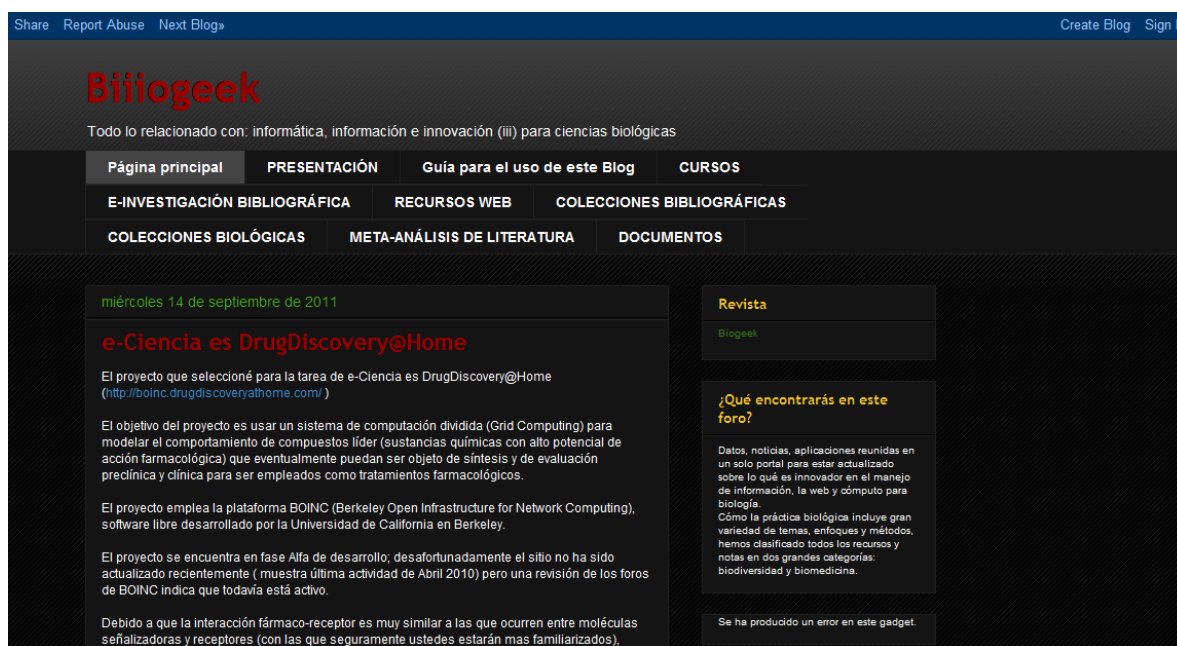


Figura 3. Blog de Biiogeek.

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 innovadores 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, Web semántica, índice de impacto, cocitación, Web 1.0, 2.0 y 3.0, redes sociales, plagio, repositorios, acceso libre (ver glosario), por mencionar los más comunes. Esto ha repercutido de manera dramática en la visión del mundo contemporánea, 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, ha 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. La sociedad científica en especial, genera y recibe información, 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 para las subdisciplinas 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 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. Pero no solo eso; 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 (*sybiomatics*), 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, al grado que se han propuesto nuevas formas de acceder a la información dejando de lado el enfoque reduccionista y se ha adoptado 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, los recursos de la Web 2.0 y 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 (ver glosario), acorde al progreso de la propia disciplina biológica.

La ambigüedad que existe en el actual entorno de la información científica es el 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, la investigación documental (o bibliográfica), en la que se localiza los documentos y se revisa uno por uno sobre un 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) a través de Internet, lo que ha favorecido el acceso, la automatización, la inmediatez y democratización de la información. Este tipo de investigación bibliográfica se le 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 han desarrollado una enorme cantidad de aplicaciones, herramientas y servicios que ofrecen la posibilidad de integrar los procesos de la recuperación de bibliografía. Uno de los principales problemas a los que se enfrentan los usuarios; es al desconocimiento de los recursos digitales disponibles en la Web, de tal manera que la mayoría de las herramientas y aplicaciones para el manejo de literatura científica más sofisticadas, generalmente se utilizan una mínima parte del potencial existente en la Web. Muchas de estas aplicaciones, herramientas, utilidades y “*mashups*” (ver glosario), 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 el 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, colaboración, 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 colaboratorios.

En la Fig. 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 que se utilizará como referencia para estructurar este libro.

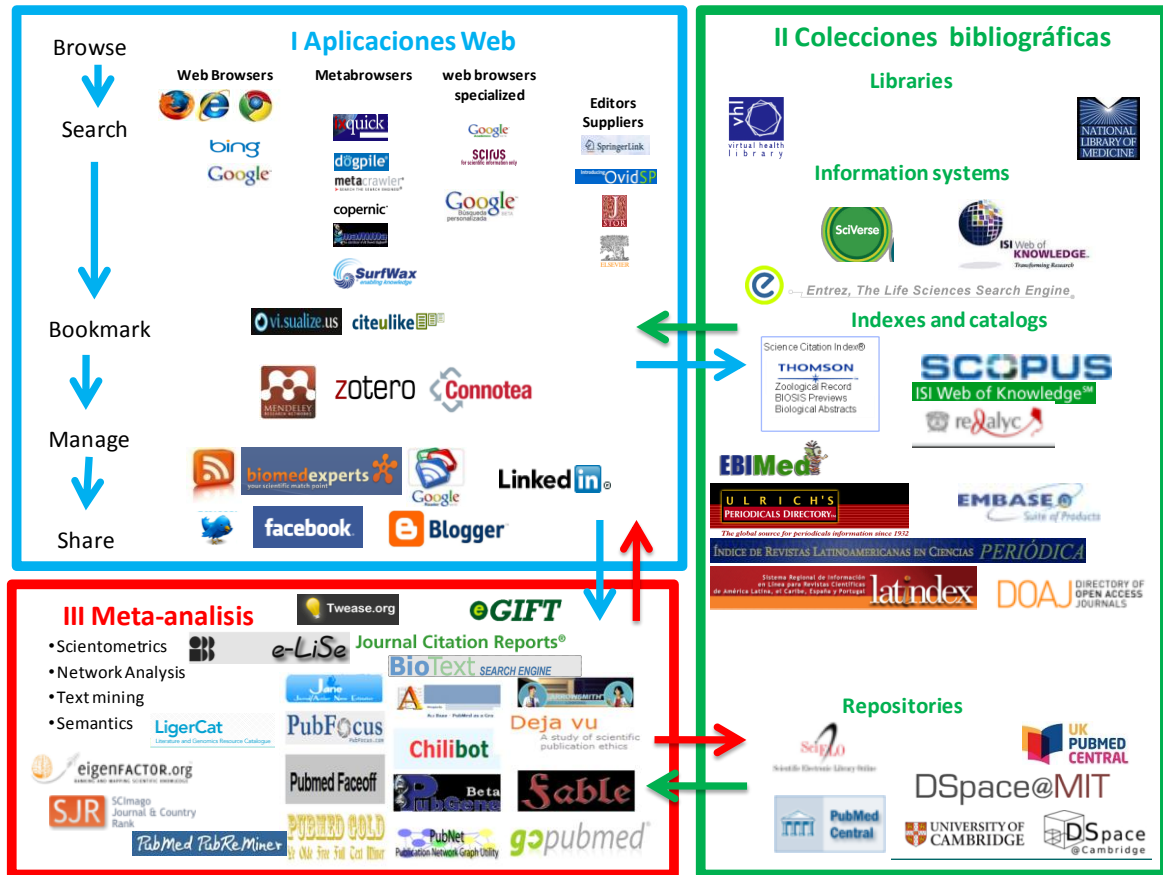


Figura 4. Principales aplicaciones presentadas en este libro, corresponden a los capítulos 1, 3 y 4.

I. LAS APLICACIONES WEB 2.0

El presente capítulo consiste en clasificar, describir y presentar los recursos Web 2.0 disponibles más relevantes para la recuperación y el manejo de literatura especializada en Biología más generales, accesibles y básicos. Cada uno de los recursos electrónicos se expondrá utilizando el orden lógico en el que se realiza la recuperación de literatura en la Web (Fig. 5).



Figura 5. Recursos o aplicaciones Web 2.0 disponibles 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. La 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 usados para transmitir los datos, conocidos como protocolos, 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, 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 a través de Internet y se basa en un modelo de información compartida por el uso del protocolo HTTP (*HiperText Transfer Protocol*). Se utiliza 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 (*hyperlinks*), los documentos pueden contener texto, gráficas, sonidos y/o video. La Web es solo una de las formas de distribuir información a través de Internet, la Web 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, 2.0 y 3.0.

El modelo de la Web 1.0 es con el que 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 lo incluido. 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.

La Web 2.0 es el término dado para describir a la segunda generación de la *World Wide Web* iniciada en el año 2000, 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, que es más organizada, se basa en el servicio de aplicaciones Web, consta de la transformación

del usuario en productor de información, fomenta la comunicación abierta con un énfasis en la conformación de comunidades de usuarios y en la posibilidad del intercambio de información. Aplicaciones distintivas de estas son los blogs, las wikis y las redes sociales (ver glosario) 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 la 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 (ver glosario y capítulo 4).

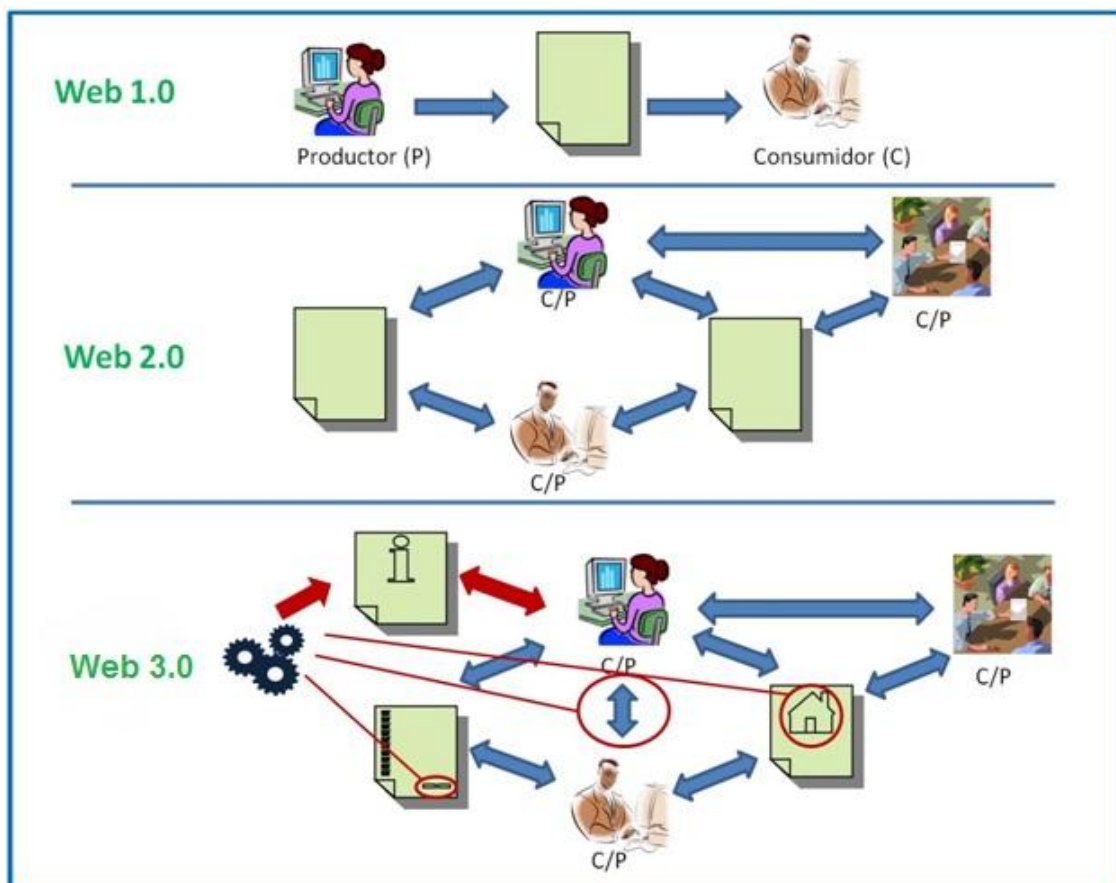




Figura 6. Evolución de la Web. (P) productor, (C) consumidor, (C/P) consumidor/productor, en círculos rojos la inclusión de ontologías para la Web semántica.

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). Se han incluido el desarrollo de nuevos sistemas de interoperabilidad que permiten interpretar meta-datos 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 la información en literatura electrónica publicada en las páginas electrónicas, *blogs* y *wikis* (ver glosario), a las cuales se accede a través de los exploradores. En la tabla 1 se presenta la clasificación y los ejemplos más representativos para los recursos disponibles de la Web 2.0, su función se describirá en los párrafos posteriores.

Recurso (clasificación)	Ejemplo
<p>Exploradores o Navegadores.</p> <p>EXPLORAR</p>	<p>Mozilla Firefox, Opera, Internet Explorer, Safari, Google Chrome, Midori, Epiphany.</p> 
<p>Buscadores (<i>search engines</i>) y metabuscadores.</p> <p>BUSCAR</p>	<p>Alexa, Ask, Bing, Exalead, Google, Wikia.</p> 







Recurso (clasificación)	Ejemplo
	Copernic, Dogpile, Ipselon, Ixquick, Kartoo, Mamma, Metacrawler, T10 (Turbo10), WebCrawler, Yippy. 
RSS o lectores de <i>feeds</i> . ACTUALIZAR	Google reader, Bloglines, Netvibes, Yahoo! . 
Marcadores. GUARDAR	Diigo, Del.icio.us, Google Bookmarks, Vi.sualze.us, StumbleUpon, Evernote, YouTube, Picasa. 
Redes Sociales. COMPARTIR	Badoo, Facebook, Hi5, MySpace, Twitter, LinkedIn, Google +. 
Manejadores de literatura GESTIONAR CITAR	BibCiter, BibMe, Citation Machine, Citeulike, Connotea, EndNoteX, Mendeley, NoodleBib, Express, Ottobib, ProCite, RefWorks, Zotero. 
Varios.	PDF-XChange Viewer, Primo PDF, 4Shared, Dropbox, Diccionario de la Real Academia de la Lengua Española, Diccionario y Tesauro Merriam's Webster Online, Logmein. 

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

Para acceder y manejar la literatura digital se pueden utilizar herramientas y aplicaciones (ver glosario) 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, por mencionar las más importantes (Fig. 7).

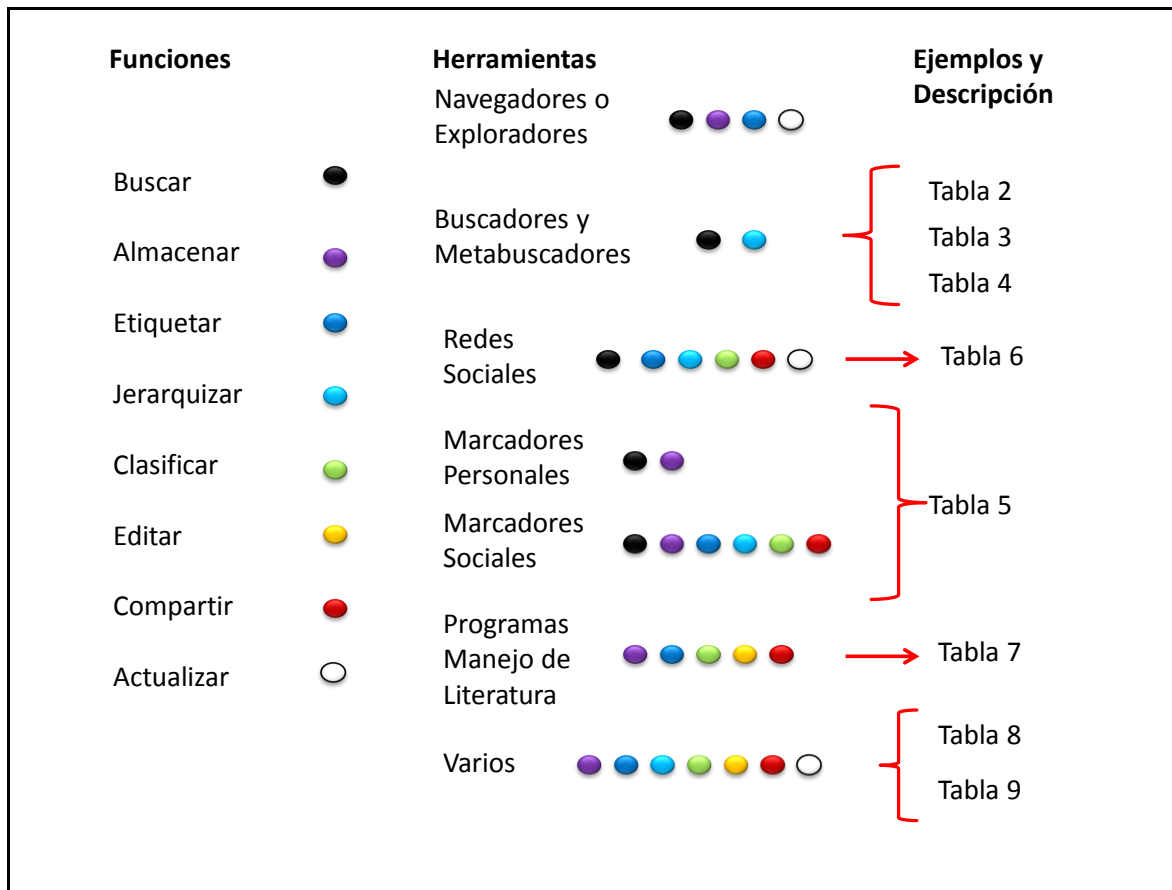


Figura 7. Herramientas de la Web 2.0 y sus funciones características, las tablas corresponden a las numeradas en este capítulo.

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

El nombre de navegador, navegador red, navegador web u hojeador (traducción literal aunque su uso es mínimo) proviene 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) interpretando 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 vía 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 (ver glosario), o buscando (ver buscadores) (Figs. 8 y 11). Los exploradores o navegadores más comunes se mencionan en la tabla 1.



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, permite aclarar dudas, añadir comentarios, hacer preguntas, publicar artículos e información, entre muchas otras cosas. Consta de entradas etiquetadas de acuerdo con su contenido, generalmente ordenadas cronológicamente (la información más reciente casi siempre se ve primero), existen millones de *Blogs* sobre cualquier tema imaginable que son actualizados periódicamente (un *blog* bien actualizado publica algo al menos una vez a la semana) en el que es posible incluir enlaces y pueden publicar varios usuarios colaboradores, siempre y cuando tenga los permisos correspondientes (Fig. 9).

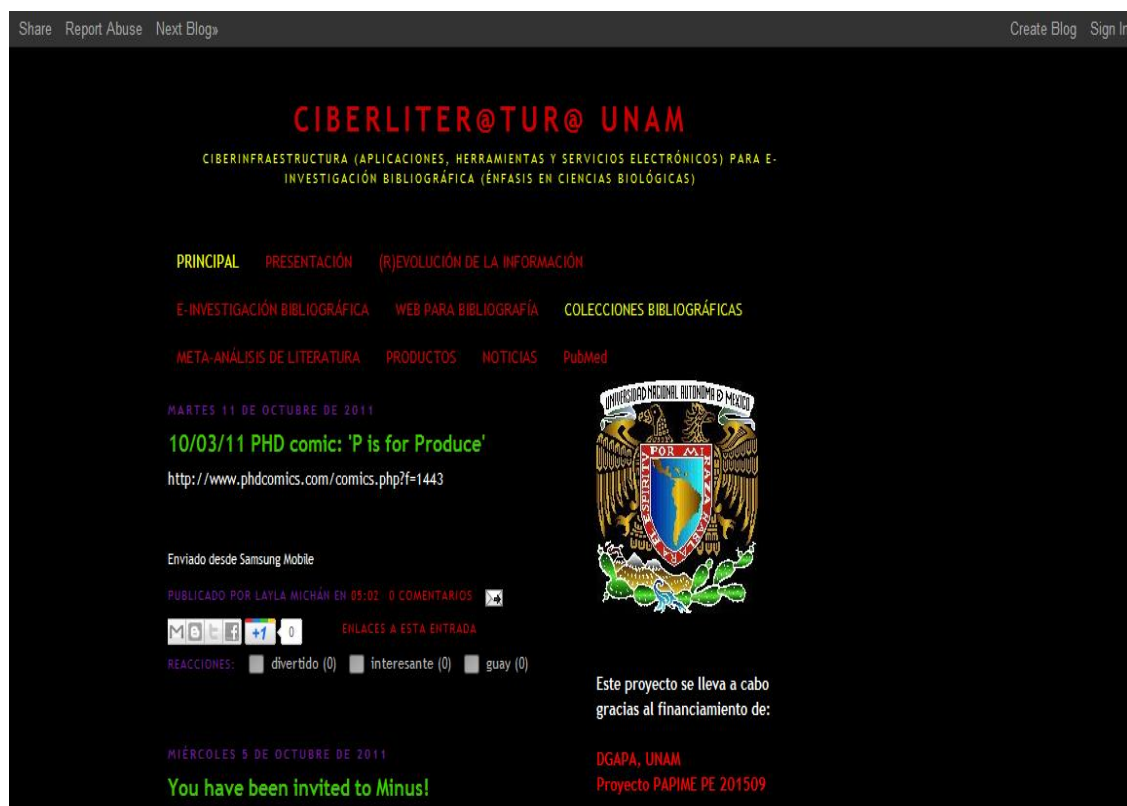


Figura 9. Estructura de un *blog*.

