

Facultad de
**Información y
Comunicación**



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Maestría en Información y Comunicación

Tesis para defender el título de la Maestría en
Información y Comunicación

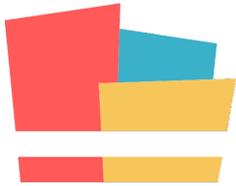
Caer en el abismo: Uso de la Deep Web en el mundo académico

Autor: Pablo Seijo

Director de tesis: Mag. Alén Pérez Casas

Montevideo

2022



Facultad de
**Información y
Comunicación**



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

El Tribunal docente integrado por los abajo firmantes, aprueba la Tesis:

“Caer en el abismo: Uso de la Deep Web en el mundo académico

Tesista: Pablo Seijo

Maestría en Información y Comunicación

Fallo:

Tribunal:

Profesor/a:

Profesor/a:

Profesor/a:

Resumen:

En el estudio que da nombre a la Web profunda, Michael Bergman, indicaba que esta constituía el 95 por ciento de la información ubicada en Internet. Veinte años después, varios estudios cuantitativos indican no solo que esa proporción no ha cambiado, sino que actualmente hay más de 30 billones de páginas inaccesibles mediante los buscadores tradicionales. En este trabajo analizaremos las razones por las cuales el contenido de interés académico puede terminar residiendo en la Deep Web. Revisaremos diferentes estudios sobre el crecimiento de redes de anonimato y las tecnologías que la constituyen, como redes Onion, uso de VPN, Secure Drop o Telegram. Profundizaremos en el problema de las restricciones del acceso al conocimiento científico mediante muros de pago y las invasiones a la privacidad y censura por parte de los gobiernos como posible generador de estos nuevos espacios. Para eso se utilizará análisis de contenido cuantitativo mediante software y un análisis cualitativo de 4 sitios: Sci Hub, Library Genesis, Freedom of the Press y Propublica para obtener indicadores mediante el método Morin-Chartier.

Palabras Clave: web profunda, censura, privacidad, muros de pago, burbujas de contenido, redes Onion, TOR, VPN, web profunda académica.

Abstract:

Michael Bergman stated in the study that gave the Deep Web its name that it constituted 95 percent of the information found on the Internet. Twenty years later, several scientometric studies indicate not only that this proportion has not changed, but that there are currently more than 30 billion pages inaccessible through traditional search engines. In this work, we will analyze the reasons why content of academic interest may end up residing on the Deep Web. We will review different studies on the growth of anonymity networks and the technologies that constitute them, such as Onion networks, use of VPN, Secure Drop, or Telegram. We will delve into the problem of restrictions on access to scientific knowledge through paywalls and invasions of privacy and censorship by governments as a possible generator of these new spaces. For this, content analysis will be used through software and a qualitative analysis of 4 sites: Sci Hub, Library Genesis, Freedom of the Press, and Propublica to obtain indicators through the Morin-Chartier method.

Keywords: deep web, censorship, privacy, paywall, filter bubble, onion routing, TOR, VPN, deep academic web

Tabla de contenido

Tabla de contenido	4
0 Definición del objeto de estudio	11
1.0 Fundamentación y Antecedentes	11
1.1 Antecedentes	11
1.2 Pertinencia y Relevancia	13
2.0 Marco Teórico Conceptual y Referencial	16
2.1 Conceptos	16
2.1.1 Información	16
2.1.2 Contenido académico	17
2.1.3 Web	17
2.1.4 Buscadores e indexado	18
2.1.5 Web profunda	20
2.1.6 Redes de anonimato y tecnologías complementarias	21
2.1.6.1 Redes Onion y Tor	21
2.1.6.2 Otras alternativas I2P, Freenet, Zero Net	24
2.1.6.3 Telegram y Durov	29
2.1.6.4 VPN	33
2.2.0 Marco Referencial: motivos y actores de la Web profunda	35
2.2.1 Muros de Pago	36
2.2.2 Aaron Swartz	37
2.2.3 Alexandra y Sci-Hub	39
2.2.4 LibGen y Researchgate	42
2.2.5 Piratería o Autonomía	45
2.2.6 Privacidad y Censura	47
2.2.7 Apagón	47
2.2.8 Aumento del uso de redes de anonimato en 2014 y 2018	53
2.2.9 Propublica y Freedom of the Press	58
2.2.9 Edward Snowden y la Desobediencia Civil	58
2.2.10 Burbujas de contenido: privacidad + censura	60

2.2.11 Exclusión	64
3.0 Objetivos y Preguntas de investigación	68
3.1 Objetivos	68
3.2 Preguntas de investigación	68
4.0 Hipótesis	68
4.1 Hipótesis directrices	68
5.0 Estrategia metodológica	68
5.1 Análisis de contenido	69
5.2 Criterios de Selección	71
5.3 Elección de unidades de muestreo	71
5.4 Categorías	72
5.5 Indicadores	73
5.5.1 Frecuencia	73
5.5.2 Parcialidad	73
5.5.3 Orientación	74
5.6 Aspectos formales del sitio	74
5.7 Análisis cualitativo de los sitios	75
6.0 Análisis de los sitios	75
6.1 Sci Hub	75
6.1.1 Análisis formal del sitio	75
6.1.2 Análisis de contenido Sci Hub - “Acerca de”	76
6.1.3 Indicadores	80
6.1.4 Inferencia del análisis:	80
6.1.5 Análisis cualitativo del sitio	80
6.1.6 Inferencia	81
6.1.7 Análisis de contenido Entrevista Alexandra Elbakyan:	82
6.1.8 Inferencia de frecuencia de términos en la entrevista:	83
6.1.9 Consideraciones adicionales	84
6.2 LibGen	85
6.2.1 Análisis formal del sitio	85
6.2.2 Análisis de contenido LibGen “Letter of Solidarity”	87
6.2.3 Indicadores	92

6.2.4 Inferencia	93
6.2.5 Análisis cualitativo del sitio	93
6.2.6 Inferencia	95
6.2.7 Análisis de contenido Entrevista administradores LiBGen	97
6.2.8 Inferencia de frecuencia de términos de la entrevista:	98
6.2.9 Consideraciones adicionales	99
6.3 Propublica	100
6.3.1 Análisis formal del sitio	100
6.3.2 Análisis de contenido Propublica “Acerca de”	101
6.3.3 Inferencia de la frecuencia de términos en la página “Acerca de:”	101
6.3.4 Análisis cualitativo del sitio	102
6.3.5 Inferencia	104
6.3.6 Análisis de contenido entrevista What Is ProPublica?: Dick Tofel presidente	104
6.3.7 Indicadores	107
6.3.8 Inferencia del análisis:	108
6.3.9 Análisis contenido entrevista: ¿why propublica join the dark web?	108
6.3.10 Indicadores	112
6.3.11 Inferencia del análisis:	113
6.3.12 Consideraciones adicionales	113
6.4 Freedom of the Press	114
6.4.1 Análisis formal del sitio	114
6.4.2 Análisis de contenido “Acerca de” Freedom of The press	115
6.4.3 Indicadores	119
6.4.4 Inferencia:	119
6.4.5 Análisis cualitativo del sitio	119
6.4.6 Inferencia	122
6.4.7 Análisis de contenido página del proyecto “Acerca de: SecureDrop”	122
6.4.8 Inferencia de frecuencia de términos en “Acerca de:” Secure Drop	124
6.4.9 Consideraciones adicionales	124
7.0 Resumen de los cuatro sitios relevados	125
7.1 Coincidencias de las categorías en el cuerpo de los sitios	126
7.2 Coincidencias de las categorías en las entrevistas a los responsables	127

7.3 Resumen de de indicadores	127
7.3.1 Parcialidad	127
7.3.2 Orientación	128
8 Respuestas a las preguntas de investigación	129
9 Conclusiones	129
10 Bibliografía	132

“Si hay una materia en la que las ciencias sociales en toda su diversidad deberán contribuir a mejorar nuestra comprensión del mundo en que vivimos, es precisamente aquella que, en el entorno académico, hemos denominado «estudios de internet”. (Castells, 2014)

0 Definición del objeto de estudio

Nuestro objeto de estudio es el contenido académico en la web profunda. Más específicamente, es la construcción conceptual abstracta que representa el fragmento que es la información relevante para las ciencias de la Información y Comunicación, que podemos encontrar en redes de anonimato y en sitios de acceso público, no indexados por los buscadores tradicionales.

1.0 Fundamentación y Antecedentes

1.1 Antecedentes

El término “Web profunda”, (traducido del inglés Deep Web) es acuñado por el estadounidense Michael Bergman en su estudio de 2001 “The Deep Web: Surfacing Hidden Value”. En este, plantea que la información accesible sin restricciones en la Deep Web era 500 veces mayor que el total de información disponible en la World Wide Web (en adelante WWW).

“Public information on the Deep Web is currently 400 to 550 times larger than the commonly defined World Wide Web” (Bergman, 2001).

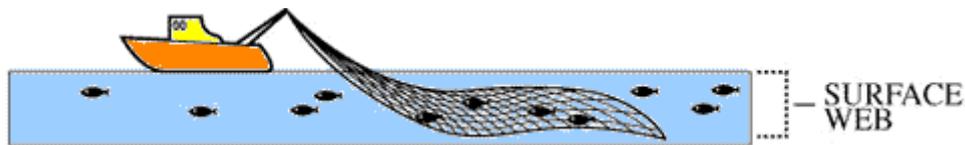
Luego del impacto de la cifra, la pregunta que surge es: ¿De qué tipo y calidad de información estamos hablando?

On average, Deep Web sites receive fifty per cent greater monthly traffic than surface sites and are more highly linked to than surface sites; however, the typical (median) Deep Web site is not well known to the Internet-searching public. The Deep Web is the largest growing category of new information on the Internet. Deep Web sites tend to be narrower, with deeper content, than conventional surface sites. (Bergman, 2001).

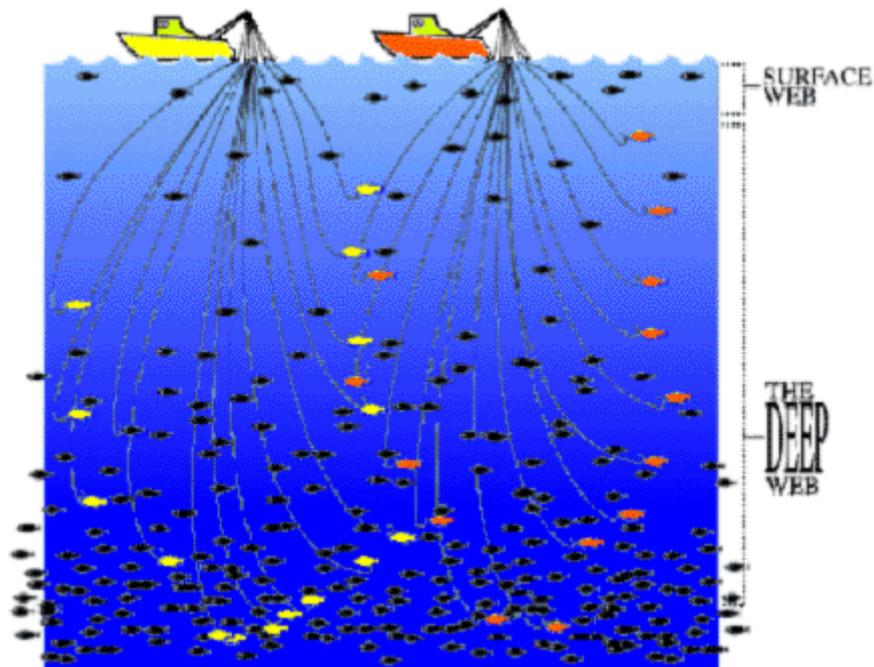
En ese momento, Bergman calificaba el contenido de la web profunda como “más preciso” y de “más calidad” que el de la web superficial. Afirmó que la mayoría del tráfico se dirigía a sitios no indexados por los buscadores y que la información que se encuentra fuera del alcance de los buscadores (será explicado más adelante) es la categoría de información con más crecimiento de toda la Internet.

Si sumamos todo el contenido de calidad de la web profunda, no solo el que está libre, sino también el que puede estar restringido a usuarios (lo veremos mejor más adelante). El contenido de la Deep Web puede ser hasta dos mil veces más grande que el contenido de la WWW.

“Total quality content of the Deep Web is 1,000 to 2,000 times greater than that of the surface Web. Deep Web content is highly relevant to every information need, market, and domain.” (Bergman, 2001).



imágenes originales del estudio inaugural de (Bergman, 2001)



La imagen del iceberg es recurrente para graficar la web. Lo que está por encima del mar sería la web superficial y un vasto bloque de hielo bajo la superficie que sería la web profunda. Pero un iceberg es impenetrable, y el hecho que la web superficial sea tan visible y la web profunda esté oculta bajo la superficie, siendo que es tan vasta y rica en contenido, probablemente ayude a perpetuar la fama de “invisibilidad” e “inaccesibilidad” de esta. Dando una nueva óptica a este concepto está la metáfora de un edificio que usan Villada y Jiménez:

Se puede afirmar que internet es un Sistema de Información distribuido en una especie de edificio y que, para acceder a la información, se debe navegar por pisos repletos de ella. La cuestión es que los navegadores tradicionales no tienen la potencia para circular por todos esos pisos y se restringen sólo a unos cuantos. Debido a esto, los usuarios no encuentran respuestas que satisfagan completamente sus búsquedas. Los primeros pisos de este edificio, que podrían considerarse los más básicos, contienen un porcentaje muy pequeño de la información en la web, y son a los que recurren los buscadores, es decir, es una navegación superficial. El resto de pisos, la mayoría, están por fuera de su alcance y el contenido permanece oculto en cada consulta. (Villada & Jiménez, 2017)

El trabajo de Bergman tiene 20 años, el avance de la tecnología de indexado, el avance de la infraestructura, las mejoras en los servidores y la masividad del acceso a la web, haría pensar que conquistar ese terreno es un objetivo posible. Sin embargo, en estudio longitudinal hecho a través de 9 años y publicado en 2016 por Van den Bosch y Bogers dice:

Our estimates do not show a monotonic growth of Web search engines indices, which was one of the hypothesized outcomes at the onset of this study in 2006. The results could be taken to indicate that the indexed Web is not growing steadily the way it did in the late 1990s. They may even be taken to indicate the indexed Web is not growing at all. Part of this may relate to the growth of the unindexed Deep Web, and a move of certain content from the indexed to the Deep Web.(Van den Bosch, A., Bogers, T., & De Kunder, M. 2016).

Según estos autores, actualmente, no solamente la brecha entre la web profunda y la web superficial no se ha acortado, sino que se ha agrandado.

¿Pero de qué crecimiento estamos hablando? ¿Cuál es el tamaño de la web indexada o superficial? “The Indexed Web contains at least 5.06 billion pages” (Worldwidewebsite, 2018).

Por lo tanto, si Bergman nos dice que la información accesible libremente en la web profunda, es 550 veces mayor que la web indexada y que el total de la información de la web profunda puede llegar a ser hasta 2 mil veces más grande la cifra es astronómica.

1.2 Pertinencia y Relevancia

Teniendo en cuenta el volumen y la calidad de la información residente en la web profunda, la pregunta que surge es: ¿cuál es su relevancia para el mundo académico?

En su trabajo Bergman nos advierte que es innegable la importancia y la “calidad” de la información residente en la Deep Web.

“Serious information seekers can no longer avoid the importance or quality of Deep Web information. But Deep Web information is only a component of total information available. Searching must evolve to encompass the complete Web”. (Bergman, 2001)

Casi diez años después, Madhavan plantea que la Web profunda es la mayor fuente de datos estructurados en toda la red.

The name Deep Web arises from the fact that such content was thought to be beyond the reach of search engines. The Deep Web is also believed to be the biggest source of structured data on the Web and hence accessing its contents has

been a long standing challenge in the data management community. (Madhavan, 2009)

Acceder a esta parte “invisible” de la web es para Villada y Jiménez un desafío informacional y una oportunidad de llegar a más de mil millones de teras que se agregan cada año.

Cuando se piensa en la web, la mayoría de personas hace referencia a la web visible, es decir, las páginas que se pueden acceder directamente a través de su enlace o desde los motores de búsqueda. Sin embargo, esta imagen de la web solamente representa el 4% de todo el contenido en internet. La web se ha convertido rápidamente en una entidad compleja y se estima que cada año se añade un zetabyte (mil millones de teras), o más, de contenido. Navegar en esta cantidad de información representa un desafío, especialmente porque a través de los motores de búsqueda tradicionales no se accede fácilmente a ella. (Villada & Jiménez, 2017)

Como Profesionales de la información urge comprender que si nos restringimos a la web superficial (indexada por buscadores, que explicaremos mejor más adelante) sólo estamos contemplando una fracción muy pequeña del contenido de calidad disponible.

Muchos usuarios piensan que cuando realizan una búsqueda a través de un motor cualquiera, van a tener acceso a toda la información en todas las páginas web. La realidad es que ese buscador solamente accede a una fracción muy pequeña de la verdadera profundidad de la web, y trae a la superficie pequeños fragmentos de la información disponible. (Villada & Jiménez, 2017)

Varde por su parte cataloga al contenido de la Deep Web como “información de la mejor calidad” y una “rica fuente de conocimiento”.

This Deep Web has been estimated to contain 500 times as much data as the Surface Web. If such precise measures are debatable, this order of magnitude has been confirmed by recent work, and it is unquestionable that with information of the best quality, the hidden Web is not only an invaluable source of information, but is also, due to its semi-structured, template nature, a rich source for knowledge discovery. (Varde et al, 2009)

La relevancia específica para el mundo académico es explorada por Lewandoski y Mayr en un estudio que se enfoca el contenido académico de la web profunda y que denominan AIW (Academic Invisible Web) los autores plantean:

Although we focused on the Academic Invisible Web, there are also parts of the visible Web relevant to libraries. The key in achieving the best experience for the library user lies in a combined approach for both types of content. We were able to show that the AIW is very large and that its size is comparable to the indices of the largest general-purpose Web search engines. Therefore, only a co-operative approach is possible. (Lewandoski, Mayr, 2006)

Según estos autores el contenido de la “web académica invisible” es comparable en

tamaño a todo el contenido indexado por los buscadores de la web superficial. En esta misma línea Villada y Jiménez en su estudio sobre la Web profunda como sistema de Información nos dicen:

“Actualmente, las probabilidades de encontrar información de calidad en internet son tres veces más probables en la Deep Web que en la web superficial, por lo que el reto para la comunidad es buscar formas de masificar el acceso a ella.” (Villada, Jiménez, 2017).

Los autores plantean la web profunda como un desafío informacional, un reto para las profesionales de la información que consiste en “masificar el acceso” a estos recursos subacuáticos, Villada y Jiménez nos dicen que es tres veces más probable encontrar información de calidad dentro de la Deep Web que en la web superficial.

Lewandowski y Mayr por otro lado afirman que el tamaño de la “web académica invisible” es igual al tamaño de toda la web y que además estudios recientes indican que la Deep Web crece en mayor proporción que la web indexada.

Table 2. "Quality" Document Retrieval, Deep vs. Surface Web

Query	Surface Web			Deep Web		
	Total	"Quality"	Yield	Total	"Quality"	Yield
Agriculture	400	20	5.0%	300	42	14.0%
Medicine	500	23	4.6%	400	50	12.5%
Finance	350	18	5.1%	600	75	12.5%
Science	700	30	4.3%	700	80	11.4%
Law	260	12	4.6%	320	38	11.9%
TOTAL	2,210	103	4.7%	2,320	285	12.3%

This table shows that there is about a threefold improved likelihood of obtaining quality results from the Deep Web as from the surface Web on average for the limited sample set. Also, the absolute number of results shows that Deep Web sites tend to return 10% more documents than surface Web sites and nearly triple the number of quality documents. (Essien, 2019)

También Essien en su estudio “Relevance of the Deep Web to Academic Research” afirma que pueden obtenerse contenido de interés académico de calidad en la Deep Web y que incluso este es mayor en un 10% que en los que puede obtenerse en la

web superficial.

Todos estos autores relevados indican la relevancia y pertinencia de la web profunda para el mundo académico por su volumen y por la calidad de su contenido, pero para entender más cabalmente a que se refiere necesitamos un marco de referencia más preciso.

2.0 Marco Teórico Conceptual y Referencial

En nuestro marco teórico conceptual o más bien referencial, trataremos de especificar a qué nos referimos en este trabajo cuando decimos “información de interés académico en la web profunda”, para esto debemos diseccionar esta articulación conceptual y definir cada término en el ámbito tecnológico y de la ciencias de la información y comunicación.

A continuación, veremos los principales conceptos y categorías que estructuran este marco referencial, que sitúan y da sentido a nuestro objeto de estudio.

2.1 Conceptos

2.1.1 Información

El término “información” en el área de las ciencias de la Información y Comunicación puede indicar un proceso o un producto. La RAE la define en su primera acepción cómo “comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada.” y la segunda acepción sería el resultado de esa acción o efecto. Por lo tanto, “información” es adquisición de conocimientos y también el resultado de ese proceso.

La información trasciende los marcos referenciales, cómo dice Nicholas Belkin autor de “el estado anómalo del conocimiento para la recuperación de información”, esta es categoría y es materia, es el trasvase de un estado del conocimiento a otro. Es un evento que sucede cuando un sujeto recibe desde un simple texto, a información científica más compleja, es ese “evento” informativo y también el contenido del mensaje.

This variety of frameworks has led to information for information science being variously considered as: a fundamental category such as matter; a property of matter; structure or organization; the probability of the occurrence of an event; reduction in the degree of uncertainty in a state of knowledge; an event which takes place when a recipient encounters a text; data of value in decision-making; public, communicated scientific information; and the message itself. (Belkin, 1978)

Claude Elwood Shannon padre de la teoría matemática de la información nos dice que existe información más allá del contenido del mensaje, la “información” existe

cuando el mensaje original es transformado en algo nuevo. Informar por lo tanto es transformar, cambiar el estado de una porción de la realidad.

For the purposes of communication theory, the "meaning" of a message is generally irrelevant; what is significant is the difficulty in transmitting the message from one point to another. From this point of view, information exists only when there is a choice of possible messages. If there were only one possible message there would be no information; no transmission system would be required in such a case, for this message could be on a record at the receiving point. (Shannon, 1953)

El filósofo estadounidense Charles Sanders Peirce nos dice que "información" es el conocimiento recolectado desde nuestra experiencia o la de otros. Información es cualquier adquisición de conocimientos que necesite algo más que el conocer el significado de las palabras.

"By information I mean all that knowledge that we collect from the experience of ourselves and of others. Now I call any acquisition of Knowledge 'information, which has logically required any other experience than experience of the meanings of words.'" (Peirce, 1910)

2.1.2 Contenido académico

La definición de información de interés académico para las ciencias de la información y comunicación es amplia y escurridiza pues excluir cualquier tipo de datos sería arriesgado e impropio.

En trabajos anteriores que han ahondado en el concepto como: "¿Cómo buscan información académica en internet los estudiantes universitarios? Lo que dicen los estudiantes y sus profesores" (Egaña et al., 2013), "Socialización del conocimiento académico con el uso de tecnologías de información y comunicación (TIC)" (Nava, 2007) y "Relevancia de la Deep Web para la investigación académica" (Essien, 2019). Podemos decir que con fines prácticos la información de interés académico es cualquier producto de la adquisición de conocimiento o cualquier evento informativo, relevante para el mundo académico. Si bien cualquier tipo de datos tiene interés, los más relevantes son: las fuentes (información original en cualquier formato), libros, papers científicos, normas técnicas, artículos periodísticos, investigaciones e informes.

2.1.3 Web

Primero. ¿Por qué web y no red? Se prefiere "web" frente a su traducción literal que es "red" ya que, si bien la Web es una red o la "red de redes" se utiliza más comúnmente el término red para las que son de menor escala como red de computadores o red académica, por ejemplo.

Web e internet. Sobre la diferencia entre web e internet. Para eso la w3c el organismo fundado por Tim Berners Lee, encargado de crear estándares para internet nos dice:

The Internet is a global system of interconnected computer networks that interchange data by packet switching using the standardized Internet Protocol Suite (TCP/IP)." Por otro lado define la web como "The World Wide Web (WWW, or simply Web) is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI)." Sacando los términos técnicos Internet es: una red de dispositivos conectados que intercambian datos bajo (digamos) ciertos protocolos. Por otro lado, la web es un espacio de información que como dice la definición determinados ítems están identificados (accesibles) por estándares web (URL, HTTP, HTML). (W3, 2018)

El sitio Techopedia hace esta comparación de los términos:

"La Web" es el nombre común de la World Wide Web, un subconjunto de Internet que consiste en las páginas a las que se puede acceder mediante un navegador web. Muchas personas suponen que la Web es la misma que Internet, y usan estos términos indistintamente. Sin embargo, el término Internet realmente se refiere a la red global de servidores que hace posible el intercambio de información que ocurre en la Web. Entonces, aunque la Web sí constituye una gran parte de Internet, no son una sola cosa. (Techopedia, 2018)

Most people tend to use the words "Internet" and "Web" interchangeably, but they are not synonyms. The Internet is an interconnection of computers and networking protocol that allows computers of all types to connect to and communicate with other computers on the Internet. The Internet's origin traced back to a project sponsored by the U.S. Defense Advanced Research Agency (DARPA) in 1969 as a means for researchers and defense contractors to share information. The Web, on the other hand, is a software protocol that allows users to easily access files stored on the Internet computers. The Web was created in 1990 by Tim Berners-Lee, a computer programmer working for the European Organization for Nuclear Research (CERN). (Essien, 2019)

En resumidas cuentas, internet es infraestructura, la Web es información.

2.1.4 Buscadores e indexado

La definición de Deep Web más canónica es: "el contenido de Internet no indexado por los buscadores", a continuación iremos más a fondo en que significa "indexar" o no hacerlo, así lo plantea Berman en su trabajo:

Search engines obtain their listings in two ways: Authors may submit their own Web pages, or the search engines "crawl" or "spider" documents by following one hypertext link to another. The latter returns the bulk of the listings. Crawlers work by recording every hypertext link in every page they index crawling. Like ripples propagating across a pond, search-engine crawlers are able to extend their indices further and further from their starting points. (Berman, 2001)

Jumpstation creado en 1993 por Jonathon Fletcher en la Universidad de Stirling en Escocia fue el primer sitio que utilizó la estructura actual de todos los motores de

búsqueda: rastreo, indexación y búsqueda (crawling, indexing, and searching).

El rastreo se hace mediante un algoritmo recursivo que busca páginas en toda la Web. A estas piezas de *software* se les llama *web crawlers*, o “*spiders*” dado que cazan *links* en la web (en inglés: tela de araña). También se les llama bots (robots) porque son programas que se ejecutan de manera autónoma.

Así se refieren Kobayashi y Takeda en otro artículo fundamental para entender el proceso de búsqueda de información en la web (Information retrieval on the web) a la variedad de terminología sobre el algoritmo de indexado:

There is some ambiguity concerning proper terminology to describe these agents. They are most commonly referred to as crawlers, but are also known as ants, automatic indexers, bots, spiders, Web robots , and worms. It appears that some of the terms were proposed by the inventors of a specific tool, and their subsequent use spread to more general applications of the same genre.(Kobayashi,2000)

Esas páginas “rastreadas” son agregadas a un índice que no es otra cosa que una base de datos. Este “*index*” o índice como plantea Baeza-Yates es el centro del desarrollo de cualquier sistema moderno de recuperación de información. Es una colección de conceptos que apuntan a una serie de documentos previamente relevados por el rastreador.

Since the volume of information eventually grew beyond a few books, it became necessary to build specialized data structures to ensure faster access to the stored information. An old and popular data structure for faster information retrieval is a collection of selected words or concepts with which are associated pointers to the related information (or documents) - the index. In one form or another, indexes are at the core of every modern information retrieval system. They provide faster access to the data and allow the query processing task to be speeded up.(Baeza-Yates, 1999)

El índice llamado también por Google “catálogo” es consultado por el rastreador para actualizar los datos o agregar nuevos enlaces. Cuando utilizamos cualquier motor de búsqueda en realidad hacemos la búsqueda sobre su catálogo y no sobre la Web en tiempo real. Por esta razón a veces no podemos acceder a un sitio que se muestra en el resultado de una búsqueda (sitio caído) pero podemos ver todos sus datos incluso acceder a una versión guardada en el servidor (del buscador) guardada en una memoria intermedia (caché).

Hay varias razones para que la información no pueda ser indexada, la primera es que se encuentre fuera de la web tradicional (lo veremos en profundidad más adelante), otra es que los creadores del sitio impidan a los rastreadores acceder al contenido de este.

El principal motor de búsqueda Google nos dice esto sobre el contenido no indexado:

Puedes impedir que una página o que un recurso aparezca en la Búsqueda de Google incluyendo una etiqueta meta o un encabezado "noindex" en la respuesta HTTP. Así, cuando Google vea la etiqueta o el encabezado al volver a rastrear la página, la retirará de los resultados de la Búsqueda de Google aunque otros sitios tengan enlaces a ella.(Google 2021)

Puedes usar un archivo robots.txt en páginas web (HTML, PDF y otros formatos no multimedia que Google pueda leer) para gestionar el tráfico de los rastreadores si crees que tu servidor se sobrecargará con solicitudes del rastreador de Google, o para evitar que se rastreen páginas sin importancia o similares de tu sitio. Es posible que acaben indexándose, aunque no se visiten, si hay otras páginas que dirigen a ella con texto descriptivo. Si quieres bloquear una página para que no aparezca en los resultados de búsqueda, usa otro método. Por ejemplo, protégela con una contraseña o utiliza una directiva "noindex". (Google 2021)

De esta manera los administradores de un determinado sitio web podrían impedir que las arañas indexen el contenido de sus páginas lo que haría que por definición estuvieran en la Deep Web aunque no estarían en redes de anonimato que es lo que veremos más adelante.

2.1.5 Web profunda

Sobre la correcta manera de referirse a la Deep Web En español: La Fundación del Español Urgente (Fundéu) nos dice:

La expresión inglesa Deep Web puede traducirse como internet profunda, donde internet se escribe con inicial minúscula o mayúscula y el adjetivo profundo siempre en minúscula y concordando en masculino o femenino" "Por último, así como en inglés hay denominaciones alternativas, como Deep Net, también en español se utilizan en ocasiones las variantes web profunda y red profunda. Además, con ligeros matices de significado, pueden encontrarse las expresiones internet oculta o internet invisible, con mucha menos difusión. (Fundéu, 2018).

Finalmente, sobre las dos maneras más habituales de referirse a este espacio de internet Web profunda o Web oscura preferimos por peso académico la forma adoptada por Berman quien acuñó el término: Deep Web en español "Web profunda".

Web "oscura" (Dark Web) a veces se utiliza para referirse a una parte de la Web asociada al delito, esto más que una definición es un juicio moral e implicaría que en otras partes de Internet no existe delincuencia o transgresiones, por ello, excluiríamos esta denominación.

Por lo tanto, la definición que aportamos de Web profunda para este trabajo es:

Información alojada en internet a la que los motores de búsqueda no tienen acceso, porque . Ya sea por bloqueo a los rastreadores impidiendo que guarden e indexen su contenido, por pertenecer a redes privadas, o a conexiones detrás de servidores

que restrinjan la entrada.

2.1.6 Redes de anonimato y tecnologías complementarias

“Por eso quizás los movimientos sociales más decisivos de nuestra época sean precisamente aquellos cuyo objetivo es conservar la libertad en Internet, tanto respecto a los gobiernos como a las empresas, labrando un espacio de autonomía para la comunicación que constituye los cimientos de un nuevo espacio público de la Era de la Información” (Castells, 2009)

Cómo plantea Castells conservar la libertad en internet es una base para un nuevo espacio, probablemente una nueva era, la de las redes autónomas, Carracedo en su libro “Seguridad en redes telemáticas” define redes de anonimato cómo: “Una tecnología que trata de conseguir que la identidad de la persona que realiza una determinada operación telemática permanezca oculta ante algunos de los actores presentes en esa operación.” (Carracedo, 2004)

El anonimato como plantea Bykov está fundamentalmente conectado a la comunicación y es de particular interés para el mundo académico interesado en los medios.

“Anonymity is fundamentally connected to communication, as it makes no sense to conceal one’s name if there is no one from whom to conceal it. Necessarily rooted in interaction, anonymity is therefore of particular interest to media scholars.” (Bykov et al, 2021)

En el trabajo de 2019 “Relevance of the Deep Web to Academic Research”, se destaca a Tor y la red Onion como las tecnologías más populares y puras de las “redes de anonimato” como indica Essien.

“The TOR (The Onion Router) maintains the most popular tool for Dark Web access. On the Tor network, internet traffic is directed through the network of random relays. The browser builds a route of encrypted connections, one-by-one. Each relay knows only the previous and the next relays, but the full connection route stays almost untraceable. The Multiple layers of encryption resemble the structure of an onion” (Essien, 2019)

A continuación, repasaremos las redes de anonimato más populares y genuinas como la red Onion, las complementarias como las redes privadas virtuales, las poco conocidas como FreeNet y las emergentes como Zero Net y Telegram que si bien esta última no es una red específica, ofrece la posibilidad de crear canales de contenido la han hecho relevante para la difusión de contenido relevante a este trabajo. A continuación las explicaremos más detalladamente.

