

¿Qué es Internet invisible o Internet profunda?

[Share on facebook](#) [Share on twitter](#) [Share on google plus](#) [one share](#) [Share on linkedin](#)

Fuente: [disminuir el tamaño de la fuente](#) [aumentar tamaño de la fuente](#)

Por [Adriana Molano](#) Enero 8, 2014

Lo que conocemos de la web es menos del 5% de todo el conocimiento alojado en ella.

El concepto de Internet invisible o Internet profunda, conocida en inglés también como Deepweb, Darkweb o Hidden web, hace referencia, básicamente, a una serie de contenidos que no son visibles para los motores de búsqueda tradicional, esto es, que Google no puede encontrar.

Lo interesante de este concepto es que, tal como lo refiere Brian Wai Fung en su investigación para el MIT sobre [algoritmos para web profunda](#), “de acuerdo con una investigación conducida por la Universidad de Berkeley en California, en 2003 la web (superficial) contenía aproximadamente 167 terabytes de información. En contraste, la Internet profunda se estima que contiene 91.000 terabytes de información, lo que corresponde a 500 veces más el tamaño de la web superficial. Esto significa que el tradicional sistema de búsqueda basado en enlaces no está rastreando el contenido del 99% de la web”.

Michael Bergman, es el más reconocido estudioso de la Web Profunda, quien en su artículo [‘Web Profunda: sacando a la luz un valor escondido’](#) (2001), acuñó el término haciendo referencia a la expresión ‘Internet invisible’ que había sido usada por Jill Ellsworth ya en 1994, momento en el cual se pronosticaba el boom del .com y el crecimiento exponencial de la web y la consecuencia ampliación del acceso a la información y el conocimiento.

En su momento, entre los hallazgos más sorprendentes en relación con la Internet Invisible se cuentan:

- La información pública sobre la web señala que actualmente (2001) ésta es entre 400 a 550 veces más grande que lo que el mundo conoce como Internet.
- Sesenta de los mayores sitios de web profunda contienen colectivamente alrededor de 750 terabytes de información - suficientes por sí mismos para superar el tamaño de la web superficial en cuarenta veces su tamaño.
- La web profunda es la categoría de mayor crecimiento en información de la nueva Internet.
- La información alojada en sitios de web profunda tienden a ser más selectivos, con contenidos más extensos que los sitios superficiales comunes.
- La relación contenido y calidad de la web profunda es de 1.000 a 2.000 veces mayor que la de la web superficial.
- Los contenidos de la web profunda son altamente relevantes para todas las necesidades de información, mercadeo, y dominio de Internet.
- Más de la mitad del contenido de la web profunda reside en bases de datos sobre temas específicos.
- Un total de noventa y cinco por ciento de la web profunda es de información accesible al público, no sujeta a cuotas o suscripciones.

Lo más interesante es que pasada una década de presentación de este artículo, las cifras siguen en aumento y se puede considerar que todavía, a través de los motores de búsqueda tradicionales, es posible acceder apenas al 5% del conocimiento alojado en la web.

- Pero, ¿por qué una web termina alojada en la Internet invisible? Entre las causas que reseña [Aprender Internet](#) se encuentran:
- Los desarrolladores de la web decidieron mantenerla oculta de los motores de búsqueda por alguna razón particular (en algunos casos llega a ser incluso ilegal).
- La página es dinámica, esto es, que el acceso a la información administrada por ella se hace únicamente a través de su propio buscador que no puede ser consultado por los motores de búsqueda tradicionales.
- Los contenidos publicados están en un formato no indexable (ilegible) por los motores de búsqueda, como .pdf, .doc, .xls, .ppt, etc.
- La página está protegida con contraseñas o sistemas de Captcha que evitan que los sistemas de rastreo de los buscadores accedan a su contenido.
- Los contenidos son accesibles a través de enlaces creados con JavaScript o similares.

Uno de los datos más curiosos respecto a la Internet invisible es que los contenidos de la Biblioteca del Congreso de los Estados Unidos, la más grande del mundo, así como el Diccionario de la Real Academia de la Lengua Española (RAE), no son indexables, lo que significa que son únicamente accesibles desde los buscadores de cada uno de estos portales, convirtiéndolos así en parte de la Internet profunda, aun cuando el propósito de ambos sitios es promover el acceso a la información que alojan.

En relación con la Sociedad del Conocimiento, lo fundamental de la web invisible es su potencial de abrir las puertas a información de alta calidad y especialidad temática. Así, las bases de datos de las grandes universidades y centros de investigación, que reúnen tesis, artículos analíticos, publicaciones de libre acceso, planos, gráficos y similares, existen y están abiertas para que sean consultadas y aprovechadas de acuerdo a sus términos de uso.

En paralelo, se ha relacionado la web profunda con la piratería y la ilegalidad, especialmente porque en ella se alojan contenidos cuya fuente no es fácilmente rastreable, y solo es accesible a través de sistemas tipo [TOR](#) (The Onion Router – El router cebolla, por sus múltiples capas), que ocultan la identidad de los visitantes. Es verdad que el Internet invisible también sirve con propósitos poco asociados a la gestión del conocimiento, pero como todo en la red, queda a discreción de los usuarios saber qué contenidos aprovechar.

Aunque su nombre resulte intimidante para algunos, la verdad es que la Internet invisible ya no está tan oculta como antes y gracias a los metabuscadores (buscadores de buscadores) es posible rastrear fuentes de bases de datos, que permiten acceder a la información alojada en la web profunda, como estos que reseña Luis Castro en '[¿Qué es Deep web?](#)':

- [Scirus](#), usado para búsquedas de información científica.
- [Infomine](#), búsquedas de material escolar de todo tipo.

- [FreeLunch](#), búsquedas de datos económicos.
- [CompletePlanet](#), búsquedas de diversos temas.
- [Archive](#), metabuscador para rastrear temas específicos.
- [Search Engine Guide](#), metabuscador que permite encontrar un buscador o base de datos relacionada con el tema de interés.