

# Una visión al mundo de la búsqueda y recuperación de la información electrónica

**Andrés Felipe Echavarría Ramírez**

Bibliotecólogo, Biblioteca J. Ernesto Vélez Escobar. Corporación Universitaria Lasallista

Correspondencia: Andrés Felipe Echavarría Ramírez. email: anechavarria@lasallista.edu.co

## A glimpse to the world of search and recovery of electronic information

### Resumen

El creciente número de sitios y recursos en Internet ha dificultado el proceso de búsqueda y recuperación de información útil para la formación e investigación. Aunque la Red cuenta con importantísimas fuentes, es desconocida por los motores de búsqueda tradicionales, a esta parte se le llama Red Profunda. Para acceder a este pequeño universo de Internet es necesario conocer los mecanismos, estrategias y herramientas que faciliten y garanticen el logro de nuestros objetivos.

**Palabras Clave:** Internet. Red Profunda. Red Superficial. Bases de datos. Bibliotecas electrónicas. Buscadores. Estrategias de búsqueda.

### Abstract

The increasing number of web sites and resources provided by internet has made the process of search and recovery of useful information for schooling and research more difficult. Even though the Net has extremely important sources, these are unknown for the majority of traditional search engines. This is what it is known as "The Deep Net". To access this small universe of internet, it is necessary to know the mechanisms, strategies and tools that facilitates and guarantee the achievement of our objectives.

**Key words:** Internet. Deep net. Shallow net. Data bases, Electronic libraries. Searchers. Search strategies.

### Introducción

Internet es considerada la biblioteca por excelencia y con mayor número de consultas del mundo moderno. Sólo se necesita mover el puntero por la pantalla, hacer clic y listo, ya estamos en la Red de redes. Pero el estar en ella no nos garantiza el que encontremos lo que se necesita, incluso en la mayoría de los casos y después de pasar más de hora y media frente al computador, no se consigue el resultado deseado y en contraposición a éste, sólo se logra comerciales de juegos en línea, pantallas con publicidad pornográfica y hasta invitaciones para entrar en comunidades de buscadores de pareja.

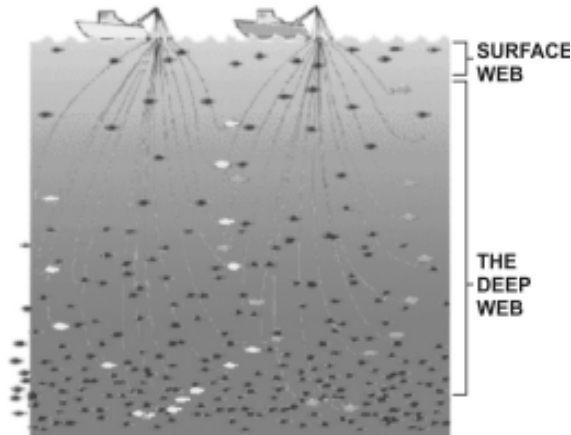
Buscar información en Internet puede convertirse en el momento más difícil y frustrante del investigador, ya que cuenta con millones de recursos y herramientas como ellas, los motores de búsqueda. El problema con los motores de búsqueda es que no solo hay que saber como usar-

los para sacarle el máximo provecho, sino porque estos apenas indexan una mínima cantidad de lo que Internet pueda ofrecer.

¿Entonces dónde se puede encontrar el resto de la información?. El resto de la información está y hace parte de lo que en 1994, la doctora Jill Ellsworth llamó «invisible Web» (Internet invisible), para referirse a la información que los motores de búsqueda tradicionales no pueden encontrar. Algunos autores prefieren llamarla "deep web" (Red profunda), ya que en realidad no es invisible, sólo hay que saber llegar a ella<sup>1</sup>.

Un estudio más reciente con datos recogidos en marzo del 2000 por la empresa BrightPlanet, dice que la información pública disponible en "deep web" es de 400 a 500 veces mayor que en el que comúnmente llamamos World Wide Web (www), y que la web profunda contiene 7500 terabytes de información comparado con 19 terabytes de la web superficial<sup>2</sup>, situación que comúnmente es

comparado con la analogía del iceberg, océano y banco de peces (ver gráfica 1). La parte superficial del iceberg es comparada con la Red superficial, y la gigantesca parte escondida es denominada Red Profunda.