2.1.6.1 Redes Onion y Tor

“Cualquier tecnología lo suficientemente avanzada es indistinguible de la magia.”
Arthur C. Clarke



Essien en su trabajo "La relevancia de la Deep Web en la investigación académica" nos plantea que La red Onion mantenida por el proyecto Tor es la manera más simple y más mayoritaria para acceder a la Dark Web. Tor es la cara más visible de la Deep Web y la más fácil de acceder para un usuario que con simplemente cambiar de navegador ya puede empezar.

“The Onion router is the major and cheapest tool used to access the Dark Web. It is beneficial for a researcher to learn how to setup and use this tool for maximal engagement on the web for academic research.” (Essien, 2019)

Es la tecnología que más claramente puede ser atribuible a la Deep Web ya que cada enlace está cifrado y es imposible acceder mediante un buscador o un navegador tradicional. Su nombre proviene de la sigla: The Onion Router, el enrutador cebolla, aludiendo a las capas de encriptación que cubren cada conexión dentro de la red.

El actual proyecto Tor está liderado por Roger Dingledine y se autodefine como una organización de divulgación científica sin fines de lucro radicada en Massachusetts. “Nuestra misión es promover los derechos humanos y las libertades mediante la creación y despliegue de tecnologías de anonimato y privacidad libres y de código abierto, el apoyo a su disponibilidad y utilización sin restricciones y el fomento de su comprensión científica y popular.” (The Tor Project, 2019)

La actual implementación de Tor (onion routing 2) es una evolución de un proyecto del Laboratorio de Investigación Naval de los Estados Unidos NRL.

En la década de los 90, la falta de seguridad en Internet y su capacidad para el seguimiento y la vigilancia se hicieron evidentes, y en 1995, David Goldschlag, Mike Reed y Paul Syverson, del US Naval Research Lab (NRL) se preguntaron si había alguna forma de crear conexiones de Internet que no revelaran quién estaba hablando con quién, ni siquiera a alguien que pudiera monitorear la red. Su respuesta fue crear y desplegar los primeros diseños de investigación y prototipos de enrutamiento de cebolla. El objetivo del enrutamiento de cebolla era que hubiera una forma de utilizar Internet con la mayor privacidad posible, y la idea era enrutar el tráfico a través de múltiples servidores y cifrarlo en cada paso del camino. Esta es

aún una sencilla explicación de cómo funciona hoy Tor. (The Tor Project, 2019)

Esta implementación comenzó a principios del 2000 cuando Dingledine recién graduado del MIT comenzó a trabajar con Paul Syverson del proyecto original del NRL y Nick Mathewson, un compañero de clase de Roger en el MIT.

A finales de 2002 Dingledine y sus colaboradores publicaron el código bajo una licencia de software libre y abierta. Pero en este momento solo estaba implementada en una docena de servidores voluntarios que enrutaban el tráfico. Para hacerlo más simple para usuarios comunes en 2005 crearon el navegador Tor.

Tor comenzó a ganar popularidad entre los activistas y los usuarios expertos en tecnología interesados en la privacidad, pero aun así era difícil para las personas con menos conocimientos técnicos, por lo que a partir de 2005 comenzó el desarrollo de herramientas más allá del proxy de Tor. El desarrollo del Navegador Tor comenzó en 2008. Al hacer que Tor fuera más accesible para los usuarios y activistas de Internet, el Navegador Tor resultó ser una herramienta instrumental durante la primavera árabe desde finales 2010. No sólo protegió la identidad de las personas en línea, sino que también les permitió acceder a recursos críticos, redes sociales y sitios web que estaban bloqueados. (The Tor Project, 2019)

Nick Mathewson cofundador de la red anónima Tor, frente a la pregunta de Sebastián Córdoba: ¿quién usa Tor? responde: “La herramienta Tor browser, es usada especialmente por exiliados políticos, refugiados, periodistas y víctimas de violencia de género” (Córdoba, 2017)

En otra entrevista Isabela Bagueros, Project Manager de TOR frente a la pregunta: ¿Cuáles son los buenos usos que te gustaría destacar?

Primero la libertad de expresión, acceder a páginas web que están censuradas. Yo soy de Brasil, en donde hubo una orden judicial que bloqueó WhatsApp y mucha gente utiliza Tor para acceder a WhatsApp. También para que los periodistas puedan tener una comunicación segura con sus fuentes. Tenemos el ejemplo de Snowden, cuando hizo contacto con los periodistas lo hizo a través de la red Tor. Muchos casos en donde hay filtración de información pasa por Tor, es muy usado en diarios como el New York Times, The Guardian y otros. Y por supuesto, es una muy buena herramienta para el uso de los ciudadanos, que quieren navegar por internet sin tener que dejar un rastro que te persiga. (TrendTic, 2017)

En la página de la organización Reporteros sin fronteras se insta a utilizar redes de anonimato especialmente Tor. Pero además están trabajando conjuntamente para ampliar servicios comunes.

Reporteros sin Fronteras y Torservers.net se asocian para crear y mantener 250 servidores adicionales dentro de la red Tor. El anonimato es esencial en la plena expresión y realización de las libertades civiles. En Internet, solo los servicios de anonimización confiables y descentralizados, como la red Tor, permite

comunicaciones seguras y sin supervisión (Reporters sans Frontieres, 2016)

El proyecto Tor y su navegador se han convertido en referentes no solo tecnológicos sino entre las organizaciones que luchan por los derechos en línea y la protección de la libertad de expresión.

The Deep Web community comprises lawyers, doctors, journalists, activists, CEOs, etc. While the name and the whole concept sound anti-Government, accessing or browsing the Deep Web is entirely legal with TOR browser. There are tons of illegal activities all over the deep network as it offers the same privacy and anonymity to the individuals involved in such rackets. Amidst all the negativity, there is a lot of positivity too. (Essien, 2019)

Por ello en 2011 la “Free Software Foundation” de Richard Stallman que ya había premiado a Wikipedia, The Internet Archive o Creative Commons premió a el proyecto Tor por su servicio a la sociedad y haber permitido a 36 millones de personas alrededor del mundo a alcanzar libertad de acceso y expresión en Internet

The Award for Projects of Social Benefit recognizes a project that intentionally and significantly benefits society through collaboration to accomplish an important social task. This year, the award went to the Tor Project. Using free software, Tor has enabled roughly 36 million people around the world to experience freedom of access and expression on the Internet while keeping them in control of their privacy and anonymity. Its network has proved pivotal in dissident movements in both Iran and more recently Egypt. (Free Software Foundation, 2011)

Nosotros, en el Proyecto Tor, luchamos todos los días para que todos tengan acceso privado a un Internet sin censura, y Tor se ha convertido en la herramienta más poderosa del mundo para la privacidad y la libertad en línea. Pero Tor es más que un simple software. Es un trabajo de amor producido por una comunidad internacional de personas dedicadas a los derechos humanos. El Proyecto Tor está profundamente comprometido con la transparencia y la seguridad de sus usuarios. (The Tor Project, 2019)

Las redes Onion y el proyecto Tor apoyados por organizaciones claves en la lucha por el acceso y la difusión del conocimiento, como Creative Commons, Free Software Foundation u otras organizaciones que luchan contra la censura como Reporters sans Frontieres y este trabajo de amor (en propias palabras de sus representantes) por salvaguardar los derechos humanos. Hacen a la complejidad de un producto tecnológico que a simple vista puede ser un color distinto en la barra del navegador pero en también, como se dice en la declaración de la Free Software Foundation: permitió a más de 36 millones de personas alrededor del mundo experimentar la libertad de acceso y expresión en Internet mientras se sentían en control de su privacidad y anonimato. Más adelante en este trabajo veremos cómo se relaciona el uso de esta tecnología en concreto con los ataques a la libertad de expresión y la censura.

2.1.6.2 Otras alternativas I2P, Freenet, Zero Net

I2P

Invisible Internet Project (i2P) es una red de anonimato que permite a los usuarios estar dentro de espacios invisibles llamados "Darknet". I2P proporciona un canal de comunicación en los que actúan varios protocolos y estándares de cifrado para mantener el anonimato del usuario. Toda comunicación está cifrada de extremo a extremo, en total hay cuatro capas de cifrado usadas cuando se envía un mensaje. Mejora el modelo de comunicación TCP / IP estándar al garantizar que los paquetes intercambiados entre los participantes siempre contengan datos cifrados. En lugar de utilizar direcciones IP para identificar hosts y enrutar tráfico de manera única, I2P introduce sus propios identificadores y lógica de enrutamiento. Mientras exista conectividad de red puede operar en completo aislamiento del resto de la infraestructura pública de Internet.



I2P EL PROYECTO DE INTERNET INVISIBLE

Descargas Acerca de Ayuda Voluntario Idioma

¿Qué hace I2P por ti?

La red I2P proporciona una gran privacidad para las comunicaciones sobre Internet. Muchas actividades que podrían poner en riesgo su privacidad en Internet pueden hacerse anónimamente dentro de I2P.

DESCARGA I2P 0.9.37

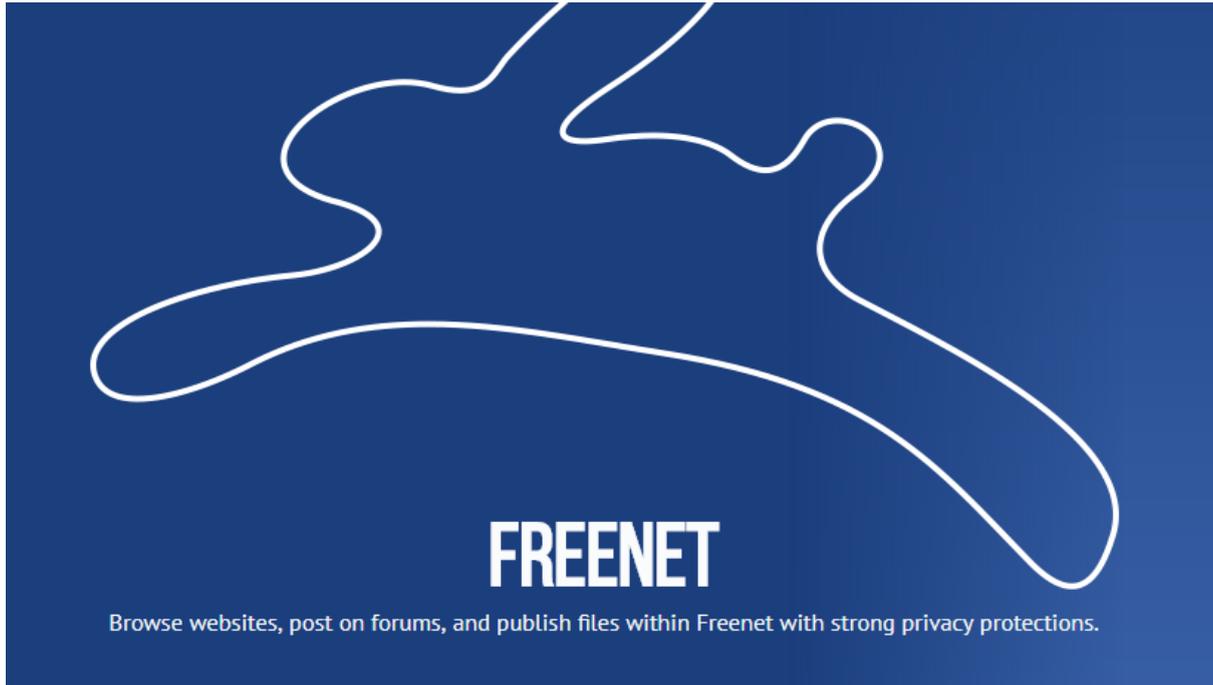
A diferencia de Tor, el navegador web necesita ser configurado para poder navegar y utilizar los *out proxies* disponibles en I2P. Para usarlo es necesario abrir los puertos 4444, para HTTP; y puerto 4445 para HTTPS e introducir valores HTTP proxy para y proxy SSL. No es tan simple como Tor pero es una red segura y confiable. está en chino

“Estas mejoras de seguridad y privacidad son particularmente valoradas por los usuarios de I2P que viven en países con políticas restrictivas que rigen el uso de Internet” (Bazli, B., Wilson, M., Hurst, W. 2017)

FreeNet

Freenet es una de las iniciativas más difundidas y antiguas relacionadas con la privacidad sus inicios datan del año 2001 y continúa teniendo vigencia en la actualidad. Su principal objetivo es facilitar las comunicaciones entre dos o más participantes de forma segura y anónima. No hay servidores que controlen o gestionen el tráfico de red y en su lugar cada usuario se conecta FreeNet aporta una

cuota de ancho de banda a la red y reserva un espacio en su disco duro para almacenar datos de otros usuarios.



Al espacio en disco se le llama “datastore”, cada usuario de esta red tiene un directorio en su disco en el que se almacenan ficheros cifrados por otros integrantes de FreeNet y por lo tanto cada usuario individual no tiene control sobre los contenidos que allí se almacenan. FreeNet es una solución altamente personalizable, se adapta a los más diversos requerimientos, adopta una filosofía y un funcionamiento muy similar a las redes I2P. Ambas redes siguen el modelo de anonimato “inproxy” limitado únicamente al contexto de una VPN sin acceso directo a Internet. A diferencia de Tor, Freenet es una red descentralizada que depende en gran medida de la cantidad de usuarios que la utilicen, ya que, entre más usuarios, será más complejo para un atacante localizar el origen de una petición determinada y por lo tanto será más robusto su anonimato.

ZeroNet

ZeroNet

Sitios abiertos, gratis y sin censura,
usando la criptografía Bitcoin y la red BitTorrent

Descargar para Windows

o [Otras plataformas y código de fuente](#)

17.8MB · Extraer · Ejecutar ZeroNet.exe

ZeroNet es una red descentralizada creada en 2015 por Tamas Kocsis que opera de forma similar a las redes P2P, es decir, que la información se transmite directamente entre usuarios y no con un servidor de por medio. Es un servicio basado en código abierto y en lugar de operar con una dirección IP para cada usuario, los identifica como en una red blockchain. Al utilizar una clave blockchain firma y publica cambios en una página web a la que accede, de forma similar a cómo lo hace cualquier criptomoneda como Bitcoin.

ZeroNet, created in 2015, is a decentralized network similar to a website and composed of peer-to-peer users. ZeroNet is a BitTorrent-based structure, which also transfers site files in a similar manner. ZeroNet is essentially a site publishing platform. Sites can be accessed through an ordinary web browser where the ZeroNet application acts as a local web host for such pages. (Wang et al., 2020)

En teoría es imposible censurar una web de ZeroNet, ya que está distribuida en diferentes ordenadores y no en un solo servidor. Tampoco hay por lo tanto costes de mantenimiento de la página en un servidor, pues se almacena en local en los ordenadores de los usuarios que quieren acceder a ella.

ZeroNet funciona en la mayoría de navegadores modernos como Chrome, Firefox, Safari u Opera. Solo necesita instalar una aplicación que se encuentra disponible para Windows, macOS o Linux. Con esta app se crea un servidor local desde el que acceder y donde almacenar las páginas web a las que se accede. Por poder se puede incluso operar sin conexión a Internet, a fin de cuentas, las webs aclarar que quieres decir con web se descargan en el ordenador y en local.

ZeroNet enabled TOR (The Onion Router) network i.e. it hides an original internet protocol address of the user by two levels of encryption. If a user sends a data over

the internet traffic it done the encryption by using elliptic curve cryptography (which immune to brute force attack) once the first level of encryption is done the encrypted data will again enter into a relay node here second level of encryption can be done using the same cryptographic algorithm and then the data will move towards the recipient now, destination knows only its previous node the origin of data can be difficult to identified. (Pavitra et al., 2018)

¿Qué diferencia a ZeroNet de Tor? Son tecnologías diferentes, aunque compartan características. La primera diferencia de todas es que ZeroNet no es anónimo en realidad, sí que es cierto que una vez el usuario está dentro puede anonimizar su conexión, y se hace precisamente utilizando TOR. TOR por su parte es anónimo desde el principio, de hecho, es su principal objetivo, anonimizar al usuario para que navegue por la red de forma descentralizada y sin ser rastreado.

Por lo tanto, TOR se centra en cifrar con diferentes capas la conexión del usuario y modificar la conexión para que no sea directa con el servidor. ZeroNet por su parte se centra en ofrecer alojamiento a páginas web de forma descentralizada mediante conexión P2P. El combo de ambas tecnologías consigue una red totalmente descentralizada y anónima.

Zeronet makes everything as a decentralized network. It changes the future internet into a decentralized network and the security for the network is provided using bitcoin cryptography. The block chain is used as a security for zero net networks. Block chain is a highly secure decentralized based database which is transparent to every user in the network. But the data cannot be edited by the user, the data can be changed only by an owner. So zeronet is a highly secure and uncensorship network which can be easily accessed by an end user.” (Pavitra et al., 2018)

Lo más innovador de Zeronet es que en lugar de utilizar direcciones IP tradicionales ubica a los usuarios detrás de una clave Blockchain.

Blockchain es una cadena de bloques que contienen información (datos) y un *hash* único que identifica cada bloque y se genera a partir de su contenido, si algo en el contenido de los datos del bloque cambia, cambia el *hash*, a su vez cada bloque tiene el hash de su predecesor de manera que todos quedan conectados en lo que se llama “cadena”. Cuando un bloque cambia el contenido y por lo tanto el hash, la cadena queda invalidada

Para mitigar esto, las cadenas de bloques tienen algo llamado prueba de trabajo. Es un mecanismo que ralentiza la creación de nuevos bloques. En el caso de Bitcoins: se tarda unos 10 minutos para calcular la prueba de trabajo requerida y agregar un nuevo bloque a la cadena. Este mecanismo hace que sea muy difícil alterar los bloques, porque si alteras 1 bloque, tendrás que volver a calcular la prueba de trabajo para todos los siguientes bloques. Entonces, la seguridad de una cadena de bloques proviene del uso creativo del *hashing* y del mecanismo de la prueba de trabajo.

Pero hay una forma más de que las cadenas de bloques se aseguren a sí mismas y eso es por ser distribuidas. En lugar de usar una entidad central para administrar la

cadena, los blockchains usan una red *peer-to-peer* (P2P) y cualquiera puede unirse. Cuando alguien se une a esta red, obtiene una copia completa de la cadena de bloques. El nodo puede usar esto para verificar que todo está en orden todavía. Ahora cuando alguien crea un nuevo bloque, ese nuevo bloque se envía a todos en la red. Luego cada nodo verifica el bloque para asegurarse de que no ha sido alterado. Si todo sale bien, cada nodo agrega este bloque a su propia cadena de bloques. Todos los nodos en esta red crean consenso. Acuerdan qué bloques son válidos y cuáles no lo son. Los bloques alterados serán rechazados por otros nodos en la red. Así que para alterar con éxito una cadena de bloques tendrás que manipular todos los bloques en la cadena, rehacer la prueba de trabajo para cada bloque y tomar el control de más del 50% de la red *peer-to-peer*.

Blockchain puede considerarse la tecnología más disruptiva del siglo XXI y puede llegar a cambiar también el funcionamiento de las redes de anonimato en el futuro.

2.1.6.3 Telegram y Durov

Telegram fue creado en 2013 por Pavel Durov y su hermano. Pavel licenciado en filología inglesa por la Universidad de San Petersburgo debutó en el mundo digital con su sitio Durov.com, que contenía una biblioteca digital donde compartía materiales de todo tipo.



Durov.com vía Wayback Machine

Esta web se popularizó dentro de la universidad y luego devino en la red social V Kontakte luego conocida como VK que a finales de 2007 contaba con 3 millones de usuarios. Durov entre polémicas y persecuciones vendió VK en 2014 y actualmente esta red social cuenta con más de 100 millones de usuarios activos. En su artículo “Pavel Durov, The live wire Disrupter No.1 Russia” nos especifica:

Durov cut his libertarian teeth feuding with the Kremlin over V Kontakte, a Facebook clone he created in 2006. After repeatedly refusing to shut down opposition blogs — responding to one government demand by tweeting a photo of a dog sticking its tongue out — he lost control of V Kontakte and fled the country (obtaining a passport from St. Kitts and Nevis) in 2014. (Politico.eu, 2021)

Entre la creación de Telegram y la posterior venta de VK. Durov comenzó a sentir la presión del gobierno Ruso por dar de baja publicaciones cuestionadas por el régimen dentro de la red social. En 2014 se exilia de Rusia obteniendo la ciudadanía del país antillano San Cristóbal y Nieves. Ermoshina y Musiani en su publicación:

“The Telegram ban: How censorship “made in Russia” faces a global Internet” nos plantean:

The encrypted messenger Telegram was founded by the Russian entrepreneurs, brothers Pavel and Nikolai Durov, in 2013. According to popular belief, Telegram was created as a tool to protect their communication in the context of political persecutions Pavel Durov was subject to after he had refused to collaborate with the Russian government and had to cede his first famous project, the social network VKontakte. (Ermoshina, 2021)

Pero los problemas no habían terminado el 16 de abril de 2018 Telegram fué bloqueado en Rusia cuando la empresa se negó a dar las claves de encriptación de su mensajería y Durov fue acusado en Irán de de que Telegram era usado para comunicarse por grupos terroristas.

Telegram fell under scrutiny of the Russian political police in the space of a few years: on 14 July 2017, the Russian Federal Security Service (FSB) requested decryption keys for all messages sent and received via Telegram, in accordance with the 2016 approved Yarovaya Law. This request concurred with another important event: a criminal case initiated against Durov in Iran, where Telegram had allegedly been used by terrorists. Telegram did not satisfy the FSB’s request, and after a second request to provide the keys in March 2018, Telegram’s lawyers explained that it was cryptographically impossible because of the way in which encryption works in Telegram (Ermoshina, 2021)

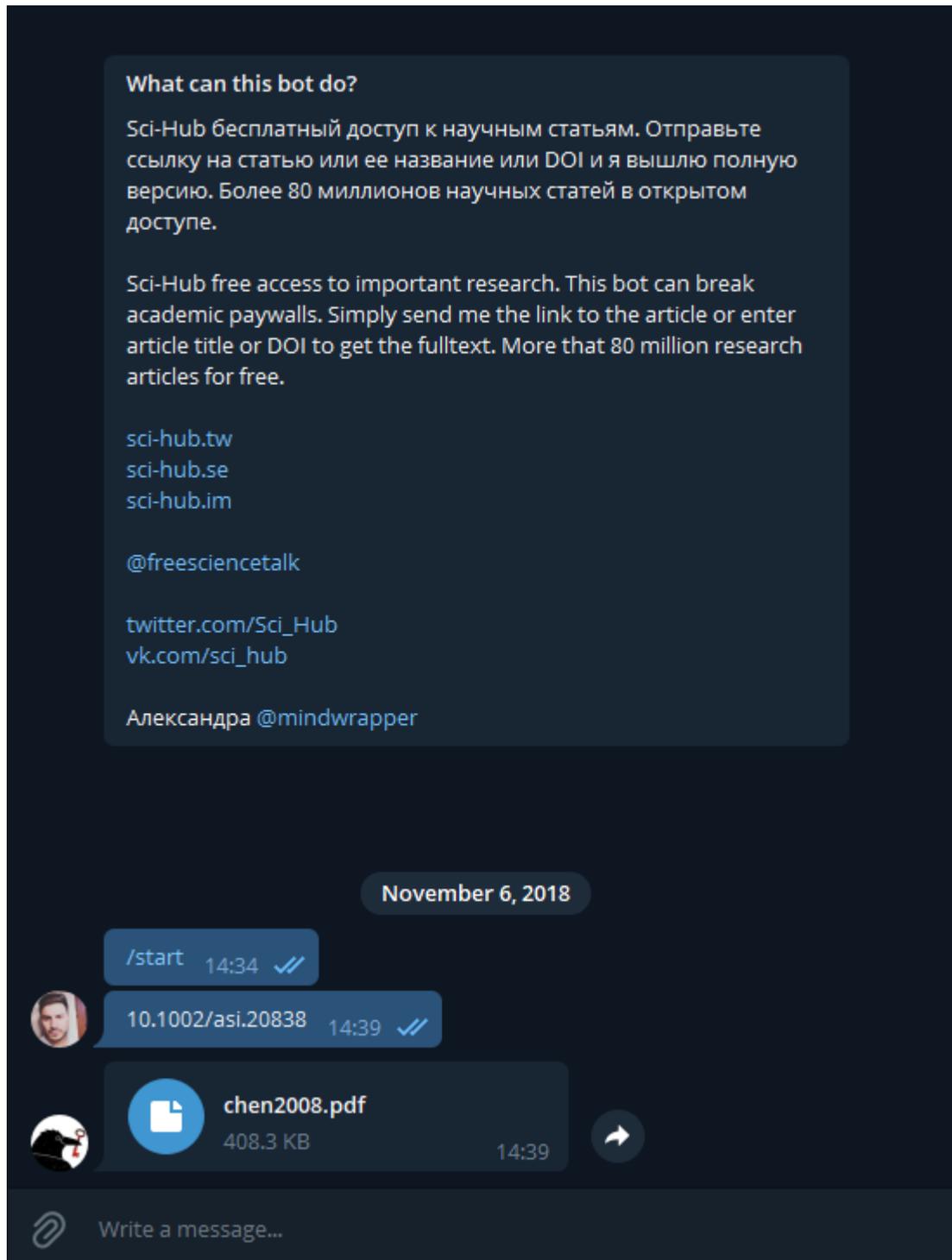
Telegram continuó siendo un lugar para la disidencia en Rusia pero también activistas políticos alrededor del planeta y hoy también como repositorio de contenido, donde los usuarios pueden compartir todo tipo de archivos de manera segura.

Beyond its initial status of “messenger” application, Telegram has become the pre-eminent circumvention tool for hundreds of censored media resources and political activists whose Web sites were blocked by RKN. Popular liberal media such as Grani or Meduza, or opposition politicians Alexey Navalny, Leonid Volkov and many more, have been actively using Telegram’s broadcasting function to continue delivering news content to their audiences regardless of the blocking of their main Web sites and blogs.(Ermoshina, 2021)

Si bien telegram no es una red de anonimato sino más bien un servicio de mensajería o una red social Telegram ha sido utilizado como cita Bykov para transmitir artículos de prensa a suscriptores de manera anónima. El principal sitio pirata de acceso a papers científicos “Sci Hub” ofrece un canal de Telegram donde uno introduce el identificador de un documento electrónico (DOI) y como respuesta obtenemos el PDF.

Through them, an anonymous user or group of people transmits messages or entire articles to subscribers who are also a group of anonymous people who subscribed to the channel’s mailing list. So, it is safe to say that Telegram provides a tool for

anonymous communication. Anonymity as a research construct provides an important avenue for addressing fundamental communication issues related to publicness and privateness. (Bykov et al, 2021)



En 2021 Telegram reportó 500 millones de usuarios activos siendo un 21 % de estos latinoamericanos. (Terzin, 2021) Si bien Telegram no es una red de anonimato pura, la única vinculación que puede hacerse es la de un nickname, por ejemplo @usuario con un teléfono de verificación no necesariamente con una identidad real.

Los canales de aplicaciones de mensajería como Telegram o Discord son en la actualidad muy utilizados para compartir contenido que podría ser eliminado de redes tradicionales, Durov y las fluctuaciones de sus producciones tecnológicas también son un ejemplo de que la disidencia, la controversia y el compartir contenido en el ámbito universitario están como veremos a lo largo de este trabajo asociados de múltiples maneras.

2.1.6.4 VPN

VPN es la sigla de Red privada Virtual, es una tecnología que permite crear conexiones privadas entre dos nodos de Internet.

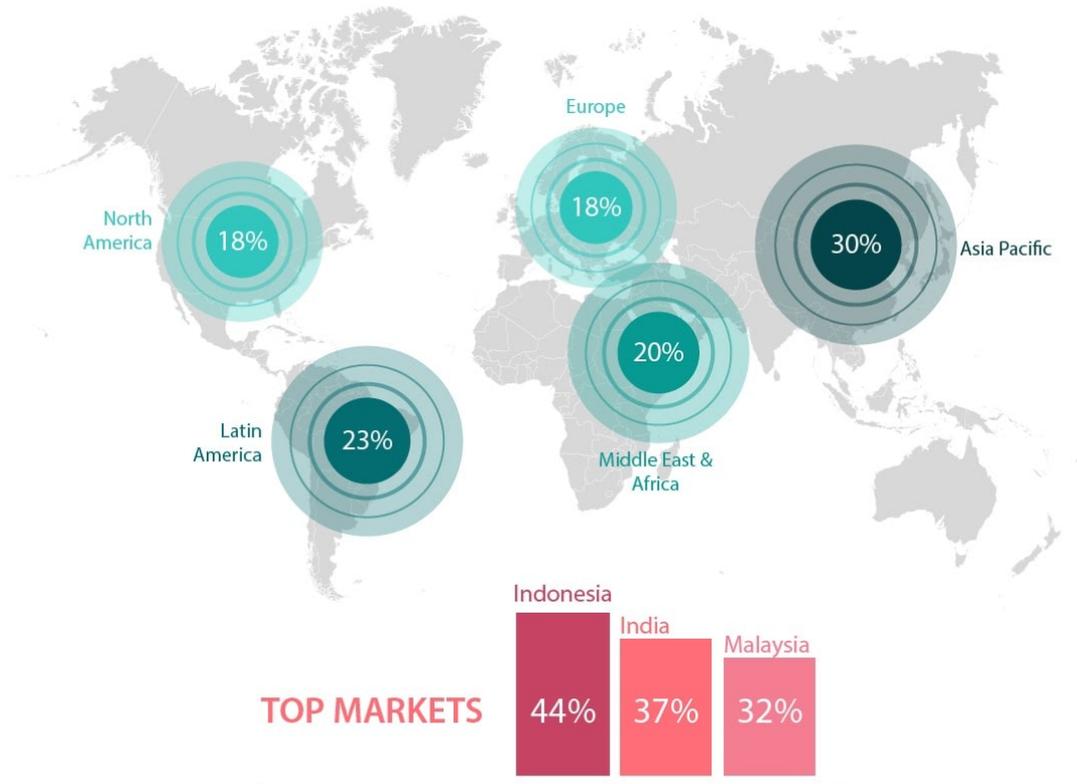
Cada vez que nos conectamos a Internet nuestro proveedor de servicio (ISP) Internet *service provider*, (por ejemplo, Antel) nos asigna una dirección IP de forma automática, como parte de esta dirección indica que nos conectamos desde Uruguay, el sitio al que accedemos puede además de identificarnos, restringir o modificar el contenido que muestra.

Como una VPN crea una extensión de la red de área local sobre una red pública la página a la que accedemos verá la dirección IP de la VPN en lugar de la que nos da el ISP.

El producto comercial VPN es un software que se instala en computadoras o celulares que crea una conexión en la que se cambia nuestra dirección IP por otra, esto permite esconder nuestra dirección IP, o que el software nos provea una dirección correspondiente a otra ubicación geográfica. A este producto de software le llamaremos “servicio VPN”.

El sitio vpnMentor muestra que Latinoamérica es la segunda región del mundo con más uso de redes privadas virtuales sólo superado por Asia.

VPN USAGE AROUND THE WORLD



Percentage of internet users who've used a VPN server within the last month, by location



Tomado de: (Fawkes, 2021)

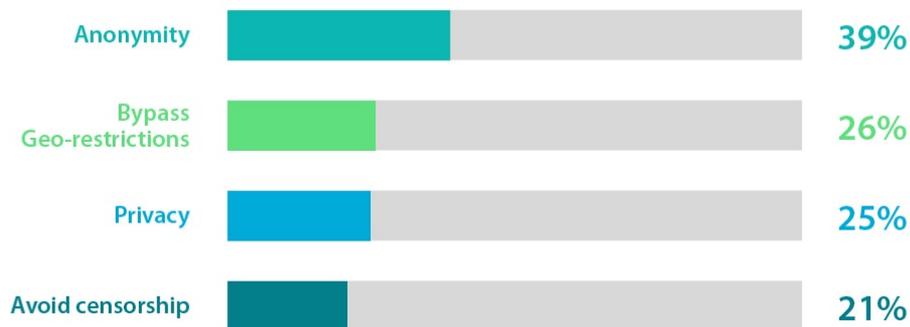
Si bien un servicio VPN no es Deep Web es muy común combinarlos con tecnologías de la Deep Web para proteger la privacidad del usuario. Como plantea Fawkes tanto las redes Onion como las I2P se usan en conjunto con la tecnología VPN para blindar la conexión y asegurar la privacidad, esto como veremos más adelante es una de las recomendaciones más habituales de sitios de periodismo de investigación al contactar con informantes o de activistas que están bajo supervisión gubernamental en regímenes donde las libertades están amenazadas.

“The Dark Web is a very small portion of what makes up the Deep Web, although among the non-tech savvy, both terms are often used interchangeably. The Onion Routing project (TOR) and Invisible Internet Project (I2P) are just two of the ways to access the Dark Web. Using a Virtual Private Network (VPN) alongside these portals is a key way to ensure your privacy is protected, but it does not mean those who use it are engaging in illegal or morally questionable activity.” (Fawkes, 2021)

Si observamos la gráfica de vpnMentor vemos que la única problemática nueva son las restricciones geográficas.

