

*Cátedra Taller en Tecnologías en Comunicación Social  
La Plata, Argentina*

Documento de la Cátedra. 2005

## ***El web profundo o Deep Web***

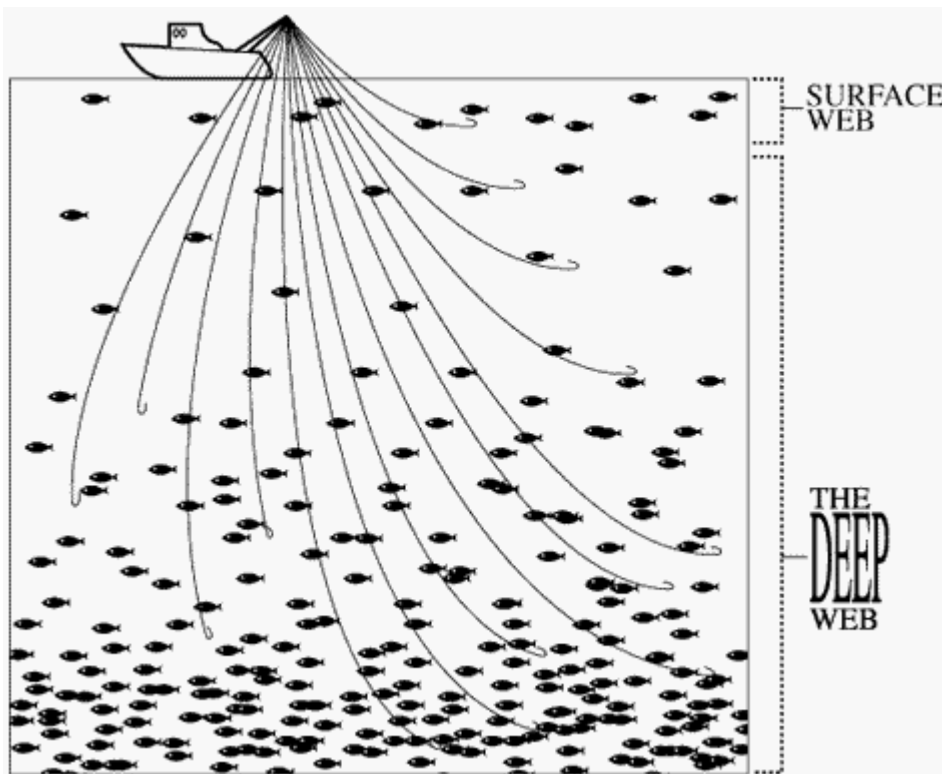
Autor: Mela Bosch. Información ha sido extraída,  
traducida y reelaborada por Mela Bosch a partir de fuentes como : Laura Cohen:  
lcohen@uamail.albany.edu; Mauro Lupi: mlupi@admaiora.com



**Palabras Clave:** TECNOLOGÍAS EN COMUNICACIÓN SOCIAL / WEB PROFUNDO

### **¿Qué es el web profundo?**

La empresa BrightPlanet (<http://www.brightplanet.com/>) inventó el término "deep Web" en contraposición al "surface Web."



El Web es una entidad compleja que contiene información de diferentes tipos. Comprende mucho más que las páginas estáticas, la parte del web no fija, la que está soportada en servers, es decir lo que se suele denominar web dinámico. Es en realidad la

parte más grande del web. Algunos llaman a esta parte Web invisible, web oculto, o deep web.

Esta definición está en los buscadores, o mejor dicho en lo que los buscadores no pueden hacer ya que entran en esta categoría materiales que no se acceden directamente y que los buscadores no pueden atrapar con los métodos usuales, porque no son páginas editadas en HTML.

