



# Deep Web

Nicolás Reinales

# Definición

- **Internet profunda**<sup>1</sup> (del inglés, *deep web*), **internet invisible** o **internet oculta** es el contenido de internet que no está indexado por los motores de búsqueda convencionales, debido a diversos factores. El término se atribuye al informático Mike Bergman. Es el opuesto al Internet superficial.



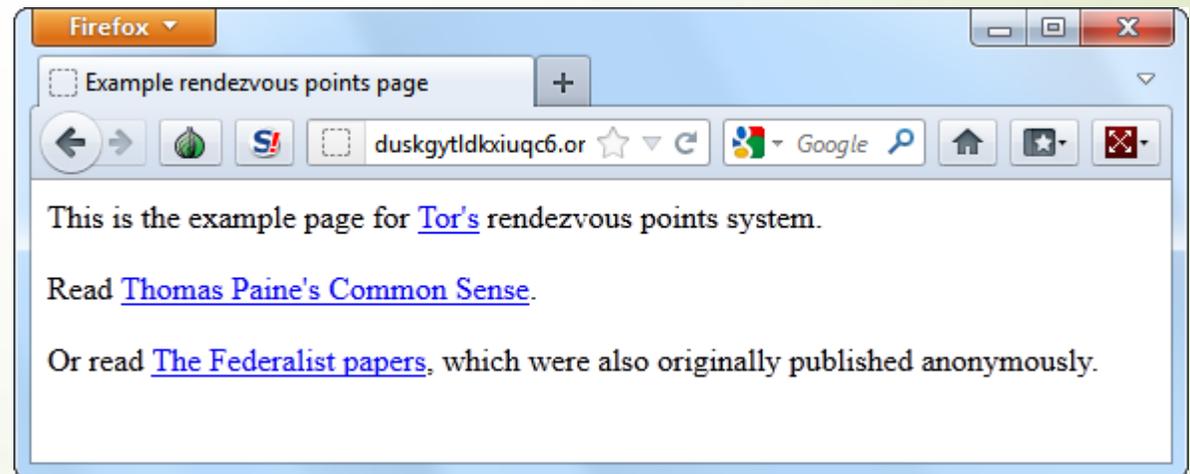


# Origen

- ▶ La principal causa de la existencia de la internet profunda es la imposibilidad de los motores de búsqueda ([Google](#), [Yahoo](#), [Bing](#), etc.) de encontrar o indexar gran parte de la información existente en Internet. Si los buscadores tuvieran la capacidad para acceder a toda la información entonces la magnitud de la «internet profunda» se reduciría casi en su totalidad. No obstante, aunque los motores de búsqueda pudieran indexar la información de la internet profunda esto no significaría que esta dejara de existir, ya que siempre existirán las [páginas](#) privadas. Los motores de búsqueda no pueden acceder a la información de estas páginas y sólo determinados usuarios, aquellos con contraseñas o códigos especiales, pueden hacerlo.

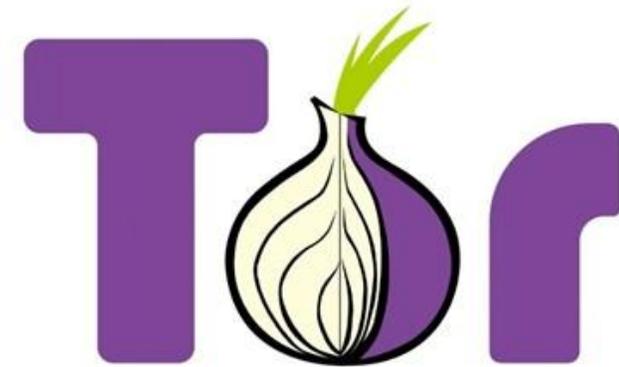
# Acceso

- ▶ Mientras que las páginas normales son identificadas mediante el protocolo UDP/IP, las páginas .onion son repetidas por el ramal al que se ha conectado el navegador especializado para tal sistema de navegación segura. El sistema de búsqueda es el mismo que usa BitTorrent. El contenido se coloca en una base de datos y se proporciona sólo cuando lo solicite el usuario.



# Tor

- **Tor** es la sigla de **The Onion Router** (en español: El Rúter Cebolla). Es un proyecto cuyo objetivo principal es el desarrollo de una red de comunicaciones distribuida de baja latencia y superpuesta sobre internet, en la que el encaminamiento de los mensajes intercambiados entre los usuarios no revela su identidad, es decir, su dirección IP (anonimato a nivel de red) y que, además, mantiene la integridad y el secreto de la información que viaja por ella. Por este motivo se dice que esta tecnología pertenece a la llamada darknet o red oscura también conocida con el nombre de deep web o web profunda.



# Criptomoneda

- ▶ Es posible encontrar sitios web que realicen transacciones ilegales como drogas, en servidores exclusivos para usuarios de Tor. Es una práctica común utilizar monedas digitales como el Bitcoin que se intercambia a través de billeteras digitales entre el usuario y el vendedor, lo que hace que sea prácticamente imposible de rastrear.