Why Do People Use the Dark Web?

Nearly 40% simply prefer to be anonymous online



*Percentage of population per country



Tomado de: (Fawkes, 2021)

2.2.0 Marco Referencial: motivos y actores de la Web profunda

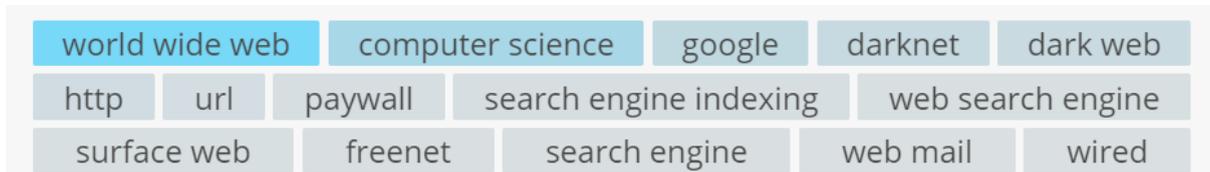
En este trabajo desarrollaremos cuatro grandes motivos para el uso de la Web Profunda: la restricción por muros de pago, la protección de la privacidad, la censura y la exclusión por parte de los buscadores. Estos conceptos y fenómenos constituyen nuestro marco referencial.

El abordaje que realizaremos de aquí en adelante será tecnológico e ideológico, donde conviven decisiones técnicas que hacen que un sitio de interés académico use tecnologías asociadas a la Deep Web como también posicionamientos ideológicos. La militancia de los referentes por causas como el acceso al conocimiento y la libertad de expresión que también creadores e impulsores de las corrientes de pensamiento que dan marco al fenómeno que analiza este trabajo, repasaremos las implicaciones políticas y morales de esos posicionamientos.

Aportaremos evidencias empíricas y descripciones de investigaciones y también pequeños resúmenes biográficos que sirvan de referencia y den contexto a los fenómenos que hacen que contenido de interés académico resida en la web profunda.

2.2.1 Muros de Pago

Cuando buscamos términos asociados a Deep Web sacando los sinónimos y los términos técnicos directamente relacionados el primero en aparecer es “muro de pago”.



tomado de: <https://relatedwords.org/relatedto/deep%20web>

Es el principal obstáculo para el mundo académico. Un muro de pago es una restricción impuesta al contenido científico (generalmente) por parte de las editoriales para restringir el acceso a los documentos. En un estudio publicado en 2018 por investigadores de EEUU, Australia y Alemania indica que la amplia mayoría del contenido académico (75 por ciento) se encuentra detrás de muros de pago.

“Recent estimates suggest paywalls on the web limit access to three-quarters of scholarly literature” (Himmelstein & Rodriguez et al, 2018)

los autores continúan:

Access to scholarly literature remains a pressing global issue. Foremost, widespread subscription access remains restricted to institutions, such as universities or medical centers. Smaller institutions or those in the developing world often have poor access to scholarly literature. As a result, only a tiny percentage of the world’s population has been able to access much of the scholarly literature, despite the fact that the underlying research was often publicly or philanthropically funded. Compounding the problem is that publications have historically been the primary, if not sole, output of scholarship. Although copyright does not apply to ideas, journals leverage the copyright covering an article’s prose, figures, and typesetting to effectively paywall its knowledge. (Himmelstein & Rodriguez et al, 2018).

El problema que se plantea es que el acceso a la literatura científica es vital para el desarrollo de cualquier actividad académica, y que no todas las universidades e institutos cuentan con los fondos necesarios para cubrirlos, incluso la paradoja es que la mayoría de estos estudios muchas veces se hacen de forma honoraria por parte de los investigadores, o son financiados por la misma universidad o por parte del gobierno.

En resumidas cuentas, los autores son obligados por las editoriales a ceder sus derechos vulnerando el derecho básico a acceder a la información y generando un

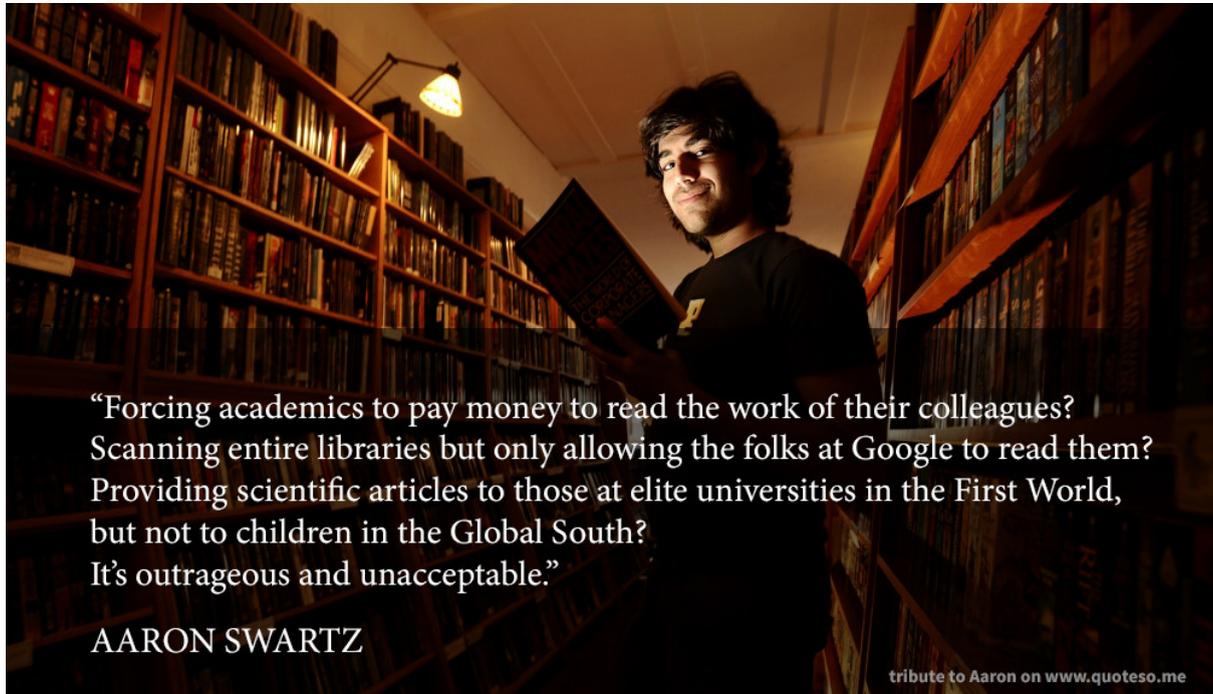
negocio lucrativo sólo para una parte:

\$25.2 billion a year that flows into for-profit academic publishers, examines the 35-40% profit margin associated with the top academic publisher Elsevier and looks at how that profit margin is often greater than some of the most profitable tech companies like Apple, Facebook and Google. Paywall: The Business of Scholarship (2018).

Este negocio mueve más de 25 mil millones de dólares y genera a las editoriales ganancias considerables que las equiparan a las empresas más poderosas e influyentes del mundo. Esta inequidad también genera la búsqueda de alternativas por parte del mundo académico que reside fuera de los sectores que acceden a altos presupuestos.

Si la investigación de alguien está detrás de un muro de pago, y este me impide hacer investigación en ese campo en mi vida, ¿cuantas más vidas tenemos que esperar para que alguien más pueda dar ese paso evolutivo? A veces, la innovación requiere que la persona adecuada, esté en el lugar correcto en el momento adecuado, y todo lo que hace un muro de pago es asegurarse de que sea mucho menos probable que la persona adecuada vaya a estar en el lugar correcto en el momento adecuado para hacer algo valioso.” Tom Callaway en Paywall: The Business of Scholarship (2018)

2.2.2 Aaron Swartz



La primera iniciativa importante y la más mediatizada la dió Aaron Swartz, un joven investigador de la Universidad de Stanford y del MIT. Aaron era una figura muy importante, una especie de niño prodigio de la programación, fue uno de los

creadores del protocolo RSS, fundador de la red Reddit (el foro de discusión más importante de Internet, (con 542 millones de visitantes al mes), cofundador de Creative Commons la principal generadora de licencias complementarias al copyright y que hoy está presente en más de mil millones de documentos a nivel global. Aaron redactó en 2008 un texto que se ha hecho célebre entre todos los que luchan por el acceso a la literatura científica el “Guerilla Open Access Manifiesto”

We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access. (Swartz, 2008).

Además de escribir textos subversivos que invitan a tomar la información “donde sea que esta esté guardada” y subirla a redes abiertas. Aaron pasó a la acción. El 6 de enero de 2011, Swartz fue arrestado por el Departamento Policía en el Instituto Tecnológico de Massachusetts bajo los cargos de fraude electrónico, fraude informático, entrada ilegal e imprudente a un ordenador protegido y daños, después de conectar una computadora a la red del MIT desde un armario sin vigilar ni cerrar, y poniéndola a descargar publicaciones académicas automáticamente desde JSTOR (base de datos científica), usando una cuenta de invitado que le proporcionó el mismo instituto. Swartz buscaba descargar toda la información de esta base de datos privativa para colocarla de manera gratuita en su sitio abierto llamado “Open Library”.

According to a federal indictment, police at the Massachusetts Institute of Technology (MIT) arrested Mr. Swartz in 2011 for allegedly using computers at MIT to gain illegal access to millions of scholarly papers kept by JSTOR, a subscription-only service for distributing scientific and literary journals. Mr. Swartz allegedly broke into computer networks at MIT by gaining entry to a utility closet on campus and leaving a laptop that he programmed to request a huge volume of articles from JSTOR. (Issa, Cummings, 2013).

Se le imputaron los cargos de fraude en línea y once violaciones al "Acta de Fraude y Abuso Computacional", imponiéndole una multa de 1 millón de dólares, 50 años de prisión, incautación de bienes, indemnización y libertad vigilada.

On July 14, 2011, federal prosecutors charged Mr. Swartz with four felony counts: Wire Fraud, Computer Fraud, Theft of Information from a Computer, and Recklessly Damaging a Computer. On September 12, 2012, prosecutors filed a superseding indictment with thirteen felony counts. It appears that prosecutors increased the felony counts by providing specific dates for each action, turning each marked date into its own felony charge, and significantly increasing Mr. Swartz's maximum criminal exposure to up to 50 years' imprisonment and \$1 million in fines. A trial was scheduled for April 2013." (Issa, Cummings, 2013).

Aaron Swartz se suicidó mientras permanecía en calidad de acusado por la

Magistratura Federal de los Estados Unidos por robo de información, una persecución que fue tildada por la familia como "el producto de un sistema criminal-judicial plagado de intimidación y enjuiciamientos excesivos".

Pese a ser considerado un criminal para el gobierno de EEUU, recibió póstumamente el premio "James Madison" de la American Library Association (ALA principal organización de bibliotecarios del mundo) El premio James Madison es un premio para quienes defienden el acceso público a la información. (ALA, 2013)

En 2013 Aaron Swartz fue incluido póstumamente en el salón de la fama de Internet por sus aportes e innovaciones. Muchos de los personajes que hemos visto en este trabajo ocupan ese sitio: Licklider, Caillou, Vince Cerf, Paul Baran y Tim Berner-Lee entre otros, pero también ese mismo año una uruguaya pionera en la red académica Ida Holz fué también incluida para orgullo del Uruguay y la Universidad de la República.

In the early 1990s, Ida Holz helped lead the group of computing pioneers whose efforts resulted in the development of the first networks of what has become the Internet in Latin America. She played a leading role in the construction of the Latin American Network Forum, the Latin American and Caribbean Internet Address Registry, the organization of Latin American and Caribbean ccTLDs and the Latin American Cooperative of Advanced Networks (RedCLARA). For over 20 years, until 2011, she directed Central Computer Services at Uruguay's University of the Republic; it was under her direction that the university installed the first node of the Internet in Uruguay." (Internet Hall of Fame, 2013)

2.2.3 Alexandra y Sci-Hub

"La página universitaria pirata más buscada en la Deep Web" (Labarca, 2016)

“Es cierto: robo a los editores para dárselo a los científicos”

Alexandra Elbakyan, de 28 años, es la fundadora de Sci-Hub, que ha puesto en la web 62 millones de artículos gratis a disposición de los investigadores



Tomado de:(Aluffi, 2017)

En una entrevista para el El País de Madrid la neurocientífica Kazaka Alexandra Elbakyan dice:

En 2009, para hacer mi tesis, tenía que consultar una treintena de estudios. Todos eran de pago, con el acostumbrado precio medio de 30 dólares cada uno. Era demasiado para mí, así que me dirigí a la Red: ya había descargado gratis libros técnicos de páginas piratas y pensaba que podría hacer lo mismo con los estudios. En realidad, no era tan fácil. Me encontré con una comunidad de piratas, donde me explicaron cómo evitar las barreras de pago. Y en un par de años lancé Sci-Hub. (Aluffi, 2017)

En 2017 más de 60 millones de estudios científicos estaban disponibles de forma gratuita para todo aquel que quiera consultarlos. Uno simplemente accede a Sci-hub introduce el DOI (número que identifica una publicación científica en formato digital) e inmediatamente puede descargar el pdf del artículo sin necesidad de registrarse ni cumplir ningún requisito.

Como muestra la imagen tomada de (Himmelstein et al, 2018) el crecimiento de Sci-Hub en el último año ha sido exponencial ya que viene a solucionar un problema clave para la comunidad académica.

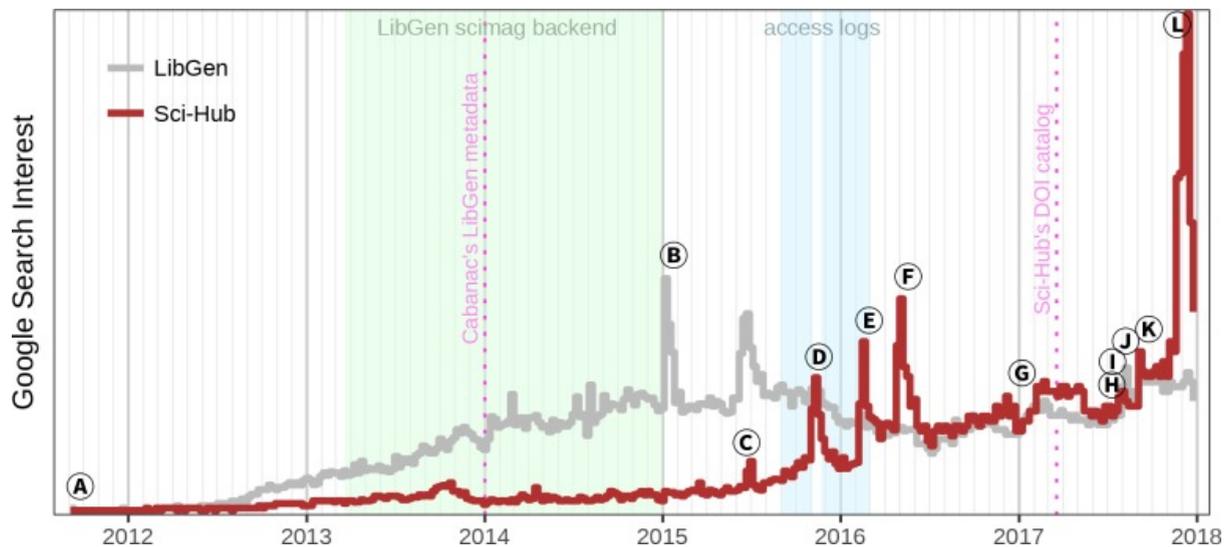
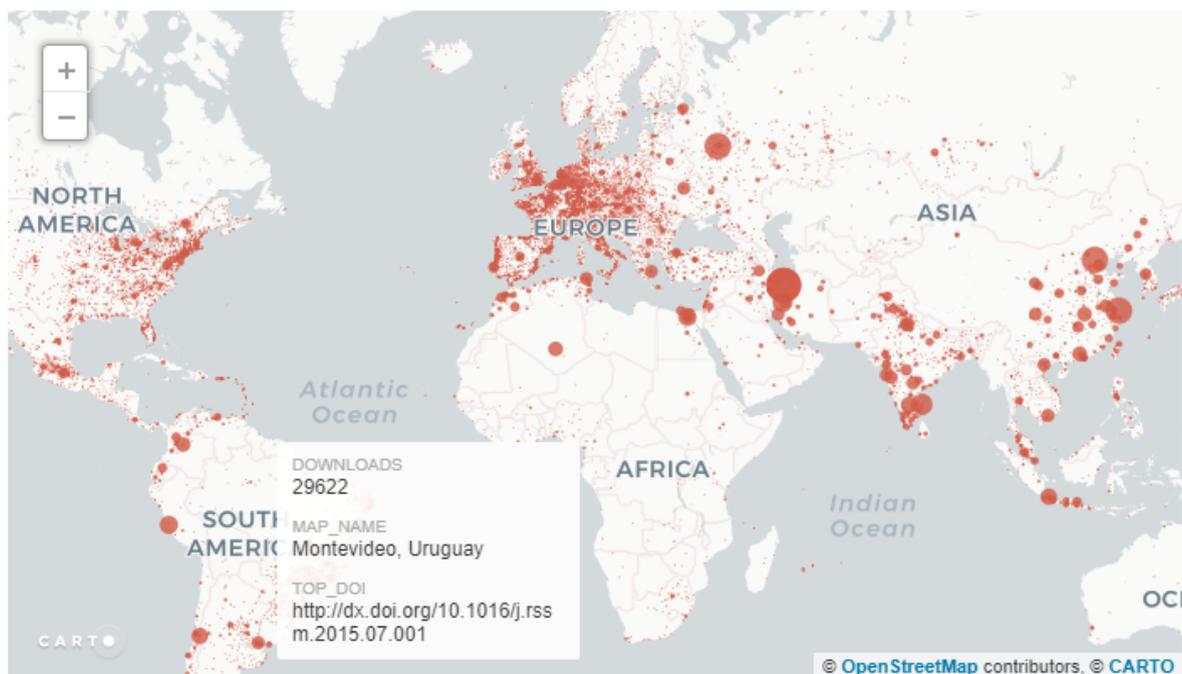


Imagen tomada de :(Himmelstein & Rodriguez et al, 2018)

En un estudio de la Revista Science del 2016 John Bohanons “Who's downloading pirated papers” revela datos de que toda la comunidad científica en general utiliza SciHub, aquí puede verse un gráfico que incluye el uso en Uruguay.

It's a Sci-Hub world

Server log data for the website Sci-Hub from September 2015 through February paint a revealing portrait of its users and their diverse interests. Sci-Hub had 28 million download requests, from all regions of the world and covering most scientific disciplines.



Tomado de: (Bohanons, 2016)

En la encuesta cerca del 60% de los encuestados dijo usar Sci-Hub mensualmente y un cuarto diaria o semanalmente, y el 88% respondieron que no estaba mal descargar documentos pirateados. (Travis, 2016)

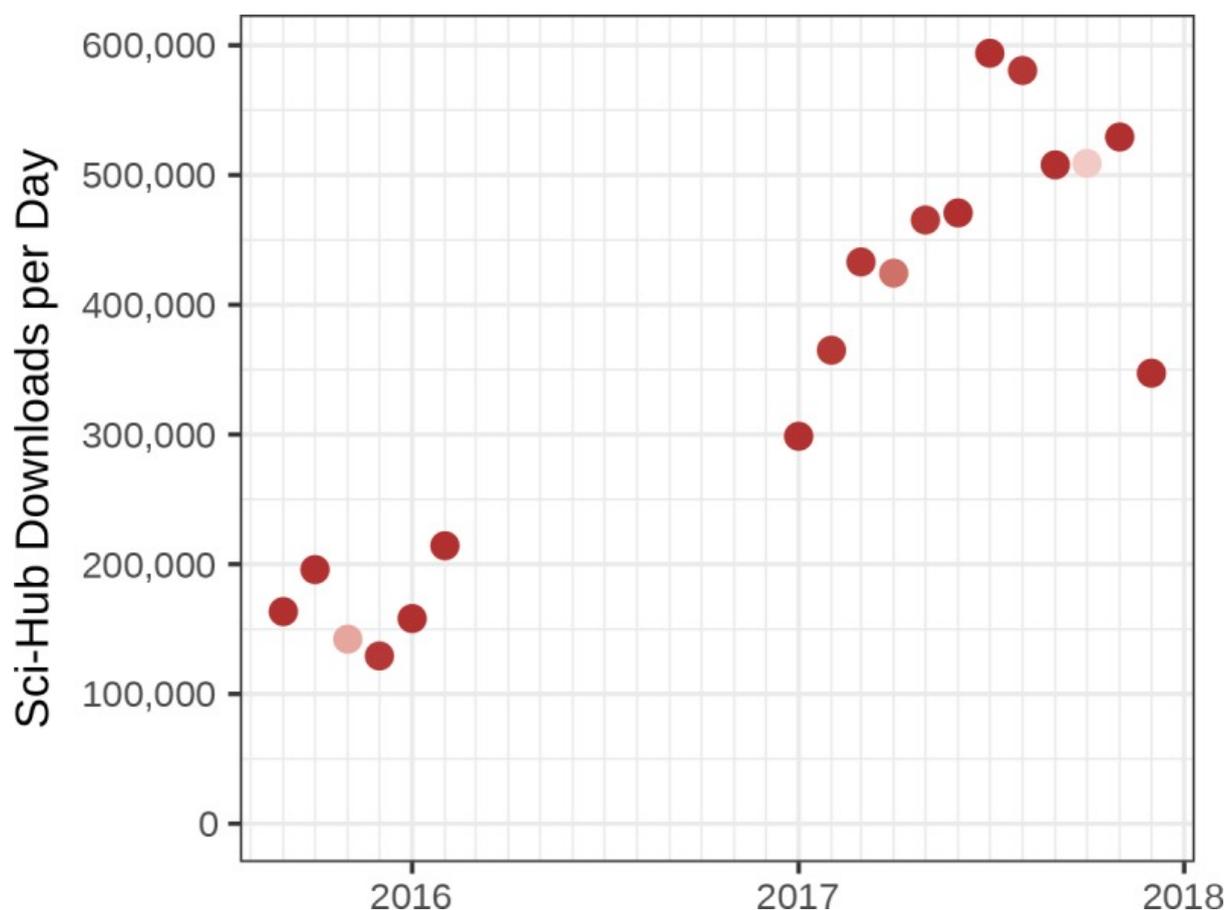


Imagen tomada de: (Himmelstein et al, 2018)

El 3 de junio de 2015 Elsevier presentó una demanda ante un tribunal en Nueva York contra Alexandra Elbakyan por infracción de derechos de copyright y por violación de la "Computer Fraud and Abuse Act". Elbakyan era la única persona nombrada en la demanda, que también era dirigida contra los sitios web Sci-Hub, Library Genesis y sus responsables. En noviembre de 2015 el juez del distrito de Nueva York Robert W. Sweet emitió una orden provisional de cierre contra Sci-Hub, haciendo que el dominio fuera inaccesible. El sitio volvió a estar accesible ese mismo mes tras cambiar el dominio de Sci-Hub a ".tw"

Sci-Hub tiene un sitio en la deepweb <http://scihub222660qcxt.onion/>

En diciembre de 2016, la Revista Nature nombró a Elbakyan como una de las diez personas más destacadas del año debido al impacto que ha tenido Sci-Hub en el mundo de la ciencia. (Nature, 2016)

Cómo Sci-Hub posee un sitio en la Deep Web será analizado en profundidad más adelante.

2.2.4 LibGen y Researchgate

Library Genesis es una biblioteca digital que permite el libre acceso a artículos científicos, libros y normas técnicas, al 5 de junio de 2018, contaba con más de 2.7 millones de libros y 58 millones de artículos.

LibGen está bloqueada por proveedores de internet en Reino Unido y en octubre de 2015 una corte del distrito de Nueva York ordenó que el dominio libgen.org fuera dado de baja, pese a lo cual el sitio sigue accesible a través de otros dominios alternativos.

In its declaratory judgment, the Court awarded Elsevier a maximum \$15 million in damages based on the intentional nature of the infringement of a representative sample of 100 works. Along with the maximum damages permitted by law, the judgment makes permanent a 2015 preliminary injunction that required U.S. domain name registries to suspend the defendants' U.S.-administered domain names, which the AAP believes will be a deterrent to those who support or do business with illegal operators. (Madigan, 2017)

LibGen ofrece un sitio en la deepweb en: <http://genotypeinczgrxr.onion/>

Author	Journal	Year	Month	Day	Volume	Issue	First page	Last page	File Size	Actions
G., T.	Collected Papers of Charles Sanders Peirce	1933	5	6	131	3314	639	639	114kB	MD5 Errors Edit BibTeX dx.doi.org Abstract
	Collected Papers of Charles Sanders Peirce	1935	1	26	135	3404	131	131	106kB	MD5 Errors Edit BibTeX dx.doi.org Abstract
G., T.	Collected Papers of Charles Sanders Peirce	1936	12	19	138	3503	1037	1037	109kB	MD5 Errors Edit BibTeX dx.doi.org Abstract

Como Library Genesis tiene un sitio en la web profunda será analizado en profundidad más adelante.

Advance your research

Discover scientific knowledge, and make your research visible.

Join for free

 Connect with LinkedIn  Connect with Facebook

"ResearchGate is changing how scientists share and advance research."

Forbes

Researchgate es una red social académica, de científicos e investigadores que permite compartir artículos, formular y responder preguntas y encontrar colaboradores. Cuenta con 13 millones de usuarios pese a que no es necesario registrarse para utilizarla. La sección "Haz preguntas, obtén respuestas" en la que los investigadores pueden encontrar posibles soluciones a sus problemas de investigación. Actualmente, cuenta con 200.000 preguntas y 1 millón de respuestas, un 75% de las cuales son contestadas en al menos 48 horas.

ResearchGate, a for-profit firm based in Berlin that was founded in 2008, is one of the largest social networking sites aimed at the academic community. It claims more than 13 million users, who can use their personal pages to upload and share a wide range of material, including published papers, book chapters, and meeting presentations. Science funders and investors have put substantial funds into the firm; it has raised more than \$87 million from the Wellcome Trust charity, Goldman Sachs, and Bill Gates personally. (Singh, 2017)

Pese a ser un sitio ampliamente utilizado en 2017 Elsevier y la American Chemical Society (ACS) presentaron una demanda en Alemania contra ResearchGate, denunciando una infracción de derechos de autor a gran escala.

As a result, two coalition members—ACS and Elsevier—have opted to go to court to try to force ResearchGate's hand. The lawsuit, filed in a German regional court, asks for "clarity and judgement" on the legality of posting such content, says James Milne, spokesperson for the Coalition for Responsible Sharing and senior vice president of ACS's journals publishing group in Oxford, U.K. (Singh, 2017)

Todas estas iniciativas para compartir contenidos académicos por parte de los autores han sido continuamente atacadas, perseguidas e ilegalizadas, las comunidades académicas y sus miembros, científicos e investigadores conviven con la sensación de estar rompiendo reglas y coqueteando con la ilegalidad en sitios tan prestigiosos como Researchgate o tan cuestionados como Sci Hub.

2.2.5 Piratería o Autonomía

“En una civilización sin barcos, los sueños se secan, el espionaje reemplaza a la aventura y la policía a los piratas.” (Foucault, 1967)

WHO'S DOWNLOADING PIRATED PAPERS? EVERYONE

Data from the controversial website Sci-Hub reveal that the whole world turns to it for journal articles

By John Bohannon

Tomado de (Bohannon, 2016)

Hemos visto repetidamente el término “pirata” o “piratería” con su connotación de robo o apropiación indebida, pero esta “cultura pirata” a medida que se extiende, pierde su connotación criminal y da un paso hacia el cambio cultural. Cambio que es resistido y condenado como dice el filósofo francés Giles Deleuze:

Las viejas sociedades de soberanía manejaban máquinas simples, palancas, poleas, relojes; pero las sociedades disciplinarias recientes se equiparon con máquinas energéticas, con el peligro pasivo de la entropía y el peligro activo del sabotaje; las sociedades de control operan sobre máquinas de tercer tipo, máquinas informáticas y ordenadores cuyo peligro pasivo es el ruido y el activo la piratería o la introducción de virus (Deleuze, 2006).

Esta nueva cultura pone en jaque al control por parte de los estados y empresas que intentan imponer reglas contra el acceso y la autonomía. La cultura pirata “saltea” como dice el sociólogo español Manuel Castells la discusión de que es legal o no, pues al entender que una importante parte de la población recurre a estos

canales alternativos para obtener conocimiento y siendo una cultura de medios, dependiente del conocimiento, la actitud frente a los impedimentos y bloqueos no es raro que sea activa y responsiva.

We do not intend to discuss whether we are dealing with legal or illegal practices; our launching point for this analysis is that, when a very significant proportion of the population is building its mediation through alternative channels of obtaining content, such behavior should be studied in order to deepen our knowledge of media cultures. Because we need a title to characterize those cultures in all their diversity (but at the same time, in their commonplaceness) we propose to call it Piracy Cultures. (Castells, 2013).

Estamos frente a un fenómeno cultural que va más allá de clases sociales, nacionalidades o edades, estamos inmersos en una cultura de redes, una cultura de compartir, una cultura que abre canales y cuestiona las normas.

Como dice el filósofo uruguayo Rafael Capurro “Las normas morales y legales están expuestas ahora al impacto de la técnica digital, y en especial de Internet” (Capurro, 2016).

Frente a las necesidades informativas no resueltas, priorizando el compartir información los actores del ámbito académico también se vuelcan a redes no convencionales, soluciones alternativas mayoritariamente dentro de la ilegalidad.

Muchos de los responsables de crear y mantener estos sitios “piratas” pese a ser perseguidos por las autoridades, son reconocidos y premiados por parte de instituciones y la propia comunidad científica.

Pero estos muchas veces llamados “delincuentes” son también actores de cambio. y como dijo Aaron Swartz la cuestión es “hacernos escuchar”.

En el antiguo sistema de broadcasting, el límite era la cantidad de espacio de las ondas radiofónicas. Sólo se podían transmitir 10 canales de televisión, e inclusive con el cable, tenías 500 canales. En Internet todos pueden tener un canal. Todos pueden tener un blog o una página en (una red social) MySpace. Todos tienen la posibilidad de expresarse. Lo que vemos ahora, no es quien tiene acceso, sino quién controla la forma de encontrar. Empiezas a ver la centralización del poder en sitios como Google, éstos guardianes que te dicen adónde debes ir en Internet. Quien puede darte información o noticias. Ya no solo un grupo de personas tiene licencia para hablar, ahora todos pueden hablar. La cuestión es quién se hace escuchar. Aaron Schwartz en: The internet own boy (2014)

Your increasingly obsolete information industries would perpetuate themselves by proposing laws, in America and elsewhere, that claim to own speech itself throughout the world. These laws would declare ideas to be another industrial product, no more noble than pig iron. In our world, whatever the human mind may create can be reproduced and distributed infinitely at no cost. The global conveyance of thought no longer requires your factories to accomplish. (Barlow, 2019)

John Perry Barlow plantea que las ideas no son “otro producto industrial”, en su famosa “declaración de independencia del ciberespacio” propone que en un entorno

digital cualquier producto de la mente humana debe ser reproducido infinitamente sin costo. Afirma que las leyes que las limitan son propuestas por “industrias de información obsoletas” que buscan perpetuarse a sí mismas declarando a las ideas como otro producto industrial “sin más nobleza que el acero crudo”.

2.2.6 Privacidad y Censura

“Me he dado cuenta de un muy interesante punto de discusión, que es, eso que la gente solía llamar libertad, y a la libertad ahora la llamamos privacidad. Y decimos, en la misma categoría, que la privacidad está muerta. Esto es algo que realmente me preocupa sobre mi generación sobre todo cuando hablamos acerca de cómo no estamos sorprendidos por nada. Creo que debemos considerar que cuando perdemos privacidad perdemos acción, perdemos la libertad misma porque ya no nos sentimos libres para expresar lo que pensamos.” Jacob Appelbaum en CitizenFour (2014)

La privacidad es otra de las razones principales por las cuales como individuos o como comunidad decidimos escapar de la visibilidad de la web auditada e indexada y como plantea Himanen es una cuestión ideológica pero también técnica.

En el futuro, la privacidad será una cuestión no sólo ética sino también tecnológica. La realización técnica de las redes electrónicas tiene un gran impacto sobre el derecho a la privacidad individual. La defensa de la privacidad pasa a ser un esfuerzo exigente de cooperación: además de salvaguardar Internet, la influencia ha de ejercerse sobre muchas otras redes que almacenan datos sobre las vidas de las personas. (Himanen, 2002)

El abordaje que hace este trabajo combinará las discusiones éticas con las discusiones tecnológicas, pues cómo vimos con los muros de pago detrás de cada decisión técnica, optar por una tecnología de conexión, o crear un espacio de libertad hay discusiones y posicionamientos ideológicos.

Muchas personas con las que he hablado han mencionado que son cuidadosos con lo que escriben en los motores de búsqueda, porque saben que les están grabando y eso limita las fronteras de su exploración intelectual. Y estoy más dispuesto a correr el riesgo de encarcelamiento o cualquier otro resultado negativo personalmente de lo que estoy dispuesto a arriesgar la restricción de mi libertad intelectual y la de los que me rodean. Edward Snowden, Citizenfour (2014)

Snowden plantea que el comportamiento informacional de las personas cambia si estas sospechan que pueden estar siendo vigiladas o los datos de sus investigaciones son monitoreados.