Comprende:

- a. El contenido de bases de datos accesibles a través de web: se trata de información estructurada en tablas de datos creados y manejadas con programas como: Access, Oracle, SQL Server, MySql. Esta información sólo puede ser presentada si se la requiere por medio de una consulta, un query. Para poder realizarlo se debe realizar un ingreso, login, a un área especial del sitio, a veces gratuito, a veces pago. Se ha estimado que el contenido del web de databases es 500 veces mayor que el del web estático. Se trata de bases de datos de empresas, organismos, instituciones, y pueden tener la forma de bases de datos de apoyo a la gestión, de catálogos para clientes e incluso de bases de datos bibliográficas especializadas en temáticas particulares: médicas, negocios, espaciales, y incluso bibliotecas virtuales de universidades y centros de estudios. Se dice que esta es información invisible, oculta o profunda porque los buscadores no pueden entrar en ellas para extraer los datos
- b. Forma parte también del web profundo el conjunto de los archivos llamados No textuales, es decir archivos multimedia, gráficos, software, y documentos en Portable Document Format (PDF). Algunos de estos sí son accesibles a través de los buscadores, siempre y cuando tengan la metadata que permite localizarlos.

Algunas cifras para comprender las dimensiones:

<b>Fuente</b>	<b>Dimensiones del archivo</b>
National Climatic Data Cnter	366.000 Gb
NASA	219.000 Gb
National Oceanographic Data Center	32.940 Gb
MP3.com	4.300 Gb
Amazon	461 Gb
Library Of Congress Online Catalog	116 Gb

## ¿Qué hay en el web profundo?

El web profundo es una realidad que no podemos ignorar.

Lo más importante es tener presente que información incluye, de manera que cuando buscamos algo de esto deberemos plantearnos otra estrategia de acceso:

- Guías y listas de teléfonos, e-mail, y todo tipo de directorios
- "people finders" es decir listas de profesionales de todas la disciplinas
- Leyes, decretos, en general información legal, aunque alguna se puede encontrar en web estático
- patentes

- Diccionarios, aunque muchos glosarios se encuentran disponibles, y no nos olvidemos que la Wikipedia tiene un poco de todo
- productos de venta a través de e-commerce
- archivos multimedia y gráfico que no tengan la metadata de palabras claves que la identifique en forma clara.
- Sitios de empresas
- Publicaciones digitales de libros y diarios
- Páginas amarillas o blancas (Yellow/White pages)
- Bibliotecas
- Librerías

Hay otra información que también forma parte del web profundo y que puede estar disponible en forma estática, pero el problema es que se mueve velozmente, es decir que cambia y se modifica a veces sin dejar rastros:

- Noticias en general, pues a veces no se dejan los históricos.
- Ofertas y puestos de trabajo
- Vuelos, ofertas de viaje, de alojamiento
- Datos financieros, de bonos, y mercados

## ¿Como acceder al web profundo?

Se puede recurrir a directorios temáticos, que nos den enlaces a los sitios donde se encuentran las bases de datos, que a veces tienen partes gratis o accesibles si se insertan algunos datos. Por ejemplo para trabajar con noticias existe:

<http://www.newsnow.co.uk/>

También a veces los sitios que se exploran con bases de datos pueden ser recuperables por medio de los buscadores: por ejemplo realizando una búsqueda en Google con las palabras claves "american newspapers" Newslink, (<http://newslink.org/>) que es una base de datos de enlaces a sitios de en todo el mundo.

Asimismo, existen otros servicios de recolección de enlaces a bases de datos que pueden servirnos:

CompletePlanet: <http://aip.completeplanet.com/>

Ofrece acceso a muchísimas bases de datos e incluye sumarios de lo que se puede encontrar.

Direct Search: <http://www.freepint.com/gary/direct.htm>

Es una gran compilación de enlaces a interfaces de una gran variedad de recursos web  
Invisible-web.net: <http://www.invisible-web.net/> Es un directorio muy bueno sobre bases de datos

ProFusion: <http://www.profusion.com/> es un meta buscador que ofrece búsquedas ten categorías temáticas

ComplePlanet: <http://www.completeplanet.com/> trae la referencia a miles de sitios la mayor parte invisible a los motores de búsqueda.

Finalmente, no debemos olvidar que los buscadores a veces tienen diferenciados los paquetes de tipos de información, Google separa por imágenes, web, etc. en tanto

AltaVista permite hacer búsquedas separadas por news, maps, jobs, auctions, items de compra.  
Acceder por imágenes u otras estrategias a veces lleva a encontrar otra información no directamente accesible.