**Gráfica1. Surface Web vs. Deep Web<sup>2</sup>**

### Contenido de la red profunda

La Red Profunda está conformada por gran parte de información pública, pero en su gran mayoría está elaborada con protocolos no indexables por los motores de búsqueda de la Web superficial, algunos de ellos son: Ftp (File Transfer Protocol), e-mail, grupos de noticias, Internet Relay Chat (IRC) ...

Según Isidro F. Aguillo, atendiendo a criterios documentales y de ciencias de la información, podríamos agrupar en la Red profunda, los siguientes documentos:

- Catálogos de bibliotecas y bases de datos bibliográficas nacionales e internacionales.
- Bases de datos no bibliográficas (Dialog, STN, Lexis-Nexis entre otras).
- Recursos de apoyo para la investigación y documentos en formatos no indexables, como documentos en pdf, word, avi, mp3 entre otros.
- Revistas electrónicas, en las que es necesario un registro previo y las que sólo se puede recu-

perar la información mediante búsquedas en su base de datos.

- Directorios de información mundial: por ejemplo periódicos o páginas especializadas.
- Obras de referencia: enciclopedias, diccionarios en las que es necesario interrogar a la base de datos para acceder al contenido.
- Información gubernamental, bursátil, comercial y educativa general<sup>3</sup>.

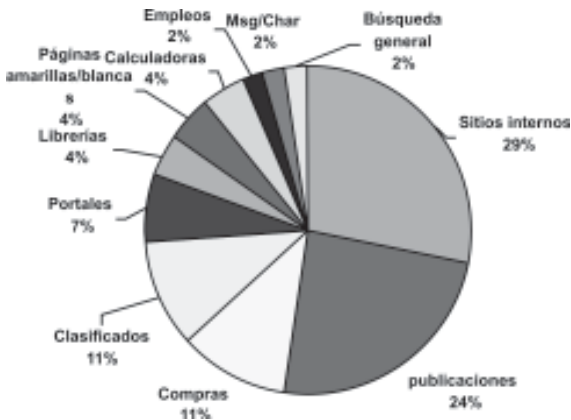
### Industria de la información electrónica

Ante el surgimiento y proliferación de las publicaciones electrónicas, aparecieron varios protagonistas en el mercado nacional e internacional, junto con una industria de la información electrónica; la cual todavía se encuentra en evolución y no es tan estable como otras industrias que tienen décadas de existencia.

Ésta es una industria altamente dinámica y, hasta cierto punto, difícil de visualizar. Sin embargo, es una industria con alto potencial y que atrae cada vez más empresas de otras industrias, ya que se está descubriendo el alto potencial de la misma.

Entre los participantes de esta industria contamos con proveedores de importantes bases de datos, así como bibliotecas electrónicas y portales de organismos científicos, pero todos bajo la siguiente estructura.

**Sistemas en línea:** son sistemas de bases de datos que permiten recuperar información de las mismas, con un lenguaje común de recuperación de información y a distancia, conformados por sistemas profesionales orientados a hombres de negocios y especialistas en información o investigadores, sumando un promedio aproximado del 54% del total de la Información en la Red Profunda<sup>4</sup>. Algunos, los más comunes en el medio por su antigüedad e importancia son: Dialog, Dialog Corporation, Questel-Orbit, STN, OVID, Lexis-Nexis, Dow Jones Interactive, Gale, Wilson entre otros. (Gráfica 2).



**Gráfica 2. Contenido de la Red Profunda <sup>2</sup>**

**Sistemas en CD-ROM:** es una variación de los sistemas en línea a un entorno local o Intranet y en su gran mayoría son los mismos proveedores, aunque existen otras empresas dedicadas sólo a este sector. Algunos son: SilverPlatter, DIALOG OnDisc, Bell & Howell Information and Learning (UMI), Gale, Wilson, entre otros.