2.2.7 Apagón

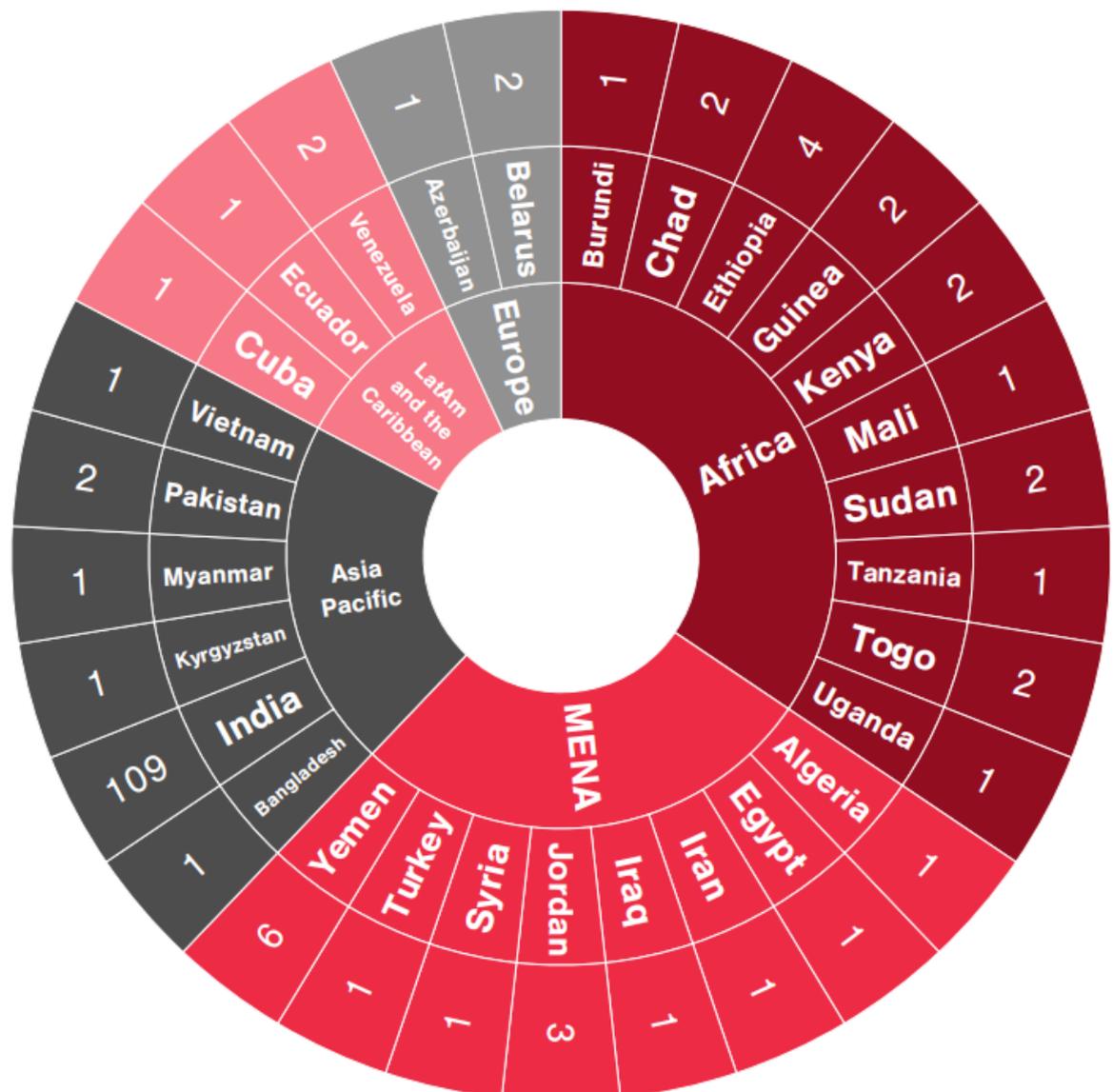
“En la era de la información la verja corporativa del monasterio está custodiada por un San Pedro informatizado que se distingue del Dios omnisciente sólo porque no muestra conmiseración.” (Himanen, 2002)

En muchos países servicios páginas web o aplicaciones han sido prohibidas y muchas veces los gobiernos recurren a cortes totales o parciales de internet.

La campaña “Keep it on” de la organización “Access Now” reúne a 243 organizaciones de 105 países para denunciar y visibilizar la censura en Internet. En su informe del 2021 “Shattered dreams and lost opportunities” reúne interrupciones, cortes y censura en Internet por parte de los gobiernos locales.

The #KeepItOn campaign unites and organizes the global effort to end internet shutdowns. The coalition is growing rapidly, and so far 243 organizations from 105 countries around the world have joined the movement, ranging from research centers to rights and advocacy groups, detection networks, foundations, and media organizations.(AccessNow,2021)

Internet shutdowns by region in 2020 ▾

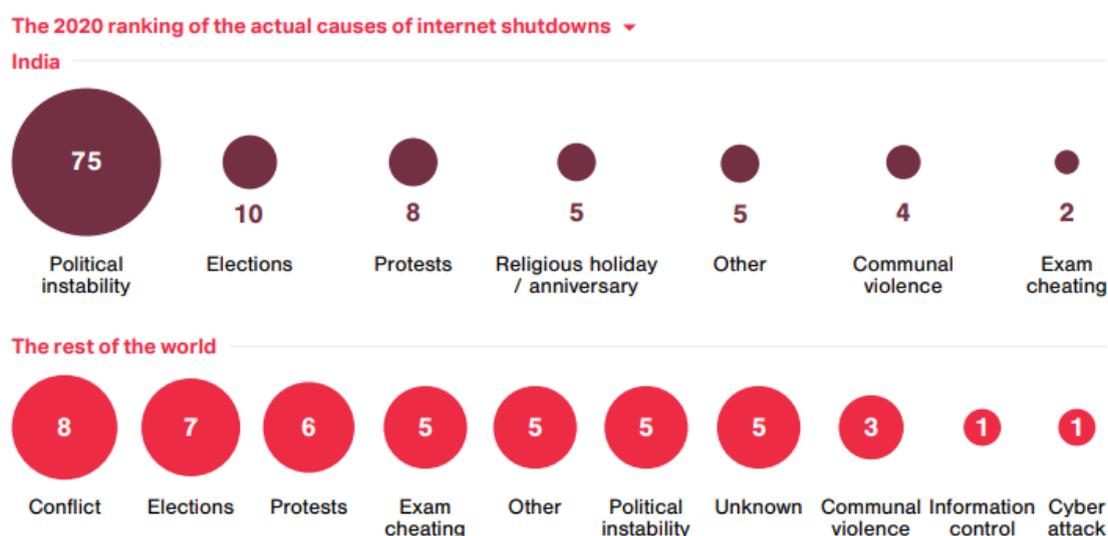


Tomado de: (AccessNow, 2021)

Este gráfico del reporte 2020 del sitio Access Now muestra los apagones de Internet en diferentes regiones de las cuales no está excluida Latinoamérica.

Continuing the pattern from 2019 and 2018, governments shut down the internet in evident attempts to hide political instability, respond to communal violence, suppress opposition groups, claim purported victory in disputed elections, thwart protests, and stop students from cheating during exams. One notable trend in 2020 is an increased number of internet shutdowns being deployed in response to ongoing violence — particularly in active conflict zones. We have long pointed out that amid conflict, shutdowns can hide human rights violations or war crimes, thwart journalism, and put people’s lives in danger. (AccessNow, 2021)

El informe marca como causa principal de las interrupciones intentar encubrir la inestabilidad política del país. Apuntan que la censura de Internet encubre violaciones de los derechos humanos, crímenes de guerra y bloqueos al periodismo.



Tomado de: (AccessNow, 2021)

Interrumpir Internet por parte de un gobierno interfiere con el derecho a acceder a la información y vulnera la libertad de expresión. En 2020 organizaciones de derechos humanos han denunciado que al menos 17 de estas interrupciones en la comunicación fueron acompañadas de violaciones flagrantes de los derechos humanos.

Not only do internet shutdowns interfere with the rights to access information, freedom of expression, and other fundamental freedoms, they are used in attempts to hide egregious human rights violations. In 2020, at least 17 incidents of internet shutdowns were accompanied by blatant rights violations. In Ethiopia, Belarus, India, Guinea, and other countries, human rights organizations and civil society groups reported patterns of abuse taking place during a shutdown. (AccessNow, 2021)

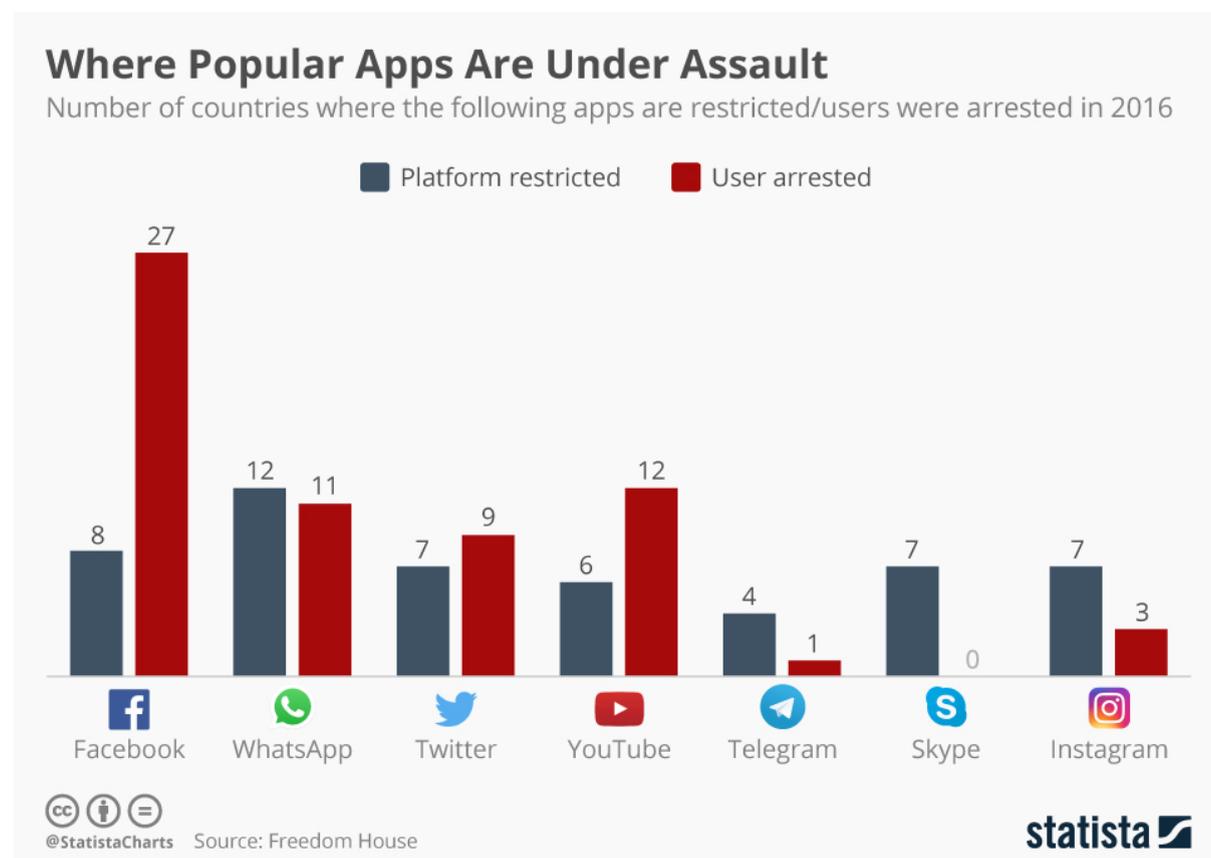
Las justificaciones por parte de los gobiernos que incurren en estas prácticas muchas veces son tan irracionales como combatir “noticias falsas” o “discursos de

odio” en redes y evitar la violencia, pero en realidad los hechos sugieren que la censura y el impedimento al acceso generan más hechos violentos de los que pudieran llegar a evitar.

An internet shutdown is an inherently disproportionate interference with the right to free expression, yet in 2020, numerous countries imposed shutdowns on the pretext of combating “fake news,” violence-inciting hate speech online, and to respond to other content moderation issues. As we note above, authorities in India shut down the internet repeatedly while citing “rumors spread on social media,” including in Manipur, where the state government cut access to the internet for three days in an attempt to prevent clashes between two villages over a land dispute. The official rationale was that social media had become useful for rumor-mongers to incite violence, and unless the use of internet and mobile data was curbed temporarily, “there is a likelihood of deterioration of law and order and communal violence in the state.” Yet as we have noted, research suggests that internet shutdowns may encourage, not discourage, violence. (AccessNow, 2021)

Existen bloqueos temporales como vimos antes pero también ciertas páginas o aplicaciones son prohibidas en muchos países con consecuencias penales para quienes tengan el coraje de usarlas, Mc Carthy en su informe “Where Popular Apps Are Under Assault” marca que durante 2016 personas de 27 países fueron arrestadas por acceder a Facebook y personas de 11 países diferentes por ser usuarios de Whatsapp.

WhatsApp is blocked in twelve countries while Facebook access has been cut off in eight. Twitter is also banned by seven nations. Users of these apps have also been the subject of the authorities' wrath and arrested for posting political, social or religious content. Facebook users have been arrested in 27 countries, while Whatsapp users have been targeted by the authorities in eleven nations. (McCarthy, N, 2016)



Datos extraídos del reporte de transparencia de Google nos muestran las prohibiciones a sus servicios:

Actualmente el servicio YouTube es bloqueado en China desde marzo del 2009, también la Búsqueda web de Google desde de junio de 2014 (en curso), Gmail está bloqueado desde el 26 de diciembre de 2014, en la República del Congo se ha bloqueado la Búsqueda web de Google desde 19 de octubre de 2015. (Google, 2018)

Otros bloqueos han sido temporales: Irán 24 de septiembre de 2012, 4 días 21 horas 53 minutos, Irán Google y Gmail. Gmail, 19 de febrero de 2012, 23 horas 59 minutos. Irán: 9 de febrero de 2012, 2 días 6 horas 12 minutos Gmail, Google, Yahoo, y sitios que usen el protocolo "HTTPS". Tayikistán: YouTube y Twitter, 25 de agosto de 2015, 658 días 21 horas 50 minutos 5 de abril de 2015, 23 horas 27 minutos. (Google, 2018)

Un estudio del sitio web Freedom House llamado "Silenciando al mensajero: aplicaciones de comunicación presionadas" indica que en 2016 la libertad en Internet ha disminuído por sexto año consecutivo, con más gobiernos interfiriendo con aplicaciones y redes sociales:

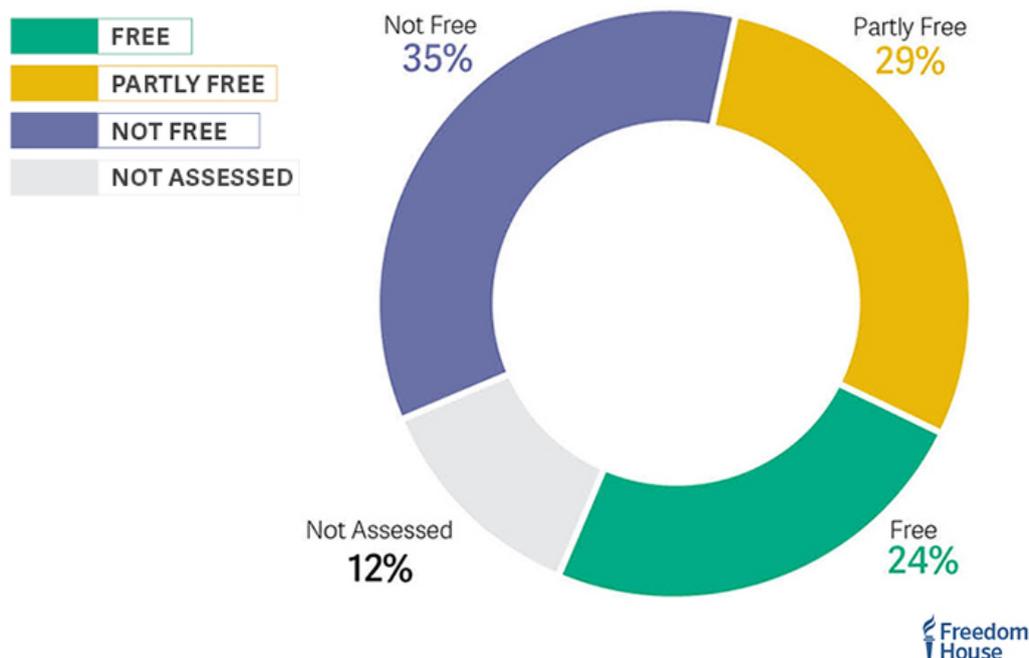
Internet freedom has declined for the sixth consecutive year, with more governments than ever before targeting social media and communication apps as a means of halting the rapid dissemination of information, particularly during anti-government protests. Public-facing social media platforms like Facebook and Twitter have been subject to growing censorship for several years, but in a new trend, governments increasingly target voice communication and messaging apps such as WhatsApp and Telegram. These services are able to spread information and connect users quickly and securely, making it more difficult for authorities to control the information landscape or conduct surveillance. (Freedom on the Net, 2016)

Este estudio indica que 2/3 de los usuarios de Internet viven en países donde la crítica al gobierno, militares, o la familia real son objeto de censura. Usuarios de Internet enfrentan condenas en 38 países y globalmente. El 27% de todos los usuarios de Internet vive en países donde una persona ha sido arrestada por publicar, compartir o simplemente dar "me gusta" en Facebook.

Two-thirds of all internet users 67 percent live in countries where criticism of the government, military, or ruling family are subject to censorship. Social media users face unprecedented penalties, as authorities in 38 countries made arrests based on social media posts over the past year. Globally, 27 percent of all internet users live in countries where people have been arrested for publishing, sharing, or merely "liking" content on Facebook. "(Freedom on the Net, 2016)

GLOBAL INTERNET POPULATION BY 2016 FOTN STATUS

FOTN assesses 88 percent of the world's internet user population.



Tomado de: (Freedom on the Net, 2016)

“Freedom on the Net is a comprehensive study of internet freedom in 65 countries around the globe, covering 88 percent of the world's internet users.” (Freedom on the Net, 2016)

Este gráfico lo aporta Freedom House, quien afirma relevar datos del 88 por ciento de los usuarios de Internet y refleja que la mayoría de los usuarios relevados no se sienten libres en Internet.

Otros datos muestran que no es solo una percepción que el hostigamiento y la persecución a nivel global han aumentado, los datos de incremento en el uso de VPN y redes Tor, estas alternativas a la conexión tradicional (que veremos en profundidad más adelante) son una respuesta a esta falta de libertad.

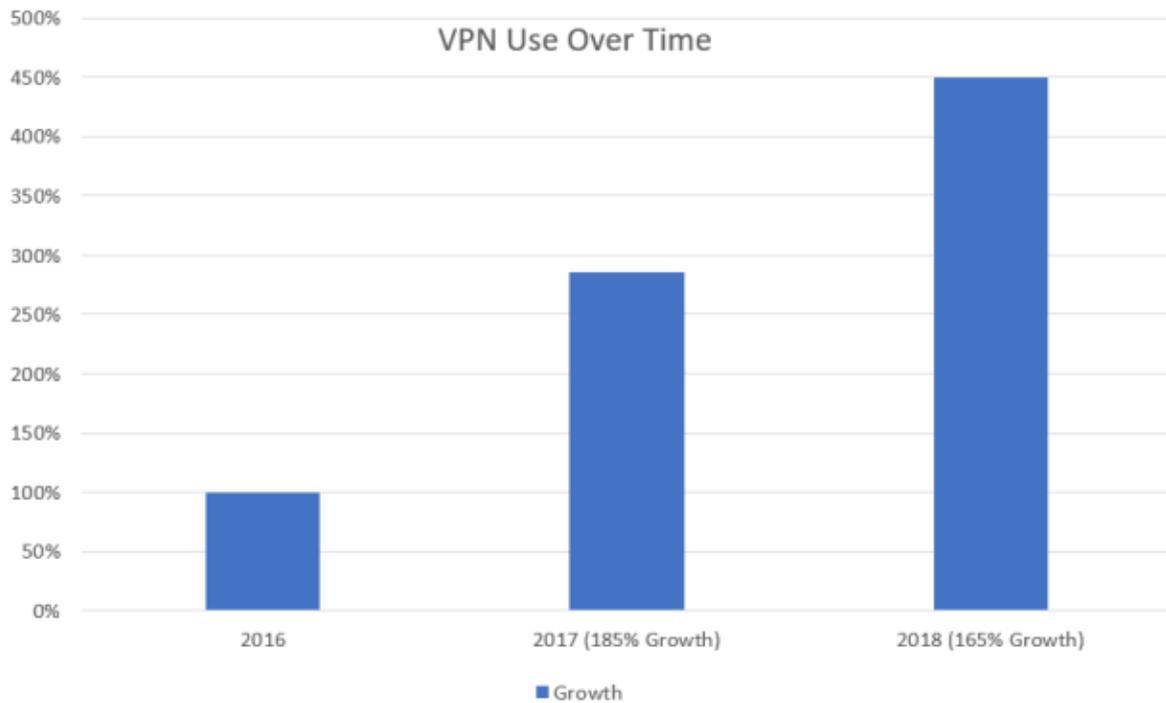


Imagen: <https://www.vpnmentor.com/blog/vpn-use-data-privacy-stats/>

Pero también las redes de anonimato sufren intentos de interrupciones por parte de los diferentes gobiernos. Tor project publica estadísticas de los países que han intentado irrumpir en sus redes de anonimato (según el sitio) buscando la censura. Los gobiernos de China e Irán más de mil veces han intentado interferir con el anonimato de las redes onion.

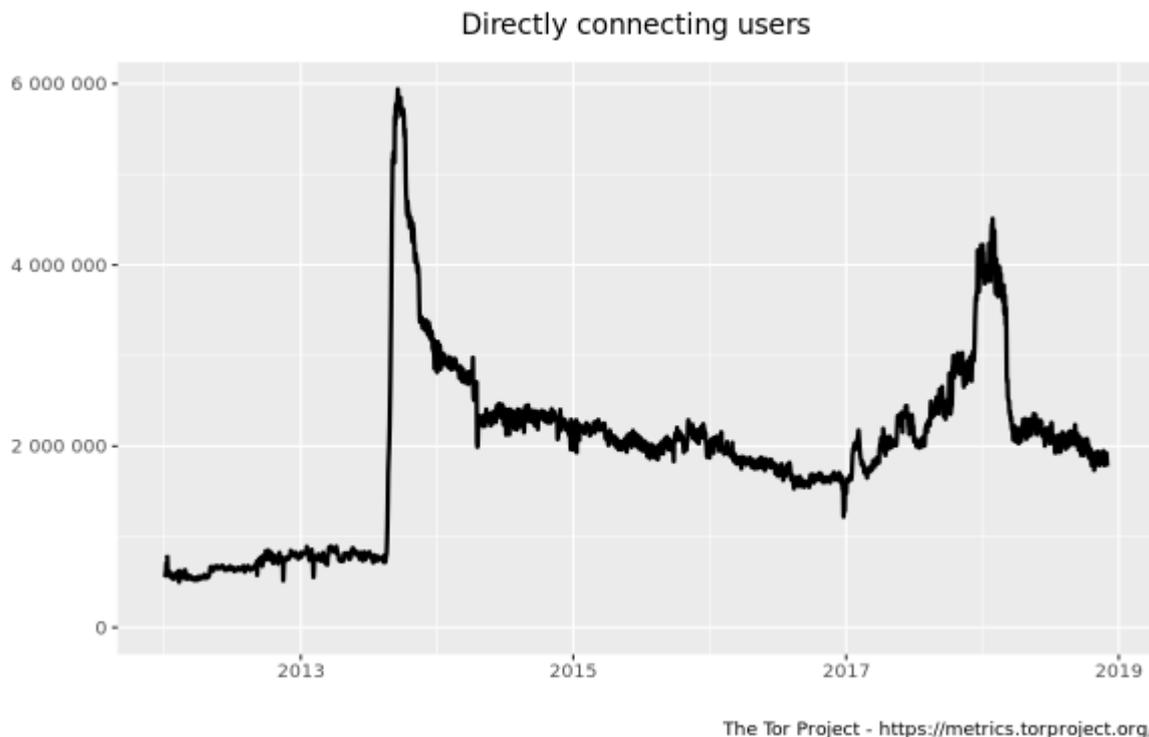
Country	Downturns	Upturns
China	374	437
Iran	240	337
Turkey	225	307
Mongolia	212	270
Lithuania	194	243
Armenia	182	245
Kyrgyzstan	171	235
Guyana	163	189
Egypt	162	215
Ethiopia	135	200

This table shows the top-10 countries by possible censorship events, as obtained from an anomaly-based censorship-detection system (for more details, see this [technical report](#)). For further details see these [questions and answers about user statistics](#).

Tomado de: (The Tor Project, 2021)

2.2.8 Aumento del uso de redes de anonimato en 2014 y 2018

Analizando las estadísticas de aumento de usuarios de la principal red de anonimato Tor descubrí que se producían dos picos muy marcados en 2014 y 2018.



Tomado de: The Tor Project. (2018)

En esta gráfica de usuarios de Tor se pueden ver dos saltos tremendamente significativos que corresponden con los dos escándalos relativos a la privacidad en la Red.

El primer salto se da a finales del 2013 aquí una Cronología de los hechos relacionados a la violación de la privacidad:

16/08/2013	Un empleado de la Agencia de Seguridad Nacional (en ese momento anónimo) y la CIA denuncia a medios de prensa que los servicios de seguridad de EEUU, vulneraron la privacidad de miles de personas residentes y en el extranjero. El documento, fechado en mayo de 2012, registra 2.776 incidentes en los 12 meses precedentes de almacenamiento o acceso a comunicaciones protegidas de los ciudadanos por correo electrónico o teléfono.
2/09/2013	El Gobierno brasileño informa que ha pedido explicaciones "rápidas" y "por escrito" a Estados Unidos, sobre la denuncia de que sus servicios de inteligencia espionaron los teléfonos y correos electrónicos de la presidenta Dilma Rousseff. México también

	solicita a Estados Unidos una "investigación exhaustiva" sobre las informaciones de un supuesto espionaje al presidente del país, Enrique Peña Nieto.
11/11/2013	La Agencia de Seguridad Nacional estadounidense reconoce que investigó indebidamente miles de llamadas telefónicas de ciudadanos estadounidenses, violando sus propias reglas. El Director Nacional de Inteligencia, James Clapper, divulgó varios documentos que así lo demuestran.
19/11/2013	Edward Snowden obtuvo los documentos secretos que revelaron el programa masivo de espionaje de Estados Unidos del servidor web interno de la Agencia de Seguridad Nacional (NSA), según informó un responsable de esta fuente gubernamental estadounidense a la cadena de radio pública NPR.
24/10/2013	El ministro de Asuntos Exteriores de Alemania, Guido Westerwelle, cita al embajador de Estados Unidos en Berlín, John B. Emerson, en relación con el supuesto espionaje del teléfono móvil de la canciller Angela Merkel por parte de los servicios secretos norteamericanos, según informa un portavoz del Ejecutivo alemán.
25/10/2013	La Agencia Nacional de Seguridad de Estados Unidos (NSA) espía las llamadas telefónicas de 35 líderes mundiales, asegura el diario británico The Guardian a partir de un documento filtrado por el exagente de ese organismo estadounidense Edward Snowden.

Hablaremos más en profundidad de Edward Snowden más adelante.

Cronología extraída de: (EFE, 2013)

El segundo salto se da cuando se da a conocer el caso de Cambridge Analytica que vulneró los datos de más de 87 millones de usuarios de Facebook.

17/03/2018	The New York Times, The Guardian y The Observer denunciaron que la empresa Cambridge Analytica estaba explotando la información personal de los usuarios de Facebook, la empresa afirmaba estar haciéndolo para fines académicos. La consultora está acusada de haber obtenido la información de millones de usuarios, atentando contra las políticas de uso de la red social y de haber utilizado esos datos para crear anuncios políticos durante las elecciones presidenciales de 2016 en Estados Unidos
------------	---

21/03/2018	Mark Zuckerberg brindó una entrevista a la especialista en tecnología de CNN, Laurie Segall, en la que admitió que Facebook cometió errores.
10/04/2018	Mark Zuckerberg testifica frente al senado de los EEUU
14/05/2018	Facebook ha suspendido 200 aplicaciones de manera preventiva por haber tenido acceso a grandes cantidades de datos, antes de que Facebook cambiara sus términos de uso en cuanto a datos privados
22/05/2018	El fundador y CEO de Facebook compareció ante el presidente del Parlamento Europeo, Antonio Tajani, y la Conferencia de Presidentes para dar explicaciones sobre el escándalo de Cambridge Analytica

Cronología tomada de (Meredith, 2018)

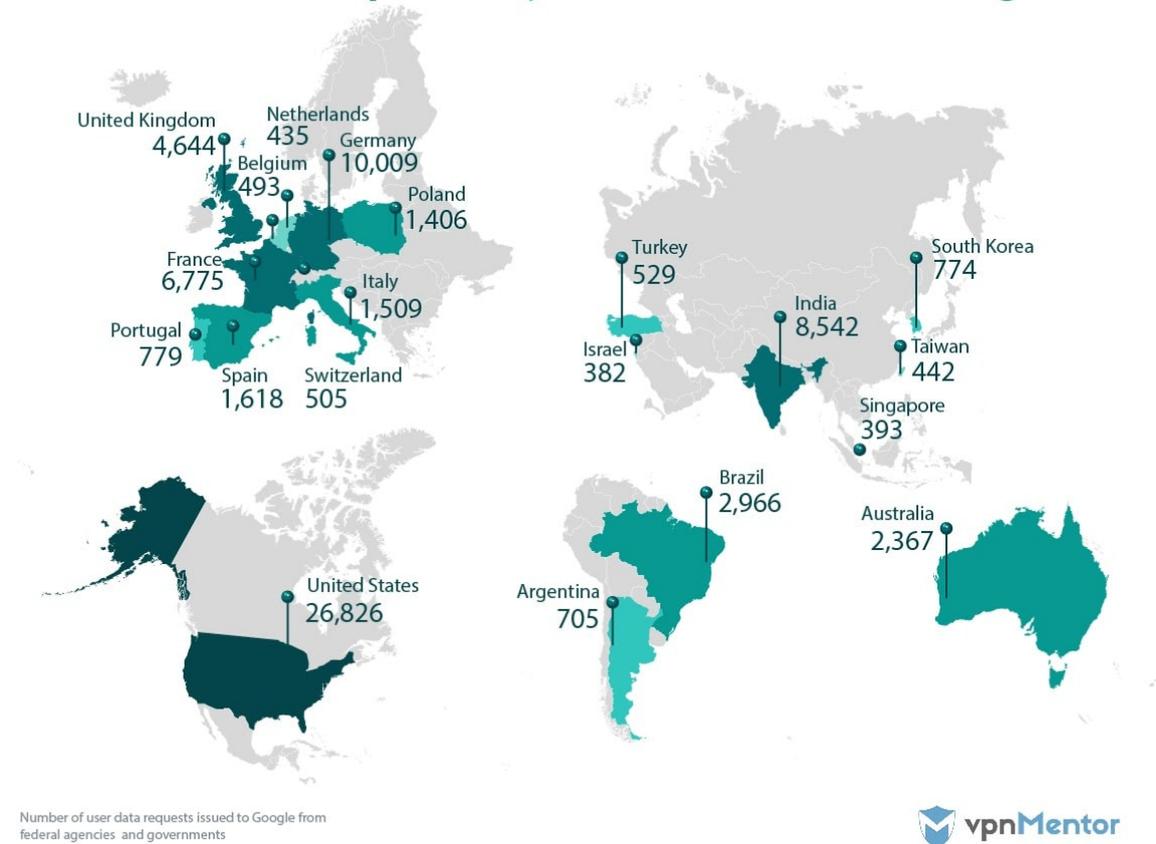
Contrastando las estadísticas de la red Anónima Tor con las cronologías de los dos hechos más relevantes de violación de la privacidad de la historia reciente se puede intuir que frente a este tipo de amenazas el tráfico en la web profunda aumenta exponencialmente.

Pero esto no es algo del pasado, Google presenta en su informe de transparencia las solicitudes de información por parte de los gobiernos sobre usuarios a nivel mundial

“Existe una variedad de leyes que permiten que las agencias gubernamentales de todo el mundo soliciten información sobre los usuarios con fines civiles, administrativos, penales y de seguridad nacional. En este informe, compartimos información sobre la cantidad y el tipo de solicitudes que recibimos de las agencias gubernamentales.” (Google, 2021)

Vpn Mentor nos muestra en el gráfico de debajo los pedidos mensuales de información por parte de los gobiernos en el cual Latinoamérica tiene una participación importante.

User Data Requests by Governments to Google



Tomado de: (Vpn Mentor, 2021)

Aunque en el gráfico anterior no se ve Uruguay si está disponible en el sitio oficial de Google con récord de pedidos a finales de 2018 y a finales de 2020.