# Motivos

Motivos por los que los motores de búsqueda no pueden indexar algunas páginas:

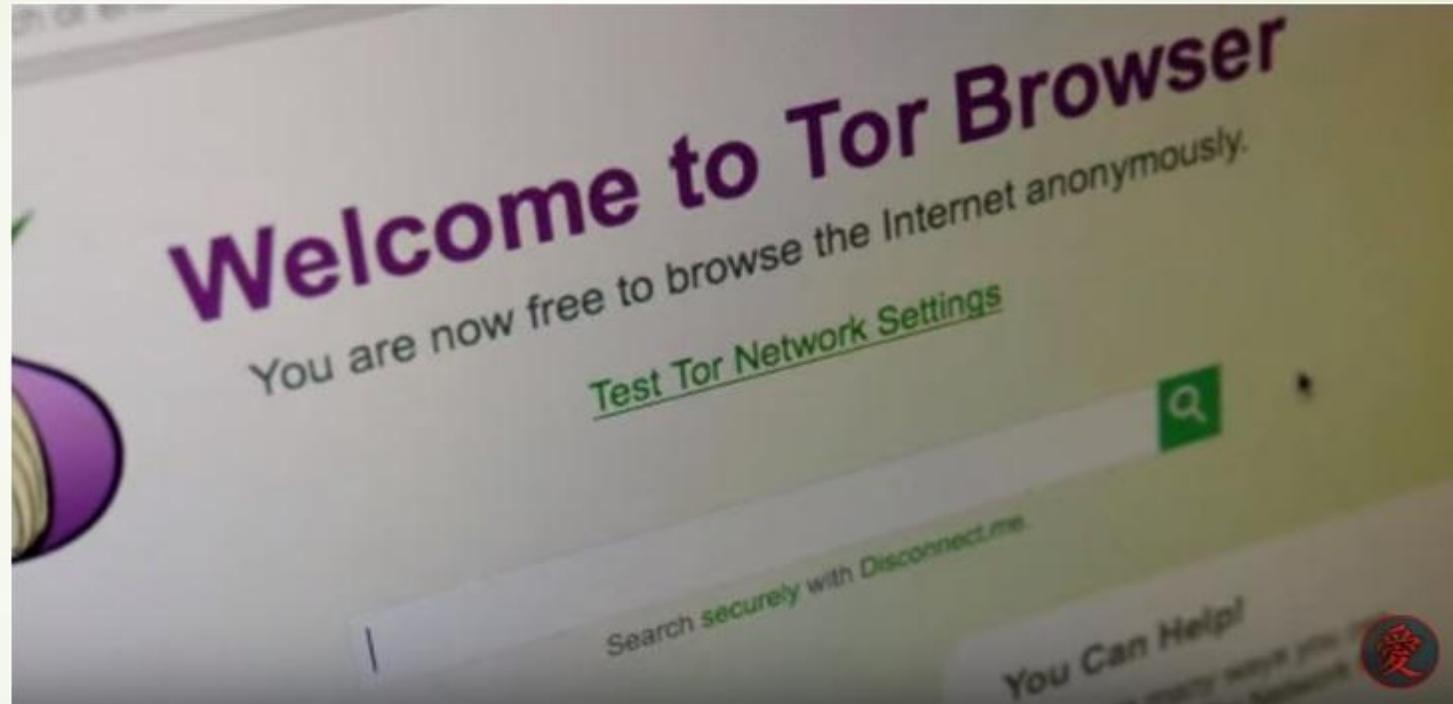
- **Documentos o información oculta:** Pdfs que no se encuentran en las paginas indexadas, listas de datos no publicas (sobre todo los ciber-criminales)...
- **Web contextual:** páginas cuyo contenido varía dependiendo del contexto (por ejemplo, la dirección IP del cliente, de las visitas anteriores, etc.).
- **Contenido dinámico:** páginas dinámicas obtenidas como respuesta a parámetros, por ejemplo, datos enviados a través de un formulario.
- **Contenido de acceso restringido:** páginas protegidas con contraseña, contenido protegido por un [Captcha](#), etc.
- **Contenido No [HTML](#):** contenido textual en archivos multimedia, otras extensiones como exe, rar, zip, etc.
- **Software:** Contenido oculto intencionadamente, que requiere un programa o protocolo específico para poder acceder (ejemplos: [Tor](#), [I2P](#), [Freenet](#))
- **Páginas no enlazadas:** páginas de cuya existencia no tienen referencia los buscadores; por ejemplo, páginas que no tienen enlaces desde otras páginas.

# Recursos de la internet profunda

Los recursos de la internet profunda pueden estar clasificados en las siguientes categorías:

- **Contenido de acceso limitado:** los sitios que limitan el acceso a sus páginas de una manera técnica (Por ejemplo, utilizando el [estándar de exclusión de robots](#) o [captcha](#), que prohíben los motores de búsqueda de la navegación por y la creación de copias en caché
- **Contenido dinámico:** las páginas dinámicas que devuelven respuesta a una pregunta presentada o acceder a través de un formulario, especialmente si se utilizan elementos de entrada en el dominio abierto como campos de texto.
- **Contenido no enlazado:** páginas que no están conectadas con otras páginas, que pueden impedir que los programas de rastreo [web](#) tengan acceso al contenido. Este material se conoce como páginas sin enlaces entrantes.
- **Contenido programado:** páginas que solo son accesibles a través de [enlaces](#) producidos por [JavaScript](#), así como el contenido descargado de manera dinámica a partir de los [servidores web](#) a través de soluciones de [Flash](#) o [Ajax](#).
- **Sin contenido HTML:** contenido textual codificado en [multimedia](#) ([imagen](#) o [video](#)) [archivos](#) o [formatos de archivo](#) específicos no tratados por los motores de búsqueda.
- **Web privada:** los sitios que requieren de registro y de una contraseña para iniciar sesión
- **Web contextual:** páginas con contenidos diferentes para diferentes contextos de acceso (por ejemplo, los rangos de [direcciones IP](#) de clientes o secuencia de navegación anterior).

# El inframundo de Internet



[www.youtube.com/watch?v=A3NCfCvWtJc&t=1s](https://www.youtube.com/watch?v=A3NCfCvWtJc&t=1s)