**Publicaciones y bibliotecas electrónicas:** es la versión más demandada y en algunos de los casos más costosa de la Red Profunda, ya que entre sus elementos de trabajo cuentan con los acervos más grandes de información electrónica en texto completo y aunque existen empresas dedicadas a este negocio como, (British Library Document Supply Centre, UnCover – Ingenta, CISTI, OCLC, EBSCO, AIP), existen alternativas mundiales dedicadas a la distribución y recuperación gratuita de información académica y científica<sup>5</sup>.

## Estrategias de búsqueda de información electrónica

No basta con saber cuales son ni que recursos tiene la Red Profunda para nuestro saber y así lograr el éxito en la búsqueda de información, un ejemplo de ello son los motores de búsqueda, los cuales han mejorado sus herramientas y sus máquinas para indexar y recuperar más y diferentes formatos de información, y aun así nos arrojan miles de resultados o en algunos casos la misma cantidad de siempre con elementos en su mayoría mediocres.

El éxito de una consulta depende de la planeación y organización previa, es como elaborar un plan

de trabajo con sus respectivos planes de contingencia y correctivos. Pensemos en una consulta en Internet como si de una empresa se tratara y si vemos de cerca esta analogía, nos daríamos cuenta que si la empresa es dirigida de forma improvisada, ya que tarde o temprano caerá en el fracaso o en el peor de los casos en la quiebra. La improvisación sin planeación es la principal fuente de frustraciones en Internet. Sólo basta con tomarnos 15 minutos con un papel en la mano y elaborar el siguiente plan:

**Entrevista:** es una reflexión inicial, pero justa para reconocer la necesidad real de información y, al mismo tiempo, analizar las estrategias pasadas. Con ella buscamos establecer el objetivo y las metas a lograr. Para su correcto desempeño el investigador debe hacerse las siguientes preguntas:

¿Requiero datos sobre?, ¿Dónde he buscado?, ¿Qué encontré?, ¿Me sirve?, ¿Cómo me sirve?, ¿Por qué me sirve?, ¿Qué más me falta?, ¿Para que lo necesito?, ¿Qué beneficio espero recibir de lo que encuentre?, ¿Para cuándo lo quiero?, ¿Cómo lo quiero? y ¿Cuánto estoy dispuesto a pagar? <sup>6</sup>.

**Tema:** a partir de la definición y respuesta a todos los interrogantes de la entrevista, es necesario visualizar el tema a tratar, ya sea por medio de una frase o una palabra. Lo que se busca es dejar claro cuál es la meta a lograr y que mejor para ello que un enunciado que me permita jugar con terminologías específicas o palabras claves de mi interés.

**Tema general y Subtema:** para lograr claridad en la consulta es necesario esclarecer las ideas y analizar nuestro objetivo, por medio de una estructura jerárquica de nuestro pensamiento, y para ello debo definir el tema general y subtema de nuestro tema de consulta. Ejemplo:

o **Tema general:** Contaminación

o **Tema:** Perturbaciones ambientales de origen Humano

o **Subtema:** Efecto Invernadero  
Lluvia ácida

**Otros temas:** en este punto debemos ampliar nuestras fronteras y revisar temas que puedan complementar o mejorar nuestro tema de consulta.

**Términos de Consulta:** en este paso se busca elegir todas las palabras y posibles combinaciones que permitan hallar la información deseada y para ello es necesario recurrir a sinónimos, siglas, terminologías específicas y lo más importante, términos en un segundo idioma.

Una vez que hemos elegido los términos adecuados es el momento de redactar nuestra estrategia de búsqueda. Para ello podemos combinar los términos elegidos con diferentes operadores y/o realizar los truncamientos que nos parezcan oportunos<sup>7</sup>.

Entre ellos tenemos, los operadores boléanos, que son un tipo de operadores lógicos que se utilizan en la elaboración de una estrategia de búsqueda y que nos permiten combinar diferentes términos de búsqueda entre ellos<sup>8</sup>. Los más conocidos son, y (AND), o (OR), no (NOT).