Uruguay ▾ Sin límite de tiempo ▾

Total de solicitudes y cuentas por periodo del informe

Período del informe	Solicitudes para la divulgación de información sobre usuarios	Cuentas	Porcentaje de solicitudes para las que se generaron datos
▶ jul. de 2020 - dic. de 2020	4	5	0 %
▶ ene. de 2020 - jun. de 2020	3	3	33 %
▶ jul. de 2019 - dic. de 2019	1	1	0 %
▶ jul. de 2018 - dic. de 2018	1	1	0 %
▶ ene. de 2018 - jun. de 2018	5	5	20 %
▶ jul. de 2017 - dic. de 2017	2	2	0 %
▶ ene. de 2017 - jun. de 2017	2	2	0 %
▶ jul. de 2016 - dic. de 2016	1	1	0 %
▶ jul. de 2015 - dic. de 2015	1	1	0 %

Tomado de: (Google, 2021)

2.2.9 Propublica y Freedom of the Press



Es una organización sin fines de lucro fundada en 2007 que produce periodismo investigativo de interés público. Fue galardonada con 4 premios Pulitzer 2010, 2011, 2016, 2017. Propublica creó su propio sitio en la web profunda en enero del 2016 para que sus usuarios puedan navegar anónimamente y para demostrar que la web profunda no es sitio oscuro que se pretende mostrar “the benefits to readers of having a more anonymous way to browse our site, and why the dark web isn’t necessarily the seedy place it’s cracked up to be.” (Gordy, 2016)



[Donate](#) [Store](#) [Contact](#) [About](#) [Twitter](#) [Facebook](#)

[CROWDFUNDING](#) [NEWS & ADVOCACY](#) [GUIDES & TRAINING](#) [PROJECTS](#)

Freedom of the Press Foundation protects and defends adversarial journalism in the 21st century. We use crowdfunding, digital security, and internet advocacy to support journalists and whistleblowers worldwide.

Freedom of the Press Foundation, es una organización sin fines de lucro fundada en 2012 para impulsar y promocionar la libertad de expresión y la libertad de prensa, entre sus directivos está incluido el informante de la NSA Edward Snowden. Es apoyada por medios masivos de todo el mundo como: The Washington Post, The Guardian o The New Yorker.

En octubre del 2013 incorporó un software desarrollado Aaron Swartz llamado “Secure Drop”. Este software permite a informantes contactar a periodistas mediante la red anónima Tor. Freedom of the net también tiene su propio sitio en la web profunda y aporta financiamiento al proyecto Tor. Freedom of the Press fue premiada con el Premio “James Madison” en 2016 por la American Library Association.

Ya que ambas tienen servicios en la Deep Web las analizaremos más adelante en este trabajo.

2.2.9 Edward Snowden y la Desobediencia Civil

“Until I want the protection of Massachusetts (gobierno de EEUU) to be extended to me in some distant Southern port, where my liberty is endangered, or until I am bent solely on building up an estate at home by peaceful enterprise, I can afford to refuse allegiance to Massachusetts, and her right to my property and life. It costs me less in every sense to incur the penalty of disobedience to the State than it would to obey” (Thoreau, 1866)

Como vimos en la primera cronología coincidente con incremento en redes de anonimato y asociado también a Freedom of the Press está el ex empleado de la CIA y la Agencia de Seguridad Nacional Edward Snowden. Este informático exiliado en Rusia se hizo conocido en 2013 cuando denunció a medios británicos y estadounidenses que el gobierno norteamericano tenía programas de vigilancia masiva llamados PRISM y XKeyscore.

William Scheurman profesor de ciencias políticas de la Universidad de Indiana plantea que su filtración sea considerada como un acto de desobediencia civil. Scheurman plantea que la “desobediencia civil” es un acto político contrario a la ley con el objetivo de generar un cambio en la ley o en las políticas del gobierno.

Definition of civil disobedience as a ‘public, nonviolent, conscientious yet political act contrary to the law usually done with the aim of bringing about a change in the law or policies of the government’, we can interpret Snowden’s revelations as falling under its rubric. (Scheurman, 2014)

Snowden plantea que sus acciones fueron en pos de lo que “él consideraba correcto” y generar una campaña para “corregir lo incorrecto”.

Accordingly, I did what I believed was right and began a campaign to correct this wrongdoing. I did not seek to enrich myself. I did not seek to sell US secrets. I did not partner with any foreign government to guarantee my safety. Instead, I took what I knew to the public, so what affects all of us can be discussed by all of us in the light of day, and I asked the world for justice. That moral decision to tell the public about spying that affects all of us has been costly but it was the right thing to do and I have no regrets. Edward Snowden en (Scheurman, 2014)

Kimberlee Brownlee filósofa canadiense dice que los actos de Snowden no solo son desobediencia civil sino un ejercicio de responsabilidad moral a un costo personal importante.

Returning once more to Edward Snowden, his acts are not only civilly disobedient, they are morally justified. He was properly sensitive to the responsibility that public officials have to exercise first-order moral reasoning about the programs they oversee. He was willing to put that moral responsibility ahead of formal expectations even though it was personally costly and meant breaking the law. It is true that he defended his acts in part on the political grounds that the public has a right to be informed about policies that lack judicial and congressional oversight. But, behind that argument lies the moral argument that personal privacy is a fundamentally important interest which we should not allow our policies to erode without our

informed consideration. (Brownlee, 2016)

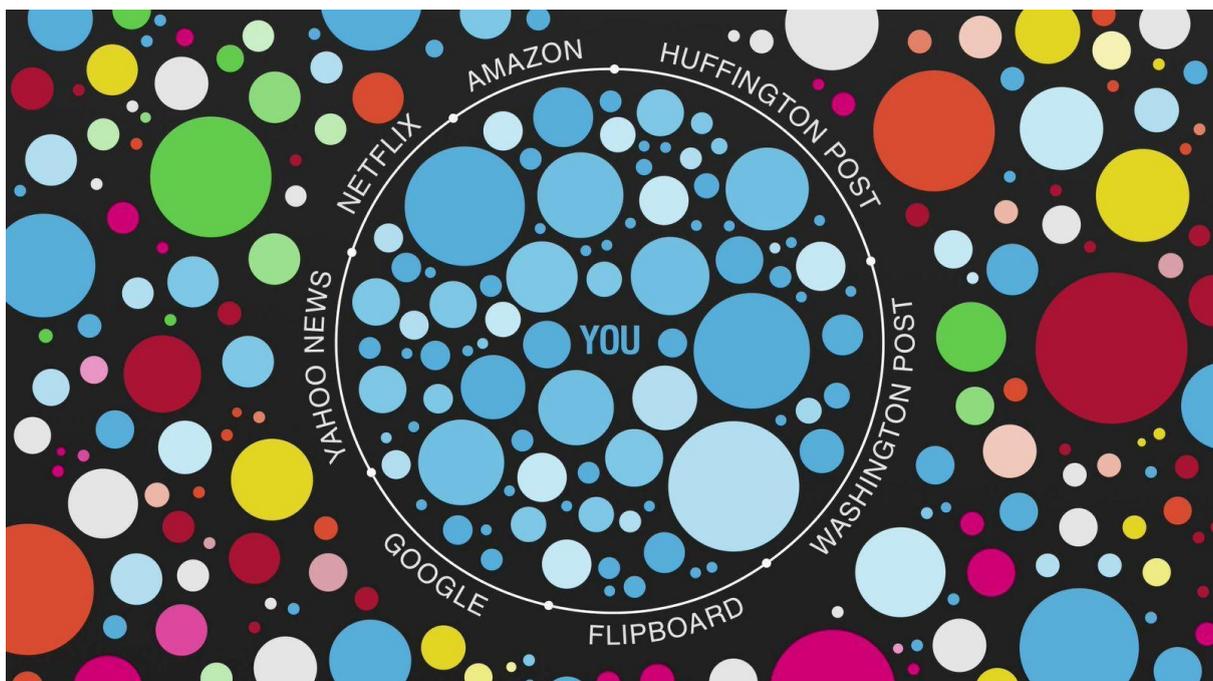
Los datos relevados para este trabajo muestran que, si el objetivo de Snowden era generar un cambio, al menos en lo que respecta a redes de anonimato y tráfico a la Deep Web esto se ha dado. Pero sin duda más que nada como plantea Himanen debemos defender nuestra privacidad si queremos llegar a tener alguna, el costo puede ser la desobediencia y el incumplimiento de disposiciones legales el beneficio puede ser nada más y nada menos que la libertad.

Debemos defender nuestra privacidad si queremos llegar a tener alguna. Tenemos que unirnos y crear sistemas que permitan las transacciones anónimas. La gente ha defendido su derecho a la privacidad durante siglos mediante susurros, penumbras, sobres, puertas cerradas, apretones de manos y mensajeros secretos. Las tecnologías del pasado no permiten una gran privacidad; en cambio, las tecnologías electrónicas sí. (Himanen, 2002)

2.2.10 Burbujas de contenido: privacidad + censura

“There's 2.1 billion people, each with their own reality. And once everybody has their own reality, it's relatively easy to manipulate them.” Roger McNamee

“La pantalla de la computadora es cada vez más una especie de espejo unidireccional que refleja tus propios intereses, mientras los analistas de los algoritmos observan todo lo que clicas” (Parisier, 2011)

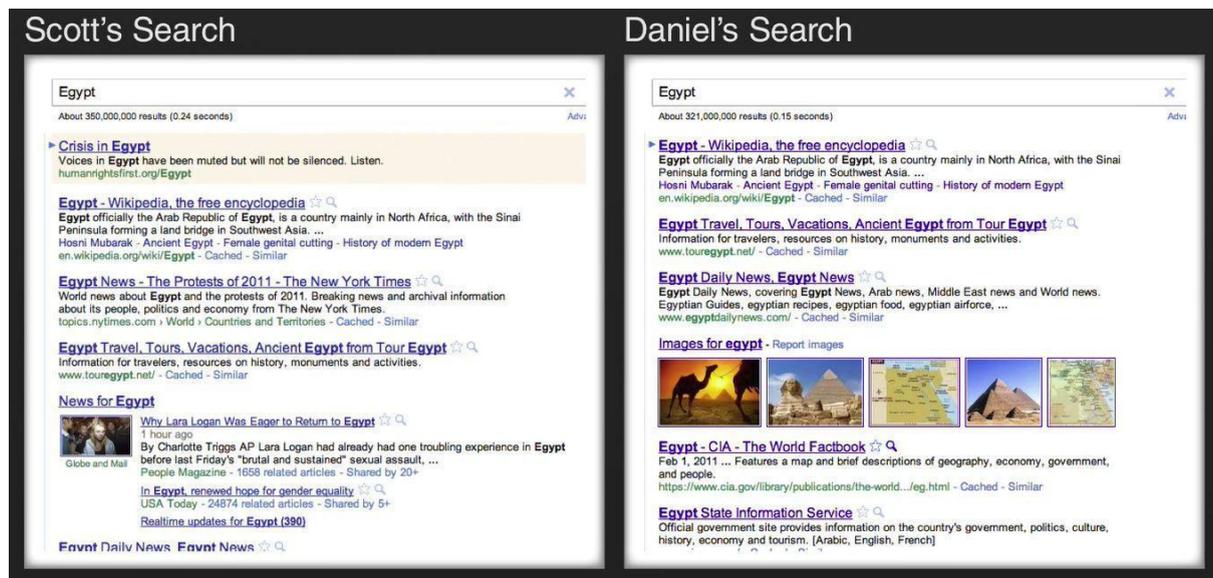


Representación gráfica de la burbuja de filtro, Eli Pariser TED Talk 2011

<https://www.youtube.com/watch?v=B8ofWFx525s>

Eli Pariser con su libro de 2011 *“The filter bubble: What the Internet is hiding from you”*, fue de los primeros investigadores en advertirnos del problema de las burbujas de filtro.

Para su charla TED del 2011 Pariser les pidió a varias personas que realizaran una misma consulta a Google, el término de búsqueda era “Egipto” entre esas personas estaba Scott y Daniel.



tomado de: Eli Pariser TED Talk 2011 <https://www.youtube.com/watch?v=B8ofWFx525s>

La búsqueda de Scott presentaba resultados relativos a las protestas en Egipto, e incluso el resultado patrocinado por Google que se remarca con un color diferente, era sobre la crisis. Por otro lado, la búsqueda de Daniel no presenta ningún resultado relacionado con política ni noticias sino más bien orientados al turismo y datos generales sobre Egipto.

Eli Pariser define en su libro, a las burbujas de filtro como: un universo de información creado por los motores de búsqueda, que fundamentalmente altera la manera en que encontramos ideas e información.

“Engines create a unique universe of information for each of us—what I’ve come to call a filter bubble—which fundamentally alters the way we encounter ideas and information.” (Pariser, 2011)

Las consecuencias y los costos de este recorte del contenido son varias, Anibal Rossi nos dice:

“Las burbujas de filtro involucran un amplio abanico de problemas como el control, la privacidad, la venta de datos, el refuerzo del sesgo de confirmación y la pérdida de iniciativa en el proceso de búsqueda” (Rossi, 2018)

Pariser por su lado remarca que las consecuencias no son solo personales sino también afectan nuestro desarrollo como sociedad.

The filter bubble's costs are both personal and cultural. There are direct consequences for those of us who use personalized filters (and soon enough, most of us will, whether we realize it or not). And there are societal consequences, which emerge when masses of people begin to live a filter-bubbled life. (Pariser, 2011)

Analicemos estas problemáticas: Las burbujas de filtro tienen un costo personal al fomentar los sesgos cognitivos especialmente el sesgo de confirmación. Hugo Saez nos dice:

Este sesgo nos lleva instintivamente a sobreestimar el valor de la información que encaja con nuestras creencias, expectativas y asunciones, lo que muchas veces nos conduce a errores. Y también nos hace subestimar e incluso ignorar la información que no coincide con lo que pensamos o creemos. Este sesgo influye mucho en la toma de decisiones, ya que si tenemos una creencia fuerte sobre lo que queremos hacer tendemos a descartar todo lo que no encaje con ella. El sesgo de confirmación es un filtro a través del cual vemos una realidad que encaja con nuestras expectativas. (Sáez, 2018)



La burbuja de contenido es por lo tanto un espacio donde son reforzadas creencias e ideas en búsqueda de generar una sensación de comodidad por parte del usuario, pero desestimula el pensamiento crítico y genera visiones sesgadas de la realidad.

Como afirma Pariser el método creado por Marchiori y que hizo famoso a Google ya no opera en pos del mejor resultado sino de uno que se ajuste al perfil que han construido con nuestros datos.

Most of us assume that when we google a term, we all see the same results—the ones that the company's famous Page Rank algorithm suggests are the most authoritative based on other pages' links. But since December 2009, this is no longer true. Now you get the result that Google's algorithm suggests is best for you in particular—and someone else may see something entirely different. In other words, there is no standard Google anymore. (Pariser, 2011)

Un estudio de la revista International Journal of Management & Information Systems

de 2011 se alinea con los que Pariser sugiere en el párrafo anterior.

Although search engines are programmed to rank websites based on their popularity and relevance, empirical studies indicate various political, economic, and social biases in the information they provide. These biases could be a direct result of economic and commercial processes (e.g., companies that advertise with a search engine can also become more popular in its organic search results), and political processes (e.g., the removal of search results in order to comply with local laws). (Seymour et al., 2011)

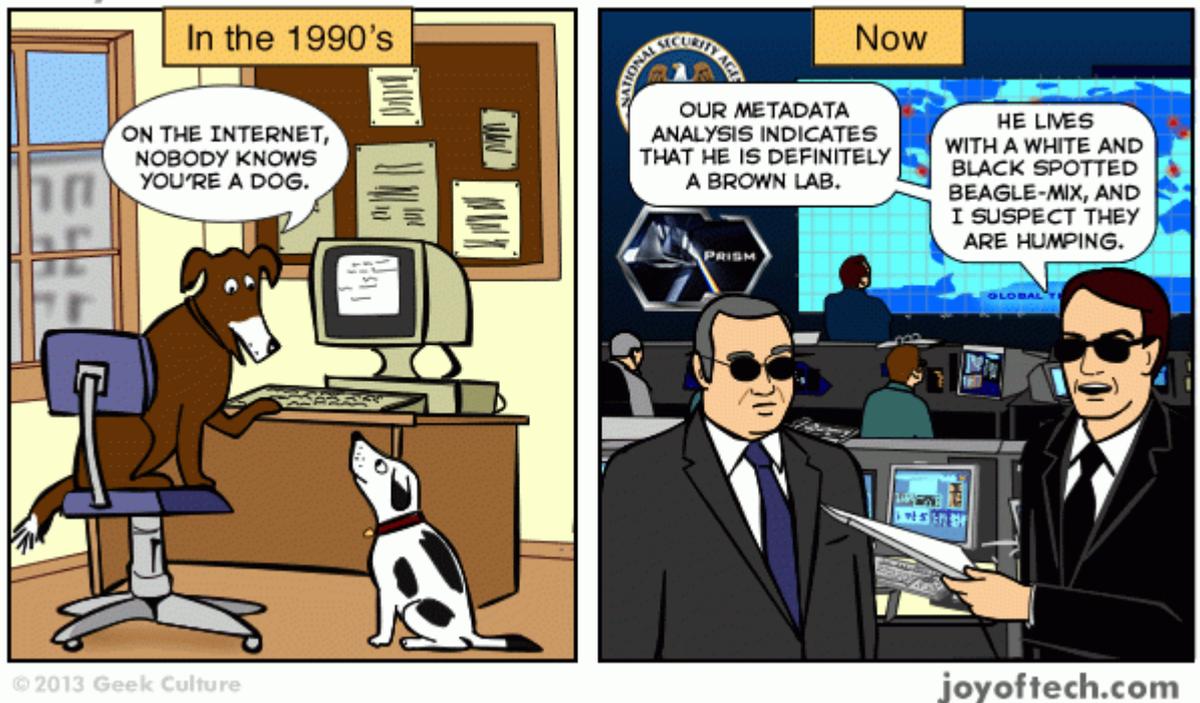
Douglas Rushkov en su libro Team Human plantea:

A filter bubble really goes back to corporate capitalism, a filter bubble is the externalization of those we don't want to see if I have a good filter bubble I don't have to see black people I just get to see white porn or whatever it is we're supposed to want and Google's really good at that right we understand that's what the search engine is for is not to let you search the web it's to let you narrow the web it's to and externalize the stuff don't nobody bring me no bad news. (Rushkoff, 2019a)

Según Rushkoff, la burbuja de filtro es el camino al capitalismo corporativo donde una corporación excluye lo que yo no quiero ver o me perturba, generando una zona de confort y mitigando mi capacidad crítica.

También es importante saber que para generar burbujas de contenido las empresas de internet están continuamente realizando “minería” sobre todos nuestros datos personales, con fines comerciales y político-empresariales. Esta dinámica de creación de contenido personalizado se alimenta de nuestros nuestros datos personales, y toda huella que dejamos, con el objetivo de construir perfiles detallados de cada uno de nosotros y de las tipologías de consumidor y ciudadano que nos clasifica.

The top fifty Internet sites, from CNN to Yahoo to MSN, install an average of 64 data-laden cookies and personal tracking beacons each. Search for a word like “depression” on Dictionary.com, and the site installs up to 223 tracking cookies and beacons on your computer so that other Websites can target you with antidepressants. Share an article about cooking on ABC News, and you may be chased around the Web by ads for Teflon-coated pots. Open—even for an instant—a page listing signs that your spouse may be cheating and prepare to be haunted with DNA paternity-test ads. The new Internet doesn't just know you're a dog; it knows your breed and wants to sell you a bowl of premium kibble. (Pariser, 2011)



2.2.11 Exclusión

La definición base de Deep Web es: páginas no indexadas por los buscadores, pero también puede suceder que determinadas páginas queden relegadas de los resultados de búsqueda y se vuelvan invisibles.

En la actualidad el principal indexador es Google (Alphabet Inc.); con indexar nos referimos a rastrear sistemáticamente el contenido de la web, servidor por servidor, analizando textos, contenido multimedia, metadatos e hipervínculos y colocándolos en una especie de índice el cual consultamos cada vez que hacemos una búsqueda.

A su vez Google los rankea, o sea los jerarquiza utilizando el algoritmo antes referido, supuestamente por asociación directa con el término introducido en el cuadro de búsqueda, sumado a datos ambientales como zona geográfica, idioma etc.

Pero, ¿qué impacto tiene esa jerarquización que hace un buscador? En un estudio realizado por “The Markup”, una agencia de noticias sin ánimo de lucro que realiza periodismo de datos, investigando “las instituciones poderosas utilizan la tecnología para cambiar nuestra sociedad”, analizando una muestra de las 15 mil búsquedas más frecuentes, descubrieron que por ejemplo:

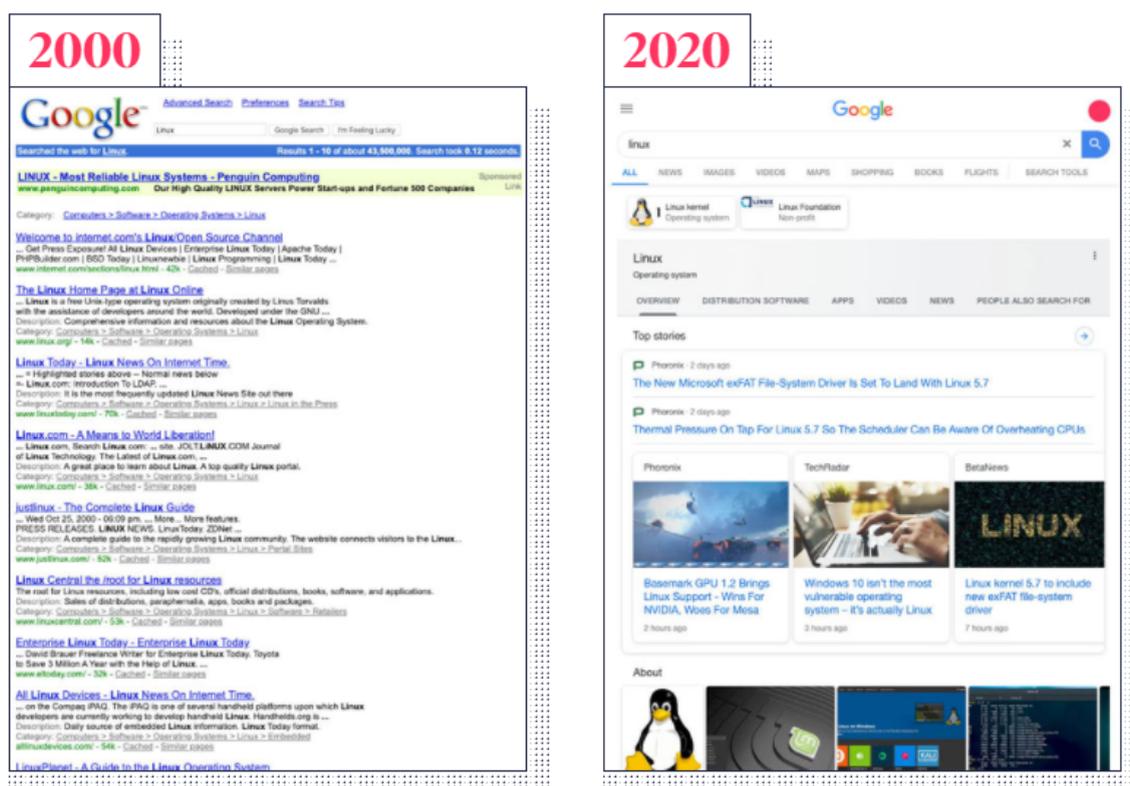
A trending search in our data for “myocardial infarction” shows how Google has piled up its products at the top. It returned: Google’s dictionary definition, a “people also ask” box that expanded to answer related questions without leaving the search results page. A knowledge panel,” which is an abridged encyclopedia entry with various links. And a “related conditions” carousel leading to various new Google

searches for other diseases. All of these appeared before search results by WebMD, Harvard University, and Medscape. In fact, a user would have to scroll nearly halfway down the page (about 42 percent) before reaching the first “organic” result in that search. (Jeffries A., Leon, Y., 2020)

Así, en una búsqueda de un término científico como “Infarto del miocardio” un usuario debe recorrer más del 40 por ciento de la página, debe scrollear hacia abajo, fuera del área inmediatamente visible de la página para encontrar resultados que efectivamente lleven a un enlace externo, asociado a ese término de búsqueda. Todos los resultados anteriores son “contenidos” propios de Google, enlaces a otros de sus productos, ediciones automatizadas de enciclopedias libres, contenidos contratados, promocionados y publicidad.

We examined more than 15,000 recent popular queries and found that Google devoted 41 percent of the first page of search results on mobile devices to its own properties and what it calls “direct answers,” which are populated with information copied from other sources, sometimes without their knowledge or consent. (Jeffries A., Leon, Y., 2020)

En la imagen de abajo puede verse cómo cambió la lista de resultados de búsqueda hoy prácticamente queda relegada del área visible en pantalla.



How a search for “Linux” has changed. In 2000, Google only returned “organic” results, ads, and a category header. Now the top of the page is full of Google products and “answers.” Source: 2000 screenshot/Archive.org

Tomado de: (Jeffries A., León, Y., 2020)

Otro ejemplo del estudio de Markup es que al hacer una consulta tipo diccionario el

resultado de Google (resultado que da la empresa y no explícita de donde lo toma) aparece antes que fuentes más reconocidas como Wikipedia o el diccionario de la lengua, cuando probablemente esas definiciones sean tomadas de esos sitios.

“In our sample, Google featured its dictionary definitions before Urban Dictionary, Cambridge Dictionary, Dictionary.com, Wikipedia, Merriam Webster, and Investopedia, among others. And searches for song titles typically returned a YouTube video in the top spot, followed by the lyrics, displayed in full on the search results page.” (Jeffries A., Leon, Y., 2020)

Pero podríamos pensar que pese a estar relegados a una posición de media página igualmente estos resultados siguen obteniendo clics. Un estudio de Chitika sobre la distribución entre las páginas de las búsquedas tomando resultados de 8 millones de búsquedas indica:

Los porcentajes de visitas que reciben a las páginas posicionados del uno al 20 son el primer resultado el 35% ya el segundo resultado el 16% el tercer resultado el 11% y por ejemplo el noveno resultado sólo recibe 2% de la visita, si vamos al resultado 20 por ejemplo es sólo un 0,29% por lo tanto no aparecer entre los 20 primeros resultados de Google es caer en un pozo de anonimato en los resultados de las búsquedas. (Chitika, 2018)

PÁGINA	VISITANTES	PORCENTAJE
1	2,834,806	34.35%
2	1,399,502	16.96%
3	942,706	11.42%
4	638,106	7.73%
5	510,721	6.19%
6	416,887	5.05%
7	331,500	4.02%
8	286,118	3.47%
9	235,197	2.85%
10	223,320	2.71%
11	91,978	1.11%
12	69,778	0.85%
13	57,952	0.70%
14	46,822	0.57%
15	39,635	0.48%
16	32,168	0.39%
17	26,933	0.33%
18	23,131	0.28%
19	22,027	0.27%
20	23,953	0.29%

Tomado de: Chitika(2018)

Este mal manejo de los resultados, el ignorar resultados orgánicos y posicionar sobre estos enlaces propios o patrocinados ha llevado a Google a sufrir importantes multas.

En 2018 la Unión Europea lo condenó a pagar 4340 millones por abuso de posición dominante en el mercado. Esta investigación fue abierta en abril del 2015.

Google ha utilizado Android como vehículo para consolidar el dominio de su motor de búsqueda. Estas prácticas han privado a sus competidores de la posibilidad de innovar y competir en función de sus méritos. Han impedido a los consumidores europeos beneficiarse de una competencia efectiva en un ámbito tan importante como el móvil. Esto es ilegal con arreglo a las normas de defensa de la competencia de la UE. (Castillo ,2018)

Un año antes también había sido objeto de otra multa gigante de 2420 millones de dólares en este caso la Unión Europea concluyó que la estrategia de Google para su servicio de compras comparativas no se limita a atraer clientes ofreciendo un producto que sea mejor que los de sus rivales, sino que Google ha abusado de su posición dominante como motor de búsqueda situando en mejor lugar su propio servicio de compras comparativas dentro de sus resultados de búsqueda y colocando en peor situación los de sus competidores. (Martín, 2017)

Este monopolio de las búsquedas en la web y el colocar parte de la Internet en la oscuridad casi puede definirse cómo una Deep Web ya no por la incapacidad de los motores de búsqueda sino por el abuso de la posición dominante de una empresa. Por eso la Deep Web tiene una dimensión técnica explicada antes pero una dimensión ética que puede ser incluso más determinante. Digamos que muchos sitios residen en redes de anonimato por una intencionalidad editorial de mantener contenido público bajo conexiones seguras.

3.0 Objetivos y Preguntas de investigación

3.1 Objetivos

Analizar información de interés académico residente en la web profunda para tratar de responder las preguntas de investigación.

Diseñar un conjunto organizado, sistemático y cuantificado, de categorías de significado que ayuden a interpretar este fenómeno complejo en particular.

Realizar un examen metódico, sistemático, objetivo y replicable de los documentos para la posterior la construcción de matrices, elaboración de indicadores e inferencias interpretativas derivadas del marco teórico referencial expuesto en los capítulos que preceden a este análisis.

3.2 Preguntas de investigación

¿Encontramos información de interés académico en la web profunda? ¿Quiénes producen esta información? ¿Por qué se encuentran allí?

4.0 Hipótesis

4.1 Hipótesis directrices

Las hipótesis directrices desarrolladas en el preanálisis son:

- La web profunda es al menos variada y compleja, alberga al menos en parte contenido académico y de interés para el mundo académico
- Estar incluido en la web profunda es también una declaración política y de principios éticos y no solo un tecnicismo.

Hipótesis: Existe contenido de interés académico para el área de información y comunicación en la web profunda o redes de anonimato. Esta información reside en redes de anonimato o en redes no indexadas por buscadores por motivos ideológicos, para escapar de la censura o evadir muros de pagos.

5.0 Estrategia metodológica

Al enfrentarse a analizar un fenómeno tan vasto y complejo como es la web o en este caso una parte de la web, es imperioso tratar de reducir al mínimo el objeto de

estudio sin que por ello deje de ser representativo de los problemas antes planteados que sí son amplios y multifactoriales.

También al ser un tópico poco frecuente en las ciencias de la comunicación y la información (al menos desde una perspectiva integradora) la contextualización y la explicación de temas técnicos puede llegar a ser extensa.

Por lo tanto, se buscó una muestra que se compone de cuatro sitios que fueron elegidos con la premisa de que la muestra sea representativa, objetiva y pertinente.

Se buscó que fuera representativa de dos de las problemáticas antes estudiadas que llevaban a determinados sitios a alojarse en la web profunda: La crisis del sistema de publicación académico-científico y el problema de la privacidad, persecución y censura en Internet. Se incluyeron sitios donde estas problemáticas se explicitaran claramente en el contenido de los sitios, en los discursos de sus creadores y administradores.

Para que sea objetiva se buscaron sitios que fueran reconocidos, populares, que sus creadores fueran instituciones o personas relevantes. Que manejan un número importante de usuarios y un tráfico de datos considerable. Descartamos sitios con objetivos ambiguos o múltiples que pudieran incluirse en más de una tipología ya sea por la descripción misma del sitio o por sus características técnicas.

Para que fuera pertinente incluimos solo objetos que pertenecieran sin ninguna duda a la Internet profunda. Se eligieron sitios que en la actualidad tengan algunos de sus servidores en la Internet profunda, tengan un sitio Tor o una dirección Onion. Se prefirió esta red porque a diferencia de otras modalidades de web profunda la red Tor brinda estadísticas de uso y de tráfico que son fundamentales para el tipo de análisis que se pretende en la metodología.

Los sitios elegidos para el análisis son:

Sitios de recursos académicos y científicos: Sci-Hub y Library Genesis

Sitios por la libertad de prensa y la libertad de expresión: Freedom of the Press, y ProPublica

5.1 Análisis de contenido

Como plantea Lawrence Bardín (Bardín, 1992) la finalidad analítica primordial que resuelven las técnicas de análisis de contenido es la identificación y explicación de las representaciones cognitivas que otorgan el sentido a todo relato comunicativo su objetivo es el tratamiento de mensajes (contenido y expresión de este contenido) para actualizar indicadores que permitan inferir de una realidad otra distinta al

mensaje. En este caso nos permitirá mediante el análisis de declaraciones de los responsables de los sitios y de los textos descriptivos que se encuentran en ellos, identificar los conceptos principales que manejamos en este trabajo como también explicar y dar sentido a estos mensajes.

Para el análisis de contenido cuantitativo del texto de estas páginas, se convertirá la descripción del sitio “acerca de” a texto plano y se procesará mediante software de análisis textual. Por ejemplo, análisis de aparición, frecuencia o ausencia de términos asociados a las categorías de análisis.

Se utilizaron para este trabajo la auto descripción de los sitios en su página normalmente llamada “Acerca de:”, pero también entrevistas a los responsables que muchas veces son personas públicas pero como veremos más adelante en algunos casos son anónimos.