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 que comparten. Los textos o páginas *wiki* tienen títulos únicos, conservan un historial de los cambios, a partir del cual, es posible recuperar fácilmente cualquier versión anterior e identificar quien 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).

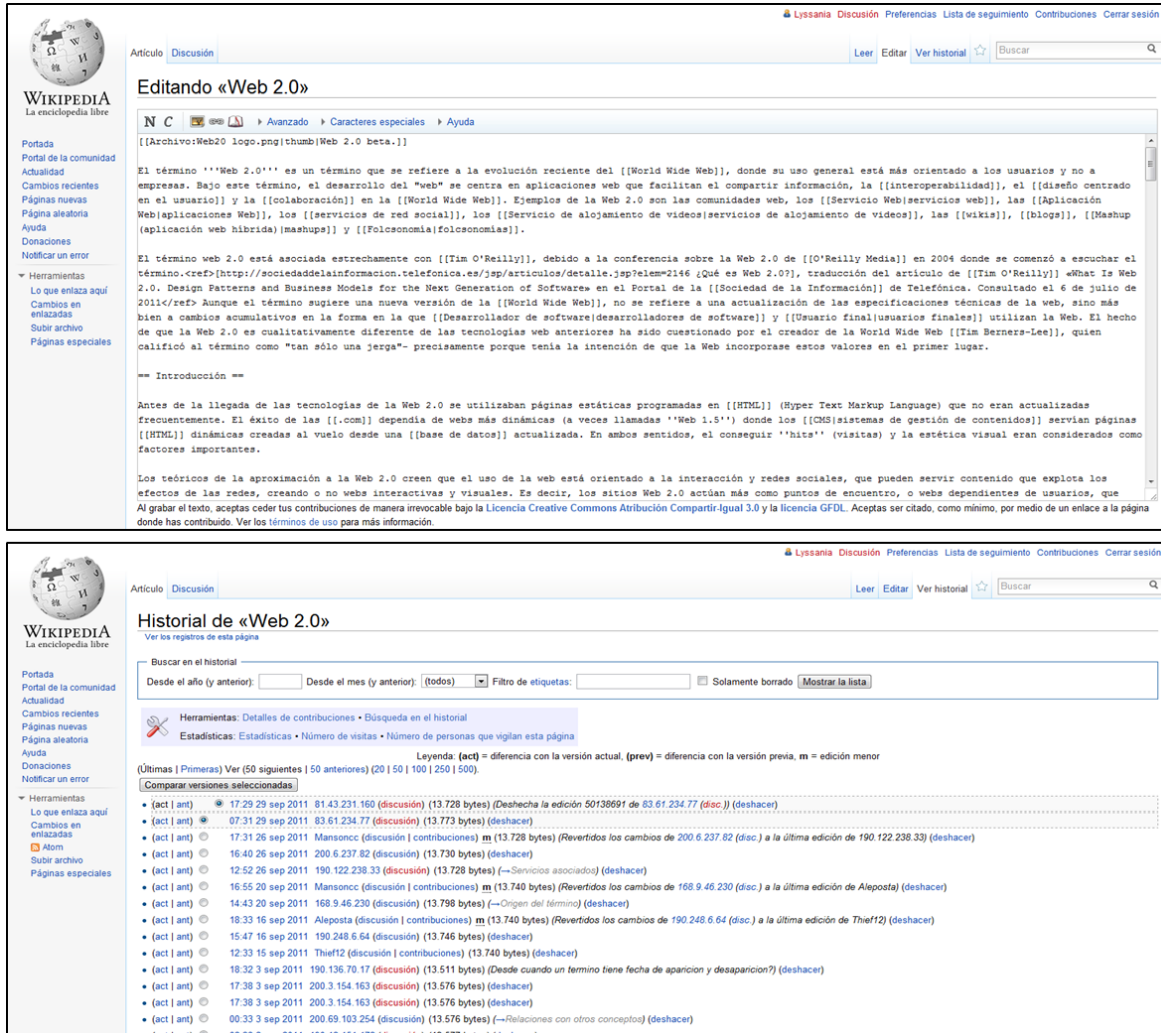


Figura 10. Edición e historial en Wikipedia.

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 que existen (una búsqueda con la letra “a” en Google actualmente resulta en más de 17 millones de páginas), para ello se utilizan programas llamados buscadores y metabuscadores.

4. Buscadores y Metabuscadores / Buscar

Un buscador (Tabla 2) es un sistema de bases de datos que incorpora automáticamente páginas web mediante robots de búsqueda en la Red, que indizan las páginas web disponibles en los servidores gracias a su *spider* (ver glosario). Las búsquedas se hacen a partir de palabras clave o por temas y pueden ser generales o especializadas (que establecen límites respecto a los campos

de búsqueda). El resultado de la búsqueda es un listado de direcciones Web en los que se mencionan las palabras clave buscadas o los temas.

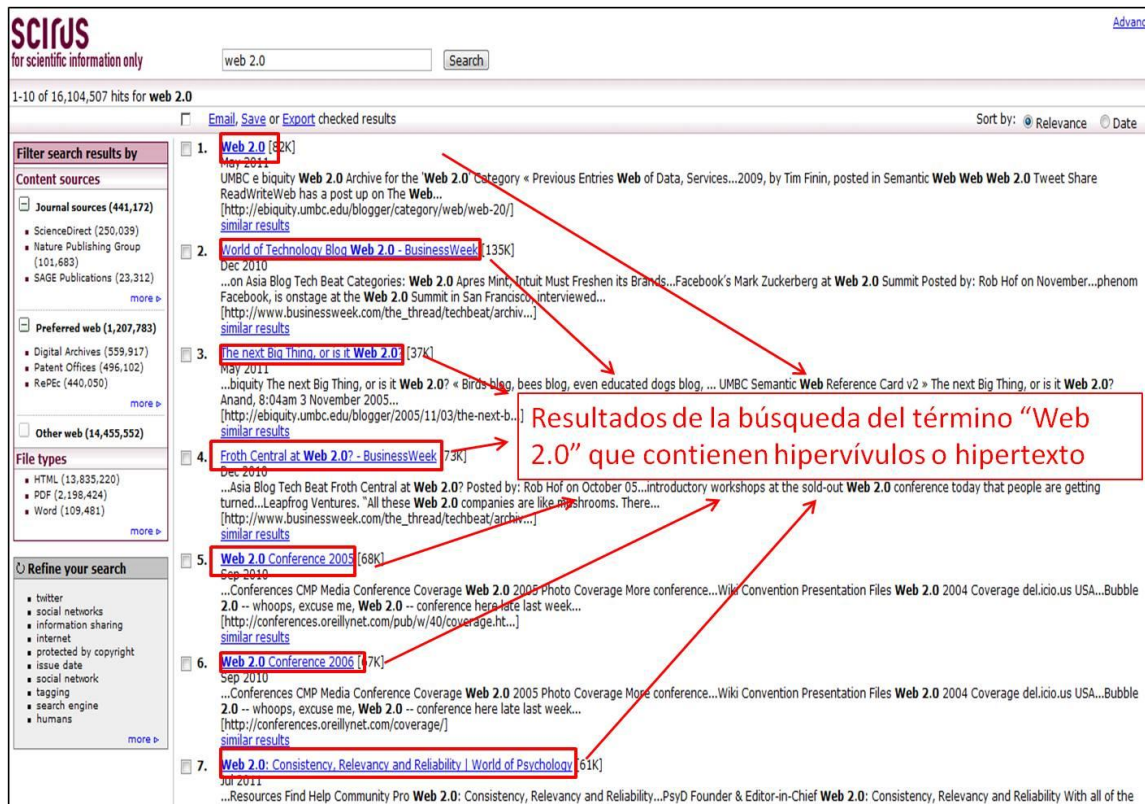


Figura 11. Resultados de una búsqueda en un buscador especializado.

Otro recurso para buscar es el metabuscador (Tabla 3) 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 presentados en las tablas 2 y 3 respectivamente, se seleccionaron según el criterio del sitio Web Alexa, que muestra información jerarquizada (*ranking*) de los sitios más populares (más consultados).

Buscador	URL	Editor	Características generales
Alexa	http://www.alexa.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 mediante 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 2009. Incluye una lista de sugerencias de búsqueda en tiempo real y una lista de las búsquedas relacionadas, 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 el correo electrónico. El 29 de julio de 2009, Microsoft y <i>Yahoo!</i> anunciaron que habían hecho un trato en el que Bing reemplazaría el motor de búsqueda de <i>Yahoo!</i>
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

Buscador	URL	Editor	Características generales
			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 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 (ver glosario).
Wikia	http://answers.wikia.com/wiki/Wikianswers	Wikia	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.

Tabla 2. Buscadores generales más utilizados y su descripción.

Meta buscador	URL	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 encuentra el filtrado, la clasificación de los resultados, sus criterios de pertinencia propios, y/o posibilidad de salvar y recuperar las búsquedas para actualizar los resultados.

Meta buscador	URL	Editor	Características Generales
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=008968604428227682493:8qmtgt-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> (ver glosario) 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, ixmoz, 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 por cada buscador. En la página principal tiene la opción para elegir cuáles buscadores usar.
Kartoo	http://www.kartoo.com/	Kartoo	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, salvarlos 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 construyendo un listado ponderado.

Meta buscador	URL	Editor	Características Generales
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).

Tabla 3. Metabuscadores más comunes y sus características distintivas.

Los buscadores, a su vez, se pueden clasificar en generales, son aquellos que tienen como intención registrar las páginas web existentes y los especializados que 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, en la tabla 4 se observan las características generales de estos buscadores, todos de acceso libre.

Buscador especializado	URL	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, por ejemplo; estudios revisados por especialistas, tesis, libros, resúmenes y artículos de fuentes como: editoriales académicas, sociedades profesionales, depósitos de impresiones

Buscador especializado	URL	Editor	Características generales
			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 temas de búsqueda 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 centra sólo en las páginas web que contienen contenido científico.
Scientific WebPlus	http://scientific.thomsonwebplus.com/BasicSearch.aspx?lbu=WebOfKnowledge&key=49jf934i20s0g4i290s0&bhcp=1	Thomson Reuters	Thomson Scientific WebPlus 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 y su descripción.

5. RSS o agregadores de feeds / Actualizar

Dentro de todos los recursos web actuales, uno de los más innovadores y útiles son 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 actualizado 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. Ésta es una tecnología representativa de la Web 2.0, ya que el usuario puede enlazar o etiquetar una página Web (Fig. 12), pero también el contenido de ésta, recibiendo notificaciones en un solo lugar cada vez que se produce una actualización, sin necesidad de consultar distintas páginas (*blogs*, revistas, etc.).

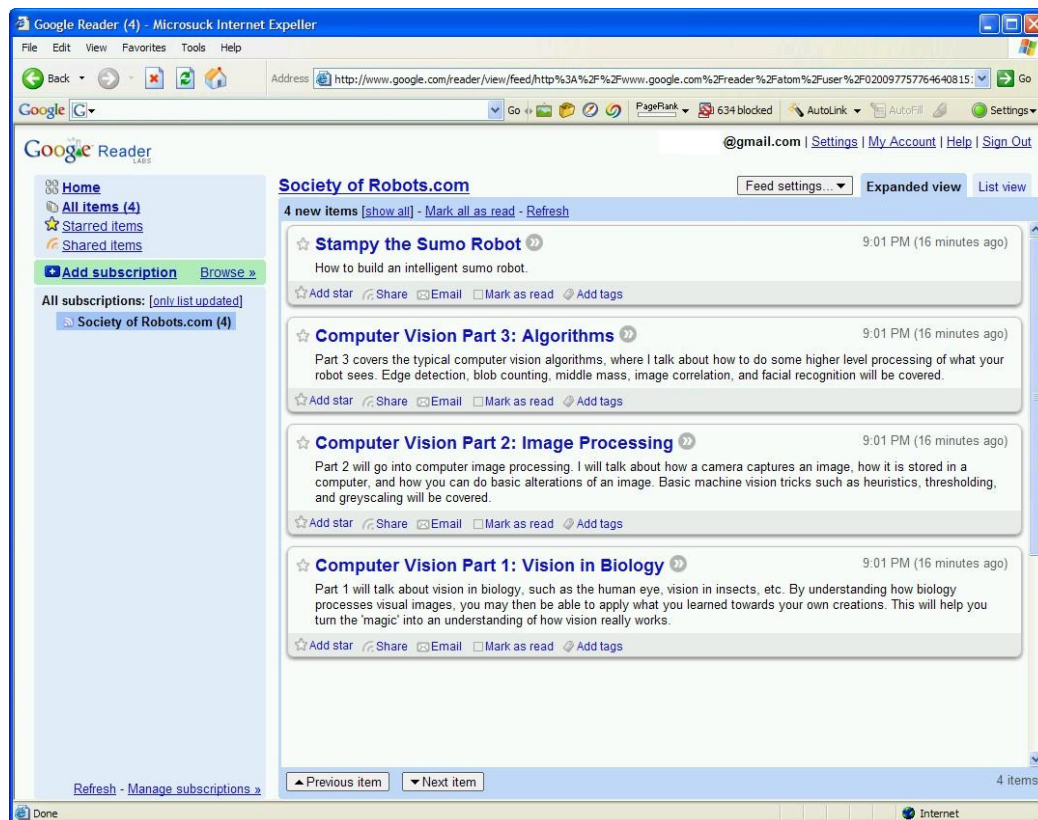


Figura 12. Ejemplo de RSS feed de Google.

6. Marcadores / Guardar

Los llamados marcadores o favoritos son herramientas que permiten guardar la dirección electrónica de una página web seleccionada (URL) (ver glosario). 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, atribuyendo mayor relevancia a los contenidos más populares (*ranking*) a partir de los cuales se crean 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: en sociales y 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 (ver tablas 2 y 5) 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 de uso solo persona, la tendencia ha sido permitir la sociabilidad. Actualmente, además 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* o *tagging*) 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. Este 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 inter-creatividad. Muchos sistemas utilizan esta organización cooperativa, al sugerir al usuario las etiquetas que ya tiene registradas un contenido 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: 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. 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 la 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 (un ejemplo característico son las opciones Beta), bajo esta idea, 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 de 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 en sus funciones 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 este modo las fuentes de información quedarán almacenadas en una cuenta. Su funcionamiento tiene un carácter colectivo pues se pueden compartir enlaces 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 pueden guardar toda las fuentes pertinentes, independiente del lugar donde se encuentre cualquiera de los integrantes del grupo.

Marcadores	URL	Características generales
Favoritos (P)	Herramienta de Internet Explorer	Almacena direcciones electrónicas (URL) en la computadora personal, no se pueden compartir.
Marcadores (P)	Herramienta de Firefox	Almacena direcciones electrónicas (URL) en la computadora personal, no se pueden compartir.
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.
Del.ici.us (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 puede usar directamente en la página web de Delicious; además de compartirlas con otros usuarios.
Diigo (S)	http://www.diigo.com/	Diigo ofrece un complemento (<i>add-on</i>) en el navegador que permite marcar la página, resaltar las partes de las páginas web 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.
Evernote (S)	http://www.evernote.com/about/intl/es/	Es una aplicación informática destinada a la organización de información personal a través de guardar notas. Existen versiones para diversos sistemas operativos que se instalan 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.
Picasa (S)	http://picasa.google.com/	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 picasaweb permitiendo colocar las fotos directamente en los álbumes.

Marcadores	URL	Características generales
StumbleUpon (S)	http://www.stumbleupon.com/	Marcador social con clasificación por temas, y formato de vídeo, imágenes y noticias, se puede compartir a través de redes sociales y agregar comentarios.
Vi.sualze.us (S)	http://vi.sualize.us/	Sitio web de marcadores sociales de contenidos visuales, permite almacenar, compartir, etiquetar y guardar las imágenes favoritas junto con dirección del documento electrónico que las contiene.
Youtube (S)	http://www.youtube.com/	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.

(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. En la tabla 6 se mencionan las redes sociales con mayor difusión y uso en la actualidad, entre sus funciones están, por ejemplo; la de agrupar a especialistas en ciencia, en muchos de ellos existen grupos de discusión sobre biología, servicios de literatura y bibliotecas.

Existen un gran número de redes sociales científicas como BiomedExperts (<http://www.biomedexperts.com/>), AuthorMapper (<http://authormapper.com/>) WorldWideScience.org (<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	URL	Características Generales
Badoo	http://www.badoo.com	Badoo (usa el 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 (glocal). 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	http://www.facebook.com	Facebook es un sitio web gratuito de redes sociales creado por Mark 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.
Hi5	http://www.hi5.com	Es una red social lanzada en 2003, es uno de los 500 sitios web más visitados del mundo según el sitio web Alexa. El sitio hi5 es famoso por su interactividad, pues hace de una simple cuenta de usuarios una especie de tarjeta de presentación virtual.
MySpace	http://www.myspace.com	Es un sitio web de interacción social formado por perfiles personales de usuarios que incluye redes de amigos, grupos, blogs, fotos, vídeos y música, una red interna de mensajería que permite comunicarse y un buscador interno.
Twitter	http://www.twitter.com	Twitter (usa el lema ¿qué estás haciendo ahora?) es un servicio gratuito de microblogging (ver glosario) innovador y en boga, que hace las veces de red social y que 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 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.
LinkedIn	http://www.linkedin.com/	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

Red Social	URL	Características Generales
		orientada a un perfil de público profesional que cuenta actualmente con más de 65 millones de usuarios y empresas. Que 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 entre otras, porque no permite almacenar o compartir fotos, videos, modificar diseños del perfil y otras cosas, debido a que está orientada al currículum y es este el que se puede subir y cambiar la foto.
Google +	https://plus.google.com/	Google +, de reciente creación 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.

Tabla 6. Redes Sociales.

8. Manejadores de Bibliografía / Citar

La investigación biológica implica de forma indispensable la revisión bibliográfica, 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 que permiten importar, organizar, exportar, editar y compartir referencias bibliográficas, así como crear bibliografías personales, dar formato e incluso algunas 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, 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 presentados en la 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, permite almacenar, compartir, ordenar y sistematizar las referencias consultadas, de hecho se utilizó para citar y poner las referencias de este libro, esta decisión se debió a las siguientes ventajas que tiene:

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 a la ficha bibliográfica el documento en formato PDF (ver glosario .
6. Sincronización con el marcador de bibliografía CiteUlike (ver tabla 7).
7. Extracción de las referencia bibliográfica de un documento automáticamente por medio de minería de textos (ver sario). Si uno importa todo un archivo que contenga documentos en PDF genera la bibliografía y la adiciona, junto con los archivos.

