Surface Web vs. Deep Web

Univ. Bryan Enrique Miranda Palacios
Universidad Mayor de San Andrés
Carrera de Informática
Análisis y Diseño de Sistemas de Información
killerbmp@hotmail.com

RESUMEN

En el presente artículo se da a conocer algunas de las diferencias existentes entre la web superficial (surface web) y la web profunda (deep web), al igual que sus ventajas y desventajas de ambas; y por último se hace mención al mal uso que le dan algunas personas a la deep web.

Palabras Clave

Deep Web, Surface Web, Web Superficial, Web Profunda.

1. INTRODUCCIÓN

Lo que comúnmente llamamos la web en realidad es solo la superficie. Debajo de ella hay un vasto océano, en su mayoría inexplorado llamado deep web [1].

Cuando un internauta navega por la web buscando información en realidad solo lo hace por la superficie, si lograra sumergirse encontraría un número incommensurable de páginas que la mayoría de la gente nunca ha visto que incluyen de todo.

La deep web es, por tanto, un recurso que deben tener muy en cuenta las personas que buscan en internet algo más que generalidades. El hecho de que Google o Yahoo! no encuentren una consulta no tiene por qué significar que la respuesta no esté en internet. Tan solo hay que bucear algo más para encontrar este mar de información muchas veces desconocido por los usuarios que se limitan al uso de buscadores comunes [2].

Los buscadores rastrean la red con programas denominados “arañas”, indizando el contenido de las páginas y saltando de una a otra siguiendo los enlaces que contienen. Sin embargo estos programas no pueden indizar o extraer el contenido de esas páginas cuando esta información se encuentra en bases de datos, cuya información aparece solo dinámicamente, en respuesta a una consulta realizada por el usuario [3].

Además de las páginas dinámicas existen otras páginas que no pueden ser indexadas como los sitios protegidos por alguna contraseña, documentos en formatos no indexables, páginas que contienen documentos secretos de estado, contenido de descarga no gratuita, etc.

Es así que los buscadores solo indizan las páginas estáticas que equivaldrían al 15 %, el 85% restante estaría dentro de la deep web. Haciendo una analogía diríamos que el internet que todos creemos conocer solo es la punta del iceberg y por debajo hay más información de la que podemos imaginar.

2. SURFACE WEB

La surface web traducida al castellano como la web superficial o también conocida como web visible es prácticamente la web que conocemos, con la que tenemos contacto diariamente.

Comprende todos aquellos sitios cuya información puede ser indexada por los robots de los buscadores convencionales y recuperada casi en su totalidad mediante una consulta a sus formularios de búsqueda [4].

Los robots también conocidos como arañas son programas inteligentes que tienen como labor la búsqueda, indización y catalogación de la información; lo hacen saltando de una página web a otra siguiendo los enlaces de hipertexto.

El problema aparece cuando la información requerida se encuentra en una página que carece de enlaces. En este caso la única forma que tiene de ser registrada en un buscador es que su autor la incluya manualmente rellenando un formulario. Caso contrario, esta web resultará invisible para todos aquellos usuarios de Internet que no conozcan la URL o dirección concreta [2].

Actualmente la posibilidad de indexar archivos no textuales como imágenes, audio, video, archivos PDF, archivos comprimidos o programas ejecutables, se ha convertido en un reto para los motores de búsqueda más convencionales. Técnicamente la mayoría de estos formatos pueden ser indexados pero muchos buscadores eligen no hacerlo porque estos formatos son más difíciles de archivar y organizar, por lo que requieren más recursos del servidor y un costo económico mayor.

2.1 Ventajas

Las ventajas más notorias de la web superficial que se pueden mencionar son:

- Ser de libre acceso.
- Su información no está contenida en bases de datos.
- No requiere la realización de un proceso de registro para acceder.

2.2 Desventaja

Como la desventaja más relevante que tiene la web superficial se puede mencionar el acceso a solo una parte pequeña de toda la información online que existe.

3. DEEP WEB

La web profunda que viene del inglés deep web también conocida por muchos otros nombres como invisible web, hidden web entre

otros, es sin duda toda esa información a la que no tenemos acceso mediante los buscadores más comunes.

La web profunda es el término utilizado para describir toda la información disponible en Internet que no es recuperada interrogando a los buscadores convencionales. Generalmente es información almacenada y accesible mediante bases de datos [4].

La deep web o internet profunda son redes que escapan a los buscadores más conocidos de la superficie, de ahí su nombre. Sus páginas, manuales, documentos, no están indexados y necesitamos usar programas específicos para poder acceder a ellas. Son las bases de datos no indexadas, son redes que no quieren mantener comunicación con la Internet común, son las entrañas de la red, los suburbios [5].

La deep web contiene millones de datos a los que solo se puede acceder con navegadores especializados, tanto para actividades legales como para cometer delitos.