Otro tipo de operadores son los operadores de proximidad, adyacencia y contexto. Estos operadores nos permiten indicar en el sistema de búsqueda la

distancia que puede haber entre un término y otro. Entre los más comunes destacan los siguientes, NEAR (NEARn), PARRAFO (.P), ADJ, ADJn,<sup>9</sup>.

**Fuente:** el gran número de bases de datos, motores de búsqueda y herramientas para la consulta hace este proceso más complejo de lo pensado, pero sólo hace falta elegir nuestra lista de herramientas y organizarlas para futuras ocasiones, y en esto podemos ayudarnos de los famosos Bookmark's o Favoritos de la barra de herramientas de nuestro navegador.

Antes de realizar la consulta se debe elegir cuál es su fuente de mayor utilidad para el logro de nuestro objetivo, sea esta un libro, artículo de revista, monografía, patente o informe técnico, lo que en documentación es llamado fuente primaria. O una bibliografía, resumen o recopilación, también llamadas fuentes secundarias.

A continuación listado de Bases de datos y motores de búsqueda especializados en la Red Profunda:

**Tabla No. 1 Bibliotecas electrónicas y Bases de Datos**

Por Suscripción	
Sitio	URL
Dialog	<a href="http://www.dialog.com/">http://www.dialog.com/</a>
British Library	<a href="http://www.bl.uk/">http://www.bl.uk/</a>
CISTI	<a href="http://cat.cisti.nrc.ca/search/">http://cat.cisti.nrc.ca/search/</a>
EBSCO (Pregunte en la Biblioteca J. Ernesto Vélez por este servicio)	<a href="http://trial.epnet.com">http://trial.epnet.com</a>
GALE	<a href="http://infotrac.galegroup.com/itweb/lasallista">http://infotrac.galegroup.com/itweb/lasallista</a>
Science Direct	<a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a>
Gratuitas	
Sitio	URL
Highwire	<a href="http://highwire.stanford.edu/">http://highwire.stanford.edu/</a>
Scielo	<a href="http://www.scielo.org/">http://www.scielo.org/</a>
DOAJ	<a href="http://www.doaj.org/">http://www.doaj.org/</a>
CompletePlanet	<a href="http://www.completeplanet.com/">http://www.completeplanet.com/</a>
Profusion	<a href="http://www.profusion.com/">http://www.profusion.com/</a>
The Invisible Web Directory	<a href="http://www.invisible-web.net/">http://www.invisible-web.net/</a>
Librarians' Index to the Internet	<a href="http://lii.org/">http://lii.org/</a>
FindLaw	<a href="http://www.findlaw.com/">http://www.findlaw.com/</a>
LookSmart's FindArticles	<a href="http://www.findarticles.com/">http://www.findarticles.com/</a>
Internet Invisible	<a href="http://www.internetinvisible.com/">http://www.internetinvisible.com/</a>

**Idioma:** un segundo idioma es una realidad y una condición tangible en Internet. El predominio del inglés en las áreas de ciencia y tecnología se debe a varios factores de tipo socioeconómico, que se traducen en un enorme potencial de recursos económicos y humanos, en la investigación y desarrollo de países como Estados Unidos, Canadá y Reino Unido, etcétera <sup>10</sup>.

Las enormes cifras de producción científica en estos países anglófonos, así como en otros de alto nivel de desarrollo que han adoptado el inglés como lengua científica, ha dado lugar a una presencia casi universal del inglés en los medios especializados de comunicación <sup>11</sup>.

**Fecha:** elegir el periodo cronológico es importantísimo para establecer parámetros de consulta y mejorar los filtros y resultados.

Finalmente y no menos importante, es la ejecución de la búsqueda, donde podemos poner en práctica cada uno de estos parámetros, pero es necesario recordar que si no cuenta con varias alternativas una consulta que puede durar diez o quince minutos se puede tornar en dos horas o más y no arrojar los resultados esperados.

## Conclusiones

A pesar de la rapidez y de la gran variedad de servicios que nos ofrece Internet, el usuario común invierte horas o días enteros intentado encontrar recursos en la Web, pero sin lograr el objetivo deseado. Esto es debido a la falta de planeación y desconocimiento de las herramientas existentes.