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 (P)	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 (P), (OL)	Bibme.org http://www.bibme.org/	No aplica, basado en un Servidor Web	Sólo 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
Citation Machine (OL)	The Landmark Project http://citationmachine.net/index2.php	No aplica, basado en un Servidor Web	Sólo crea bibliografía manualmente y de un solo documento a la vez	No tiene	APA, MLA, Chicago/Turabian	HTML	No, solo permite copiar y pegar en Word	No
Citeulike (P), (OL)	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	ArXiv, CiteSeer, IEEE Xplore, PubMed, Otros, Amazon
Connotea (P), (OL)	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
EndNoteWeb (C)	Thomson Corporation http://www.myendnoteweb.co	Windows, Mac OS X	CSA, Endnote/Refer/BibIX, ISI, Medline,	BibTeX, RIS, Otros (CSV)	APA, Chicago/Turabian, Harvard, MLA, Otros, Método de	HTML, RTF, Texto plano, Otros (clipboard,	Microsoft Word, OpenOffice.org/Writer	PubMed, Otros

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
	m/EndNoteWeb.html		Ovid, PubMed, RIS, SciFinder, Otros.		extensión: Formato EndNote a través de la interface gráfica de usuario.	XML)	(parcial)	
Mendeley (P), (OL)	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
NoodleBib Express (P), (OL)	NoodleTools, Inc. 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
Ottobib (OL)	Jonathan Otto 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
ProCite (C)	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, Other (varios)
RefWorks (C), (P), (OL)	RefWorks 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 interface	RTF, Texto plano, RSS, Otros (DOC, ODT)	Microsoft Word	PubMed, Otros

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
					gráfica del usuario.			
Zotero (P), (OL)	Center for History and New Media http://www.zotero.org/	Windows, Mac OS X, Linux, BSD, Unix	BibTeX, Endnote/Refer/BibIX, ISI (parcial), MODS XML, Ovid (parcial), PubMed (parcial, directamente de página web, no desde archivo guardado), RIS, Otros(COinS, MARC, RDF, unAPI)	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

(C): Costo por licencia, (P): Requiere password, (OL): Versión en línea.

Tabla 7. Comparación de algunas 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, particularidad ya que comparten un gran número de aplicaciones disponibles en 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 una 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. En la tabla 9 se mencionan y describen brevemente algunos de los más útiles para la recuperación de la literatura digital. En la Fig. 12 se muestran las aplicaciones y complementos que ofrece Google y en la Fig. 13 se presenta una imagen con una pantalla de *Google Chrome* personalizada con algunos de los complementos mencionados.

Complemento	URL	Descripción
4Shared Almacenamiento en línea	http://www.4shared.com	4shared tiene un solo propósito: almacenar y comparte 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. Se puede proteger con una contraseña a los archivos, para tener más seguridad. Este rasgo es muy útil para estudiantes que desean hacer sus proyectos exclusivos para un grupo y para profesionales, 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, de ahí el compromiso adquirido 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

Complemento	URL	Descripción
		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 les permitirá contemplar la nueva versión. Ofrece instalar un botón en la barra de herramientas de Google que permite el acceso inmediato al diccionario.
Diccionario y Tesauro Merriam's Webster Online Diccionario y Tesauro Español-Inglés-Español	http://www.merriam-webster.com	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, mas enlaces a palabras y definiciones relacionadas. El tesauro Merriam-Webster contiene más de 340, 000 sinónimos, antónimos, 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 el tesauro. Se pueden instalar complementos para el navegador Firefox, botones en la barra de Google y barra de herramientas para Internet Explorer.
Dropbox Almacenamiento en línea	http://www.getdropbox.com	Dropbox es un servicio de almacenamiento online 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 en forma gratuita y se asocia uno o varios equipos a la cuenta, se colocan datos arrastrando o copiándolos a la carpeta Dropbox creada por el programa. Éstos se sincronizan automáticamente con la cuenta online, y por ende ya están al alcance de todos los equipos asociados a la cuenta. 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

Complemento	URL	Descripción
		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.
Logmein Acceso remoto a la PC.	https://secure.logmein.com	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 una posición; y una pizarra, para escribir directamente sobre el escritorio de la computadora remota. Su completo 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.
Notasuno Notas Post-It Virtuales para la PC	http://notasuno.uptodown.com	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.
PDF-XChange Viewer Visor de documentos PDF	http://pdf-xchange-viewer.uptodown.com	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.
Publish or Perish	http://www.harzing.com	Es un complemento que permite buscar, visualizar y ordenar los resultados de Google Académico, presenta un software que realiza análisis de citas.
Primo PDF Convertidor de documentos	http://www.primopdf.com	PrimoPDF es un programa que funciona instalando una impresora virtual en el sistema. Así, se puede convertir a PDF cualquier

Complemento	URL	Descripción
(.txt, .doc, html, ect.) a formato PDF.		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 un archivo PDF de forma totalmente gratuita: ni limitaciones, ni marcas de agua, ni frases publicitarias al final de los documentos.
Complementos (Add-ons) para el navegador Firefox	https://addons.mozilla.org/es-ES/firefox	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. Fáciles de instalar, y envío de avisos cuando 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.
Chrome Web Store	https://chrome.google.com/webstore	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.
Barra de Google	http://www.google.com/toolbar/firefox/toolbar/FT2/intl/es	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.

Complemento	URL	Descripción
iGoogle	http://www.google.com.mx/ig	Un servicio de Google, es una página de inicio personalizable. Sus características incluyen la capacidad de añadir extensiones de páginas web y <i>gadgets</i> de Google, especialmente desarrollados para mostrar el contenido a un usuario de la página. Estos pueden colocarse en distintas pestañas dentro de iGoogle.

Tabla 8. Algunos 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.
Impresionante captura de pantalla	A través de un clic permite capturar, recortar o seleccionar cualquier parte de una página web que se desea. Anotar con rectángulos, círculos, flechas, líneas y texto. permite editar, ocultar información y compartir en la Web.
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, las contraseñas de todos los sitios y páginas web a las que se accede frecuentemente.
Plugs 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.

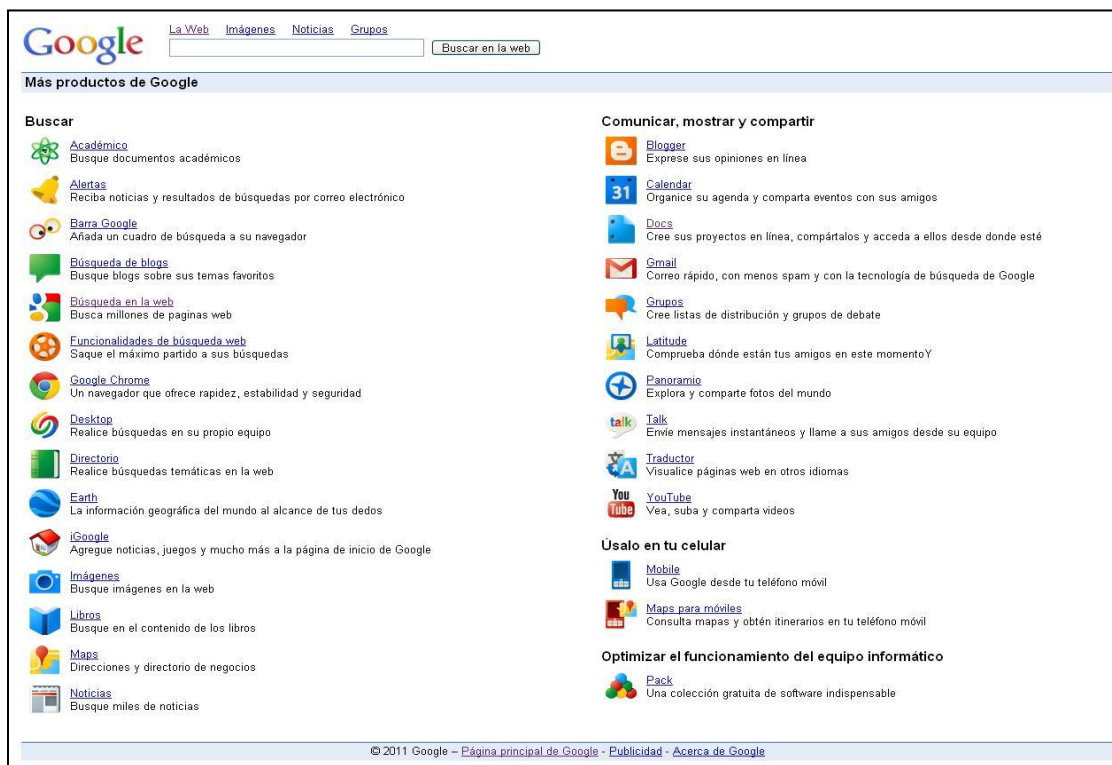


Figura 12. Algunas de las aplicaciones de Google.



Figura 13. Complementos para Google Chrome.

II. LA RECUPERACIÓN DE INFORMACIÓN EN LA WEB

Las actividades científicas y sus productos crecen de forma exponencial: el número de los investigadores y de sus publicaciones se duplica aproximadamente cada veinte años, viven actualmente entre un 80% y un 90% de los científicos que han existido, y cada año se publican más de dos millones de artículos, se editan de miles de revistas científicas y se conceden más de un millón de patentes. Es decir, estamos inmersos en la era de la información y en la sociedad del conocimiento, donde la información constituye en la actualidad 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 ha producido el formato digital y la existencia de la Web, ha trascendido la forma en la que se almacenan, sistematizan, recuperan y difunden los datos, implica que conocer sobre un tema ya no dependa únicamente de los documentos impresos disponibles en las librerías o bibliotecas, sino también de su existencia en catálogos electrónicos, formato en el que hoy se encuentra la mayor cantidad de información en línea, lo que 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, para esto se requiere reflexionar y analizar a profundidad el objeto de estudio y definir claramente una pregunta documental, para ello 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?

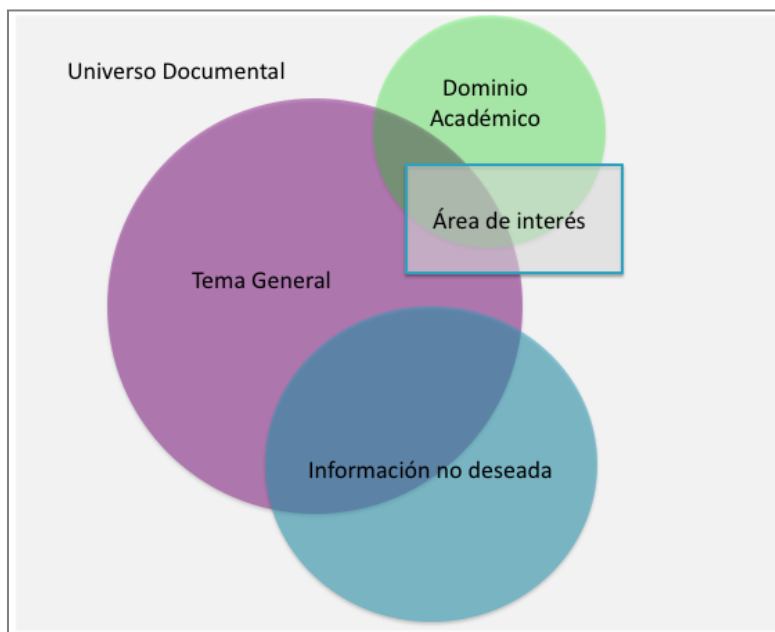


Figura 14. Planteamiento y definición de la necesidad informativa.

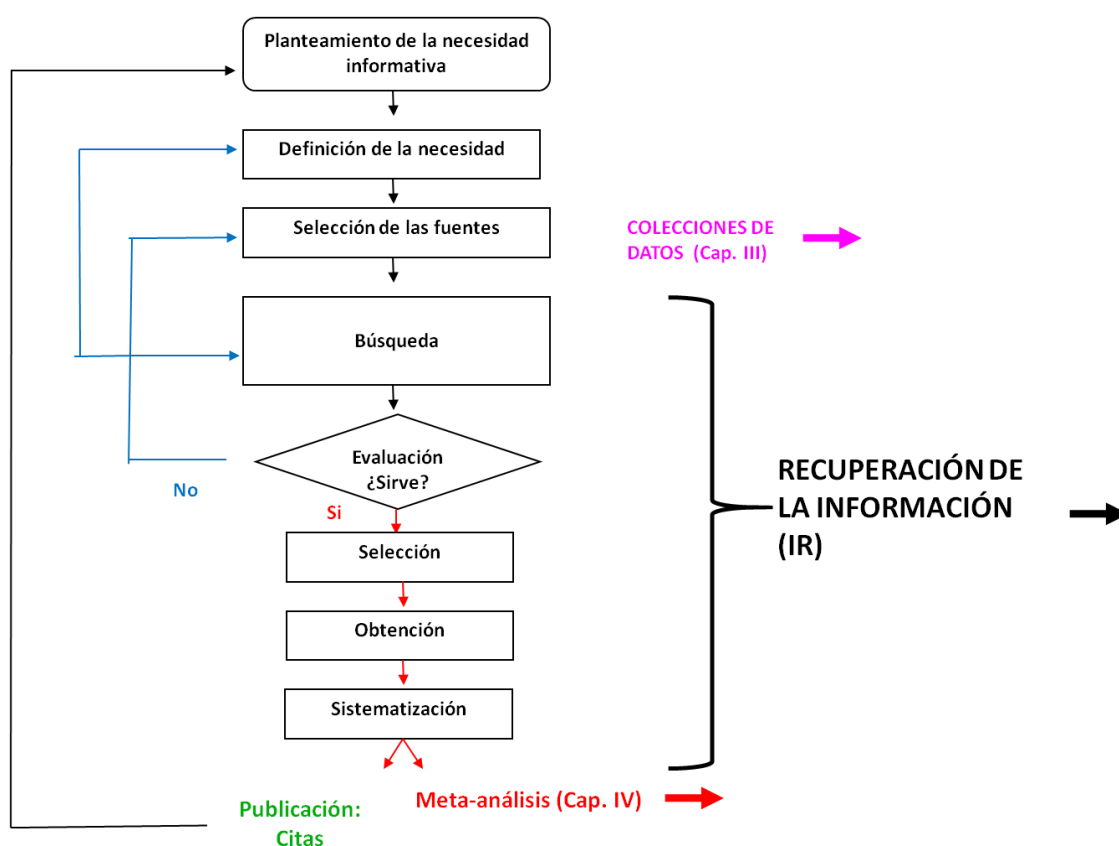


Figura 15. Proceso de la recuperación de información en la Web, se agrega a cada etapa el capítulo de este libro en el que se puede ampliar la información.

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

	General	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 Autor Revista
Uso de filtros para refinar la búsqueda	No	Si
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 una búsqueda general y una avanzada.

La búsqueda general es rápida y básica, se usa cuando se desea obtener un resultado general y consta de introducir uno o varios términos elegidos. La búsqueda avanzada consta de explorar toda la información y hacer pruebas por medio de distintas consultas utilizando los términos, los operadores indicados y estableciendo los límites correspondientes, lo cual permite aplicar criterios más específicos que arrojan resultados más exactos; se hace por medio de consultas (*queries*) que se basan en tres elementos: 1) términos, 2) operadores y 3) límites.

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, 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 (por ejemplo Mesh <http://www.nlm.nih.gov/mesh/MBrowser.html>, Fig. 16) de palabras clave de las colecciones que muestran la relación jerárquica que existe entre los diversos descriptores y en la mayoría de los casos incluyen una definición.

NATIONAL LIBRARY of MEDICINE **MEDICAL SUBJECT HEADINGS** **MeSH**

MeSH Home | Contact NLM | Site Index | Search Our Web Site | NLM Home
Health Information | Library Services | Research Programs | New & Noteworthy | General Information

MeSH Browser (2011 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:

- ☐ Main Headings
- ☐ Qualifiers
- ☐ Supplementary Concepts
- ☒ All of the Above
- ☐ Search as MeSH Unique ID
- ☐ Search as text words in Annotation & Scope Note

☐ Search in these fields of chemicals:

- ☐ Heading Mapped To (HM) (Supplementary List)
- ☐ Indexing Information (II) (Supplementary List)
- ☐ Pharmacological Action (PA)
- ☐ CAS Registry/EC Number (RN)
- ☐ Related CAS Registry Number (RR)

[MeSH vocabulary](#)

[About MeSH Browser](#) | [MeSH Home Page](#) | [Questions or Comments](#)
[NLM Classification, the scheme used to categorize and organize books, audiovisuals, and similar materials.](#)

Figura 16. Tesauro Mesh.

El segundo elemento corresponde al uso de operadores lógicos que 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); éstos 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
<i>Boleanos</i>	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
<i>Truncamiento</i>	?	Sustituye una letra o número en cualquier posición de una palabra o serie de números	mexican?.- mexicana, mexicano
	*	Sustituye una o más letras o números en cualquier posición de la palabra o serie de números	mexican*.- mexicana, mexicano, mexicanas, mexicanos
<i>Proximidad</i>	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 realiza 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

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 Go

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

WEB OF KNOWLEDGESM | DISCOVERY STARTS HERE

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](#)

Search 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: Save As My Defaults

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

Booleans: AND, OR, NOT, SAME, NEAR

Field Tags:

TS= Topic
TI= Title
AU= Author
RIID= ResearcherID
GP= Group Author
ED= Editor
SO= Publication Name
DO= DOI
PY= Year Published
CF= Conference
AD= Address
OG= Organization

SG= Suborganization
SA= Street Address
CI= City
PS= Province/State
CU= Country
ZP= Zip/Postal Code
FO= Funding Agency
FG= Grant Number
FT= Funding Text

Figura17. Búsqueda avanzada e historial de búsquedas.

2. Las etapas siguientes

La recuperación no termina cuando se realiza la primera búsqueda, es un proceso que se debe repetir varias veces hasta localizar lo que se busca, 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 puede realizarse de manera simultánea.

La evaluación consiste en la revisión de los documentos con detenimiento y detalle de manera sistémica (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; se pueden extraer para proceder al análisis de sus contenidos, como la aplicación de indicadores bibliométricos o minería de textos (*text mining*) (capítulo 4).

Por ende, es en este proceso donde se determina si nuestra necesidad informativa ha sido satisfecha; en caso de que no se cumpla con ésta, es preciso sesgar la lista refinando la búsqueda o generando una nueva consulta, también se puede elegir otras fuentes y/o redefinir la necesidad informativa hasta que se obtengan resultados favorables.

La siguiente etapa es la extracción, en este proceso la mayoría de las bases de datos permiten extraer (exportar) los documentos a una lista en distintos formatos, pueden ser transferidos a archivos personales como: documentos, bases de datos o incluso directamente agregar el registro bibliográfico a un software de administración de referencias (capítulo 1), para consultarlos con detenimiento una vez que se hayan obtenido suficientes documentos; a partir de éstos se pueden 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 de 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 2, 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 que se produce.

III. 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 las 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 el *U.S. Bureau of Census* diseñó la primera base de datos formal. En la década de los 70’s 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, solo 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 indización de documentos produjo 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's 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 legales de las bases de datos. Durante esta época los editores de periódicos y revistas habían adoptado a la computadora para la fotocomposición en sus procesos de impresión y algunos empezaron a utilizarlas para el manejo de su información.

Las bases de datos se pueden clasificar de acuerdo con varios criterios, 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 del inventario de cualquier almacén o la de un repositorio institucional (ver adelante).

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 su vez gráfica, permite que cada campo nodo tenga varios ascendientes, este tipo es muy poco utilizado y 3) Relacionales; están compuestas por una colección de tablas, cada una con campos en común que los relacionan entre sí, formando un conjunto de ellos y no hay relevancia en la manera en que se almacenen (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, 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 la 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.

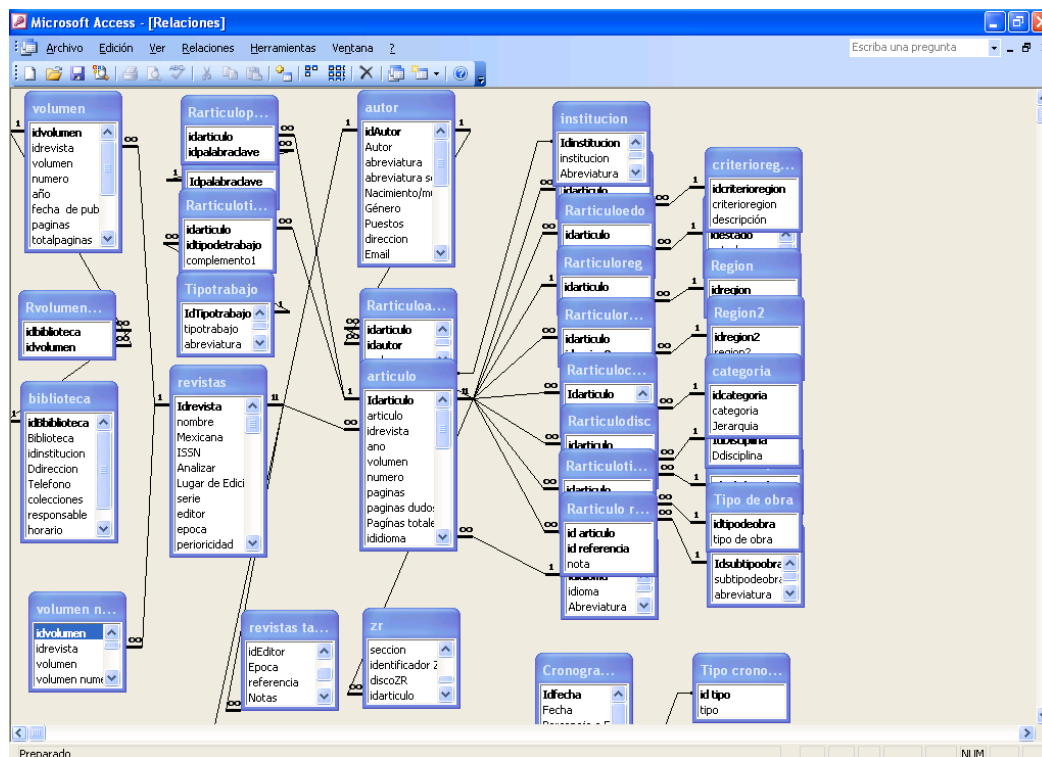


Figura 18. Estructura relacional de la Base de datos bibliográfica TaXMeXX, contiene artículos taxonómicos publicados en México durante el siglo XX disponible en línea en <http://repositorio.fcencias.unam.mx:8080/xmlui/handle/123456789/56181>.