Se dice que en la deep web no se navega, se bucea, y es muy común sentirse perdido las primeras veces; se trata de un ejercicio nuevo en un elemento conocido y eso genera cierta confusión y tendencia a abandonar el intento. Los principales problemas con los que tropezamos son: el desconocer los enlaces de interés y el no saber cómo llegar a ellos.

También es muy común que al intentar acceder a un servicio de la deep web, este se encuentre caído, perdido o muerto. Hay que tener en cuenta que el mantenimiento de esas páginas no es tan habitual como el de una página web comercial, la de un periódico digital o un blog famoso, pero como todo en esta vida, no tenemos más que seguir buscando para encontrar los que realmente son buenos enlaces.

3.1 Ventajas

Las ventajas de la deep web sobre la surface web claramente oscilan entre el mayor volumen de información y el tipo de contenido de esta información difícil de conseguir. De este último podríamos destacar algunos contenidos ansiados por miles de personas:

- Revistas electrónicas.
- Tesis Doctorales.
- Estadísticas e informes.
- Diccionarios y enciclopedias.
- Materiales discográficos.

3.2 Desventajas

Las desventajas que llaman más la atención son:

- Los buscadores convencionales no pueden indizar su contenido.
- Tener que instalar un buscador especial como ser TOR, o para dispositivos móviles ORBOT.
- Contenido que en muchos casos solo está disponible por poco tiempo.
- Búsqueda difícil al no saber por dónde empezar y por dónde continuar.
- Mal uso que le dan algunas personas.

De este último punto se hará una explicación más detallada en el siguiente apartado.

3.3 Mal uso de la Deep Web

Al ser un lugar de difícil acceso y desconocido por la mayoría de la gente la deep web ha sido utilizada por muchas personas para realizar toda clase de actos fuera de la ley.

Se realizan tráfico de órganos, de personas y de armas. Se vende todo tipo de drogas, hay infinidad de sitios que contienen pornografía infantil, inclusive se dice que hay sitios donde se puede contratar a sicarios e ingresar a chats donde se planean secuestros, robos y asesinatos.

De esta manera es que la deep web se convierte en un suburbio, aunque el anonimato de las personas que frecuentan estos sitios detestables parece ser una ventaja para ellos, existen personas que buscan estos sitios para denunciarlos y ayudar a capturar a los responsables de estas páginas.

Con esto no quiero decir que todo lo que se hace en los suburbios sea bueno o malo, quiero decir que el anonimato y sus posibilidades son una herramienta más, no un modo moderno de ser impune [5].

4. DIFERENCIAS ENTRE SURFACE WEB Y DEEP WEB

A continuación se presenta una tabla (ver tabla 1) extraída de [4] que presenta las diferencias existentes entre la web superficial y la web profunda:

Tabla 1. Diferencias entre Surface Web y Deep Web

Surface Web	Deep Web
Información disponible en Internet que es recuperada interrogando a los buscadores convencionales.	Información disponible en Internet que no es recuperada interrogando a los buscadores convencionales.
Está formada por páginas web estáticas, es decir páginas o archivos con una URL fija.	Páginas que no tienen una URL fija y que se construye en el mismo instante.
Los documentos aparecen referenciados o ligados entre otros.	Los documentos no tienen un link desde otro documento.
Es de libre acceso. No requiere la realización de un proceso de registro para acceder.	Páginas en las que es necesario registrarse para tener acceso al contenido, ya sea de forma gratuita o pagada.
Es indizable.	No es indizable.

5. CONCLUSIONES

A pesar de que la web superficial que conocemos contiene mucha información no es suficiente para satisfacer a las personas en su búsqueda del conocimiento es por eso que intentamos sumergirnos en la web profunda para ampliar nuestra búsqueda.

Ahora bien, al momento de sumergirnos en la web profunda hay que saber dónde buscar y hacerlo con la mayor responsabilidad posible, es mejor evitar los sitios donde hay toda clase de

ilegalidades, ya que al hombre todo lo prohibido le genera gran interés aunque diga lo contrario solo por no ser rechazado por la sociedad. Una frase que resume todo es “Si juegas con fuego te puedes quemar”.

Es posible que en un futuro casi toda la información que se encuentra en la web profunda esté disponible para todos por la evolución que tienen los motores de búsqueda tradicionales como Google, Yahoo!, AltaVista, entre otros.

6. REFERENCIAS

- [1] <http://www.cnnexpansion.com/tecnologia/2014/03/10/las-profundidades-del-mar-de-internet>
- [2] http://elpais.com/diario/2005/10/20/ciberpais/1129772426_850215.html
- [3] <http://www.sociedadelainformacion.com/20011103/invisible/internetprofundo.htm>
- [4] <http://encontrandotesorosnlared.wikispaces.com/Web+Superficial+y+Web+Profunda>
- [5] <http://www.rebellion.org/docs/162798.pdf>