El análisis de contenido de estas entrevistas nos mostrará cómo estos actores lidian con la legalidad, que saben, sienten, creen, cuál es su posicionamiento discursivo, lo que nos aporta profundidad del tipo autoral.

Brinda al análisis marcos de referencia de quienes produjeron la comunicación, sus intenciones, sus representaciones sociales, presupuestos, valores, estados de ánimo, contextos históricos y biográficos.

Como está en nuestros objetivos, trataremos de describir y cuantificar los contenidos manifiestos de los documentos elegidos.

Para ello utilizaremos herramientas informáticas que nos permitan analizar, codificar y categorizar, interpretar e inferir a partir de esos datos.

Ejecutar procesos de análisis mediante software nos permite ser transparentes, públicos, rigurosos y verificables.

Las herramientas utilizadas para eso son:

Para la creación de matrices de frecuencia en texto se utilizó Atlas.ti disponible en: <https://atlasti.com/>

Para la creación de matrices de frecuencia en formato de audio se utilizó el software Transana disponible en: <https://www.transana.com/>

Para realizar el Kwic que permite ubicar una palabra coincidente con las categorías dentro de una frase de sentido completo, se utilizó: Simple Concordance Program <http://www.textworld.com/scp/>

“El análisis de contenido es una técnica de investigación que permite hacer inferencias, válidas y replicables, de los datos en su contexto “. (Krippendorff, 1990)

Los análisis por software serán complementados con un análisis cualitativo de la presentación de los textos en la web, tamaño del texto, ubicación, remarcado o destaque de estos para su ponderación dentro del sitio estudiado cómo un todo. Dar cómo dice Krippendorff un contexto.

5.2 Criterios de Selección

1- Sitios alojados en la Deep Web preferentemente en una red onion ya que esta es por lo explicado en capítulos anteriores Web profunda.

2- Sitios que representen alguna de las problemáticas vistas en el pre análisis, discutidas y explicadas en el marco teórico: la restricción por muros de pago, violación de la privacidad y censura.

Tomamos el sitio web como el medio del mensaje, soporte material del código analizado. Atendiendo a sus distintas partes como su estructura y su contexto.

5.3 Elección de unidades de muestreo

Si entendemos que el análisis formal consiste en una descripción estructural de los sitios , que enumera y analiza características de estos. Para el análisis de contenido mediante software la unidad de análisis es la palabra de los creadores de los sitios, directivos, o impulsores de estos también la auto descripción (Acerca de:) dada en el sitio mismo.

Para el análisis formal la unidad de análisis es el sitio como un todo, sus estructura y organización sus características y peculiaridades.

“la unidad de registro puede considerarse como la parte de la unidad de Muestreo que es posible analizar de forma aislada“, Esta se entiende “como el segmento específico de contenido que se caracteriza al situarlo en una categoría dada” (Bardin, 1996)

1: Unidad de muestreo: el texto del “acerca de” o páginas de descripción de los sitios.

Corpus: texto plano extraído del código HTML

2: Unidad de muestreo: título y subtítulo de la página

Corpus: texto plano extraído del código HTML

3: Unidad de muestreo: texto de entrevistas a los propietarios o impulsores de los sitios

Corpus: texto plano extraído del código HTML, u obtenido por OCR o por pasaje de audio a texto.

4: Unidad de muestreo: texto de entrevistas o textos de los detractores

Corpus: texto plano extraído del código HTML, u obtenido por OCR o por pasaje de audio a texto.

5: Unidad de registro: término

6: Unidad de contexto para la unidad de registro: la frase.

7: Unidad de contexto para toda la unidad de muestreo: el sitio web en su totalidad (análisis formal).

5.4 Categorías

Se han elaborado categorías y subcategorías que responden a las problemáticas detectadas en el pre análisis que según nuestras hipótesis directrices pueden llevar a sitios con contenido de interés académico a incluirse en la web profunda.

Categoría libertad de expresión

- a) código - invasión de la privacidad
- b) código - censura
- c) código - persecución política

Categoría acceso

- a) código - muros de pago
- b) código - copyright

Categoría publicación académica

- a) código - artículos
- b) código - libros
- c) código - editoriales

Categoría problemas legales

- a) código - piratería
- b) código - hackeo

Categoría términos asociados a la web profunda

- a) código - tor
- b) código - onion
- c) código - motor de búsqueda
- d) código - indexación
- e) código - ip
- f) código - vpn

Categoría política, activismo y principios éticos

- a) código - redes
- b) código - grupos de usuarios
- c) código - comunidades
- d) código - libertad de expresión
- e) código - ideología

5.5 Indicadores

Para la elección de los indicadores se toma como referencia el método Morin-Chartier descrito en (Leray, 2008). El método (Morin-Chartier) permite reconciliar los análisis cualitativos y cuantitativos mediante la realización de una clasificación estadística de los datos recogidos en un corpus (Leray, 2008)

“The Morin-Chartier method, enriched by university research effected for over forty years, is a content analysis method which enables both quantitative and qualitative research to be carried out in studying the intensity, neutrality, partiality and orientation of the content of a news item or a speech, for example. This analytical method enables the penetration levels of a message to be measured (the media, written press, political speeches, etc.) in order to feed different types of research in social sciences, public relations and communication marketing or in management sciences.” University of Liège website

Indicadores elegidos:

5.5.1 Frecuencia

Según (Bardín, 1992) la importancia de una unidad de registro crece con su frecuencia de aparición. La frecuencia es el indicador más básico del análisis de contenido.

Frecuencia (código) = $\Sigma [UI (\text{código})] / \Sigma [UI (\text{corpus})] \times 100$ - (Leray, 2008)

2 - Frecuencia ponderada o direccionada

$\Sigma UI (+)$ = frecuencia positiva

$\Sigma UI (-)$ = frecuencia negativa

La dirección: la ponderación de la frecuencia traduce un carácter cuantitativo (intensidad) o cualitativo: la dirección. La dirección puede ser favorable, desfavorable o neutra. (Bardin, 1991).

5.5.2 Parcialidad

5.1.7.2 - Parcialidad = $[\Sigma UI (+) + \Sigma UI (-)] / (\Sigma UI) \times 100$ - (Leray, 2008)

La tasa de parcialidad permite medir la pasión con la que los medios de comunicación trataron un tema. El alto volumen de unidades orientadas da testimonio de un animado debate, pasión, o al menos interés por parte del medio de comunicación sobre un tema, lo que siempre es un dato muy relevante en un estudio. Por ejemplo: una tasa de parcialidad del 30% significa que el 70% del contenido es neutral (Leray, 2008)

5.5.3 Orientación

La medida de la orientación permite calificar el contenido y proporcionar una evaluación cuantificada de cualquier cobertura y sus componentes. La fórmula mide el "índice de favorabilidad" de cada uno de los códigos trazados. La escala va desde 100% - hasta 100% + una orientación del 0% indica que el tratamiento de los medios es neutral (Leray, 2008)

$$\text{Orientación} = [\Sigma \text{UI (+)} - \Sigma \text{UI (-)}] / (\Sigma \text{UI}) \times 100$$

5.6 Aspectos formales del sitio

A su vez se evaluarán aspectos técnicos del sitio que se relacionen a las problemáticas detectadas en el preanálisis

Servidor:

- a) web
- b) web profunda

Dominio:

- a) alojado en un país con leyes de copyright laxas
- b) alojado en un dominio tradicional
- c) múltiples dominios para prevenir bloqueos

Indexación

- a) Permite la indexación de contenido del tipo usado por los buscadores tradicionales
- b) no lo permite

Acceso a contenido académico o de interés académico:

- a) Contiene contenido académico (artículos de investigación, libros técnicos, normas etc.)
- b) Contiene contenido de interés académico. (artículos periodísticos, informes técnicos etc.)
- c) No contiene contenido académico

Explicitación de ideales:

- a) Contiene o no una sección donde explicitan su visión, objetivo del proyecto, posición política.
- b) No contiene una sección de este tipo.

Explicitación de responsables:

- a) El sitio deja claro quien o quienes son sus responsables e impulsores.
- b) El sitio es anónimo.

Incluye publicidad de terceros para monetizar

- a) Incluye publicidad
- b) No incluye publicidad

Donaciones y contribuciones

- a) Incluye en el cuerpo de la página un botón o una sección para donaciones

- b) No incluye en el cuerpo de la página un botón o una sección para donaciones

5.7 Análisis cualitativo de los sitios

“Análisis de contenido cualitativo es un método de análisis controlado del proceso de comunicación entre el texto y el contexto, estableciendo un conjunto de reglas de análisis, paso a paso, que les separe de ciertas precipitaciones cuantificadoras.” (Abela, 2002)

Se analizará la página como un todo, pero también la relación entre sus elementos, que textos que se destacan por encima de los demás ya sea por el formato, el color o la ubicación en el cuerpo. Los destaques, colores utilizados y la distribución de los elementos de los sitios más allá de los bloques de texto. Por último, el estilo del sitio, originalidad o no dentro de su caracterización. Qué busca comunicar tanto por el contenido de sus textos como por imágenes o links a otros sitios, cuál es su público objetivo si es que lo refiere, cuáles son sus funcionalidades y sus objetivos.

A continuación empezaremos a trabajar con los sitios elegidos.

6.0 Análisis de los sitios

6.1 Sci Hub

6.1.1 Análisis formal del sitio

Servidor:

Cuenta con dos servidores

- a) servidor en la web superficial <https://sci-hub.se/>
- b) servidor en la web profunda <https://scihub22266oqcxt.onion/>

Dominio:

El dominio pertenece a Suecia, pero su servidor: IP: 186.2.163.219 está alojado en Belice.

Indexación

- b) no la permite

Acceso a contenido académico o de interés académico:

- a) Contiene contenido académico.

Explicitación de ideales:

a) Contiene una sección donde explicitan su visión, objetivo del proyecto, posición política.

Explicitación de responsables:

a) Alexandra Elbakyan.

Incluye publicidad de terceros para monetizar:

b) No incluye publicidad

Donaciones y contribuciones

a) Incluye en el cuerpo de la página un botón o una sección para donaciones

Análisis título y subtítulo:

Título: Sci-Hub

Subtítulo: Removing all the barriers in the way of science

Inferencia de título y subtítulo:

El nombre “sci hub” refiere a un lugar donde se concentra la ciencia, se llama “hub” la parte central de algo, su centro neurálgico. El subtítulo pone en primer plano el problema de los muros de pago llamándoles “barreras” y afirma derribarlas en pos del camino de la ciencia, denotando ya desde el título una posición de activismo político.

6.1.2 Análisis de contenido Sci Hub - “Acerca de”

Tabla de Frecuencia

Término	Ocurrencias	Ranking
access	9	1
research	8	2
sci	8	2
hub	8	2
knowledge	7	3
papers	6	4
copyright	6	4

laws	5	5
world	5	5
open	5	5

Codificación

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

código (acceso) 9 ocurrencias

código (copyright) 6 ocurrencias

Categoría publicación académica

código (investigación) 8 ocurrencias

código (conocimiento) 7 ocurrencias

código (papers) 6 ocurrencias

código (científico) 4 ocurrencias

código (ciencia) 3 ocurrencias

Categoría problemas legales

código (piratería) 1 ocurrencia

código (leyes) 5 ocurrencias

Categoría términos asociados a la web profunda

0 ocurrencias

Categoría - política, activismo y principios éticos

código (mundo) 5 ocurrencias

código (desigualdad) 3 ocurrencias

código (lucha) 2 ocurrencias

código (injusticia) 1 ocurrencia

código (barreras) 1 ocurrencia

Kwic: Análisis del término en el contexto de la frase para detectar la orientación.

Access

1 to provide mass & public **access** to research papers/enter (+1)

6 to provide mass and public **access** to tens of millions of research (+1)

- 10 the moment, Sci-Hub provides **access** to hundreds of thousands (0)
 15 to all no copyright open **access**/We fight inequality in knowledge (+1)
 16 fight inequality in knowledge **access** across the world. The scientific (+1)
 20 Sci-Hub project supports the Open **Access** movement in science. Research
 (+1)
 20 should be published in open **access**, i.e. be free to read. / (+1)
 21 e. be free to read. //The Open **Access** is a new and advanced form (+1)
 24 Imagine the world with free **access** to knowledge for everyone (+1)

En las 9 apariciones del término acceso en 8 está asociado a términos positivos (abierto (4), público (2), libre (1), conocimiento (1) sólo encontramos una asociación neutra con el término “proveer” y ninguna negativa.

Término	Asociación mayoritaria (+)
acceso	abierto x4 público x 2 libre al conocimiento

Research

- 1 to provide mass & public access to **research** papers/enter URL, PMID / DOI or (+1)
 6 public access to tens of millions of **research** papers///A research paper is a special (+1)
 7 A **research** paper is a special (+1) publication written by scientists to be read by other **researchers** (0)
 7 Papers are primary sources necessary for **research** – for example, they contain detailed description of new results and experiments. (0)
 10 the widest possible distribution of **research** papers, as well as of other scientific (+1)
 10 access to hundreds of thousands **research** papers every day, effectively bypassing any paywalls and restrictions (+1)
 11 If it were not for Sci-Hub – I wouldn't be able to do my thesis in Materials Science (**research** related to the structure formation in aluminum alloys) (0)
 20 The Sci-Hub project supports the Open Access movement in science. **Research** should be published in open access, i.e. be free to read. (+1)

En las 9 apariciones del término (investigación) en 6 está asociado a términos positivos (proveer acceso público y masivo, amplia distribución) 5 veces y (especial), en tres encontramos una asociación neutra, (para ser leído por otros investigadores), (fuente para) y (relativos a) y ninguna negativa.

Término	Asociación mayoritaria (+)
investigación	acceso público y masivo, amplia distribución x6

Copyright

10 At this time the widest possible distribution of research papers, as well as of other scientific or educational sources, is artificially restricted by **copyright** laws. (-1)

15 /knowledge to all //no **copyright** // open access/We fight (-1)

18 We advocate for cancellation of intellectual property, or **copyright** laws, for scientific and educational resources (-1)

19 **Copyright** laws render the operation of most online libraries illegal (-1)

19 The **copyright** fosters increase of both informational and economic inequality. (-1)

24 to the battle against **copyright** laws and information (-1)

En las 6 apariciones del término copyright siempre está asociado a términos negativos (restringido), (no a él), (abogamos por su cancelación), (hace ilegal nuestro funcionamiento), (incrementa la inequidad), (damos batalla contra él) no encontramos ninguna asociación neutra y ninguna positiva.

Término	Asociación mayoritaria (-)
copyright	restringido no al abogamos por la cancelación de hacen ilegal nuestro funcionamiento incrementa la inequidad batalla contra

Código	Positivo	Negativo	Neutro	Total
acceso	8	0	1	9
investigación	6	0	3	9
copyright	0	6	0	6

6.1.3 Indicadores

Corpus = 429

Frecuencia (acceso) = $(9 / 429) 100 = 2,1$

Frecuencia (investigación) = $(9 / 429) 100 = 2,1$

Frecuencia (copyright) = $(6 / 429) 100 = 1,4$

Parcialidad (acceso) = $((8 + 0) / 9) \times 100 = 89\%$

Parcialidad (investigación) = $((6 + 0) / 9) \times 100 = 67\%$

Parcialidad (copyright) = $((0 + 6) / 6) \times 100 = 100\%$

Orientación (acceso) = $((8 - 0) / 9) \times 100 = 89\% +$

Orientación (investigación) = $((6 - 0) / 9) \times 100 = 66,6\% +$

Orientación (copyright) = $((0 - 6) / 6) \times 100 = 100\% -$

6.1.4 Inferencia del análisis:

De los 3 términos más frecuentes (acceso), (investigación) y (copyright) respectivamente todos son tratados de una manera bastante parcial. Totalmente parcial en el caso de (copyright) 100%, altamente parcial (acceso) 89% y bastante parcial (copyright) 67%.

Esto indica un debate acalorado con respecto al tema más que una opinión objetiva con respecto a los tres temas, pero sobre todo a acceso y copyright.

En cuanto a la orientación: (copyright) es totalmente parcial con una intensidad total hacia lo negativo, (acceso) es fuertemente tratado de manera parcial con intensidad casi completa hacia lo positivo y finalmente (investigación) es moderadamente parcial con tendencia hacia lo positivo.

Se observa una tendencia casi absoluta a ver el acceso (a literatura científica) y por ende un facilitador de la investigación como ampliamente positivo, una orientación marcadisima a ver el copyright cómo un problema y algo contra lo que luchar.

6.1.5 Análisis cualitativo del sitio

ABOUT

Sci-Hub

the first pirate website in the world to provide mass and public access to tens of millions of research papers

A research paper is a special publication written by scientists to be read by other researchers. Papers are *primary sources* necessary for research – for example, they contain detailed description of new results and experiments.

papers in Sci-Hub library:
more than **74,000,000** and growing

At this time the widest possible distribution of research papers, as well as of other scientific or educational sources, is artificially restricted by copyright laws. Such laws effectively slow down the development of science in human society. The Sci-Hub project, running from 5th September 2011, is challenging the status quo. At the moment, Sci-Hub provides access to *hundreds of thousands research papers every day*, effectively bypassing any paywalls and restrictions.

El texto de mayor tamaño y color rojo es el nombre del sitio, el segundo texto en tamaño y diferenciado por color son los 74 millones de papers con los que cuenta el sitio. Luego resaltado está “el primer sitio pirata del mundo en proveer acceso a decenas de millones de artículos de investigación”. En el fondo se aprecia la silueta de un mapa dando énfasis al carácter global de la propuesta.

6.1.6 Inferencia

El sitio hace énfasis en el volúmen de información que posee, algo que podría verse en un sitio comercial del estilo de valorización o promoción. Por otro lado el término “pirata” asociado al robo y la transgresión de las leyes es algo que nunca veríamos en un sitio con fines académicos o comerciales, puede ser para marcar diferencias, para establecer una posición política con respecto al término o simplemente porque desde la mirada mayoritaria de la industria del conocimiento es lo que son.

El sitio deja bien claro que su objetivo es la comunidad académica al llamar a su material “research papers” un término que solo se usa en ese ambiente.



Al final de la página se destaca el texto “únete a la lucha” este mismo texto y unas siluetas que recuerdan una manifestación están de color rojo. En las siluetas se distinguen manos alzadas, pero también cerca del centro puños cerrados que simbolizan la lucha. El sitio hace una fuerte convocatoria política a acompañar el movimiento por la democratización del conocimiento, lo reconoce como una pelea de muchos contra unos pocos, pero muy poderosos. Nos pide que ayudemos a romper las barreras, que suponemos son los muros del pago y el copyright restrictivo en pos del conocimiento como un bien común e innegociable.

6.1.7 Análisis de contenido Entrevista Alexandra Elbakyan:

<https://www.theverge.com/2018/2/8/16985666/alexandra-elbakyan-sci-hub-open-access-science-papers-lawsuit>

Tabla de Frecuencia

Término	Ocurrencias	Ranking
sci	83	1
elbakyan	81	1
access	74	2
hub	64	3

open	45	4
research	40	5
elsevier	35	6
publishers	28	7
scientists	26	7
journals	23	8

Codificación

Analizando una entrevista a Alexandra Elbakian la creadora de Sci Hub vemos dentro de los principales términos empleados que muchos se corresponden a las categorías

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

código (acceso) 74 ocurrencias

Categoría publicación académica

código (investigación) 40 ocurrencias

código (científicos) 26 ocurrencias

código (editoriales) + elsevier 75 ocurrencias

código (journals) 23 ocurrencias

Categoría problemas legales

0 ocurrencias

Categoría términos asociados a la web profunda

código (motor de búsqueda) 5 ocurrencias

Categoría política, activismo y principios éticos

código (el kremlin) 7 ocurrencias

código (el gobierno) 5 ocurrencias

6.1.8 Inferencia de frecuencia de términos en la entrevista:

La mayor frecuencia es de “acceso”, siendo que el acceso pago a publicaciones científicas es un problema para muchos usuarios y que este sitio brinda acceso gratuito e ilimitado a una parte muy grande del universo de las publicaciones

académicas que se encuentran en formato digital. El segundo término con más apariciones es “open” muchas veces ligado a “access” o a lo que deberían ser para Alexandra los sitios académicos y las bases de datos. El tercer término con mayor frecuencia es “research”, término que define la búsqueda académica y científica en contraposición a cómo podría denominarse una búsqueda habitual en internet para lo cual se utilizará el término “search” que no aparece prácticamente en la entrevista. Otros términos asociados fuertemente al mundo académico son “scientists” y “journal” siendo el público objetivo, productor y consumidor del sitio, “journals” por otro lado es cómo se denomina en este ámbito a las publicaciones periódicas.

El término que aparece luego con mayor frecuencia es “Elsevier”, nombre de la editorial que demandó legalmente a Alexandra, y pidió formalmente dar de baja el sitio. Casi igualado en frecuencia tenemos al término “access” (como dijimos antes) muchas veces ligado al término “abierto” pero siempre como una demanda, como lo que el movimiento que representa Sci Hub reclama. También se utilizan términos relacionados con política, gobierno aparece 5 veces y el término “Kremlin” en referencia al gobierno ruso otras 7 veces lo que denota que en el discurso de la cara visible del sitio están muy unidos el acceso, el mundo académico y la política.

Alexandra, en representación de Sci Hub, da en su entrevista una gran importancia a los temas de acceso al material académico, apunta al mundo de la investigación científica, y ataca el copyright restrictivo por parte de las editoriales.

6.1.9 Consideraciones adicionales

1- Sci Hub tiene una característica única entre los 4 sitios analizados, tiene una cara visible Alexandra. Si bien ella se autodefine como “desarrolladora de software” su formación es en neurociencia y en mi investigación no he podido encontrar otros desarrollos además de este sitio. En conclusión, es muy difícil pensar (aunque tampoco ella nunca ha manifestado que no haya más personas) que Alexandra es la única persona detrás de Sci Hub. Por lo tanto, la figura de Alexandra Elbakyan cómo símbolo y un referente en lo que refiere a la lucha por la democratización del conocimiento científico es válida, pero Sci Hub como organización en mi opinión debe ser más amplia.

Sci Hub es el sitio más popular, más académico y más difundido incluso uno de los más longevos en su formato, como interfaz, como producto de software también es muy eficiente y efectivo. Está bien programado y sus servidores funcionan perfecto a pesar de la masividad de las consultas lo que refuerza la teoría que hay más recursos detrás de este proyecto. El hecho de la persecución y las demandas legales que sufre Elbakyan es de suponer que hayan hecho a los posibles miembros u organizaciones detrás del proyecto permanecer en el anonimato.

Sci Hub es un sitio exitoso, principalmente como dijimos antes por permitir resolver una necesidad informativa de manera rápida y eficiente. Además, esto lo hace sin pedirnos que nos registremos, sin publicidad y obviamente sin costos para el usuario. Estas son características muy propias de la Deep Web y extremadamente raras en sitios académicos de la web superficial.

Pero si bien puede considerarse un sitio de descarga similar a PirateBay, por poner el ejemplo más paradigmático, lo hace sin renunciar un minuto a la militancia y a la reivindicación de los principios que surgen claramente en el análisis de contenido y en el análisis formal. Nunca estos sitios desde Napster hace 30 años hasta los sitios más actuales de *torrent* han tenido dentro del cuerpo de contenido nada de lo que poder inferir reivindicaciones sociales y políticas.

Alexandra en sus entrevistas siempre hace referencia a la comunidad científica y esta comunidad al menos en su concepción más clásica, y puede que un poco idealista, es una comunidad orientada a compartir el conocimiento a moverse en bloques, a consolidar alianzas.

Es difícil afirmar que el éxito de Sci Hub tiene que ver con su foco en la comunidad científica académica, Sci Hub no deja de ser un unicornio, no por ser el único sitio de su tipo sino, como expresamos antes, es un sitio extremadamente popular y muy efectivo tanto así que probablemente pueda competir y ganar en usabilidad y rapidez con cualquier sitio de las mejores editoriales.

Llegamos a los contenidos sin tener que pasar por ningún registro, pantalla de bienvenida ni menús de ningún tipo, directamente ponemos el DOI y el artículo es nuestro. Sci Hub toma el minimalismo que hizo exitosa a la barra de Google, toma elementos propios de la Deep Web y los combina en un producto que probablemente no tenga competidor en ninguna de las dos.

6.2 LibGen

6.2.1 Análisis formal del sitio

Servidor:

Cuenta con dos servidores

- a) servidor en la web superficial: <https://libgen.is/>

- b) servidor en la web profunda: <http://genotypeinczgrxr.onion/>

Dominio:

b) Su Dominio pertenece a Islandia, pero su servidor IP :193.218.118.42 está alojado en Kiev

Indexación

- b) no la permite

Acceso a contenido académico o de interés académico:

- a) Contiene contenido académico

Explicitación de ideales:

- a) Contiene una sección donde explicitan su visión, objetivo del proyecto, posición política.

Explicitación de responsables:

- b) El sitio es anónimo.

Incluye publicidad de terceros para monetizar

- b) No incluye publicidad

Donaciones y contribuciones

b) No incluye en el cuerpo de la página un botón o una sección para donaciones

Análisis título y subtítulo:

Título: Library Genesis

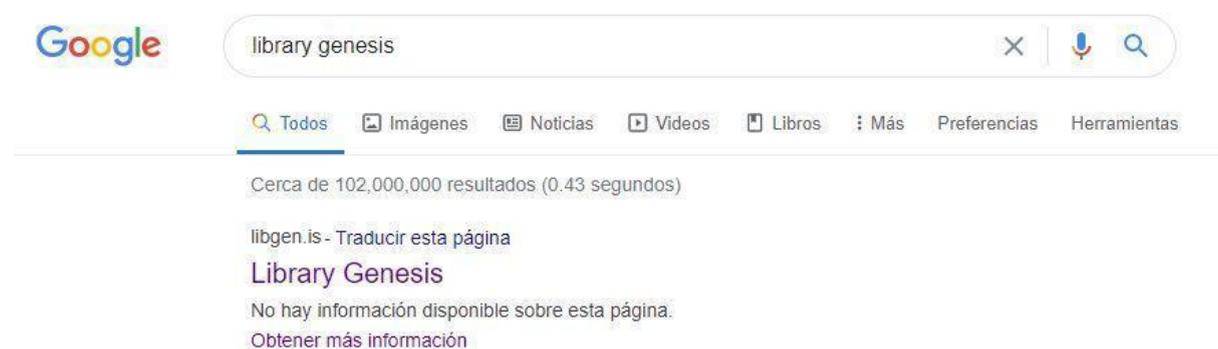
Subtítulo: no tiene

Inferencia de título y subtítulo:

La palabra Library evidentemente hace referencia a que son una colección ordenada de recursos de información. El término Génesis puede estar asociado a un comienzo.

Consideraciones adicionales

1 - Library Genesis además de no permitir el indexado de su contenido tampoco permite el indexado de su sitio. No le permite a los motores de búsqueda acceder a la información de su estructura en general, lo que puede llegar a incluir su sitio en la web superficial cómo web opaca e incluso profunda.



Esto es mediante el archivo robots.txt que hemos explicado anteriormente.

No hay información de la página en los resultados de búsqueda

Es posible que en un resultado de búsqueda de Google se indique que no hay información disponible sobre la página, tal como se muestra a continuación:

Search result title
<https://www.example.com>

Si se muestra este mensaje, el sitio web ha impedido crear la descripción de la página, pero no la ha ocultado de Google.

Si eres el propietario del sitio web, puedes mejorar este resultado de dos formas: dejando que Google lea la página y cree una buena descripción, u ocultándola completamente de los resultados de la Búsqueda. A continuación te explicamos cómo hacerlo.

Solucionar el problema

Si no es tu sitio web

Si es tu sitio web

Probablemente veas ese resultado por uno de estos motivos:

El archivo robots.txt ha bloqueado la página

2- Formalmente LibGen no tiene una página “Acerca de” es básicamente un repositorio de documentos. Si hurgamos en el mapa del sitio encontramos dos partes con contenido posible de analizar. Una se llama “Letter of solidarity” que es una carta firmada por varias personas expresando solidaridad frente a la demanda hecha por Elsevier al sitio por infringir el derecho de autor.

Finalmente, en el “sitemap” aparece: lo que dieron a llamar el “código” de Library Genesis escrito por alguien que se define cómo el bibliotecario y cuyo nickname es “book warrior”. En él, en varios puntos se describen los objetivos, prioridades y afiliaciones del sitio.

6.2.2 Análisis de contenido LibGen “Letter of Solidarity”

Tabla de Frecuencia

Término	Ocurrencias	Ranking
knowledge	10	1
access	7	2

elsevier	6	3
public	6	3
library	6	3
academic	5	4
share	5	4
little	4	5
world	4	5
commons	4	5
before	4	5
libraries	4	5
publishers	4	5
custodians	4	5
science	3	6
system	3	6
write	3	6
articles	3	6
digitize	3	6
stars	3	6

Codificación

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

código (acceso) 7 ocurrencias

Categoría publicación académica

código (conocimiento) 10 ocurrencias

código (biblioteca/s) 10 ocurrencias

código (elsevier)+ (editoriales) 10 ocurrencias

código (académico) + (ciencia) 8 ocurrencias

código (artículos) 3 ocurrencias

Categoría problemas legales

0 ocurrencias

Categoría términos asociados a la web profunda

(Darknet) 1 ocurrencia

Categoría política, activismo y principios éticos

código (compartir) + (común) 9 ocurrencias

código (publico) 6 ocurrencias

Kwic: Análisis del término en el contexto de la frase para detectar la orientación.

Knowledge

2 //There are many businessmen who own **knowledge** today. Consider Elsevier, the largest scholarly publisher (-1)

5 //We have the means and methods to make knowledge accessible to everyone, with no economic barrier to access and at a much lower cost to society (+1)

6 the public does not have the right to **knowledge**."We demonstrate daily, and on a massive scale, that the system is broken (-1)

7 other side of 37% profit margins: our **knowledge** commons grow in the fault lines of a broken system (-1)

7 broken system. We are all custodians of **knowledge**, custodians of the same infrastructures (+1)

7 that we depend on for producing **knowledge**, custodians of our fertile but (+1)

7 of use to, not to make property of, our **knowledge** commons. //More than seven years (-1)

8 message opposing the privatization of **knowledge** — we'll make it a thing of the (-1)

9 that the very existence of our massive **knowledge** commons is an act of collective i civil disobedience (-1)

10 Don't let our **knowledge** be crushed. Care for the libraries (-1)

En las 10 apariciones del término (conocimiento) en 7 está asociado términos negativos (apropiado por gente de negocios), (la gente no tiene derecho a él), (sistema roto) 2 veces, (privatizar el conocimiento), (desobediencia civil) y (aplastado). En 3 ocasiones a términos positivos (accesible a todos), (producir) y (ser custodios).

Término	Asociación mayoritaria (-)
conocimiento	apropiado por gente de negocios la gente no tiene derecho a él sistema roto privatizado desobediencia civil

	aplastado
--	-----------

Academic

2 owns some of the largest databases of **academic** material, which are licensed at (0)

2 they prohibit access to science to many **academics** - and all non-academics - across (-1)

3 sites are the only viable source of **academic** materials. The social media, mailing (0)

4 fees it charges to authors and their **academic** institutions. As we write these (-1)

5 But closed access's monopoly over **academic** publishing, its spectacular profits (-1)

5 its central role in the allocation of **academic** prestige trump the public interest (+1)

En las 6 apariciones del término académico 3 veces asociado términos negativos (impuestos y gravámenes), (monopolio de la publicación), (ganancia desmesurada), 2 veces a términos neutros: (fuente) y (material) y 1 positivo (prestigio).

Término	Asociación mayoritaria (-)
académico	impuestos y gravámenes monopolio ganancia desmesurada

Access

2 are priced such that they prohibit **access** to science to many academics and (-1)

4 refusal by Elsevier to go open **access** and give up on the high fees it charges (-1)

4 our service to the public, it denies us **access**. (-1)

5 to make knowledge **accessible** to everyone, with no economic (+1)

5 with no economic barrier to **access** and at a much lower cost to society (+1)

5 much lower cost to society. But closed **access's** monopoly over academic publishing (-1)

5 publishers effectively impede open **access**, criminalize us, prosecute our heroes (-1)

7 our publishers, circumvent paywalls to **access** articles and publications,

digitize (0)

7 to maintain libraries, to make them **accessible**. It is to be of use to, not to (0)

8 We need to fight for Guerilla Open **Access**. With enough of us, around the world (-1)

En las 10 apariciones del término acceso 6 veces asociado términos negativos (prohibir), (rechazar), (denegar), (cerrado, monopólico), (impedir, criminalizar), (combatir), 2 veces a términos neutros: (hacerlos accesibles) y (acceder a artículos) y 2 positivos (sin barreras económicas) y (accesible para todos).

Término	Asociación mayoritaria (-)
acceso	prohibir rechazar impedir combatir

Publishers

5 trump the public interest. Commercial **publishers** effectively impede open access (-1)

7 secretly behind the backs of our **publishers**, circumvent paywalls to access (-1)

2 by public money benefiting scholarly **publishers**, particularly the peer review (-1)

4 threatened to go the way of other small **publishers** that are being rolled over by (-1)

En las 4 apariciones del término editoriales aparece 3 veces asociado términos negativos (impedir acceso), (muros de pago) y (amenaza) y una a términos positivos, pero se refiere a las editoriales académicas (scholarly publishers).