2. 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, 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 *The 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; *Machine Readable 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) 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, 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 con la información básica sobre el 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, en las más complejas generalmente está organizada de la siguiente manera: 1) Los documentos, fundamentalmente artículos, contienen elementos como título, tipo de documento, idioma, palabra clave, descriptores, volumen, número y páginas; 2) La fuente de donde provienen (principalmente revistas), contienen título, año de publicación y tema; y 3) La autoría, esto es, que incluyan 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), los 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. En las figuras 19 y 20 se muestran dos tablas básicas en una base de datos bibliográfica, la de tipo de documentos y la de revistas.

Introducción a la Recuperación de Literatura Digital en Ciencias Biológicas
Layla Michán, Jack Guillén, Eduardo Alvarez, Lyssania Macías e Itzel Pedraza

Microsoft Access - [revistas : Tabla]

Idref	nombre	editor	epoca	ISSN	tipc
2	Acta Botánica Mexicana	Instituto de Ecología	1988	01877151	Cor
115	Acta Científica Potosina	Universidad Autónoma de San Luis Potosí	1957-1994	05152771 0185-1152?	Alg
122	Acta de Chapultepec	Museo de Historia Natural de la Cd. de México,	0		Alg
121	Acta Politécnica Mexicana	IPN. Escuela Superior de Medicina	1959	0515-3085	Alg
100	Acta Zoológica Mexicana	1955-1962 vol I-VI(1)Comité de publicación de inves	1955-1971	00651737	Cor
3	Acta Zoológica Mexicana nueva serie	Instituto de Ecología, A. C	1984	01855476	Cor
123	Agrociencia	El Colegio de Postgraduados	1966-	1405-3195	Alg
4	Anales de la Escuela Nacional de Ciencias Biológicas	Escuela Nacional de Ciencias Biológicas Instituto f	1938	01850946 03651932 che	Cor
5	Anales del Instituto de Biología	Instituto de Biología, U.N.A.M.	1930	00767174	Cor
6	Anales del Instituto de Biología Serie Botánica	Instituto de Biología, U.N.A.M.	1967	0374551	Cor
7	Anales del Instituto de Biología Serie Ciencias del Mar y Limnología	Instituto de Biología, U.N.A.M.	1967-1974	03688305	Cor
8	Anales del Instituto de Biología Serie Zoológica	Instituto de Biología, U.N.A.M.	1967	03688720	Cor
9	Anales del Instituto de Ciencias del Mar y Limnología	Instituto de Ciencias del Mar y Limnología	1974-1995	01853287 y 0186-3434	Cor
124	Avacient	Instituto Tecnológico de Chetumal	1990		Alg
118	Biológicas, Mor	Facultad de Biología, Universidad Michoacana de E	0		Alg
119	BIOTAM	Instituto de Ecología en Alimentos Universidad Aut	1989	01878476	Alg
10	Biótica	Instituto Nacional de Investigaciones sobre Recurs	1976-1993	01850326	Cor
111	Boletín Amaranto	Asociación Mexicana de Jardines Botánicos A. C.	1988	0188-8862	Alg
34	Boletín de la Dirección General de Agricultura	Dirección General de Agricultura	1911-1917	01870718	Alg
89	Boletín de la Sociedad Agrícola Mexicana	Boletín de la Sociedad Agrícola Mexicana	1979-1914	0187540X	Alg
12	Boletín de la Sociedad Botánica de México	Sociedad Botánica de México A. C.	1944	01853619	Cor
11	Boletín de la Sociedad Herpetológica Mexicana	Sociedad Mexicana de Herpetología	1989	0187988X	Cor
117	Boletín de la Sociedad Mexicana de Entomología	Sociedad Mexicana de Entomología	1968-1973, 198		Cor
29	Boletín de la Sociedad Mexicana de Micología	Sociedad Mexicana de Micología, A. C.	1968-1987	0085-6223	Cor
90	Boletín de Parasitología Agrícola	Comisión de Parasitología Agrícola	1900-1908 190		Alg
116	Boletín del Instituto de Botánica	Universidad de Guadalajara	1978	01877054	Alg
132	Boletín Informativo de la Sociedad Mexicana de Lepidopterología	Sociedad Mexicana de Lepidopterología A. C.	1975-1982		Cor
120	Boletín Nakari	Sociedad Jaliscience de Cactología	1990		Alg
134	Boletín Sociedad Veracruzana Zoología		0		Alg
13	Cactáceas y Suculentas Mexicanas	Sociedad Mexicana de Cactología, A. C	1955	0526717X	Cor
130	Chapingo	Universidad Autónoma de Chapingo	1927	01867075	Alg
98	Ciencia	Academia Mexicana de Ciencias	1940	14056550	Cor
125	Ciencia Nicolaita	Coordinación de la Investigación Científica, UMSNH	1992		Alg

Campos

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

Microsoft Access - [revistas : Tabla]

Idref	nombre	editor	epoca	ISSN	tipc
2	Acta Botánica Mexicana	Instituto de Ecología	1988	01877151	Cor
115	Acta Científica Potosina	Universidad Autónoma de San Luis Potosí	1957-1994	05152771 0185-1152?	Alg
122	Acta de Chapultepec	Museo de Historia Natural de la Cd. de México,	0		Alg
121	Acta Politécnica Mexicana	IPN. Escuela Superior de Medicina	1959	0515-3085	Alg
100	Acta Zoológica Mexicana	1955-1962 vol I-VI(1)Comité de publicación de inves	1955-1971	00651737	Cor
3	Acta Zoológica Mexicana nueva serie	Instituto de Ecología, A. C	1984	01855476	Cor
123	Agrociencia	El Colegio de Postgraduados	1966-	1405-3195	Alg
4	Anales de la Escuela Nacional de Ciencias Biológicas	Escuela Nacional de Ciencias Biológicas Instituto f	1938	01850946 03651932 che	Cor
5	Anales del Instituto de Biología	Instituto de Biología, U.N.A.M.	1930	00767174	Cor
6	Anales del Instituto de Biología Serie Botánica	Instituto de Biología, U.N.A.M.	1967	0374551	Cor
7	Anales del Instituto de Biología Serie Ciencias del Mar y Limnología	Instituto de Biología, U.N.A.M.	1967-1974	03688305	Cor
8	Anales del Instituto de Biología Serie Zoológica	Instituto de Biología, U.N.A.M.	1967	03688720	Cor
9	Anales del Instituto de Ciencias del Mar y Limnología	Instituto de Ciencias del Mar y Limnología	1974-1995	01853287 y 0186-3434	Cor
124	Avacient	Instituto Tecnológico de Chetumal	1990		Alg
118	Biológicas, Morelia, Michoacán	Facultad de Biología, Universidad Michoacana de E	0		Alg
119	BIOTAM	Instituto de Ecología en Alimentos Universidad Aut	1989	01878476	Alg
10	Biótica	Instituto Nacional de Investigaciones sobre Recurs	1976-1993	01850326	Cor
111	Boletín Amaranto	Asociación Mexicana de Jardines Botánicos A. C.	1988	0188-8862	Alg
34	Boletín de la Dirección General de Agricultura	Dirección General de Agricultura	1911-1917	01870718	Alg
89	Boletín de la Sociedad Agrícola Mexicana	Boletín de la Sociedad Agrícola Mexicana	1979-1914	0187540X	Alg
12	Boletín de la Sociedad Botánica de México	Sociedad Botánica de México A. C.	1944	01853619	Cor
11	Boletín de la Sociedad Herpetológica Mexicana	Sociedad Mexicana de Herpetología	1989	0187988X	Cor
117	Boletín de la Sociedad Mexicana de Entomología	Sociedad Mexicana de Entomología	1968-1973, 198		Cor
29	Boletín de la Sociedad Mexicana de Micología	Sociedad Mexicana de Micología, A. C.	1968-1987	0085-6223	Cor
90	Boletín de Parasitología Agrícola	Comisión de Parasitología Agrícola	1900-1908 190		Alg
116	Boletín del Instituto de Botánica	Universidad de Guadalajara	1978	01877054	Alg
132	Boletín Informativo de la Sociedad Mexicana de Lepidopterología	Sociedad Mexicana de Lepidopterología A. C.	1975-1982		Cor
120	Boletín Nakari	Sociedad Jaliscience de Cactología	1990		Alg
134	Boletín Sociedad Veracruzana Zoología		0		Alg
13	Cactáceas y Suculentas Mexicanas	Sociedad Mexicana de Cactología, A. C	1955	0526717X	Cor
130	Chapingo	Universidad Autónoma de Chapingo	1927	01867075	Alg
98	Ciencia	Academia Mexicana de Ciencias	1940	14056550	Cor
125	Ciencia Nicolaita	Coordinación de la Investigación Científica, UMSNH	1992		Alg

Registros

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

3. Características de las colecciones bibliográficas

Entre las características más importantes a considerar para elegir cuál o cuáles bases de datos documentales se deben utilizar para recuperar literatura están:

- El tipo de instancia (compañía o institución) que la produce y mantiene los datos y la que provee el acceso, está define en gran medida las características 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 contienen. Los más usados son los documentos y las revistas indizadas. En algunos casos también se usa como indicador la velocidad en la que adicionan nuevos documentos (Fig. 21).

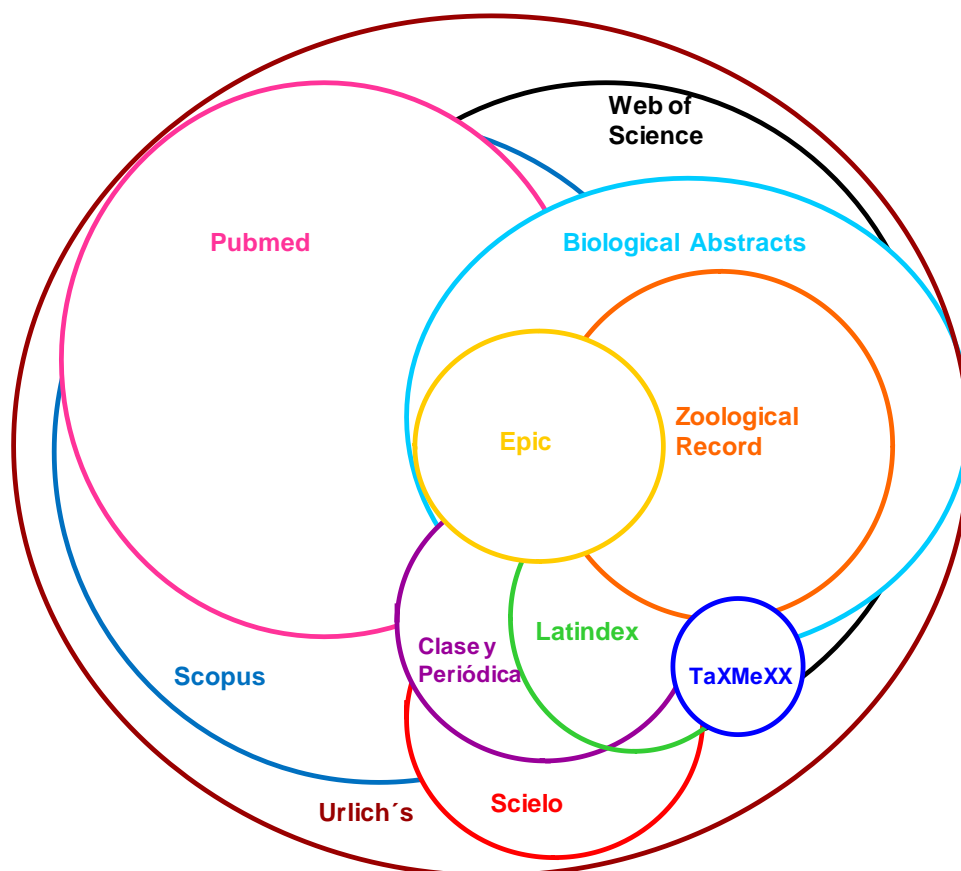


Figura 21. Representación gráfica del tamaño de las colecciones bibliográficas más utilizadas en biología, se representa también el traslape aproximado de las revistas que comparten.

- 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 con su propósito, determina los criterios que seguirá para elegir los documentos que incluirá en su colección.
- La cobertura disciplinar y temática. 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. Las segundas serían, zoología, botánica, taxonomía o genética, por ejemplo (Fig. 22).

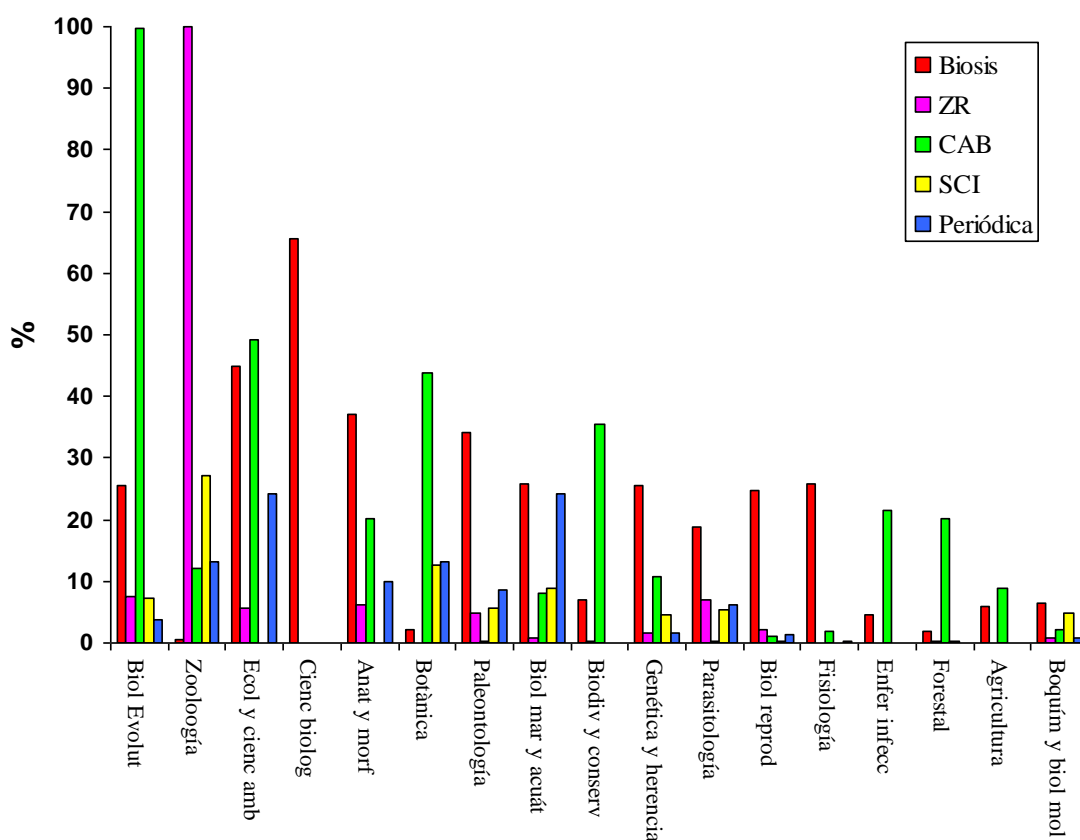


Figura 22. Temas más frecuentes en cinco de las colecciones bibliográficas con más documentos sobre biología. Se representan búsquedas sobre taxonomía de América Latina (consulta realizada en el año 2009).

- El tipo de campos. Son las entidades que se capturan para cada registro (título, autor, revista, idioma, palabras clave, resumen, referencias y citas, por ejemplo).
- El tipo de acceso a la base de datos. Este puede ser libre o restringido; el segundo se refiere a que sólo se puede acceder previo una contratación y pago, que generalmente se hacen de forma institucional, pues el precio es demasiado alto (miles de dólares) para hacerlo de manera individual. Por lo tanto, estos servicios se pueden consultar, generalmente, en las bibliotecas de instancias académicas o gubernamentales.
- Cobertura idiomática. Los idiomas de los documentos que se registran, la mayoría utilizan el inglés por ser el idioma que se utiliza más frecuentemente en las publicaciones científicas (Fig. 23).

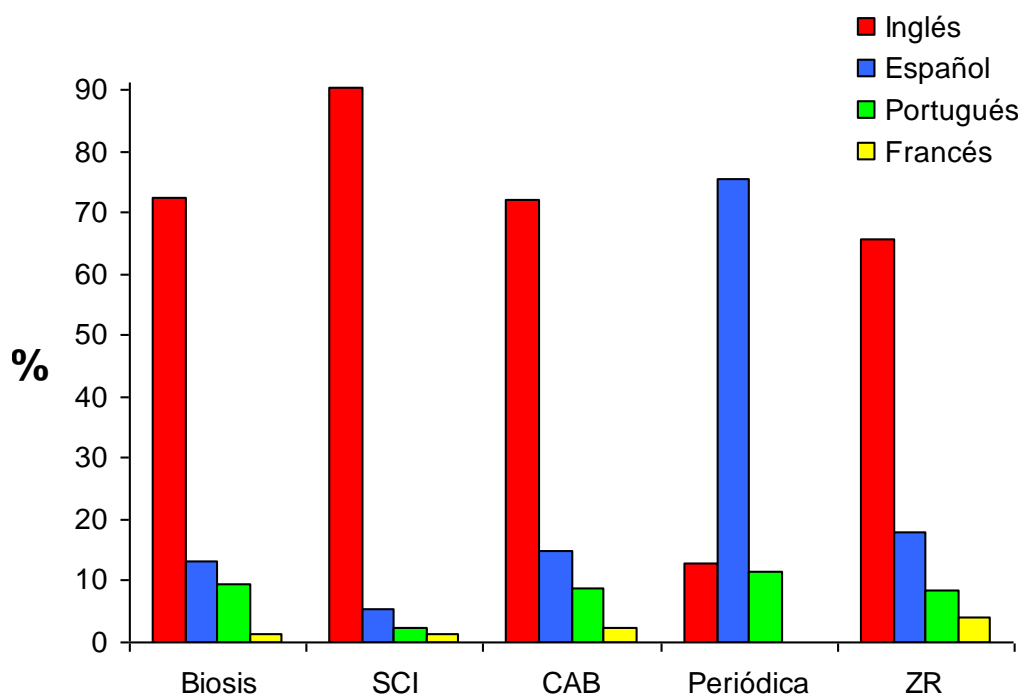


Figura 23. Idiomas más representados en cinco de las colecciones bibliográficas con más documentos sobre biología respecto a taxonomía de América Latina (consulta realizada en el 2009).

- Cobertura geográfica. Corresponde al lugar de edición de los documentos (generalmente las revistas) que se indizan. Esta puede ser mundial, si 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).