Internet es un océano de información, que actualmente es clasificado en dos partes por los teóricos más importantes en el mundo académico, Red Superficial, que es la zona de la red con información comercial y de poco interés académico, y Red Profunda que contiene toda la información invisible o no indexable para los motores de búsqueda tradicionales, en su mayoría es información académico-científica, de negocios y gubernamental.

La planeación es un paso fundamental y una garantía para el éxito de una consulta, y para llevarla a su ejecución es necesario separar diez o quince minutos con papel y lápiz en mano, antes de

sentarnos frente al computador. Este proceso facilita un guión de trabajo y realizar planes de contingencia, que nos permita cambiar de dirección cuando estemos perdidos en la Red sin perder más tiempo que el usado para digitar una combinación de palabras.

Aunque una estrategia de búsqueda es una garantía para el éxito en la búsqueda de información, no es absolutamente infalible, se debe realizar una evaluación conciente de la calidad, veracidad, actualidad y precisión de la información.

La Biblioteca J. Ernesto Vélez Escobar de la Corporación Universitaria Lasallista, ha integrado entre sus herramientas y recursos como apoyo a la actividad formadora-investigativa, bases de datos y recursos de información electrónica nacionales e internacionales y muy pronto contará con la Base de Datos EBSCO, para la cual se programarán charlas, y cursos de capacitación, dirigidos a todos los públicos de la institución. Recuerde consultar en la Biblioteca por estos cursos y recursos.

## Referencias

1. ALFONSO, Enrique et al. Encontrar lo que se busca: sumergiéndose en las profundidades [on line]. s.l. : redcom.org, s.f. [Citado 14 de noviembre de 2004]. Disponible en [http://www.redcom.org/tp/parcial/g03/g03\\_parcial\\_base.htm](http://www.redcom.org/tp/parcial/g03/g03_parcial_base.htm)
2. BERGMAN, Michael K. The Deep Web: Surfacing Hidden Value [on line]. s.l. : BrightPlanet, 2001. p.1. [Citado el 14 de noviembre de 2004]. Disponible en <http://beta.brightplanet.com/deepcontent/tutorials/DeepWeb/index.asp>
3. AGUILLO, Isidro F. Internet invisible o Infranet: Definición, clasificación y evaluación. En : JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN. (7 : 2000 : Bilabao). Ponencias VII Jornadas Españolas de Documentación. Bilbao: Universidad del País Vasco, 2000. 249-269.
4. BERGMAN. Op. Cit., p.10. [Citado el 14 de noviembre de 2004]. Disponible en <http://beta.brightplanet.com/deepcontent/tutorials/DeepWeb/index.asp>

5. TURNBULL M., Federico. Modulo de capacitación: Búsqueda de información. México : BANCOMEXT, 2002. Modulo 2.
6. CORTEZ, Jacqueline. La estrategia de búsqueda. México : Infoestratégica, 2002. p.2.
7. EL PROCESO de búsqueda y recuperación de la información [on line]. País Vasco : Universidad del País Vasco, s.f. p.3. [Citado 10 noviembre de 2004] Disponible en <http://www.sc.ehu.es/seweb/webcentro/cas/publica/numeros/n11/5.pdf>
8. Ibid, p.3.
9. Ibid, p.4.
10. CENTRO VIRTUAL CERVANTES. Anuario 1999: Conclusiones y comentarios [on line]. España : Instituto Cervantes, 1999. [Citado 10 noviembre de 2004] Disponible en [http://cvc.cervantes.es/obref/anuario/anuario\\_99/cindoc/p07.htm](http://cvc.cervantes.es/obref/anuario/anuario_99/cindoc/p07.htm)
11. Ibid, [Citado 10 noviembre de 2004] Disponible en [http://cvc.cervantes.es/obref/anuario/anuario\\_99/cindoc/p07.htm](http://cvc.cervantes.es/obref/anuario/anuario_99/cindoc/p07.htm)