Término	Asociación mayoritaria (-)
editoriales	impedir acceso muros de pago amenaza

Elsevier

2 There are many businessmen who own knowledge today. Consider **Elsevier**,

the (-1) largest scholarly publisher

Elsevier owns some of the largest databases of academic material (0)

3 **Elsevier** has recently filed a copyright infringement suit in New York against Science Hub and Library Genesis claiming millions of dollars in damages. (-1)

4 the refusal by **Elsevier** to go open access and give up on the high fees it charges to authors and their academic institutions. (-1)

6 In **Elsevier's** case against Sci-Hub and Library Genesis, the judge said: "simply making copyrighted content available for free via a foreign website, disservices the public interest (0)

6 If **Elsevier** manages to shut down our projects or force them into the darknet, that will demonstrate an important idea: that the public does not have the right to knowledge." (-1)

En las 6 apariciones del término Elsevier aparece las 4 veces asociado términos negativos (hombres de negocios que se apropian del conocimiento), (demanda legal en nuestra contra), (negarse al acceso abierto) y (destruir nuestros proyectos) y 2 a términos neutros: (propietario de la mayor base de contenido académico), (la demanda en nuestra contra merecer el interés público)

Término	Asociación mayoritaria (-)
Elsevier	apropiación del conocimiento demandas legales negarse al acceso abierto destruir proyectos

Código	Positivo	Negativo	Neutro	Total
conocimiento	3	7	0	10
académico	1	3	2	6
acceso	2	6	2	10
editoriales	1	3	0	4
Elsevier	0	4	1	5

6.2.3 Indicadores

Corpus = 1064

Frecuencia (conocimiento) = $(10/1064) \times 100 = 0,94$

Frecuencia (académico) = $(6/1064) \times 100 = 0,56$

Frecuencia (acceso) = $(10/1064) \times 100 = 0,94$

Frecuencia (editoriales) = $(4/1064) \times 100 = 0,37$

Frecuencia (Elsevier) = $(5/1064) \times 100 = 0,47$

Parcialidad (conocimiento) = $((3 + 7)/10) \times 100 = 100\%$
Parcialidad (académico) = $((1+3)/6) \times 100 = 67\%$
Parcialidad (acceso) = $((2+6)/10) \times 100 = 80\%$
Parcialidad (editoriales) = $((1+3)/4) \times 100 = 100\%$
Parcialidad (Elsevier) = $((0 + 4)/5) \times 100 = 80\%$

Orientación (conocimiento) = $((3 - 7)/10) \times 100 = -40\%$
Orientación (académico) = $((1-3)/6) \times 100 = -33\%$
Orientación (acceso) = $((2-6)/10) \times 100 = -40\%$
Orientación (editoriales) = $((1-3)/4) \times 100 = -50\%$
Orientación (Elsevier) = $((0 - 4)/5) \times 100 = -80\%$

6.2.4 Inferencia

De los 5 términos más frecuentes (conocimiento), (académico), (acceso), (editoriales) y (Elsevier), los más frecuentes son (acceso) y (conocimiento), en general todos son tratados de manera bastante parcial entre el 70 y 100 % de las veces. Lo que indica un debate acalorado y apasionado más que una exposición objetiva. En el caso de (editoriales) y (conocimiento) una total imparcialidad 100%. Siendo el término (académico) donde el trato es moderadamente imparcial 67%. En cuanto a la orientación también todos los términos con una intensidad altísima hacia lo negativo. Igualmente, los otros términos tienen una parcialidad similar siendo (editoriales) el menos imparcial de los 4 y académico el más parcial. Pero todos los términos tienen una connotación negativa el que más se acerca a la neutralidad es (académico) y el menos neutral es (Elsevier) donde la orientación negativa llega a un 80%.

El tema recurrente en el texto es el acceso sin restricciones al contenido académico y el impedimento por parte de las editoriales a que esto pase. Es muy marcada la preocupación por el problema del acceso y el fuerte rechazo a las editoriales y principalmente a Elsevier.

6.2.5 Análisis cualitativo del sitio

La página principal es extremadamente simple parecida al buscador de Google. Solo tiene el nombre del sitio, la barra de búsqueda y opciones avanzadas de búsqueda. No tiene textos para analizar y el link más prominente es a la "Letter of Solidarity"

Library Genesis^{2M}

Introducing Libgen Desktop application!

Letter of Solidarity

Search in :

- LibGen (Sci-Tech) Scientific articles Fiction
 Comics Standards Magazines

LibGen Search options:

Download type: ▼

View results: Simple Detailed

Results per page ▼

Search with mask (word*): No Yes

Search in fields The column set default Title Author(s) Series
 Publisher Year ISBN Language MD5 Tags Extension

“Letter of solidarity”

En esta carta de solidaridad a raíz de una demanda de la editorial Elsevier. La carta está firmada por varios activistas y académicos además cuenta con una bibliografía, algo que sería muy extraño y atípico para un sitio de descargas pero que es algo habitual y mandatorio en el mundo académico.

Share this letter - read it in public - leave it in the printer. Share your writing - digitize a book - upload your files. Don't let our knowledge be crushed. Care for the libraries - care for the metadata - care for the backup. Water the flowers - clean the volcanoes.

30 November 2015

Dušan Barok, Josephine Berry, Bodó Balázs, Sean Dockray, Kenneth Goldsmith, Anthony Iles, Lawrence Liang, Sebastian Lütgert, Pauline van Mourik Broekman, Marcell Mars, spideralex, Tomislav Medak, Dubravka Sekulić, Femke Snelting...

-
1. Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. "[The Oligopoly of Academic Publishers in the Digital Era](#)." PLoS ONE 10, no. 6 (June 10, 2015): e0127502. doi:10.1371/journal.pone.0127502., "[The Obscene Profits of Commercial Scholarly Publishers](#)." svpow.com. Accessed November 30, 2015. [↪](#)
 2. Sample, Ian. "[Harvard University Says It Can't Afford Journal Publishers' Prices](#)." The Guardian, April 24, 2012, sec. Science. theguardian.com. [↪](#)
 3. "[Academic Paywalls Mean Publish and Perish - Al Jazeera English](#)." Accessed November 30, 2015. aljazeera.com. [↪](#)
 4. "[Sci-Hub Tears Down Academia's 'Illegal' Copyright Paywalls](#)." TorrentFreak. Accessed November 30, 2015. torrentfreak.com. [↪](#)
 5. "[Save Ashgate Publishing](#)." Change.org. Accessed November 30, 2015. change.org. [↪](#)
 6. "[The Cost of Knowledge](#)." Accessed November 30, 2015. thecostofknowledge.com. [↪](#)
 7. In fact, with the TPP and TTIP being rushed through the legislative process, no domain registrar, ISP provider, host or human rights organization will be able to prevent copyright industries and courts from criminalizing and shutting down websites "expeditiously". [↪](#)
 8. "[Court Orders Shutdown of Libgen, Bookfi and Sci-Hub](#)." TorrentFreak. Accessed November 30, 2015. torrentfreak.com. [↪](#)
 9. "[Guerilla Open Access Manifesto](#)." Internet Archive. Accessed November 30, 2015. archive.org. [↪](#)
-

El texto final dice: "Compartan esta carta, leanla en público, dejanla en la imprenta. Compartan sus escritos, digitalicen un libro, suban sus archivos, No dejen que destruyan el conocimiento. Ocupense de las bibliotecas, ocupense del metadata, cuiden el backup. Rieguen las plantas, limpien los volcanes."

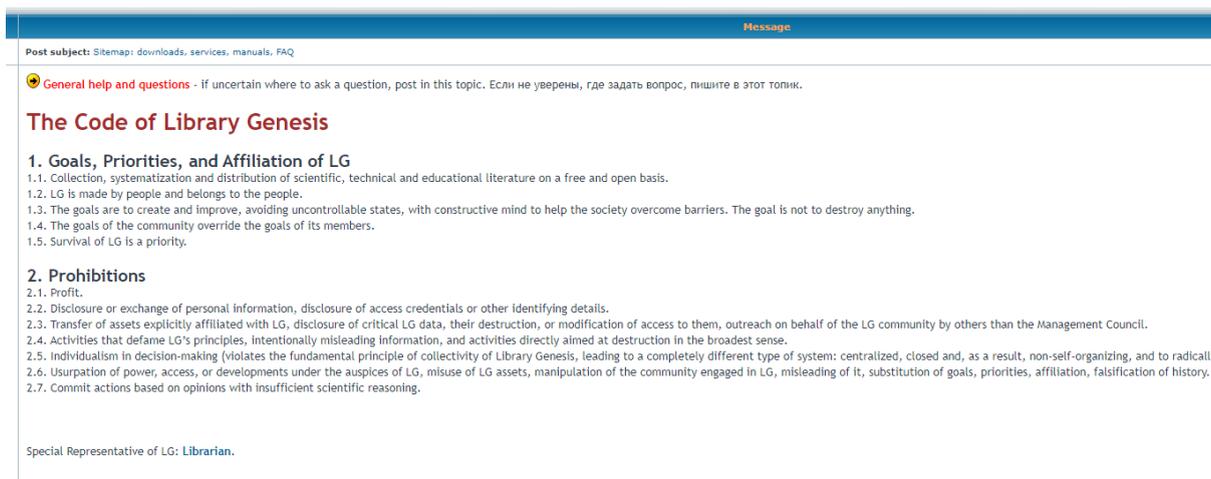
Aquí puede verse claramente el espíritu del sitio, una parte que puede ser confusa es la de preocuparse por los metadatos y el backup. Todo bibliotecario sabe de la importancia de los metadatos al momento de recuperar la información lo que muestra conocimiento técnico y también es algo raro para un sitio de descargas. El cuidar los backups es porque toda la colección de LIBGen se encuentran en dumps alojados en varios servidores y en la Deep Web pero que pueden ser descargados por cualquiera en caso que un servidor caiga y quiera levantarse otro.

En la bibliografía cita a Aaron Swartz y a su "manifiesto guerrilla". Lo que se puede inferir claramente son los objetivos del sitio y su posición política.

6.2.6 Inferencia

En esta parte del sitio sí puede verse el pensamiento político de los impulsores del sitio y nos deja claro que es más que un buscador más que una colección de contenido pirata. “The Code of Library Genesis”

Si en la página principal damos click donde dice “Foro” aparece lo que se ha dado a llamar “el código de Library Genesis” escrito con una fuente roja de mayor tamaño. Allí aparecen metas, prioridades y afiliaciones



The screenshot shows a forum post with a blue header bar containing the word "Message". Below the header, the post subject is "Sitemap: downloads, services, manuals, FAQ". A yellow speech bubble icon is followed by the text "General help and questions - if uncertain where to ask a question, post in this topic. Если не уверены, где задать вопрос, пишите в этот топик." The main title of the post is "The Code of Library Genesis" in red. The content is organized into two numbered sections: "1. Goals, Priorities, and Affiliation of LG" and "2. Prohibitions". Each section contains a list of numbered sub-points. At the bottom of the post, it says "Special Representative of LG: Librarian."

Aparecen como objetivos la distribución de materiales científicos entre otros, el hacerlo de manera gratuita y abierta. Plantea que Library Genesis fue hecha para la gente y “pertenece” a la gente, marca como objetivo ayudar a la sociedad a superar sus barreras y que los objetivos de la comunidad son más importantes que los de sus miembros.

Textualmente aparece en el sitio:

“1. Goals, Priorities, and Affiliation of LG

- 1.1. Collection, systematization and distribution of scientific, technical and educational literature on a free and open basis.
- 1.2. LG is made by people and belongs to the people.
- 1.3. The goals are to create and improve, avoiding uncontrollable states, with a constructive mind to help the society overcome barriers. The goal is not to destroy anything.
- 1.4. The goals of the community override the goals of its members.
- 1.5. Survival of LG is a priority.”

También aparecen prohibiciones entre estas el lucro, divulgar o intercambiar información personal o difundir detalles que puedan hacer identificar a las personas que participan del proyecto.

“2. Prohibitions

2.1. Profit.

2.2. Disclosure or exchange of personal information, disclosure of access credentials or other identifying details...”

Inferencias: Hay una clara dimensión política y ética de los contenidos que aparecen en el sitio. Se nota el activismo y el compromiso de la página con valores como la difusión del conocimiento y el acceso igualitario.

6.2.7 Análisis de contenido Entrevista administradores LiBGen

Consideraciones: Dado que los administradores se mantienen anónimos fue muy difícil conseguir entrevistas o declaraciones, esta entrevista fue extraída del texto de Bodó “The Genesis of Library Genesis”.

Tabla de Frecuencia

Término	Ocurrencias	Ranking
books	9	1
libgen	8	2
library	5	3
donations	4	4
support	4	4
access	3	5
downloadable	3	5
anything	3	5

world	3	5
should	3	5
people	3	5

Codificación

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

código (acceso) 3 ocurrencias

Categoría publicación académica

código (libros) 9 ocurrencias

código (biblioteca) 5 ocurrencias

código (universidad) 2 ocurrencias

Categoría problemas legales

0 ocurrencias

Categoría términos asociados a la web profunda

0 ocurrencias

Categoría - política, activismo y principios éticos

código (personas) 3 ocurrencias

código (comunidad) 2 ocurrencias

código (apoyo) 4 ocurrencias

código (donaciones) 4 ocurrencias

6.2.8 Inferencia de frecuencia de términos de la entrevista:

El término que aparece con más frecuencia es “books” pues si bien LibGen cuenta con varios tipos de documentos el sitio presenta una afinidad especial a los libros, a diferencia de Sci Hub que se basa mayoritariamente en “research papers”.

Lo sigue “library” que además de ser el nombre del sitio denota un carácter diferente a otras páginas de descarga, en LibGen tanto los filtros de búsqueda como la información de cada libro (isbn, edición, idioma) están siempre visibles tal cual lo estaría en una base de datos de cualquier biblioteca especializada. Luego los términos “apoyo” y “donaciones” con 8 ocurrencias entre los dos muestran el carácter colaborativo y no comercial del sitio. Finalmente, el término “acceso” está omnipresente en todos los sitios que luchan para la democratización del conocimiento y “people” que podría traducirse muchas veces cómo “pueblo” en el sentido de comunidad y de ciudadanía. Acceso para cualquier “persona” apoyo de la comunidad son de los tópicos recurrentes en la entrevista a los responsables de LibGen. Un sitio donde los libros y el carácter “bibliotecario” de sus administradores

se desprende de sus declaraciones.

6.2.9 Consideraciones adicionales

LibGen es totalmente anónimo y muy impersonal comparado con Sci Hub. Sus bases son totalmente descargables y su código es abierto y está disponible fácilmente. Por lo tanto, cualquier persona con mínimos conocimientos podría replicar LibGen tantas veces como quiera. De hecho, LibGen ha heredado parte de su colección de un sitio anterior, Library.nu.

LibGen es un sitio de colaboración, subir archivos es muy fácil, es muy difícil percibir la mano de un administrador, puede existir un equipo detrás, pero también podría ser un sitio casi autónomo. Cuando uno sube un archivo, se deben campos de la base de datos sin demasiado control, sin embargo, no es un sitio caótico ni desordenado. Buscar por autor, por idioma, filtrar por idioma o por edición no son tareas imposibles. LibGen es un sitio que para el que está acostumbrado a navegar por catálogos bibliográficos es un lugar que nos resulta muy conocido. Esa sensación de orden no se percibe en ningún sitio de piratería o descarga de archivos.

Si no fuera que al descargar un archivo junto con datos como isbn o una imagen de la portada vemos los múltiples servidores y la posibilidad de bajarlo en *torrent* o P2P podríamos confundirla con cualquier base de datos bibliográfica.

Pero también como encontramos en todos los sitios analizados tiene un código de ética y una carta de solidaridad llena de reivindicaciones. Lo que demuestra que diferentes aproximaciones y métodos llevan la misma mochila de defensa de derechos y lucha por el acceso al conocimiento. LibGen no tiene en su página un banner que diga que son el sitio pirata número uno. Pero desde hace casi diez años se han convertido en un sitio más que solvente a la hora de obtener recursos de información. Ha demostrado ser simple, efectivo y blindado. Técnicamente es una muy buena amalgama de tecnologías ya que sus recursos están en la web profunda, en redes P2P y torrent. Como dice (Bodó, 2018) “LibGen inherited the responsibility of serving a larger academic community beyond the boundaries of Russian-speaking academia”. Puede afirmarse que LibGen, nacida en el ámbito científico académico de una comunidad en concreto, logró con unas directrices ambiciosas

Collect valuable science/technology/math/medical/humanities academic literature. That is, to collect humanity 's valuable knowledge in digital form. Avoid junky books. Ignore bestsellers. Build a community of people who share knowledge, improve the quality of books, find good and valuable books, and correct errors. Share the files freely, spreading the knowledge altruistically, not trying to make money, not charging money for knowledge (Bodó, 2018)

Evitó los bestsellers y pudo difundir contenido de manera altruista, sobrevivió la demanda de 2015 por parte de Elsevier y al día de hoy es otro sitio sin publicidad

sin registro y que nos permite obtener las obras completas de Pierce en el idioma y formato que queramos y si lo quisiéramos armar nuestro propio clon de LibGen.

6.3 Propublica

6.3.1 Análisis formal del sitio

Servidor:

Cuenta con dos servidores

- a) web : <https://www.propublica.org/>
- b) web profunda: <https://www.propub3r6espa33w.onion/>

Dominio:

- a) alojado en un dominio tradicional: Está alojado en la ciudad de Aushburn en EEUU

Indexación

- a) Permite la indexación de contenido del tipo usado por los buscadores tradicionales

Acceso a contenido académico o de interés académico:

- a) Contiene contenido de interés académico. (artículos periodísticos, informes técnicos etc.)

Explicitación de ideales:

- a) Contiene o no una sección donde explicitan su visión, objetivo del proyecto, posición política.

Explicitación de responsables:

- a) El sitio deja claro quien o quienes son sus responsables e impulsores

Incluye publicidad de terceros para monetizar:

- b) No incluye publicidad

Donaciones y contribuciones

- a) Incluye en el cuerpo de la página un botón o una sección para donaciones

Análisis título y subtítulo:

Título: ProPublica

Subtítulo: Journalism in the Public Interest

Inferencia de título y subtítulo:

El nombre ProPublica y su subtítulo Periodismo de interés público hace referencia al espíritu del sitio que es poner en conocimiento del público en general los abusos de poder, posicionarse del lado de la gente, usando cómo herramienta el periodismo de investigación.

6.3.2 Análisis de contenido Propublica “Acerca de”

Tabla de Frecuencia

Término	Ocurrencias	Ranking
investigative journalism	5	1
power	4	2
Spur change	3	3
reporting	3	3
Public interest	3	3
abuses of power	2	4

Codificación

Categoría libertad de expresión

código (poder) 4 ocurrencias

código (abuso de poder) 2 ocurrencias

Categoría acceso

0 ocurrencias

Categoría publicación académica

código (periodismo de investigación) 5 ocurrencias

código (reportar - informar) 3 ocurrencias

Categoría problemas legales

0 ocurrencias

Categoría términos asociados a la web profunda

0 ocurrencias

Categoría - política, activismo y principios éticos

código (interés público) 3 ocurrencias

código (impulsar el cambio) 3 ocurrencias

6.3.3 Inferencia de la frecuencia de términos en la página “Acerca de:”

El término que aparece más frecuentemente es “periodismo de investigación” central a la identidad del sitio. Apareciendo ya desde el título y repitiéndose en el cuerpo más que ningún otro. En segundo lugar de frecuencia aparece el término “poder” referenciando a todo tipo de abuso merecedor de denuncias por el sitio. En tercer lugar, “interés público” como fin del tipo de periodismo investigativo que

propone el sitio, pro Pueblo, para la gente, Propublica. Luego “informar” y “reportar” que pueden ser sinónimos en algunos contextos, pero que en este cuerpo refieren a las acciones más habituales de los periodistas y por lo tanto de un sitio de periodismo investigativo. Al mismo nivel de frecuencia aparece “impulsar el cambio” algo que se repite también desde el título de la página, la organización busca no solo informar, sino busca que la información y el conocimiento sean motor de cambio “spur change”, el término “espolear” en español, en el sentido de pinchar a un caballo para que aumente su velocidad o arranque a andar si está en reposo, ejemplifica el sentido que da ProPublica al periodismo de investigación. La información de interés público y la denuncia de los abusos de poder por parte de gobiernos y empresas son esas espuelas que se clavan en los costados del animal adormilado que es el pueblo en su ignorancia de las problemáticas que los aquejan como sociedad.

Consideraciones: Dado que en la página “Acerca de” la extensión del corpus es muy acotada, se dificulta obtener indicadores confiables. Por lo tanto, analizaremos esta página, pero los indicadores los obtendremos de dos entrevistas de manera independiente: la primera una entrevista al presidente de Propublica Dick Tofel y la segunda una entrevista a Mike Tigas programador de la página de Propublica en la Deep Web.

6.3.4 Análisis cualitativo del sitio

La página es bastante acotada en texto, básicamente contiene la misión, que nos habla en términos bastante contundentes: “Exponer los abusos de poder y las traiciones a la confianza pública por parte del gobierno y empresarios”, “Usar la fuerza moral del periodismo de investigativo para impulsar reformas poniendo el foco en quien hace mal las cosas.”

About Us

- [About](#)
- [Board and Advisors](#)
- [Officers and Staff](#)
- [Diversity](#)
- [Reports](#)
- [Impact](#)
- [Awards](#)

The Mission

To expose abuses of power and betrayals of the public trust by government, business, and other institutions, using the moral force of investigative journalism to spur reform through the sustained spotlighting of wrongdoing.

Incluso la página contiene igualmente que LibGen tiene un “código de ética”.

Code of Ethics

- [About](#)
- [Board and Advisors](#)
- [Officers and Staff](#)
- [Diversity](#)
- [Reports](#)
- [Impact](#)
- [Awards](#)
- [Partners](#)
- [Supporters](#)
- [Media Center](#)
- [Press Releases](#)
- [Advertising Policy](#)
- [Jobs](#)
- [Fellowships](#)
- [Code of Ethics](#)

This Code is designed to supplement ProPublica’s Conflicts of Interest Policy (required by the Internal Revenue Service), and set out our expectations and aspirations for the conduct of our newsroom.

The mission of ProPublica is to practice and promote investigative journalism in the public interest. All of the values stated here, and the rules set out here, are intended to contribute to that mission. Much of the language below draws proudly on similar policies in place at distinguished American news organizations, including Dow Jones & Company, the Associated Press, the Washington Post and Time Inc. We do this because, while our entity is new, and our business model somewhat innovative, our ethics are neither. They reflect what we and others have learned over many years. At the same time, however, this Code is not immutable. Most of it consists of guidelines; exceptional circumstances may require exceptions to these rules. We expect to continue to learn, and, as we do so, to revise this document in light of further insight and experience.

6.3.5 Inferencia

El sitio deja muy claras sus convicciones sobre exponer al gobierno y empresarios que incurran en actos contra las personas en su conjunto, y afirman que el periodismo de investigación es el arma para impulsar el cambio hacia una sociedad mejor. Su misión dice: “Exponer los abusos de poder” por lo tanto Propublica se pone del lado del menos fuerte, en este caso el ciudadano común y “las traiciones a la confianza pública” refieren a “gobierno y empresarios ” que han sido elegidos en un voto de confianza ya sea por medios democráticos en el caso de los gobiernos o en un voto de confianza de los usuarios en el caso de las empresas. Propublica ve la información como un nivelador en esta desigualdad y la capacidad de alcanzar a más ciudadanía la fuerza de palanca necesaria para equiparar las desventajas del ciudadano común frente a las organizaciones que abusan de su posición. La segunda parte de la misión dice: “Usar la fuerza moral del periodismo investigativo para impulsar reformas poniendo el foco en quien hace mal las cosas.” La fuerza moral del periodismo investigativo no es otra que la fuerza de la verdad, este es un sitio sin fines de lucro, o sea que las investigaciones que aparecen no tienen una función comercial, son (o deberían ser) puramente herramientas para como dicen textualmente “el interés público”.

También poseen como hemos visto en varios de los sitios analizados un “código de ética” con las expectativas de cómo deben comportarse los miembros de nuevo con referencias a una “moral” y a los “valores” y a respetar la misión antes analizada.

Algunos puntos remarcables del código de Ética:

“Our facts are accurate and fairly presented, Our analyses represent our best independent judgments rather than our preferences, or those of our sources”, “There are no hidden agendas in any of our journalistic undertakings”, “We don’t pay for interviews”, “We don’t plagiarize”, “ProPublica does not contribute, directly or indirectly, to political campaigns or to political parties or groups seeking to raise money for political campaigns or parties”.

El énfasis está en el carácter no comercial, anti corporativo, imparcialidad y seriedad de las investigaciones. El desprejarse de periodismo partisano y de agenda al menos desde su perspectiva y alineado con estos valores éticos y morales representados en un texto bastante extenso.

6.3.6 Análisis de contenido entrevista What Is ProPublica?: Dick Tofel presidente

https://www.youtube.com/watch?v=o6Hs_EivKAE

Tabla de Frecuencia

Término	Ocurrencias	Ranking
investigative	6	1
change	6	1
investigative journalism	5	2
journalistic	4	3
spur change	4	3
big data	2	4
creative commons	2	4
databases	2	4
nonprofit	2	4
platforms	2	4
public interest	2	4

Codificación

Categoría libertad de expresión

sin ocurrencias

Categoría acceso

código (sin fines de lucro) 2 ocurrencias

Categoría publicación académica

código (periodismo de investigación) 4 ocurrencias

código (periodística) 3 ocurrencias

Categoría problemas legales

código (creative commons) 2 ocurrencias

5- Categoría - términos asociados a la web profunda

código (big data) 2 ocurrencias

código (plataformas) 2 ocurrencias

código (bases de datos) 2 ocurrencias

Categoría - política, activismo y principios éticos

código (impulsar el cambio) 3 ocurrencias

código (interés público) 2 ocurrencias

Kwic: Análisis del término en el contexto de la frase para detectar la orientación.

Investigative journalism / investigative news

4 legacy media expensive things like investigative journalism are being pushed down (-1)

8 we've created a nonprofit that does investigative journalism in the public interest (+1)

14 the mission of ProPublica is to do investigative journalism that spurs change (+1)

109 thousands of donors each year help us produce our investigative journalism in the public interest (+1)

113 produce impact and a spur change through investigative journalism at ProPublica (+1)

134 ProPublica is a relatively new kind of animal a non-profit investigative news organization (+1)

En las 6 ocasiones que aparece el término (periodismo investigativo) en 5 está asociado a términos positivos (interés público) 2 veces, (impulsar el cambio), (producir impacto), (sin fines de lucro/recién llegado (fresco)) y en una a términos negativos (dejado de lado) pero se refiere al periodismo tradicional no al que promueve ProPublica.

Término	Asociación mayoritaria (+)
periodismo investigativo	interés público impulsar el cambio producir impacto sin fines de lucro

Change/ spur change

11 the sole mission of which is to spur a forum to produce real change through journalistic means (+1)

15 investigative journalism that spurs change that creates reform (+1)

26 new kinds of journalism that use big data for instance but also by simply doing more

Stories so that we can produce change today and be in a position to produce even more tomorrow (+1)

60 ProPublica is focused on impact on spurring change through journalistic means (+1)

80 ProPublica our mission is impact through journalistic means spurring change spurring reform that's what we look for but beyond that we also look to engage our audience to help them better

Understand the world they live in and the stories that we and others are telling (+1)

112 providing ProPublica the chance to have impact and a spur change through investigative journalism at ProPublica we judge our accomplishments through the kind of impact our journalism has (+1)

El término (cambio) aparece 6 veces todas asociadas a términos positivos (verdadero significado del periodismo), (crear reformas), (producir más historias hoy y continuar creciendo) y (producir impacto)

Término	Asociación mayoritaria (+)
cambio	verdadero significado del periodismo crear reformas producir más historias producir impacto

Código	Positivo	Negativo	Neutro	Total
investigative	5	1	0	6
change	6	0	0	6

6.3.7 Indicadores

Corpus = 1129

Frecuencia (investigative) = $6/1129 \times 100 = 0,53$

Frecuencia (change) = $6/1129 \times 100 = 0,53$

Parcialidad (investigative) = $((5+1)/6) \times 100 = 100 \%$

Parcialidad (change) = $((6+0)/6) \times 100 = 100 \%$

Orientación (investigative) = $((5-1)/6) \times 100 = 66,6 \%$

Orientación (change) = $((6-0)/6) \times 100 = 100\%+$

6.3.8 Inferencia del análisis:

Los dos términos más frecuentes son (investigative) y (change). De los cuales se es totalmente parcial 100%. Esto muestra que el texto es una apasionada declaración acerca de estos dos términos. En cuanto a la orientación tiene una intensidad total hacia lo positivo en (change) y altamente orientada hacia lo positivo en (investigative).

Es un texto que claramente aboga por el cambio en la sociedad por parte del periodismo investigativo sin fines de lucro. Manifiestan que su objetivo es “impulsar el cambio” utilizando todas las herramientas tecnológicas disponibles en beneficio del público en general.

6.3.9 Análisis contenido entrevista: ¿why propublica join the dark web? Julia Angwin y Mike Tigas

<https://soundcloud.com/propublica/why-propublica-joined-the-dark-web>

Tabla de Frecuencia

Término	Ocurrencias	Ranking
web	24	1
tor	18	2
dark	18	2
know	16	3
website	10	4
hidden	9	5

Codificación

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

0 ocurrencias

Categoría publicación académica

0 ocurrencias

Categoría problemas legales

0 ocurrencias

Categoría - términos asociados a la web profunda

código (tor) 19 veces

código (dark web) 16 veces

código (hidden) 9 veces

Categoría política, activismo y principios éticos

0 ocurrencias

Kwic: Análisis del término en el contexto de la frase para detectar la orientación.