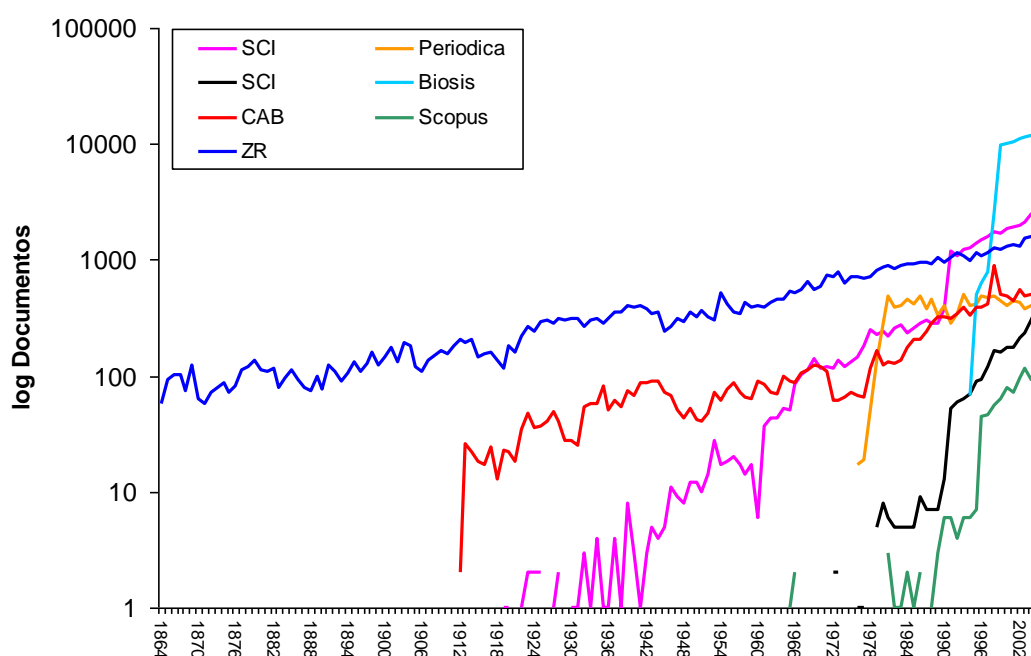


Figura 24. Años que abarcan los documentos en siete de las colecciones bibliográficas con más documentos sobre biología. Se representan documentos sobre taxonomía de América Latina (consulta realizada en el año 2009).

- 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.
- Tesoros. El uso de un catálogo controlado para la clasificación de los documentos con base en descriptores. Ésta 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 completa representación 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, si contiene la producción de una instancia académica como la Facultad de Ciencias de la UNAM se denomina institucional (Fig. 26).



Figura 25. Algunos nombres distintos utilizados frecuentemente en la Web para referirse a las colecciones bibliográficas.

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, <http://repositorio.fciencias.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 sus aplicaciones Web, con la finalidad de comparar las más extensas y destacadas accesibles en línea a través de la web y diseñadas para fines académicos (tabla 12).

Nombre	Editor-Productor	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/Pwebrecon.cgi
AGRIS 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.ntis.gov/products/ntisdb.aspx
Algology mycology and protozoology 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.aspx?id=1427
Aqualine	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos publicados	1960	Calidad del agua, legislación sobre el agua, análisis y los desechos del agua y su uso, afluentes industriales.	Mundial	Restringido	http://www.emeraldinsight.com/Insight/ac_browse.do?hdAction=link_browse_title&currentChar=A

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
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.ncsu.edu/searchcollection/database/s/more_info.php?database=6632
BIOBASE	Elsevier	Conferencias, simposios, reuniones, artículos de revista	1994	Microbiología aplicada y biotecnología, investigación 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	Mundial	Restringido	http://www.biobase-international.com/index.php?id=214
BioBusiness	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.acm.org/citation.cfm?id=15357
BioCommerce data's	BioCommerce Data Ltd.	Directorios y registros bibliográficos	1981	Bioquímica, control biológico, conversión de biomasa, biopesticidas, hibridación celular, genómica, horticultura, inmunología, microbiología, biología molecular, recombinación de RNA, transgénicos, tratamiento de residuos	Mundial	Restringido	http://www.stn-international.de/stn_news_message_04.html?&cHash=bf7915e011&tx_ttnews[pointer]=1&tx_ttnews[tt_news]=419&tx_ttnews[backPid]=5152

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
BioEngineering Abstracts	Cambridge Scientific Abstracts	Libros y monografías, conferencias, simposios, reuniones, artículos publicados, patentes	1991	Aplicaciones marinas, aplicaciones medioambientales, bioinformática, acuicultura y piscicultura.	Mundial	Restringido UNAM	http://www.lib.ncsu.edu/searchcollection/database/s/more_info.php?database=28972
Biological Abstracts	Thomson Reuters	Revistas, libros, memorias y reportes (T)	1926	Botánica, microbiología y farmacología	Mundial	Restringido UNAM	http://www.ovid.com/site/catalog/DataBase/24.jsp
Biological Digest	Cambridge Scientific Abstracts	Compilación de Abstracts e Índices de literatura nacional e internacional	1989	Ciencias biológicas	Mundial	Restringido UNAM	http://portal.acm.org/citation.cfm?id=607994
BioOne	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/
BIOSIS Previews	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,	Mundial	Restringido	http://www.ovid.com/site/catalog/DataBase/26.jsp

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
				sistemática, toxicología, ciencia veterinaria, virología			
BIOSIS 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.ovid.com/site/catalog/DataBase/26.jsp
Bireme	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 del Caribe	Restringido UNAM	http://bases.bvsalud.org/public/scripts/php/page_show_main.php?home=true&lang=en&form=simple
CAB	CABI	Revistas, libros, actas de congresos(T)	1910	Agricultura, ciencias ambientales, en particular; la ecología y el cambio 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	Mundial	Libre	http://www.cabi.org/datapage.asp?iDocID=165

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
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/cascontent/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, los artículos de periódicos	1975	Contaminación del aire, diseño ambiental y ecología urbana, educación ambiental, uso de la tierra, contaminación, océanos y estuarios, recursos renovables terrestre, agua, toxicología ambiental y seguridad, gestión de residuos, contaminación del agua, fauna	Mundial	Restringido	http://goliath.ecnxt.com/coms2/gi_0199-7670750/Enviroline-Ft-Lauderdale-FL-has.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/
Environmental Science	Cambridge Scientific Abstracts	Revistas científicas, actas de congresos, informes, monografías, libros y	1967	Biotecnología agrícola y ambiental, contaminación acuática y calidad ambiental, bacteriología microbiología ecología, declaraciones de impacto ambiental, ingeniería	Mundial	Restringido UNAM	http://www.ovid.com/site/catalog/DataBase/50.jsp

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		publicaciones gubernamentales		ambiental, manejo de recursos naturales, microbiología, 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/firstsearch/databases/dbdetails/details/GenSciAbs.htm
Google Scholar (Académico)	Google inc	Documentos académicos	2007	Multidisciplinario	Mundial	Libre	http://scholar.google.com.mx/scholar_preferences?hl=es
Industrial and Applied Microbiology Abstracts	Cambridge Scientific Abstracts	Artículos de revistas (T)	1982	Prácticas en la agricultura, alimentos y bebidas, químicos y las industrias farmacéuticas	Mundial	Restringido UNAM	http://www.lib.polyu.edu.hk/node/711
IPA Toxicology	Thomson Scientific	Artículos de revistas	1970	Cuestiones como la toxicidad relacionados con la industria farmacéutica en general	Mundial	Restringido	http://www.cas.org/support/stngen/dbss/index.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	Mundial	Restringido	http://lib.stanford.edu/falconer-biology-library/databases-life-sciences

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
				molecular, micología, neurociencias, ácidos nucleicos, oncogénesis protozoología, toxicología, virología, zoología			
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, la contaminación marina, recursos marinos, desalinización, biología marina	Mundial	Restringido UNAM	https://login.heli-con.vuw.ac.nz/login?url=http%3a%2f%2fwww.csa.com%2fhtbin%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	América Latina y el Caribe*	Libre	http://132.248.9.1:8991/F/-/?func=find-b-0&local_base=PER01
Pollution Abstracts	Cambridge Scientific Abstracts	Libros, documentos de conferencias, procedimientos, revistas, documentos de	1981	Contaminación del aire, contaminación de agua dulce, contaminación terrestre, contaminación marina y tratamiento de aguas residuales	Mundial	Restringido	http://www.newcastle.edu.au/service/library/database/pollution.html

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		investigación, y los informes técnicos					
PubMed	U.S. National Library of Medicine	Artículos y revistas en ciencias de la vida con una concentración en la biomedicina (T)	1950	Ciencias en enfermería, odontología, medicina veterinaria, farmacia, aliadas a la salud, pre-clínica y ciencias afines.	Mundial	Libre	http://www.bdbi.osciences.com/support/search_pubMed.shtml
REDALyC	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	Iberoamérica*	Libre	http://redalyc.uaemex.mx/
SciELO	FAPESP, UNAM	Revistas, números, artículos, citas	1997	Ciencias	Iberoamérica*	Restringido UNAM	http://www.scielo.org/php/index.php?lang=es
Science Citation Index*	Thomson Reuters	Artículos, bibliografías, reseñas de libros, correcciones y agregados, debates, editoriales, artículos de revisión y	1899	Ciencia y tecnología	Mundial	Restringido UNAM	http://wok.mimas.ac.uk/

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
		reseñas de software y hardware de computadora y bases de datos					
SciSearch	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.stn-international.com/scisearch.html
Scopus	Elsevier	Registro bibliográfico, resumen, conferencias, citas e indicadores bibliométricos	1869	Multidisciplinaria	Mundial	Acceso Restringido . UNAM	http://www.scopus.com/home.url
Water Resources Abstracts	Cambridge Scientific Abstracts	Informes, libros monografías, conferencias, simposios, reuniones, artículos de revistas, patentes, casos de corte	1967	Las aguas subterráneas, lagos, estuarios, la erosión y la sedimentación, suministro de agua y la conservación, desalinización	Mundial	Acceso UNAM	http://www.newcastle.edu.au/service/library/database/water.html
Wildlife & ecology studies worldwide	EBSCO	Artículos de revistas	1935	Ciencias de la vida	Mundial	Acceso UNAM	http://www.nisc.com/Frame/NISC_products-f_Modified.htm

Nombre	Editor-Productor	Cobertura Tipológica	Cobertura temporal	Cobertura temática	Cobertura geográfica	Acceso	Sitio web
Wilson 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.ovid.com/site/catalog/DataBase/923.jsp
Zoological Record Online	Thomson Reuters	Revistas, boletines, monografías, libros, revistas y actas de congresos (<i>T</i>)	1864	Comportamiento, biodiversidad, conservación y ciencias ambientales, biología marina y agua dulce, paleontología, parasitología, sistemática y taxonomía, técnicas, zoogeografía	Mundial	Acceso UNAM	http://www.ovid.com/site/catalog/DataBase/200.jsp

(*T*) Se refiere a que la base de datos tiene un tesoro.

UNAM. Se refiere a que la UNAM cuenta con actualmente con suscripción a este servicio.

*La colección privilegia documentos en español.

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

Cada base de datos tiene características diferentes, para elegir cuáles y cuántas es conveniente usar en cada caso, es necesario reconocer las particularidades de cada una. A continuación se presentan más detalles sobre las más utilizadas.

Entre las colecciones bibliográficas de documentos (principalmente artículos) más grandes están, *Google Scholar* con cientos de millones de registros, *Scopus* (de Elsevier) <http://www.scopus.com/home.url> y *Web of Science* (de Thomson Reuters) http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?highlighted_tab=WOS&product=WOS&last_prod=WOS&SID=4D11Ojic8APL3c@ompC&search_mode=GeneralSearch y *Chemical Abstracts* (de la American Chemical Society) <http://www.cas.org/> con más de 30 millones cada una, *Biosis Previews* (de Thomson Reuters) <http://science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=BP> y *PubMed* (de la U.S. National Library of Medicine) <http://www.ncbi.nlm.nih.gov/pubmed/> con aproximadamente 18 millones. Además del volumen de información, la cobertura disciplinar temática son básicas 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) <http://scholar.google.com.mx/>, *Scopus* <http://www.scopus.com/home.url> y *Science Citation Index* <http://www.thomsonscientific.com/cgi-bin/jrnlst/jloptions.cgi?PC=K>. 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* http://apps.webofknowledge.com/BIOABS_GeneralSearch_input.do?last_prod=BIOABS&SID=4D11Ojic8APL3c%40ompC&product=BIOABS&highlighted_tab=BI [OABS&search_mode=GeneralSearch](http://apps.webofknowledge.com/BIOABS_GeneralSearch_input.do?last_prod=BIOABS&SID=4D11Ojic8APL3c%40ompC&product=BIOABS&highlighted_tab=BI), *Zoological Record* http://apps.webofknowledge.com/ZOOREC_GeneralSearch_input.do?highlighted_tab=ZOOREC&product=ZOOREC&last_prod=ZOOREC&search_mode=GeneralSearch&SID=4D11Ojic8APL3c@ompC y *PubMed* <http://www.ncbi.nlm.nih.gov/pubmed/>.

En la actualidad hay un debate entre la comunidad científica sobre si Google Académico (<http://scholar.google.com.mx/>) es una fuente confiable de información o no, a causa de sus deficiencias, entre las que resaltan, que muchos de sus registros son de origen desconocido, que no se conocen los criterios de inclusión de documentos ni las coberturas que tienen, requisitos indispensables para saber a qué tipo de información se está teniendo acceso.

Scopus <http://www.scopus.com/home.url> 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* <http://www.thomsonscientific.com/cgi-bin/jrnlst/jloptions.cgi?PC=K> 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).

ISI Web of Knowledge™ Take the next step

Sign In | 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 Cited Reference Search Advanced Search Search History Marked List (0)

Web of Science®

Search for: in Topic

Example: oil spill* mediterranean

AND in Author

Example: O'Brian C* OR O'Brian C*

Need help finding papers by an author? Use Author Finder.

AND in Publication Name

Example: Cancer* OR Journal of Cancer Research and Clinical Oncology

Add Another Field >>

Search Clear

Current Limits: [Hide Limits and Settings](#) (To save these permanently, sign in or register)

Timespan: ☒ All Years (updated 2009-07-18) ☐ From 1900-1914 to 2009 (default is all years)

Citation Databases: ☒ Science Citation Index Expanded (SCI-EXPANDED)-1900-present ☒ Social Sciences Citation Index (SSCI)-1956-present ☒ Arts & Humanities Citation Index (A&HCI)-1975-present

View in: [PDF](#) [HTML](#) [English](#)

Please give us your feedback on using ISI Web of Knowledge.

Acceptable Use Policy Copyright © 2009 Thomson Reuters

THOMSON REUTERS Published by Thomson Reuters

Looking for ISI Proceedings? It is now searchable from within Web of Science as the Conference Proceedings Citation Index. More information. Note: Times Cited counts now include proceedings citations. More information.

Discover Web of Science Explore the world's leading citation database with multidisciplinary coverage of over 10,000 high-impact journals in the sciences, social sciences, and arts and humanities, as well as international proceedings coverage for over 120,000 conferences. Powerful tools include cited reference searching, Citation Maps, and the Analysis Tool.

- Want to know more?
- Training in multiple languages

Customize Your Experience [Sign In](#) | [Register](#)

- Save and manage your references online with EndNote Web – freely available and fully integrated.
- Save and run searches
- Create alerts and RSS feeds
- Choose your start page
- Want to know more?

My ResearcherID

- What is ResearcherID?
- Sign In to ISI Web of Knowledge to get your ResearcherID.

Further Information

- What's New? 2009-04-19
- Training and Support
- Help Desk
- Provide Feedback
- Search the Web with Scientific WebPlus

Figura 27. Aplicación de Web of Science.

PubMed <http://www.ncbi.nlm.nih.gov/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, 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 *Medline* antiguo) 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, su gran volumen de información, las relaciones que hay de muchos de los documentos a una gran cantidad de colecciones e información biológica a través de la plataforma llamada Entrez (<http://www.ncbi.nlm.nih.gov/pubmed/>) (Fig. 28), la actualización constante de los recursos y el uso de descriptores para cada documento con base en el tesauro del *MeSH* que permite hacer búsquedas más detalladas (Figs. 29 y 30).

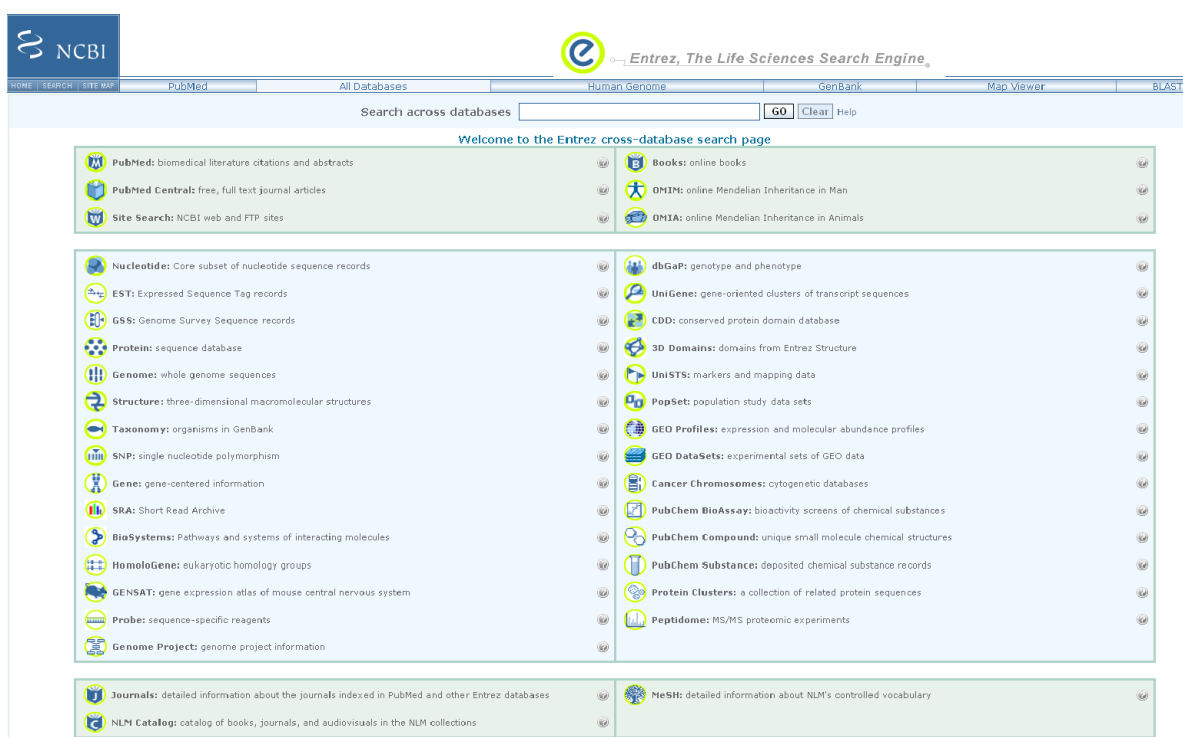


Figura 28. Portal de Entrez (<http://www.ncbi.nlm.nih.gov/sites/gquery>).

NATIONAL LIBRARY of MEDICINE | **MEDICAL SUBJECT HEADINGS** | **MeSH**

MeSH Home | Contact NLM | Site Index | Search Our Web Site | NLM Home

Health Information | Library Services | Research Programs | New & Noteworthy | General Information

MeSH Browser (2009 MeSH):
The files are updated every week on Sunday.
[Go to 2008 MeSH](#)

Enter term or the beginning of any root fragments: or

Search for these record types: ☐ Search in these fields of chemicals:

☐ Main Headings ☐ Heading Mapped To (HM) (Supplementary List)

☐ Qualifiers ☐ Indexing Information (II) (Supplementary List)

☐ Supplementary Concepts ☐ Pharmacological Action (PA)

☒ All of the Above ☐ CAS Registry/EC Number (RNT)

☐ Search as MeSH Unique ID ☐ Related CAS Registry Number (RR)

☐ Search as text words in Annotation & Scope Note

[MeSH vocabulary suggestions](#)

[About MeSH Browser](#) | [MeSH Home Page](#) | [Questions or Comments](#)
[NLM Classification](#), the scheme used to categorize and organize books, audiovisuals, and similar materials.

U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health
Department of Health & Human Services
Copyright and Privacy Policy
Last updated: 15 October 2008

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

National Library of Medicine - Medical Subject Headings	
2011 MeSH	
MeSH Descriptor Data	
Return to Entry Page	
Standard View: Go to Concept View : Go to Expanded Concept View	
MeSH Heading	Computational Biology
Tree Number	H01.158.273.180
Annotation	SPEC; SPEC qualif; coord IM with specific biol spec (IM); DF: COMPUTATIONAL BIOL
Scope Note	A field of biology concerned with the development of techniques for the collection and manipulation of biological data, and the use of such data to make biological discoveries or predictions. This field encompasses all computational methods and theories applicable to MOLECULAR BIOLOGY and areas of computer-based techniques for solving biological problems including manipulation of models and datasets.
Entry Term	Bio-Informatics
Entry Term	Bioinformatics
Entry Term	Biology, Computational
Entry Term	Computational Molecular Biology
Entry Term	Molecular Biology, Computational
See Also	Medical Informatics
Allowable Qualifiers	CL EC ED ES HL IS LJ MA MT OG SN ST TD
Entry Version	COMPUTATIONAL BIOL
Previous Indexing	Molecular Biology (1992-1996)
History Note	97
Date of Entry	19960613
Unique ID	D019295

Figura 30. Ejemplo de un descriptor del Mesh.

Otro indicador interesante de las bases de datos tiene que ver con la velocidad de captura de los documentos, ésta es cada vez más rápida, 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 en las colecciones más comunes son, por mucho, los artículos de revistas (de investigación y revisiones), estas 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 una gran cantidad de libros, este buscador junto con Amazon son 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 de 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 Universidad Nacional Autónoma de México (UNAM) paga suscripciones para que la comunidad tenga acceso a muchas de las bases de datos de acceso restringido más relevantes, es necesario que estas colecciones sean conocidas y aprovechadas en todo su potencial. De ahí la importancia de trabajos como éste, en el que se difundan los recursos disponibles para que sean utilizados por los interesados.

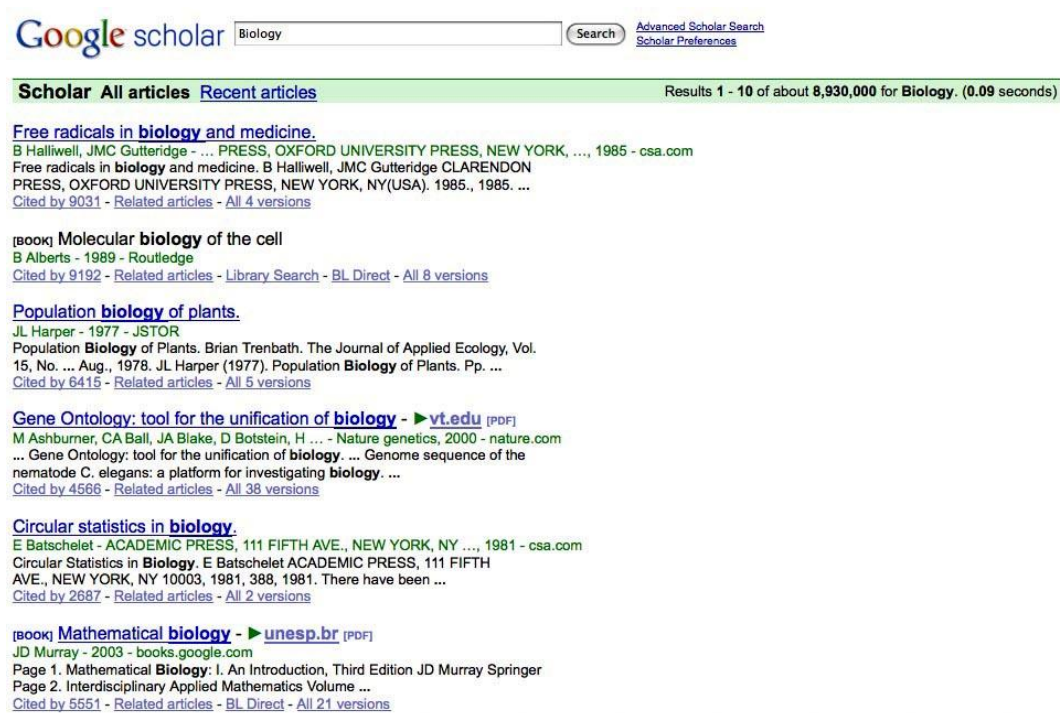


Figura 31. Portal de Google Académico.

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 contienen. 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 cuáles 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 http://132.248.9.1:8991/F/-/?func=find-b-0&local_base=CLA01 y Periódica http://132.248.9.1:8991/F/-/?func=find-b-0&local_base=PER01, SciELO <http://www.scielo.org/php/index.php>, Redalyc <http://redalyc.uaemex.mx/> o BVS <http://regional.bvsalud.org/php/index.php?lang=es>, LILACS <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&base=LILACS&lang=i&form=F>.

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 que se capturan, pero también se da el crecimiento basado en la recuperación de registros antiguos, los editores de muchas bases de datos están realizando el rescate de los archivos publicados en papel, a través de un amplio 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 por ejemplo, *Biodiversity Heritage Library* (<http://www.biodiversitylibrary.org>) (Fig. 32).

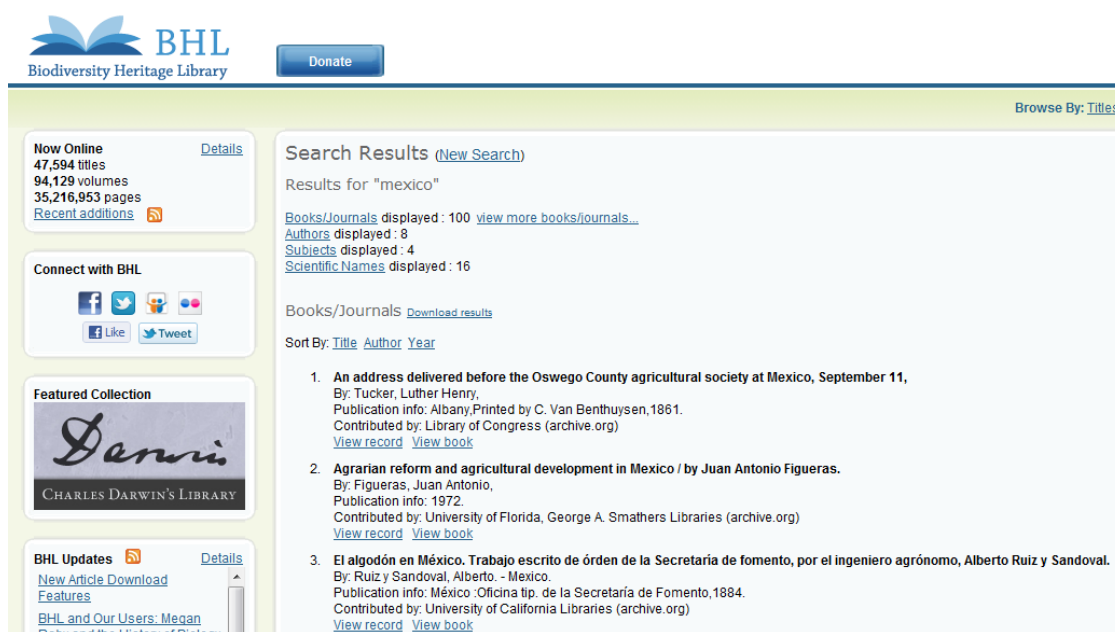


Figura 32. Portal de la biblioteca *Biodiversity Heritage*, una de las iniciativas más importantes para digitalización de documentos sobre biodiversidad, en todos los casos contiene el texto completo al que se refiere.

Web of Knowledge (ISI *Web of Knowledge* la aplicación que usa *Thomson Reuters* para acceder a sus bases de datos) es la que presenta la plataforma más completa (Fig. 27), le sigue de cerca *Scopus* (<http://www.scopus.com/home.url>) (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 indicadores bibliométricos como: cantidad de citas e índice H, Factor de Impacto, Vida Media, Índice de Inmediatez, *Eigenfactor*

<http://www.eigenfactor.org/>, por mencionar las más relevantes. Thomson Reuters presenta todos estos recursos en Web of Science http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?last_prod=WOS&SID=4D11Ojic8APL3c%40ompC&product=WOS&highlighted_tab=WOS&search_mode=GeneralSearch y Journal Citation Reports http://admin-apps.webofknowledge.com/JCR/JCR?SID=4D11Ojic8APL3c@ompC&locale=en_US que se publica en junio de cada año con acceso restringido. Para Scopus algunos análisis se acceden desde la propia plataforma de Scopus <http://www.scopus.com/home.url> o a través del portal se Scimago <http://www.scimagojr.com/> que es libre.

The screenshot displays the Scopus search interface. At the top, there are navigation links: Search, Sources, Analytics, My Alerts, My List, and My Profile. Below these is a banner asking about mobile device usage. The main section is titled 'Basic Search' and includes a search bar with a dropdown menu for 'Article Title, Abstract, Keywords'. Below the search bar, there are options to limit the search by 'Date Range' (Published or Added to Scopus) and 'Document Type' (All, Life Sciences, Physical Sciences, Social Sciences). There are also checkboxes for 'Subject Areas'. At the bottom, there is a 'Search History' section with a table showing search results, source, and actions. The interface is clean and professional, with a blue and white color scheme.

Figura 33. Aplicación de 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 mantenidos 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, que registra gran cantidad de meta-información asociada a la literatura (genes, sustancias químicas, proteínas, especies, etc), con la ventaja de que distribuye toda su información de manera libre, por lo que existen una gran cantidad de aplicaciones gratuitas (cientos) diseñadas por diversos grupos de

investigación para extraer, manejar y analizar la información de manera automatizada, inmediata y sistematizada basado en la Web 2 que aplican métodos innovadores de análisis como la bibliometría, las redes y el descubrimiento basado en literatura (minería de textos) (ver capítulo 4). Una lista preliminar se puede consultar en <http://www.diigo.com/user/lmichan/pubmed>.

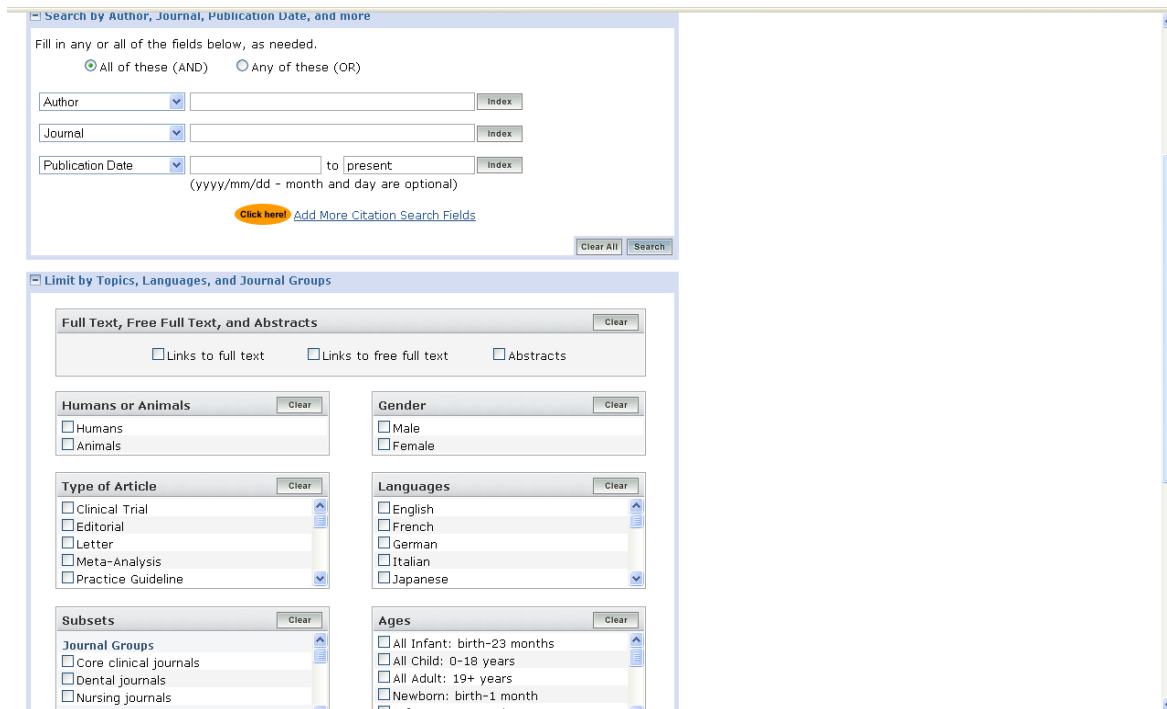


Figura 34. Búsqueda avanzada en la aplicación de PubMed
(<http://www.ncbi.nlm.nih.gov/pubmed/advanced>).

Los portales y los recursos que ofrecen las bases de datos de América Latina como SciELO <http://www.scielo.org/php/index.php> (Fig. 35), BIREME <http://regional.bvsalud.org/local/Site/bireme/E/homepage.htm> (Biblioteca Virtual en Salud) (Fig. 36), Redalyc <http://redalyc.uaemex.mx/> (Fig. 37) y Periódica http://132.248.9.1:8991/F/-/?func=find-b-0&local_base=PER01 (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.

Introducción a la Recuperación de Literatura Digital en Ciencias Biológicas
Layla Michán, Jack Guillén, Eduardo Alvarez, Lyssania Macías e Itzel Pedraza

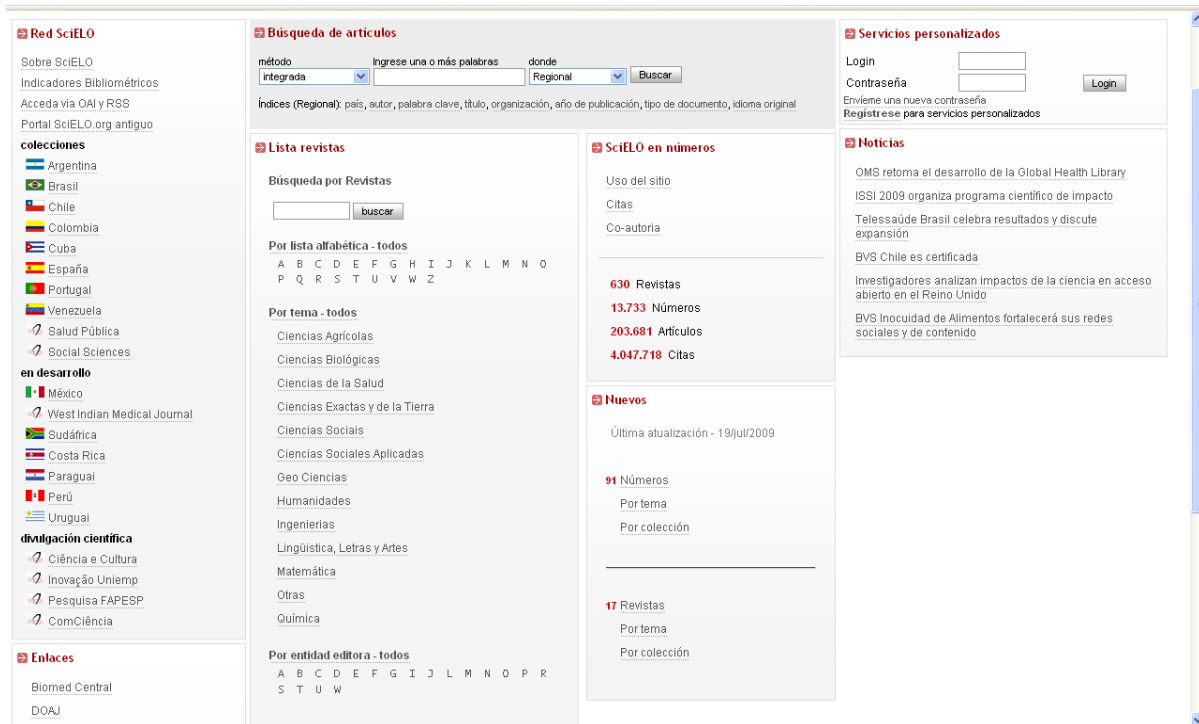


Figura 35. Portal de SciELO
(<http://www.scielo.org/php/index.php?lang=es>).



Figura 36. Sitio de la Biblioteca Virtual en Salud (BVS)
(<http://www.bireme.br/php/index.php?lang=es>).

Introducción a la Recuperación de Literatura Digital en Ciencias Biológicas
Layla Michán, Jack Guillén, Eduardo Alvarez, Lyssania Macías e Itzel Pedraza

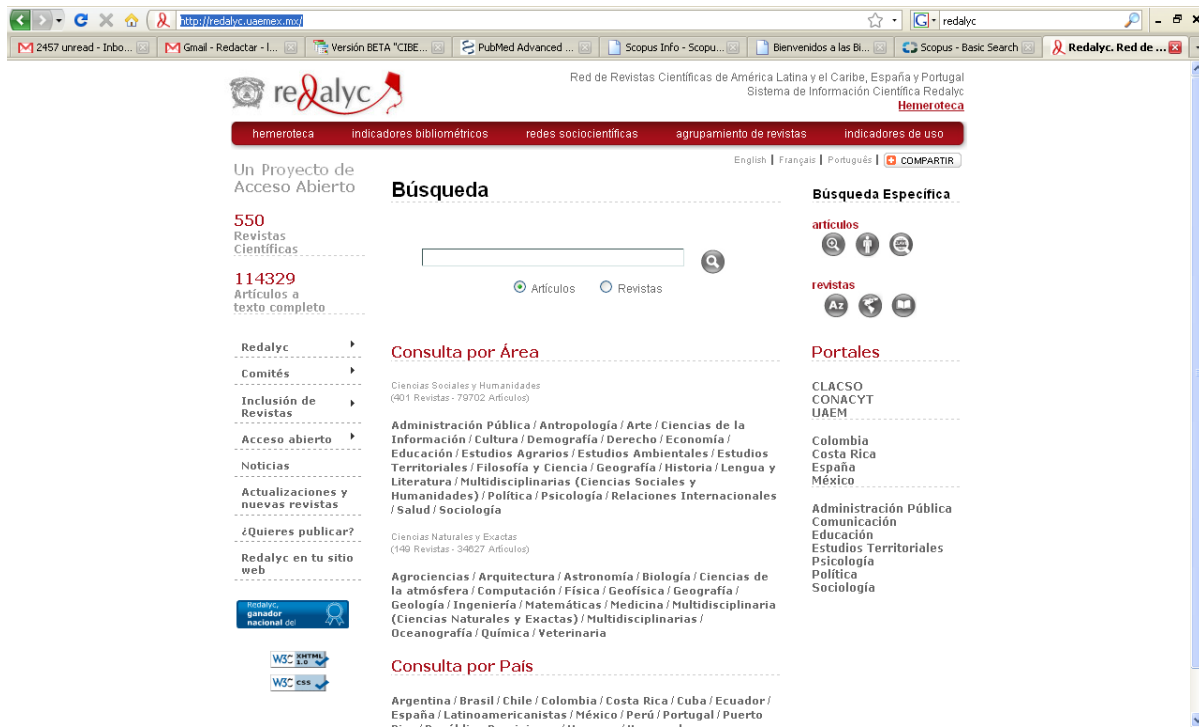


Figura 37. Servicio de REDALyC (<http://redalyc.uaemex.mx/>).

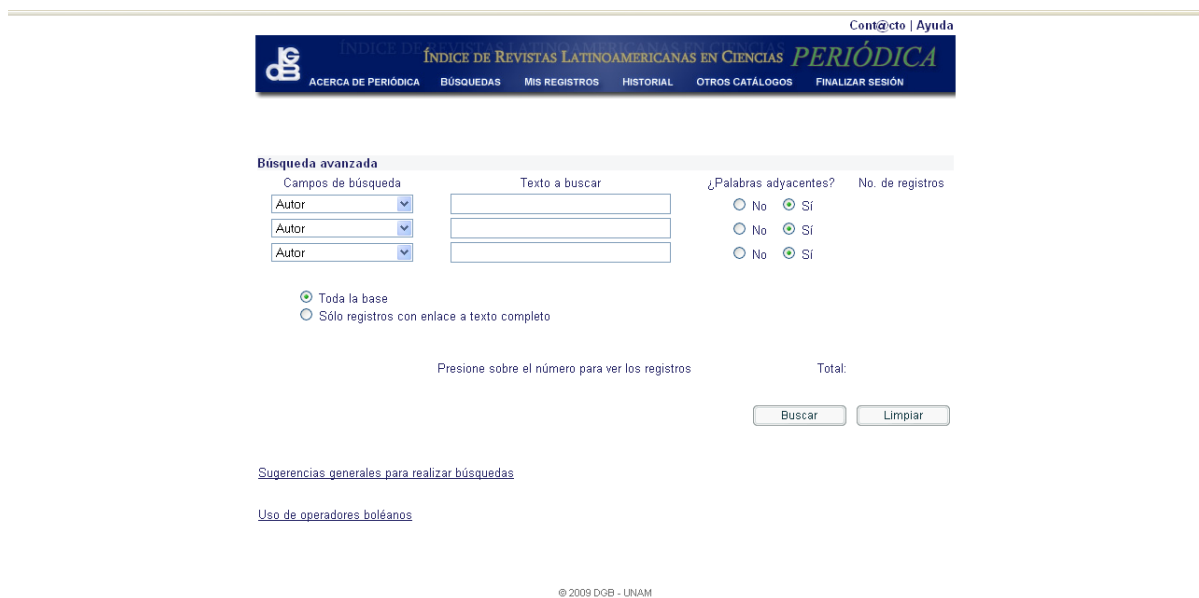


Figura 38. Aplicación de Periódica
(http://132.248.9.1:8991/F/-/?func=find-b-0&local_base=PER01).

Muchas de las herramientas presentadas en el Capítulo 1, como los marcadores sociales, las folkosomias o los que permiten comentar y compartir documentos, en especial en formato PDF, están disponibles a partir de 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 ésta 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 (Fig. 39) con 136, 900 (<http://www.cabi.org/cabthesaurus/>).

The screenshot shows the CAB Thesaurus 2011 website. The header is green with the title 'CAB Thesaurus 2011' and the CAB logo. Below the header is a navigation bar with 'Home', 'All terms A-Z', and 'Browse by subject / category'. The main content area is divided into three columns. The left column contains a search form with fields for 'Search Thesaurus', 'Language' (set to English), 'Search type' (set to 'terms begin with text'), 'Results format' (set to 'simple term list'), 'Terms per page' (set to 15), 'Match terms' (set to 'Equal or Greater'), and 'Term type' (set to 'Any'). There is a 'search' button and a note: '* Available only to search type "term begins with text"'. The middle column contains 'About CAB Thesaurus' text, a list of strengths, and an 'About CABI' link. The right column contains a 'Search string' section with a text box, a 'clear search' link, and buttons for 'Send to CAB Direct', 'Send to Google', and 'Send to Yahoo!'. The CAB logo is in the top right corner.

Figura 39. Tesoro.

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 será utilizar distintas combinaciones que satisfagan las necesidades informativas, para elegir qué bases de datos son las más adecuadas (Fig. 40). Por ejemplo para recuperar artículos sobre la investigación acerca de la biodiversidad de América Latina, sería idóneo utilizar al menos tres bases de datos: una general sobre temas biológicos, la más extensa y que tiene bien representados los últimos 30 años Biosis; 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.



Figura 40. 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. 41).

Figura 41. Line@, Base de datos con detalles de las Colecciones Bibliográficas para investigación biológica relacionadas y afines
<http://repositorio.fcencias.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*, estas son 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. Esta información es indispensable, para las instituciones gubernamentales y académicas, las cuales realizan un gran esfuerzo para mantener reservorios propios, en la mayoría de los casos de acceso abierto para el uso 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 las bases de

datos en la actualidad se llevan a cabo en el sector privado y su uso es estrictamente el comercial (pago por servicio), en tanto, los esfuerzos de instituciones académicas, y gubernamentales van a paso lento, frenadas por la burocracia y la falta de recursos.

Los registros de la literatura en las bases de datos aumentan aceleradamente debido a mejores y más completos procesos de recuperación, digitalización y sistematización de las publicaciones. Cabe destacar que el crecimiento no es casual, lo impulsan grandes editoras para su beneficio y de esa manera lo pueden acaparar, crecen con rapidez, y son las que recuperan con más celeridad la mayor cantidad de información documental de sus archivos antiguos o las que logran conseguir documentos ante la ingenua visión de algunas instituciones académicas; lo que nos muestra, el interés por acopiar datos, libros, publicaciones, memorias, tesis, patentes. Su restricción indudablemente finca fronteras a las sociedades del conocimiento.

Agradecimiento

Agradecemos la participación de Eloy Valtierra en este capítulo, su aportación fue invaluable.

IV. 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 constituido 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 en *The American Education Research Association* en 1976, el término meta-análisis fue acuñado por Gene V. Glass en 1976, 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 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 méta-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 incluidos en dicho 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 al cuál se quiera estudiar y/o la hipótesis que se plantee.

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 de estas proporcionadas por

las propias colecciones bibliográficas como *Web of Science* (http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?SID=1A8llBe53%40kfdBffH4O&product=WOS&search_mode=GeneralSearch), *Scopus* (<http://www.scopus.com/home.url>) o *SciELO* (<http://www.scielo.org/php/index.php?lang=es>) (ver capítulo 3) y otras diseñadas por terceros (empresas o instituciones académicas), que utilizan colecciones para procesar información como *Scimago* o *Eigenfactor* (<http://www.scimagojr.com/>, <http://www.eigenfactor.org/>). Los análisis pueden ser incluso en tiempo real, como los que hacen *PubMedreminer* (<http://bioinfo.amc.uva.nl/human-genetics/pubreminer/>) y *Ligercat* (<http://ligercat.ubio.org/>) a partir de la base de datos de *PubMed*, (<http://www.ncbi.nlm.nih.gov/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. Están disponibles para el uso de la comunidad académica en general, no especializada en las áreas de ciencias computaciones o versados 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, por ejemplo: 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 se han clasificado en cuatro tipos, que se pueden combinar; 1) bibliometría, 2) redes, 3) minería y 4) semántica.

1. Bibliometría

La bibliometría conceptualizada por Derek de Solla Price en 1963, en su más grande obra titulada “*Little Science, Big Science*” donde de manera gráfica explica el alcance de la ciencia en cuanto a la producción por años, misma que describió 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 uno denominado de la siguiente manera Cienciometría, Infometría, Cibermetría 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.

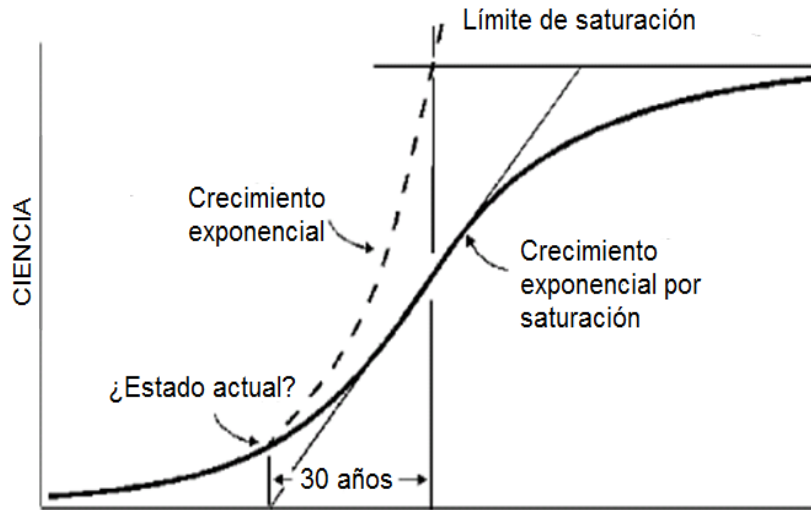


Figura 42. Representación gráfica del modelo sobre el crecimiento exponencial de la ciencia según de Solla Price en 1963.

Cienciometría	<ul style="list-style-type: none">• Propuesta en 1963• Análisis de la ciencia y la investigación científica
Bibliometría	<ul style="list-style-type: none">• Propuesta en 1969• Análisis de documentos
Infometría	<ul style="list-style-type: none">• Propuesta en 1979• Análisis a la información
Cibermetría	<ul style="list-style-type: none">• Propuesta en 2003• Análisis de páginas Web o blogs
Alt-metría	<ul style="list-style-type: none">• Propuesta en 2010• Web social con fines académicos

Figura 43. Métodos distintos para realizar métricas de la ciencia, según el objeto a estudiar.

Se emplea comúnmente, para la evaluación de la ciencia (tanto a 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

examina 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 en 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://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/) y en la aplicación de algoritmos de relevancia tipo *Page Rank* de Google. Proporciona datos anuales que abarcan desde 1995 a 2010 acerca de las 7, 000 revistas incluidas en JCR y de otras muchas fuentes citadas por ellas. Presenta *rankings* temáticos 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), es una propuesta de la biblioteca de biología es parte del proyecto de envejecimiento (*Biology of Aging project at the MBLWHOI Library at the Woods Hole Marine Biological Laboratory* http://mbl.edu/news/press_releases/2011_pr_01_27.html) es una herramienta de búsqueda y análisis de resultados de PubMed (<http://www.ncbi.nlm.nih.gov/pmc/>) 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).

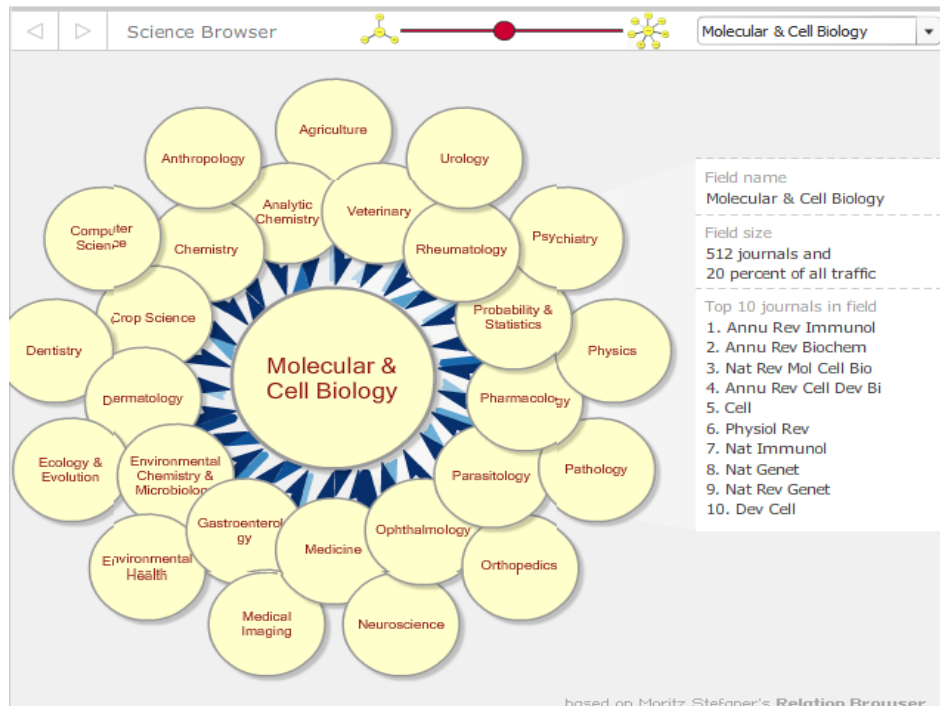


Figura 44. Interfaz Web de *Eigenfactor* (<http://eigenfactor.org/>) con un ejemplo de análisis del área de biología molecular y las disciplinas relacionadas, representadas en forma de red.

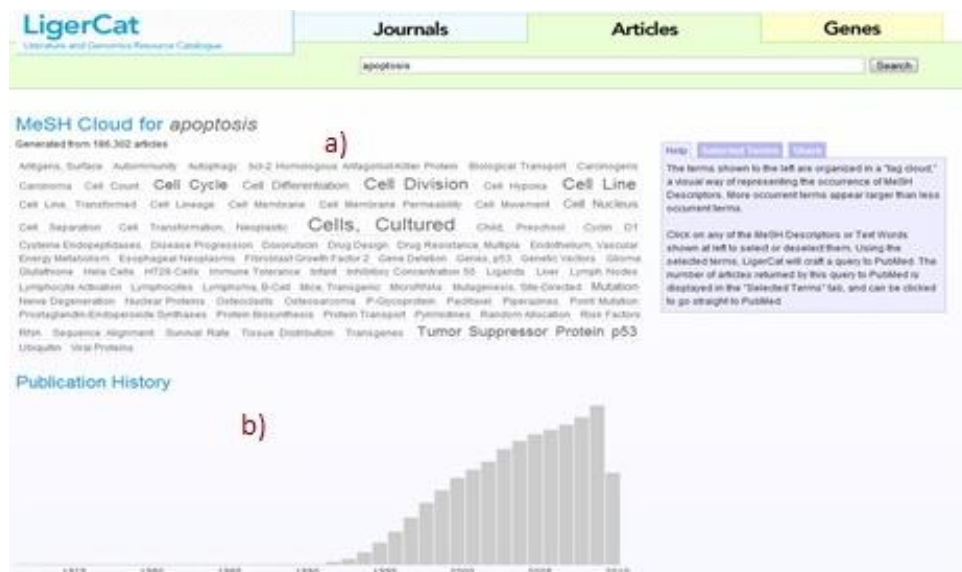


Figura 45. Interfaz Web de *LigerCat* con la búsqueda apoptosis.
a) nube de términos relevantes, b) Histograma con el número de artículos publicado por año.

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 de su trabajo actual (mediante el escaneo de las revistas más utilizadas de 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 un autor).

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

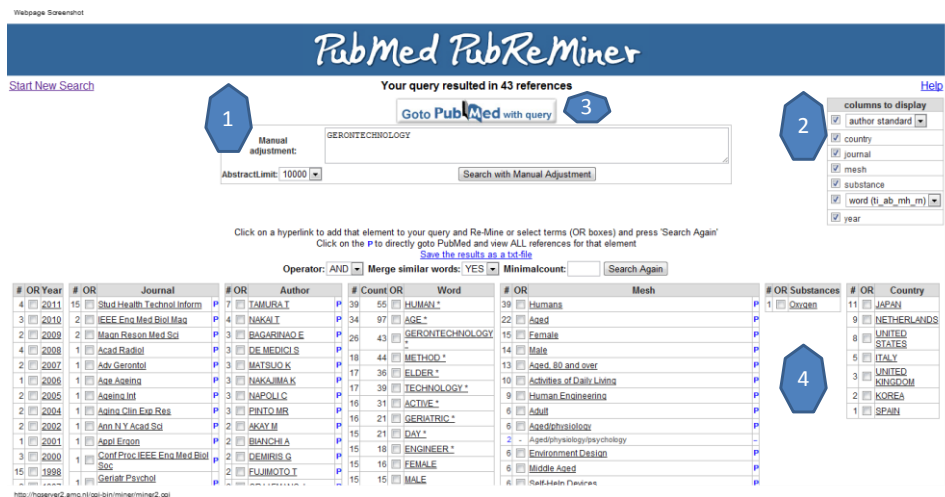


Figura 46. PubMed PubReminer (<http://bioinfo.amc.uva.nl/human-genetics/pubreminer/>), plataforma para análisis bibliométricos en el área biomédica. 1) consulta, 2) indicadores bibliométricos, 3) Ir a la consulta en PubMed y 4) análisis realizado para la consulta Gerontechnology (<http://www.gerontechnology.info/Journal/>).

El Laboratorio de Cibermetría (que pertenece al Centro de Ciencias Humanas y Sociales (CCHS) que es parte del mayor centro nacional de investigación de España, el CSIC) 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. 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 (<http://www.webometrics.info/>) de la Universidades del Mundo, creado y mantenido por el laboratorio de Cibermetría Español (Fig. 47).

Recientemente se ha desarrollado un nuevo método para estudiar las medidas sociales del impacto académico a través de las publicaciones con base en nuevos "indicadores sociales" midiendo el uso de *blogs*, plataformas de *microblogging*, sistemas de producción entre pares, 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 *University of North Carolina-Chapel Hill*, Dario Taraborelli de *Wikimedia Foundation*, Paul Groth de *VU University Amsterdam* y Cameron Neylon de *Science and Technology Facilities Council* y promete ser muy utilizado en los años futuros (Fig. 48).

Ranking Web de Universidades del Mundo

Julio 2011

inicio

países del mundo

ranking mundial

ranking por país

ranking europeo

ranking latino americano

> inicio > top Latino América

Datos del Ranking

Acerca de Nosotros

Acerca del Ranking

Top 12000 Universidades

Top USA y Canadá/Regional

Top Latino Americano

Top Europa

Top Europa Cent. y del Este

Top Asia

Top Sudeste Asiático

Top Subcontinente Indio

Top Mundo Árabe

Top Oceanía

Top África

Top Latino América

Universidades 1 a 100 de 100

RANKING CONTINENTAL	UNIVERSIDAD	PAÍS	RANKING MUNDIAL	POSICIÓN			
				TAMAÑO	VISIBILIDAD	FICHEROS RICOS	SCHOLAR
1	Universidade de São Paulo *		43	13	111	21	4
2	Universidad Nacional Autónoma de México *		49	23	108	38	10
3	Universidade Federal do Rio Grande do Sul		150	123	337	99	23
4	Universidade Estadual de Campinas *		158	103	341	68	79
5	Universidade Federal do Rio de Janeiro		170	190	286	108	86

Figura 47. Ranking Web de Universidades del Mundo en su Top Latino Americano (http://www.webometrics.info/top100_continent_es.asp?cont=latin_america).



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, Priem *et al.*, 2010).

2. 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 contactos 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, o entre familias 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 World Wide Web.

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 <http://www.ncbi.nlm.nih.gov/pubmed/> (ver capítulo 3) (Fig. 50) son:

Tamaño redes	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 los más relevantes entre tantos nodos y relaciones, sin perder las características del grafo principal.

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

- **Ali Baba:** (<http://alibaba.informatik.hu-berlin.de/>): Muestra la conexión de términos que aparecen en los documentos como células, fármacos, tejidos, enfermedades, reacciones, enzimas y compuestos; una vez que los ha identificado muestra los vínculos de los artículos hallados como interacciones proteína-proteína, localización de proteínas, nutrientes y genes.
- **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, dependiendo de los valores obtenidos, existe la posibilidad de re-graficar los resultados, generado de PubMed 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* y crea la red correspondiente. Los nodos representan autores, artículos, genes de *Gene Bank* o proteínas de *Swiss Prot*.

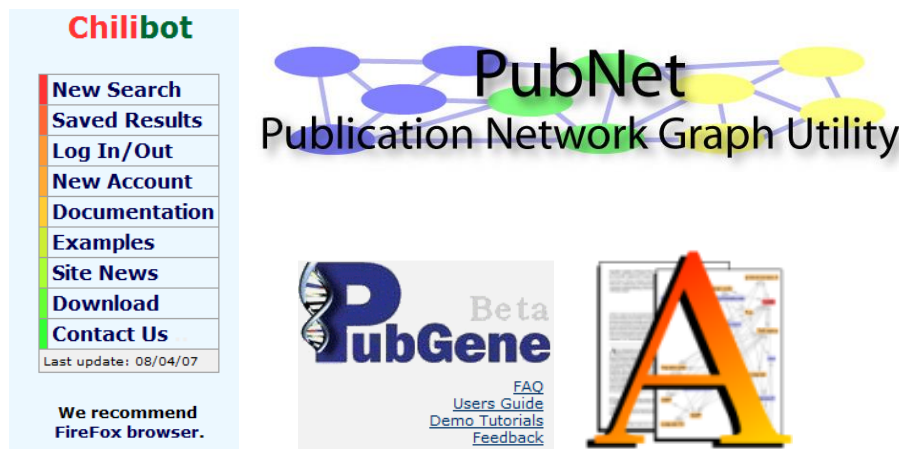


Figura 50. Cuatro aplicaciones para análisis de redes en el área biomédica. Chilibot <http://www.chilibot.net/>, Pubnet <http://www.pubnet.org/>, PubGene <http://www.pubgene.org/> y Ali Baba <http://alibaba.informatik.hu-berlin.de/>.

3. Minería de textos

La minería de textos es un proceso que se utiliza para la generación de nuevo conocimiento 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, y de acuerdo con esta representación, será la clase de patrones por descubrir. Posteriormente el método consiste en relacionar, dos o más conceptos dentro para descubrir patrones interesantes que ayuden a generar nuevos conocimientos.

Existen miles de aplicaciones para aplicar esta técnica a diversos conjuntos de información, de tal forma que se pueden categorizar según el tipo de datos que se analicen, por ejemplo, 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 solo en términos. Si el proceso anterior se realiza utilizando en lugar de términos el significado (también llamado semántica ver adelante) entonces se denomina minería de textos, 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. *The National Center for Text Mining* es un centro de investigación y desarrollo mantenido por la Universidad de Manchester en colaboración con la universidad de Tokio especializado en minería.



Figura 51. Distintos tipos de minería desarrollados en NaCTeM
<http://www.nactem.ac.uk/>, institución de mayor prestigio en este tipo de análisis.

4. 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, para esto se usan los formatos RDF y OWL factibles de leer automáticamente por máquinas inteligentes. El objetivo es mejorar el Internet, ampliando la interoperabilidad entre los

sistemas informáticos 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) que representa un proceso determinado en el que se expresa de manera explícita el significado de cada uno de los datos, las propiedades, los objetos y las relaciones que lo componen 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 y 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 *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://gopubmed.org/web/gopubmed/1?WEB10000h00100090000>, es un buscador semántico (Fig. 53) que realiza consultas con base en el significado, presenta al usuario un breve análisis bibliométrico por año, por autor o por país, 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 <http://www.ncbi.nlm.nih.gov/pubmed/>, que a través del uso de las ontologías procesa de manera lógica y semántica las consultas realizadas.

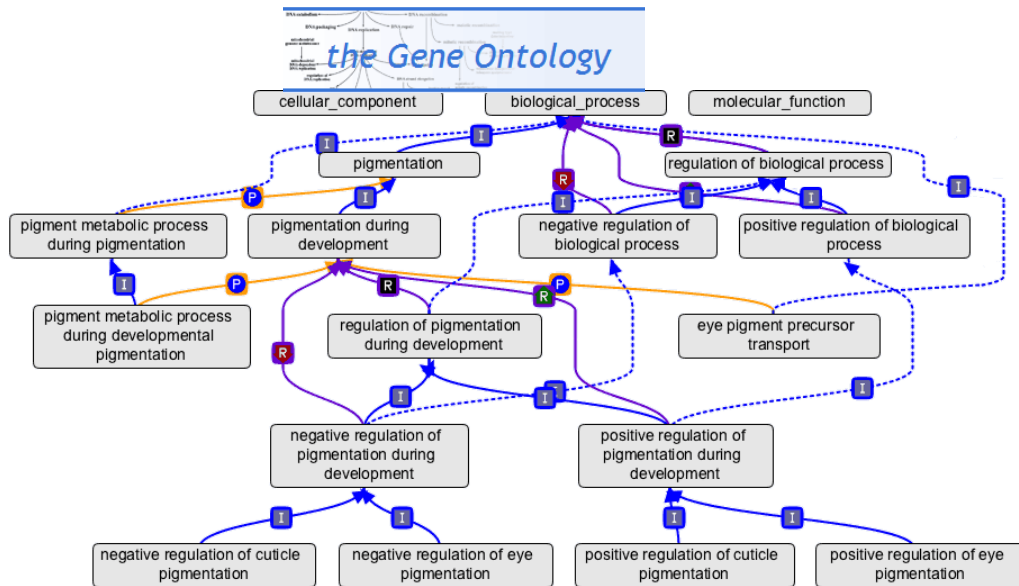


Figura 52. Proyecto *the Gene Ontology* (<http://www.geneontology.org/>) y la relación entre su base de datos, su vocabulario de términos controlados y su interfaz de análisis.

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 de *Journal of Biomedical Semantics* <http://www.jbiomedsem.com/>.

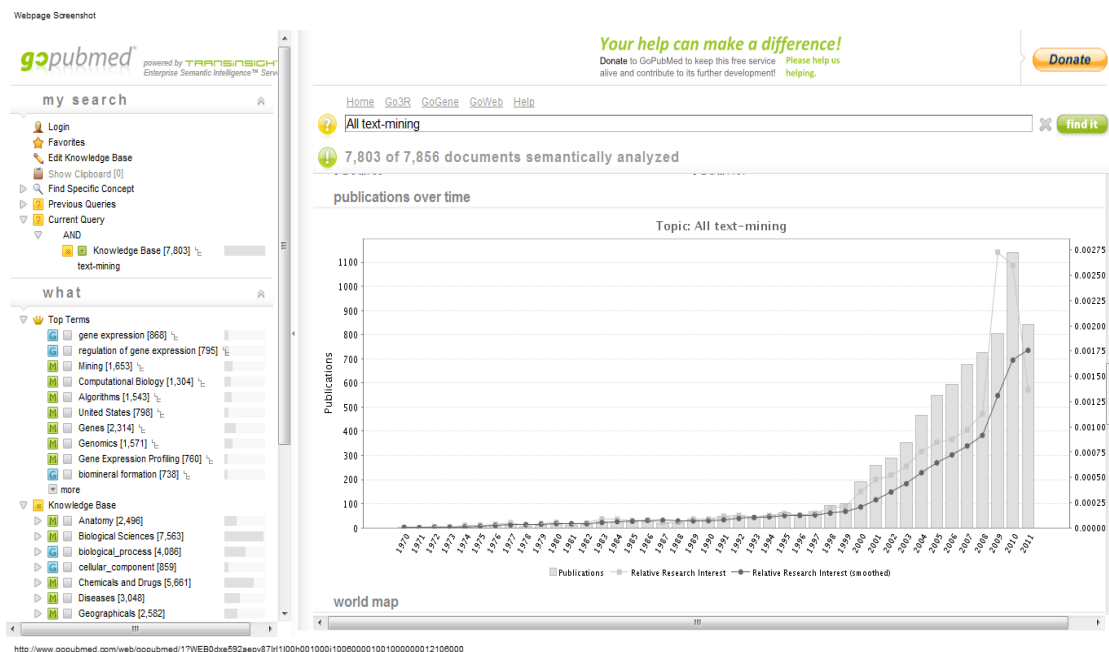


Figura 53. Interfaz Web de GOpubMed, presentando un histograma con los artículos por año procesados por minería de textos.

V. DISCUSIONES ACTUALES

En este capítulo se presentarán algunas 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 BBB” por las declaraciones producto de tres conferencias realizadas en las ciudades: Budapest, Berlín y Bethesda.

La Declaración de Budapest (BOAI, 2002) define al acceso abierto como: La disponibilidad gratuita de la información en Internet para que cualquier usuario pueda leer, descargar, copiar, distribuir e imprimir, buscar o citar toda esa información, exportarla para la selección exhaustiva, usarla como datos para software, o utilizarla para cualquiera otro propósito legal, sin barreras financieras, legales o técnicas, distintas de la idea fundamental de ganar acceso al propio Internet.

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

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, por medio del uso de licencias (ver más adelante) para copiarlo, usarlo, distribuirlo, transmitirlo y exhibirlo públicamente, para hacer y distribuir los trabajos derivativos, en cualquier medio digital con un propósito responsable, lo mismo que el derecho de efectuar copias impresas en pequeño número para su uso personal. Todo ello sujeto al reconocimiento apropiado de la autoría correspondiente.

2. El depósito en línea de una versión completa del trabajo y todos sus materiales complementarios, en un formato electrónico estándar, que incluya una copia de la licencia, y esté publicado, en al menos, un repositorio de una institución académica, sociedad erudita, agencia gubernamental, u organización establecida, que busque implementar el acceso abierto, distribución irrestricta, interoperabilidad y capacidad archivística a largo plazo.

La declaración de Bethesda (2003) define a la investigación científica y sus objetivos de la siguiente forma: La investigación científica es un proceso interdependiente, donde cada experimento es informado por el resultado de otros. Los científicos que hacen investigación y las sociedades profesionales que los representan, tienen gran interés en asegurarse que los resultados de sus investigaciones sean difundidas lo más inmediata, amplia y efectivamente posible. Las publicaciones electrónicas ofrecen la oportunidad y la obligación de compartir los resultados de investigación, ideas y descubrimientos libremente entre la comunidad científica y el público en general.

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 <http://www.ncbi.nlm.nih.gov/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.fcencias.unam.mx:8080/xmlui>) (Fig. 54).

Peter Súber 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), investigador en filosofía en el Earlham College, director del Open Access Project de Public Knowledge y autor de la [SPARC Open Access Newsletter](#).



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

2. Creative Commons

Se refiere al espectro de permisos que definen 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”.

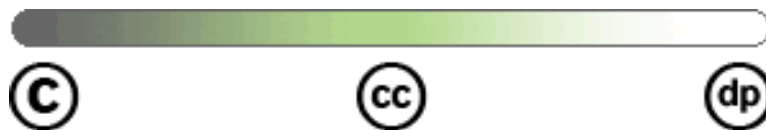


Figura 55. Esquema de las licencias Creative Commons: © Todos los derechos reservados (cc) Dominio público de Creative Commons y (dp) ningún derecho reservado.
Tomado y modificado de <http://creativecommons.org.mx/que/>.

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, 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.

3. Plagio

Uno de los principales problemas de la publicación científica en la era digital son: el plagio y la duplicación de artículos. Según el diccionario de la real academia española <http://www.rae.es/rae.html> (RAE) define Plagiar (Del lat. *plagiāre*) como copiar en lo sustancial obras ajenas, dándolas como propias.

La mayoría de los científicos se enfrentan 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. Con este objetivo se utiliza generalmente la base de datos libres más grande de esta área, PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). El ejemplo más interesante de estas herramientas es *Déjà vu 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 el documento de otro) y 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 la aplicación para detectar plagio en las publicaciones sobre el área biomédica.

4. Computo en la nube

La computación en nube (*cloud computing*) se refiere, tanto a las aplicaciones de la entrega de servicios a través de Internet, como al *hardware* y el *software* de los sistemas que proporcionan estos servicios. El *hardware* del centro de datos y el *software*, es 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 pondrá a disposición del público en general. Algunos ejemplos de aplicaciones en la nube son: [eyeOS](#), el cual 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. [Dropbox](#), 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 implicar un costo.

5. 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 tales 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 otros. Aún hay muy pocos metadatos de preservación con sus políticas o estrategias asociadas (estabilidad, integridad y autenticidad).

Iniciativas	Liga electrónica
<i>Metadatos para la descripción</i>	
DC: Dublin Core Metadata Initiative.	http://dublincore.org/
METS: Metadata Encoding and Transmission Standard.	http://www.loc.gov/standards/mets/

Iniciativas	Liga electrónica
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	
	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. (AGLS).	http://www.agls.gov.au/
Metadatos geoespaciales	
Content Standard for Digital Geospatial Metadata.	http://www.fgdc.gov/metadata/contstan.html
Metadatos generale	
W3C Metadata activity	http://www.w3.org/Metadata/Activity.html
W3C Semantic Web activity	http://www.w3.org/2001/sw/

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

6. Tipo de formatos de documentos electrónicos

El proceso de preservación digital no está todavía completamente entendido y valorado, ya que depende directamente del uso de la información sistematizada, en especial, para el caso de la biología, ya no es necesario solo a 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*) fue diseñado únicamente para mostrar los datos en una página Web (Fig. 57).

Un documento XML está formado por el prólogo y el cuerpo del documento conformado por etiquetas y su texto correspondiente, de tal forma que se puede elegir específicamente cuál 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, hace hincapié en la estructura de las relaciones entre los contenidos a través del uso de ligas entre URI's (Uniform Resource Identifier). Así es posible crear la descripción de una persona utilizando las ligas a los recursos Web correspondientes a través de la asignación de URI's al actor central (Fig. 58). 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.

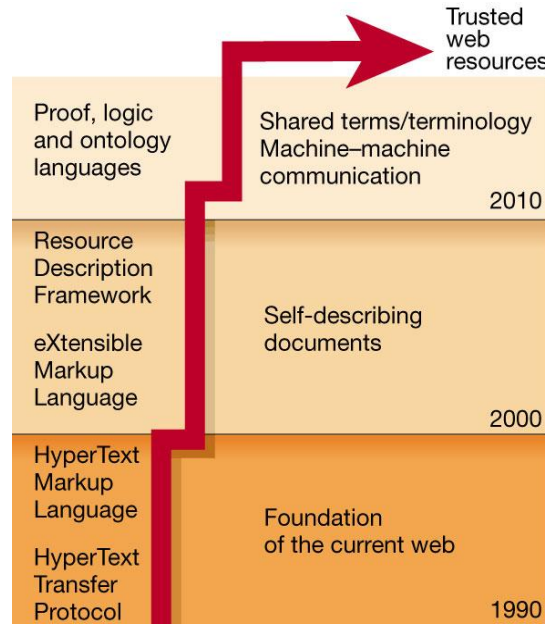


Figura 57. Evolución de los estándares Web. De: Berners-Lee T, Hendler J: *Publishing on the semantic web*. *Nature* 2001, 410:102 3-4.



Figura 58. Grafo RDF que describe algunas características de la persona “Erick Miller” a través de URIs.
 Tomado y modificado de <http://www.hipertexto.info/documentos/rdf.htm>.

Otro estándar usado para web semántica es el OWL (*Web Ontology Language*) o lenguaje de ontologías para la Web, 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 4). En realidad, OWL es una extensión del lenguaje RDF, aunque es un lenguaje con más poder expresivo que éste.

VI. GLOSARIO

Add-on: extensión o tema que realza 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. Aplicación 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, 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, ya sean geográficos, temáticos, institucionales e incluso individuales.

Bases de datos: Sistema informático de registros con el propósito de almacenar, mantener y generar información, la cual está sistematizada en tablas con un orden establecido. Programas que permiten guardar, ordenar, procesar y presentar los datos.

Bibliometría (*Bibliometrics*): Evaluación cuantitativa de la literatura, describiendo el impacto de la investigación y la institucionalización de la misma.

Blog (contracción del término *weblog*): Bitácora, es un tipo de sitio Web, pero las entradas se muestran en orden cronológico inverso. "Blog" puede también utilizarse como un verbo, lo que significa 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 son principalmente textuales, aunque algunos se centran en arte (*artlog*), fotografías (*fotoblog*), dibujos (*sketchblog*), vídeos (*vlog*), música (*MP3 blog*) o audio (*podcasts*), 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 140, es el caso de *twitter*).

blogger: en inglés, autor de un **weblog**.

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

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.

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

Chat (*mensagger*): Aplicación que permite la comunicación, fundamentalmente escrita, en tiempo real entre dos o más usuarios vía.

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 por los equipos interdisciplinarios de profesionales que son responsables de su desarrollo, divulgación y uso en propuestas transformadoras para el descubrimiento científico y la enseñanza.

Cibermetría (*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, 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 juntos 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 ya existentes ni complicar el desarrollo del programa principal.

Computación en nube (*Cloud computing*): Se refiere tanto a las aplicaciones como a los servicios a través de Internet, la información se almacena de manera permanente en servidores en Internet 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 pueden adjuntar archivos de datos, imágenes, vídeo o música. El correo electrónico es uno de los servicios más populares de Internet.

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 web que informen de su sitio para que sea incluido, y después los editores autorizados revisan esas solicitudes antes de incluir sus enlaces para comprobar que se adecúan a los requisitos de aceptación determinados por el directorio Web. 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 puesto que ponen al alcance del usuario la posibilidad de descubrir proveedores que desconocía que existieran y con esto resolver una necesidad de compra.

E-ciencia: Es el proceso de investigación realizado en el espacio virtual. Este proceso tiene las siguientes características: colaboración internacional entre los investigadores (Colaboratorios), la interconexión de computadoras, la aplicación de la arquitectura grid y el uso de reservorios de datos. La posibilidad de visualizar los datos comunes, desarrollo de herramientas y procedimientos basados en Internet, construcción de estructuras de organización virtual en donde se lleva a cabo la investigación, distribución y publicación electrónica de los resultados en línea.

Enlace de Hipertexto (*hypertext, link, hyperlink*): El enlace entre dos documentos de texto o de una marca de texto a una posición denotada.

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 es el de los portales Flickr para la gestión de imágenes o del.icio.us para la gestión de marcadores.

Etiquetar: Asignar categorías a través del uso de términos, las etiquetas son en consecuencia 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 a través de ordenar contenidos mediante etiquetas. Las categorías usadas no obedecen a una lógica jerárquica sino a las decisiones de etiquetado de los usuarios.

Gadgets: Es un dispositivo que tiene un propósito y una función específica, generalmente de pequeñas proporciones, práctico y a la vez novedoso.

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, sin ser necesaria para el funcionamiento del procesador, le aumenta 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 a 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*): EL HTML, 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, Mozilla, Firefox el HTML se ha convertido en uno de los formatos más populares que existen para la construcción de documentos.

Indización: 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 de 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 almacenan direcciones de páginas Web que el usuario encontró útiles y a las que se quiere acceder fácilmente. Conocidos también como favoritos.

Mensajería instantánea: Son un conjunto de programas que utilizan el protocolo TCP/IP que 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.

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. La técnica hace hincapié en los resultados de múltiples estudios en comparación con los resultados a partir de una sola investigación.

Metabuscador: Sitios web especializados en consultar varios motores de búsqueda simultáneamente que permiten 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 del 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 ejemplo de servicios de microblogging.

Minería de textos (*Text-mining*): La minería de texto es la más reciente área de investigación del procesamiento de textos. Ella se define como el proceso de descubrimiento de patrones interesantes y nuevos conocimientos en una colección de textos, es decir, la minería de texto 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. Estos son navegadores gráficos, lo cual quiere decir que pueden mostrar gráficos y texto. La mayoría de navegadores actuales pueden presentar información multimedia, incluyendo sonido y vídeo.

Obsolescencia (*Obsolescence*): Se refiere a los documentos que están decayendo en desuso.

Ontología (*Ontology*): Una ontología es una base de datos que describe los conceptos generales o sobre un dominio, algunas de sus propiedades y cómo los conceptos 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, al que se accede mediante enlaces de hipertexto.

PDF: acrónimo del inglés *Portable Document Format*, (formato de documento portátil) es un formato de almacenamiento de documentos, desarrollado por la empresa Adobe Systems. Este formato 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.

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

El concepto de portal de internet básicamente define a un punto de entrada común a una colección de recursos electrónicos integrados, donde se ofrecen un serie de servicios complementarios, tales como búsqueda interna, personalización, herramientas de comunicación, servicios de información u otros específicos asociados al tipo de portal generales, temáticos ya sea públicos o privados.

Preservación digital (*digital preservation*): 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 pese a los rápidos cambios tecnológicos.

Plagio (*Plagiarism*): Copiar en lo sustancial obras ajenas, dándolas como propias.

Plug-in: Ver complemento.

Ranking: Es una relación entre un conjunto de elementos tales que, para uno o varios criterios, el primero de ellos presenta un valor superior al segundo, este a su vez 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.

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.

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

Sistema de votación (*Ranking*): Clasificación. En el entorno de Internet, se dice del orden de presentación de los resultados de una búsqueda con un programa WAIS (*Wide Area Information Server*). Este orden se establece según el grado de probabilidad de mayor a menor.

Sitios web (*Website*): Conjunto de páginas web que comparten un mismo tema e intención y que generalmente se encuentra en un sólo servidor. 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.

Software: Programa o programas de computadora, en contraste con el equipo físico en que se ejecutan estos.

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.

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

URL: acrónimo de *Uniform Resource Locator*, dirección de una página web. Dirección global de documentos y de otros recursos en la World Wide 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 vía el 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 enfocadas al usuario final.

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 ampliando la interoperabilidad entre los sistemas informáticos y reducir la necesaria mediación de operadores humanos.

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 los de 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 a 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.

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.

XML: Estándar para transportar y almacenar datos.

VII. LITERATURA RECOMENDADA

- Alierta, I. C. (2008). *La Sociedad de la Información en España 2008*, España: Fundación Telefónica, Editorial Ariel. 529pp.
- Berners-Lee, T. (1996). En: Simplicity, Standards, and Intercreativity. The W3C Team World Wide Web consortium. Journal 3. URL <http://rugmd4.chem.rug.nl/hoesel/tbl-int.html>
- 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>
- Cobo, R. C. y Pardo, K. H. (2007). *Planeta web 2.0, Inteligencia Colectiva o Medios Fast Food*. Version 1. Barcelona / México DF.: Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flasco México.
- Giménez, L. M. y Tramullas, S. J. (2007). *Evaluación de software libre para la gestión de bibliografía*. En: *IX Jornadas Españolas de Documentación*. Santiago de Compostela, España.
- 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>
- Lyman, P. y Hal, V. (2003). *How much information*, Berkeley, California. URL <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/index.htm>
- Macías, L., y Michán, L. (2009). Los recursos de la web 2.0 para el manejo de información académica. *Fuente*, 1 (1), 18-27. URL http://fuente.uan.edu.mx/publicaciones/01-01/los_recursos_de_la_web_2.0_para_el_manejo_de_informacion_academica.pdf
- 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.fcencias.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.
- NSF. (2007). *Cyberinfrastructure vision for 21st century discovery*. Arlington, VA: National Science Foundation. 57pp.
- 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.
- 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>
- Specia, L. y Motta, E. (2007). *Integrating Folksonomies with the Semantic Web*. In The Semantic Web: Research and Applications, 639, 624.
- Truell, A. D. (2003). Use of Internet Tools for Survey Research. *Information Technology, Learning, and Performance Journal*, 21(1), 31-37.

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

AUTORES

Layla Michán
Jack Guillén
Eduardo Alvarez
Lyssania Macías
Itzel Pedraza

[<http://www.fciencias.unam.mx/xxxx/>]
Universidad Nacional Autónoma de México
Facultad de Ciencias
Primera edición
México, 2012



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