Tor

- 23 Using **Tor**, the website was able to sort of go undetected (+1)
- 25 So basically, if a regular person wants to go on the dark web, the way I understand it is that they would download the **Tor** Web browser (0)
- 29 It just happens to be a special type of website that you can only access through **Tor** browser (0)
- 33 I started experimenting with just using **Tor** around ProPublica related things to see if maybe we could protect ourselves from being censored (+1)
- 35 and people in China routinely use the **Tor** browser and these types of dark websites. (0)
- 39 nobody knows what it is unless you're using **tor**. Our content is, you know, freely available Right. To evade that censorship? Mike Tigas Yeah (+1)
- 39 just installed a special version of **Tor** for, you know, Web servers and we have (0)
- 39 have a special Web server listening to **Tor** connections that will serve our website (0)
- 49 I've known that not many places use **Tor** hidden services yet, but I hadn't really (-1)
- 51 might be because of the reputation that **Tor** has and that hidden services and these onion sites have (-1)
- 57 was that a landmark moment for **Tor** Hidden Services? ////Mike Tiggers//I think so (+1)
- 59 did it was because a lot of hackers use **Tor** to like attack normal websites. And so (-1)
- 59 And so by providing their own hidden service standpoint, they can tell **Tor** users use this. (0)
- 61 differently than if you're using **Tor** and just going to Facebook.com through (0)
- 61 It was actually it's it's sort of to help users who use **Tor** because they do know that there are countries that block Facebook (+1)
- 77 easy to remember names the exact **tor** hidden service like domain name for our site (0)

77 They're generated through an encryption technology inside **Tor** and so they're usually impossible to remember (0)

79 So people who are listening, you should try it. Download the Tor browser type in our hidden link and so you can tell your friends you've been on the dark web and it wasn't so bad (+1)

En las 18 apariciones del término Tor en 9 aparece asociados a términos neutros (bajar), (acceder), (usar rutinariamente), (instalar), (conexiones), (usuarios de), (usar), (servicios)y (tecnología interna de), en 6 asociado a términos positivos (no ser detectado), (protegerse de ser censurado), (evadir la censura), (momento histórico para), (ayudar a los usuarios), y (deberían probarlo) y en 3 ocasiones a términos negativos (no muy usado todavía), (mala reputación) y (usado para atacar sitios)

Término	Asociación mayoritaria (0)
Tor	bajar acceder usar rutinariamente instalar conexiones usuarios usar servicios tecnología

Dark web/ website

13 You may have heard of the **Dark Web** as a place where drugs and guns (-1)

15 But actually, **dark web** is just a way to browse the Internet (0)

17 to talk more about how we built our **dark Web** site is Mike Tiga (0)

19 So, Mike, tell me, what is the **dark Web**? (0)

19 Explain to listeners what it means to browse on the **Dark Web**. /Mike Tiga It really is just a way of accessing a Web site, you know, completely anonymous and securely. (+1)

21 The website can't see where you're coming from and you can't see where the **Dark Web** website is (0)

25 if a regular person wants to go on the **dark web**, the way I understand it is that they would download the Tor Web browser, which is just an operates like a normal Web browser (0)

31 And so what made you decide to build a **dark Web** site for ProPublica? (0)

35 use the Tor browser and these types of **dark websites**. Right. To evade that censorship (+1)

37 So how hard is it to build a **dark website**? /Mike Tiga: The way

ProPublica works, I mean, is we just have this special Web server that sits in front of our existing website (0)

43 all the traffic that comes in through our **dark web** hidden service site looks all the same to us. (0)

47 I think that both you and I were surprised when Wired came out with this story saying we were the first major news site on the **Dark Web** (0)

53 Facebook is the first big website to have gone on to the **Dark Web** (0)

55 they're one of the first pioneers, I guess, of an existing large, well-used website, having sort of a dual presence on the normal Internet and in the **dark Web**. (+1)

63 it is amazing because I think just a few years ago I myself would have thought like the **dark web**, like I don't know if I need to go there, but now it feels like there are so many people watching you when you're browsing the Web (0)

73 it's a little the weird thing about the **Dark Web** is you have to know already where you're going (-1)

79 tell your friends you've been on the **dark web** and it wasn't so bad. (+1)

De las 16 apariciones del término (Dark web) en 11 aparece asociado a términos neutros (una manera más de navegar por internet), (construir) 3 veces, (¿que es?), (ir a) 2 veces, (tener un pensamiento acerca de) (llegar a través de) y (estar en) en 3 ocasiones a términos positivos: (estar seguro), (evadir la censura), (ser pioneros) y 2 veces a términos negativos: (lugar de drogas y armas) y (extraño)

Término	Asociación mayoritaria (0)
Dark Web	navegar construir x3 ser ir x2 pensar llegar a través estar

Hidden

39 It's not like Silk Road where there is a secret thing that's completely **hidden** and nobody knows what it is unless you're using tor. Our content is, you know, freely available (+1)

43 So all the traffic that comes in through our dark web **hidden** service site looks all the same to us (0)

45 Your traffic is all mixed in with the traffic of everybody else who uses our **hidden** service (0)

51 To some extent, it might be because of the reputation that Tor has and that **hidden** services and these onion sites have, you know, places like Silk Road and all kinds of weird underground communities that exist in this space all sort of have this weird reputation for being places where you could just buy drugs and peddle child porn and things like that (-1)

57 was that a landmark moment for Tor **Hidden** Services? /I think so (+1)

59 And so by providing their own **hidden** service standpoint, they can tell Tor users use this (0)

77 easy to remember names the exact tor **hidden** service like domain name (0)

77 for our hidden service like domain name for our **hidden** services, (0)

79 Download the Tor browser type in our **hidden** link and so you can tell your friends you've been on the dark web and it wasn't so bad (+1)

El término (Hidden) en sus 9 apariciones 5 veces está asociado a término neutros: (iguales a otros servicios), (usar), (proveer servicios), (nombre de dominio) 2 veces, aparece 3 veces asociado a términos positivos:(no como sitios criminales “Silk Road”), (momento histórico), (cuéntale a tus amigos que no fue tan malo) y una vez a términos negativos (sitios dedicados a las drogas y la pedofilia)

Término	Asociación mayoritaria (0)
Hidden	igual a otros servicios usar proveer servicios nombre de dominio x2

Código	Positivo	Negativo	Neutro	Total
Tor	6	3	9	18
Dark Web	3	2	11	16
acceso	3	1	5	9

6.3.10 Indicadores

Corpus = 1412

Frecuencia (Tor) = $(18/1412) \times 100 = 1,3$

Frecuencia (Dark Web) = $(16/1412) \times 100 = 1,13$

Frecuencia (acceso) = $(9/1412) \times 100 = 0,64$

Parcialidad (Tor) = $((6+3)/18) \times 100 = 50\%$
Parcialidad (Dark Web) = $((3+2)/16) \times 100 = 67\%$
Parcialidad (acceso) = $((3+1)/9) \times 100 = 44\%$

Orientación (Tor) = $((6-3)/18) \times 100 = 16,6\%+$
Orientación (Dark Web) = $((3-2)/16) \times 100 = 6,25+$
Orientación (acceso) = $((3-1)/9) \times 100 = 22,2\%+$

6.3.11 Inferencia del análisis:

Los dos términos más frecuentes son (Tor) y (Dark Web) tratados de manera moderadamente parcial un poco más para (Dark Web) lo que indica una manera poco apasionada de acercarse a estos temas. El término que aparece en tercer lugar es (acceso) con la parcialidad más baja casi imparcial. En cuanto a la orientación en los tres términos es moderada hacia lo positivo siendo (acceso) la que es más intensamente positiva y (Dark Web) la que menos,

Es remarcable que al hablar de términos tan polémicos como la Dark Web o Tor el debate no sea acalorado o parcial. Esto se debe a que ProPublica ve a estos como herramientas para cumplir un objetivo que es que las personas puedan acceder a periodismo de investigación sin ser rastreadas y que su libertad sea comprometida.

6.3.12 Consideraciones adicionales

ProPublica es un sitio de noticias, ellos se definen como una agencia de noticias independiente y sin fines de lucro, no el típico sitio de noticias con temáticas variadas, aquí el papel principal lo juegan las denuncias de abusos de poder por parte de gobiernos y empresas, es un sitio activista y que propone la vigilancia ciudadana. El sitio existe desde 2008 y hasta este momento ha ganado seis premios Pulitzer. Lo que significa que es un sitio reconocido y que el nivel de sus investigaciones es muy alto. Se centran principalmente en la denuncia de los abusos de poder y los ataques a la libertad de expresión de cualquier tipo. ProPublica al igual que LibGen también ofrece sus datos para generar publicaciones o informes de terceros. Tiene secciones de su sitio con Datasets públicos descargables, estos son conjuntos de datos recaudados y organizados por ProPublica disponibles para quien sea que le interese hacer minería o análisis, al momento de consultar tenía datasets de investigaciones en salud, crímenes, educación, políticas, medio ambiente, etc. También cuenta con variadas APIs, que son interfaces de aplicación, estas interfaces permiten a productos de terceros sea aplicaciones móviles o páginas web manejar datos en tiempo real de los diferentes tópicos que investiga el portal. Para ProPublica la Deep Web es un refugio, es un lugar donde como lo manifiestan en la entrevista sus lectores o usuarios pueden acceder a su sitio sin tener que preocuparse por ningún tipo de represalia. Pero

también para Propublica la Deep Web es una declaración, al crear su sitio onion, lo publicitaron en podcasts y entrevistas en revistas super populares como Wired usando la Deep Web cómo alternativa, como herramienta, pero sobre todo como bandera.

Propublica es un sitio de periodismo moderno, proactivo que aprovecha la tecnología disponible para cumplir con sus objetivos, pero también es un sitio que desarrolla tecnología.

Aunque no sea una web de descarga como LibGen y Sci Hub descargas de software y datos. A diferencia de todos los demás sitios que analizaremos sus miembros son personajes reconocidos no tanto por ser actores sociales ni defensores de ideales sino por ser periodistas con grueso currículum, trayectoria y múltiples premios. Siendo el sitio que se toma más en serio la labor periodística no deja de tener el perfil utilitario de todos los sitios que analizaremos. No deja de tener un código ético fuerte, en el que se rechaza todo tipo de interés económico y el fuerte rechazo al sistema.

6.4 Freedom of the Press

6.4.1 Análisis formal del sitio

Servidor:

Cuenta con dos servidores

- a) servidor en la web superficial: <https://freedom.press/>
- b) servidor en la web profunda:

Consideración: No tiene su página principal en la Deep Web, pero si su servicio de intercambio de información Secure Drop:

<https://secdrop5wyphb5x.onion>

Dominio:

- a) alojado en un dominio tradicional en Montreal Canadá en la nube de CloudFlare

Indexación

- a) Permite la indexación

Acceso a contenido académico o de interés académico:

- a) Contiene contenido de interés académico. (artículos periodísticos, informes técnicos etc.)

Explicitación de ideales:

- a) Contiene una sección donde explicitan su visión, objetivo del proyecto, posición política.

No sólo tiene su misión y visión, también tiene un código de conducta.

6- Explicitación de responsables:

- a) El sitio deja claro quien o quienes son sus responsables e impulsores.

Incluye publicidad de terceros para monetizar:

- b) No incluye publicidad

Donaciones y contribuciones

- a) Incluye en el cuerpo de la página un botón o una sección para donaciones

Análisis título y subtítulo:

Título: Freedom of the Press

Subtítulo: Freedom of the Press Foundation protects, defends, and empowers public-interest journalism in the 21st century.

Inferencia de título y subtítulo:

La denominación Freedom of the Press deja claro el énfasis en la Libertad de prensa y de la prensa, en el subtítulo también aluden a la defensa y el empoderamiento del periodismo de interés público con un enfoque de los tiempos que corren.

6.4.2 Análisis de contenido “Acerca de” Freedom of The press

Tabla de Frecuencia

Término	Ocurrencias	Ranking
security	8	1
journalists	7	2
news	6	3
journalism	6	3
tools	5	4
organizations	5	4
digital	5	4
to protect	5	4
digital security	5	4

security tools	3	5
protect journalists	3	5
open source	3	5
right to know	2	6
public interest journalism	2	6
press freedom violations	2	6
press freedom tracker	2	6
freedom of information	2	6
critical press freedom	2	6

Codificación

Categoría libertad de expresión

código (libertad de información) 2 ocurrencias

código (libertad de prensa crítica) 2 ocurrencias

código (violaciones a la libertad de prensa) 2 ocurrencias

código (proteger al periodismo) 2 ocurrencias

Categoría acceso

0 ocurrencias

Categoría publicación académica

código (prensa crítica) 2 ocurrencias

código (periodismo y periodistas) 13 ocurrencias

código (noticias) 6 ocurrencias

Categoría problemas legales / criminales

código (open source) 2 ocurrencias

Categoría - términos asociados a la web profunda

código (seguridad digital) 5 ocurrencias

código (seguridad) 8 ocurrencias

Categoría - política, activismo y principios éticos

código (proteger) 5 ocurrencias

código (proteger a los periodistas) 3 ocurrencias

código (periodismo de interés público) 2 ocurrencias

código (derecho a saber) 2 ocurrencias

Kwic: Análisis del término en el contexto de la frase para detectar la orientación.

Protect

19 a non-profit organization that **protects**, defends, and empowers public-interest (+1)

25 give them the best available technologies to **protect** themselves and their sources (+1)

26 In partnership with the Committee to **Protect** Journalists, we run the US Press (+1)

27 technology projects that aim to **protect** journalists or promote transparency (+1)

28 excessive government secrecy, the **protection** of whistleblowers, the surveillance (+1)

31 Freedom of the Press Foundation aims to **protect** and promote the basic human right (+1)

32 freedom issue in the 21st century. To **protect** journalists, their sources and their audiences (+1)

El término (proteger) aparece 7 veces todas asociado a términos positivos (empoderar), (darles lo mejor), (proteger periodistas), (promover la transparencia), (proteger los informantes), (promover los derechos humanos), (proteger fuentes, proteger la audiencia)

Término	Asociaciones mayoritarias (+)
proteger	empoderar darles lo mejor proteger al periodismo promover la transparencia proteger los informantes promover los derechos humanos proteger fuentes

Journalism

19 defends, and empowers public-interest **journalism** in the 21st century. (+1)

21 that this kind of transparency **journalism** — from publishing the

Pentagon (+1)

29 open security tools are increasingly a critical component of the **journalism** landscape. (0)

29 it easy for people to support the best **journalism** from an array of organizations (+1)

31 Public interest **journalism** (+1) the kind of **journalism** (0) that holds power accountable and defends human rights is under threat everywhere

El término (periodismo) aparece 6 veces: 4 asociado términos positivos (defender y empoderar), (transparencia), (apoyar), (defender los derechos humanos) y 2 a términos neutros (panorama) y (tipo de)

Término	Asociaciones mayoritarias (+)
periodismo	empoderar transparencia apoyar defender los derechos humanos

Tools

20 including the development of encryption **tools**, documentation of attacks on the press, training newsrooms on digital security practices, and advocating for the public's right to know (+1)

25 help teach them about digital security **tools** and offer them the best available technologies (+1)

27 Encryption Tools & Special Projects. We also build a variety of experimental technology projects that aim to protect journalists or promote transparency (+1)

29 organizations, and open-source security **tools** are increasingly a critical component (+1)

32 newsrooms use best-available security **tools** and practices, including encryption of sensitive communications and materials, anonymization of sources, and distribution of news through secure and censorship-resistant channels (+1)

El término herramientas aparece asociado a términos positivos las 5 veces que aparece: (derecho a saber), (ayudar), (proteger), (abierto), (seguro)

Término	asociaciones mayoritarias (+)
herramientas	derecho a saber ayudar proteger abierto seguro

Código	Positivo	Negativo	Neutro	Total
protect	7	0	0	7
journalism	4	0	2	6
tools	5	0	0	5

6.4.3 Indicadores

Corpus= 536

Frecuencia (protect) = $(7 / 536) 100 = 1,3$

Frecuencia (journalism) = $(6 / 536) 100 = 1,11$

Frecuencia (tools) = $(5 / 536) 100 = 0,93$

Parcialidad (protect) = $((7 + 0) / 7) \times 100 = 100 \%$

Parcialidad (journalism) = $((4 + 0) / 6) \times 100 = 67\%$

Parcialidad (tools) = $((5 + 0) / 5) \times 100 = 100\%$

Orientación (protect) = $((7 - 0) / 7) \times 100 = 100 \%\+$

Orientación (journalism) = $((4 - 0) / 6) \times 100 = 67 \%\+$

Orientación (tools) = $((5 - 0) / 5) \times 100 = 100 \%\+$

6.4.4 Inferencia

Los tres términos más frecuentes (proteger), (periodistas) y son tratados de manera bastante poco parcial. Siendo en (herramientas) y (proteger) totalmente parcial 100% y en (periodismo) altamente parcial 67%. En cuanto a la orientación es totalmente positivo en (herramientas) y (proteger) y altamente positivo para (periodistas).

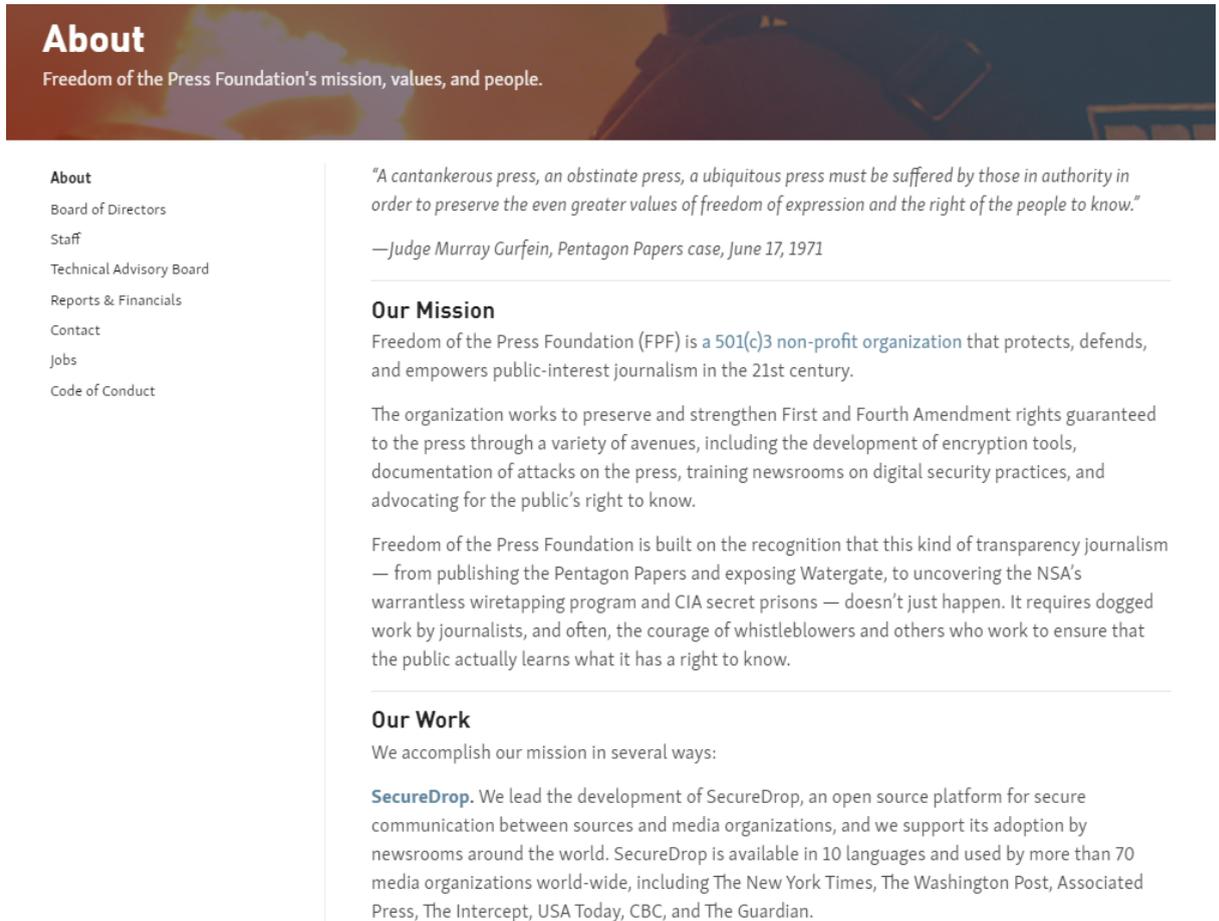
El tema recurrente en el texto son las herramientas para proteger a los periodistas. En nuestro caso una herramienta de la Deep Web como Securedrop y la explicitación clara para utilizar Tor. Freedom of the Press confía en estas herramientas para empoderar, apoyar a los periodistas, aumentar la transparencia y defender los derechos humanos.

6.4.5 Análisis cualitativo del sitio

Lo primero que aparece es una frase de Murray Gurfein: "la prensa cascarrabias,

obstinada, omnipresente debe soportar los ataques de la autoridad para lograr preservar los valores de libertad de expresión y el derecho de la gente a saber.”

Luego en la misión se plantea que Freedom of the Press es una organización sin fines de lucro que protege, defiende y empodera el periodismo en el siglo XXI.



The screenshot shows the 'About' page of the Freedom of the Press Foundation. The header features the word 'About' in a large, bold, white font against a dark background with a faint map of the United States. Below the header, a navigation menu lists various sections: About, Board of Directors, Staff, Technical Advisory Board, Reports & Financials, Contact, Jobs, and Code of Conduct. The main content area is divided into three sections: a quote from Judge Murray Gurfein, 'Our Mission', and 'Our Work'. The quote states: "A cantankerous press, an obstinate press, a ubiquitous press must be suffered by those in authority in order to preserve the even greater values of freedom of expression and the right of the people to know." The 'Our Mission' section describes the Freedom of the Press Foundation (FPF) as a 501(c)3 non-profit organization that protects, defends, and empowers public-interest journalism in the 21st century. The 'Our Work' section mentions the development of SecureDrop, an open source platform for secure communication between sources and media organizations.

Después de la misión lo primero en aparecer fue el proyecto Securedrop donde más de 70 medios internacionales para recibir denuncias sin que los informantes puedan ser identificados.

En el apartado de los responsables, el primer responsable en aparecer en la página web es Edward Snowden, estudiado anteriormente y un símbolo de lo que significa el activismo por develar las violaciones a la privacidad por parte de los gobiernos.

Board of Directors

About

Board of Directors

Staff

Technical Advisory Board

Reports & Financials

Contact

Jobs

Code of Conduct



Edward Snowden

Edward Snowden is a former intelligence officer who served the CIA, NSA, and DIA for nearly a decade as a subject matter expert on technology and cybersecurity. In 2013, he revealed the NSA was unconstitutionally seizing the private records of billions of individuals who had not been suspected of any wrongdoing, resulting in the largest debate about reforms to US surveillance policy since 1978. He has received awards for courage, integrity, and public service, and was named the top global thinker of 2013 by Foreign Policy magazine. Today, he works on methods of enforcing human rights through the application and development of new technologies. He joined the board of Freedom of the Press Foundation in February 2014 and was named the president of the board in 2016.



John Cusack

John Cusack is an actor, director, producer, and screenwriter who has appeared in over 60 films. He's also a political activist and regularly speaks out and writes on issues of human rights, government transparency, and accountability—amongst other things.



Daniel Ellsberg

Daniel Ellsberg is a co-founder of the Freedom of the Press Foundation. He is best known as the whistleblower who gave the Pentagon Papers to the *New York Times* in 1971. Ellsberg is also the author of three books: *Papers on the War* (1971), *Secrets: A Memoir of Vietnam and the Pentagon Papers* (2002), and *Risk, Ambiguity and Decision* (2001). In December 2006, he won the Right Livelihood Award, known as the "Alternative Nobel Prize," in Stockholm, Sweden, "for putting peace and truth first, at considerable personal risk, and dedicating his life to inspiring others to follow his example."

En la página de Securedrop en rojo y encima de todo directamente dice: Baje e instale "Tor". Una muestra clara de que la Deep Web en este ámbito es una herramienta muy útil e imprescindible.

Have a document to share?

You are not anonymous while using this browser!

Here's what you need to do to download Tor Browser to be anonymous online:

1. Go to <https://www.torproject.org>.
2. Download and install Tor Browser.
3. Launch Tor Browser and return to this site. You are now anonymous online.

También aparece que el proyecto fue desarrollado inicialmente por Aaron Swartz referido anteriormente y vuelto a nombrar en la página LibGen. Evidentemente el nombre de Aaron Swartz es el del máximo referente en la lucha por el acceso al conocimiento científico y la lucha contra el copyright y colocarlo indica una toma de posición por parte de Freedom of the Press.



SecureDrop is an open-source whistleblower submission system that media organizations can install to securely accept documents from anonymous sources. It was originally coded by the late Aaron Swartz and is now managed by [Freedom of the Press Foundation](#).

This page contains a list of active instances. Is your SecureDrop instance missing? Please [submit it here](#). Provided that your setup meets reasonable security standards, we are happy to add it to the directory.

This list is also available for programmatic access at our [API endpoint](#).

6.4.6 Inferencia

El sitio contiene una fuerte carga política y activista. Directamente insta a usar tecnologías de la Deep Web y resalta colaboradores como Aaron Swartz y Edward Snowden, dos de los activistas más renombrados por el acceso y democratización del conocimiento y la protección de los datos personales y la vulneración de la privacidad por parte de los gobiernos.

6.4.7 Análisis de contenido página del proyecto “Acerca de: SecureDrop”

<https://securedrop.org/directory/>

Tabla de Frecuencia

Término	Ocurrencias	Ranking
government	10	1
news	9	2

justice	7	3
business	7	3
technology	6	4
corruption	5	5
national security	5	5
tor	4	6
liberties	4	6
donate	4	6
contribute	4	6
browser	4	6
anonymous	4	6
social justice	4	6
civil liberties	4	6

Codificación

Categoría libertad de expresión

0 ocurrencias

Categoría acceso

0 ocurrencias

Categoría publicación académica

código (noticias) 9 ocurrencias

Categoría problemas legales

código (ciberseguridad) 3 ocurrencias

Categoría - términos asociados a la web profunda

código (tecnología) 6 ocurrencias

código (tor) 4 ocurrencias

código (anonimato) 4 ocurrencias

Categoría - política, activismo y principios éticos

código (donaciones y contribuciones) 8 ocurrencias

código (justicia) 7 ocurrencias

código (corrupción) 5 ocurrencias

código (libertades civiles) 4 ocurrencias

6.4.8 Inferencia de frecuencia de términos en “Acerca de:” Secure Drop

La página se enfoca mayoritariamente en la tecnología que permite a las personas divulgar datos del gobierno de manera anónima. De cómo denunciar la corrupción, violaciones de las libertades civiles.

También hacen énfasis en que es un sitio de noticias sin fines de lucro que depende de donaciones y contribuciones. El término que parece con mayor frecuencia es “noticias” algo totalmente natural en un sitio de prensa cuya misión es comunicar. En segundo lugar, aparece “justicia” como el objetivo de este proyecto justicia social al permitir el acceso a la información. En tercer lugar “donaciones” y “contribuciones” destacando el carácter colaborativo y sin fines de lucro de la organización, conminando a cada persona a colaborar y apoyar el proyecto repetidas veces dentro del cuerpo del texto. En cuarto lugar “tecnología” ya que SecureDrop es un producto tecnológico, una herramienta para que los informantes puedan llegar a los periodistas sin ser interceptados, ya sea por parte de los gobiernos o las organizaciones denunciadas. En quinto lugar, aparece el término “corrupción” tema de muchas de las denuncias que llegan a todos los sitios de noticias, pero en especial a los que buscan difundir abusos por parte de gobiernos o empresas. Ya en sexto lugar todos con cuatro ocurrencias “anonimato”, siendo esto lo que busca preservar la deepweb y que es fundamental para los informantes, “Tor” el buscador que permite acceder a la web profunda y “libertades civiles” entendidas en este contexto cómo el derecho a la privacidad y a no ser monitoreado ni rastreado eventualmente.

6.4.9 Consideraciones adicionales

Freedom of the Press es una organización internacional sin fines de lucro cuyo objetivo principal es defender la libertad de expresión y la libertad de prensa.

Entre la junta de directores de la fundación tenemos a Daniel Ellsberg, Edward Snowden y John Barlow. El primero, exanalista de las fuerzas armadas de EEUU, famoso por filtrar más de 7000 páginas información al New York Times sobre la guerra de Vietnam. Estos sucesos fueron conocidos en 1971 como los Pentagon Papers, Edward Snowden del que ya hablamos antes, es un experto en seguridad informática ex empleado de la CIA y la NSA que en 2001 filtró información sobre vigilancia masiva hacia civiles por varias organizaciones gubernamentales de varios países. John Barlow es un reconocido activista de internet y autor de la “declaración de independencia del ciberespacio”. Freedom of the Press Foundation colabora

directamente con el proyecto Tor, tiene servicios en la Deep Web y lleva adelante el proyecto “Secure drop” desarrollado por Aaron Schwartz que permite a los informantes transferir información de forma segura mediante la Deep Web. Secure Drop hace lo mismo que hizo Schwartz desde el MIT en 2001 con las colecciones de la editorial JSTOR, algo que le costó la persecución por parte del gobierno de EEUU y eventualmente su vida. Secure Drop utiliza los TOR *hidden services* para resguardar el contenido. En ambos casos se trasvasa web superficial a la Deep Web, donde las editoriales no pueden ejercer la misma presión que en el caso de Schwartz y donde los informantes no pueden ser arrestados por compartir documentos en el caso de Freedom of the Press. Algo que está más que claro para Snowden y Ellsberg siendo que el periodista pudo reivindicarse, pero Snowden aún no ha podido volver a su país. El fuerte mandato a usar Tor por parte de una organización cuyo nombre empieza con la palabra Libertad nos da mucho para reflexionar. Hemos visto que la Deep Web es un espacio de supervivencia para alguno de los sitios estudiados, una declaración para otros y aparentemente para Freedom of the press un espacio para cumplir sus objetivos. Barlow uno de los directores de Freedom of the Press en su declaración de independencia del ciberespacio nos dice:

“Declaro el espacio social global que estamos construyendo independiente por naturaleza de las tiranías que estáis buscando imponernos. No tenéis ningún derecho moral a gobernarnos ni poseéis métodos para hacernos cumplir vuestra ley que debemos temer verdaderamente” (Barlow, J. 1996)

7.0 Resumen de los cuatro sitios relevados

Aspectos Formales

Servidor:

3 tienen sus sitios principales alojados tanto en la web superficial como en la web profunda, y uno ofrece servicios en la Deep Web.

Dominio:

Tanto ScHub como Lib Gen al haber sido objeto de múltiples demandas e intentos de dar de baja el sitio están alojados en múltiples dominios y en países con leyes de copyright laxas, los sitios de periodismo de investigación tienen sus dominios principales en países del primer mundo.

Indexación

Tanto a Sci Hub como LibGen no les sirve que se sepa qué contenido albergan en sus sitios así que no permiten que los buscadores indexen su contenido. Es el caso opuesto con los sitios de periodismo de investigación

Acceso a contenido académico o de interés académico:

Todos estos sitios contienen contenido académico o de interés académico. Sci Hub y Libgen directamente publicaciones académicas y Propublica y Freedom of the Press informes técnicos y notas de periodismo de investigación

Explicitación de ideales:

Todos los sitios dejan muy claro su visión, sus ideales políticos e incluso dos explicitan códigos de conducta responsable.

Explicitación de responsables:

Sin ser LibGen todos los sitios tienen responsables claramente identificados.

Publicidad

Ninguno de los sitios monetiza mediante publicidad de terceros.

Donaciones y Contribuciones

Sin ser LibGen todos los sitios tienen un botón que permite colaborar con los sitios.

7.1 Coincidencias de las categorías en el cuerpo de los sitios

Categorías	SciHub	LibGen	ProPublica	Freedom of the Press
Libertad de Expresión	0	0	6	8
Acceso	2	7	0	0
Publicación Académica	5	41	8	21
Problemas Legales	6	0	0	13
Términos	0	1	0	2

asociados a web profunda				
Política, activismo y principios éticos	12	15	6	13

7.2 Coincidencias de las categorías en las entrevistas a los responsables

Categorías	SciHub	LibGen	ProPublica	Freedom of the Press
Libertad de Expresión	0	0	0	0
Acceso	74	3	0	0
Publicación Académica	164	16	0	9
Problemas Legales	0	0	0	3
Términos asociados a web profunda	5	0	44	14
Política, activismo y principios éticos	12	13	0	24

7.3 Resumen de de indicadores

7.3.1 Parcialidad

Término	SciHub	LibGen	ProPublica	Freedom of the Press
acceso	89%	80%	44%	
investigación	67%		100%	
copyright	100%			

conocimiento		100%		
académico		67%		
editoriales		100%		
Elsevier		80%		
cambio			100%	
Tor			50%	
Dark Web			67%	
proteger				100%
periodismo				67%
herramientas				100%

7.3.2 Orientación

Término	SciHub	LibGen	ProPublica	Freedom of the Press
acceso	89% +	40% -	22% +	
investigación	67%+		67% +	
copyright	100% -			
conocimiento		40% -		
académico		33% -		
editoriales		50% -		
Elsevier		80% -		
cambio			100% +	
Tor			17% +	
Dark Web			6% +	
proteger				100% +
periodismo				67% +
herramientas				100% +

8 Respuestas a las preguntas de investigación

1- ¿Encontramos información académica en la web profunda?

Encontramos información académica, investigaciones científicas, obras originales y normas técnicas. Nos lo indica el enorme acervo que tienen colecciones como la de Sci Hub o Library Genesis en servidores onion (Himmelstein et al, 2018). Sitios periodísticos como Freedom of the Press o Propública nos dan acceso a artículos periodísticos e investigaciones a salvo de la censura y las violaciones a la libertad de expresión.

2- ¿Quiénes producen esta información?

Académicos investigadores que buscan compartir recursos, usuarios preocupados por proteger su privacidad. (McCarthy, 2016), (Miguel, 2017).

Periodistas como lo indica el sitio "Reporters sans frontières" o ciudadanos que consideran que sus derechos son vulnerados o que no pueden acceder a aplicaciones o sitios bloqueados por políticas gubernamentales o empresariales. (Freedom on the Net, 2016), (McCarthy, 2016), (Reporters Sans Frontiers, 2018)

¿Por qué se encuentran allí?

Siempre que haya persecución va a haber lugares donde las personas busquen comunicarse y compartir, la Web no es la excepción. La prohibición de aplicaciones en países con regímenes autoritarios y la invasión de la privacidad en general, han hecho que software de anonimato, redes privadas y otras herramientas de la web profunda proliferen. La injusta apropiación del conocimiento por parte de las editoriales genera movimientos y soluciones informáticas para contrarrestar las inequidades que esto produce, por otra parte, el continuo ataque a estos sitios que buscan compartir la información académica, ha hecho que para sobrevivir migren a servidores alternativos y redes ocultas.

Lo antes expuesto y la sinergia que esto genera, hace que la web profunda continúe creciendo con actores cada vez más definidos y comprometidos con el problema, estos a su vez participan del desarrollo de nuevas herramientas tecnológicas y nuevos espacios de libertad y militancia.

9 Conclusiones

"La locura de confundir una hipálage con una prueba, un vómito de palabras con una fuente de verdades capitales, y a uno mismo con un oráculo, es un mal que nace junto con

nosotros.” Paul Valéry, *Introduction à la méthode de Léonard de Vinci* (1895)

Si queremos ser parte de una sociedad que haga énfasis en la adquisición del conocimiento y la libertad de expresión, la web profunda, las tecnologías de anonimato y protección de la privacidad, no deben ser vistas como tópicos alejados de nuestro quehacer universitario. Sino que debemos empezar a entenderlas como armas a nuestro favor, partes importantes de nuestro presente y posibles vehículos hacia una sociedad más igualitaria, pero sobre todo más libre.

En los últimos años la realidad nacional ha mostrado que los problemas estudiados en este trabajo: los ataques a la libertad de expresión y las dificultades de acceso a la literatura científica se hacen cada día más frecuentes e ineludibles. (Asociación de Bibliotecólogos del Uruguay, 2021), (Facultad de Información y Comunicación, 2021), (El Observador, 2021), (Centro de Archivo y Acceso a la Información Pública, 2022), (Reporters Sans Frontiers, 2021)

El teólogo alemán Dietrich Bonhoeffer famoso por sus disquisiciones entre “maldad” y “estupidez” (y finalmente ejecutado por el régimen Nazi), nos plantea en su *Ética*: (Bonhoeffer, 2008) “Responsibility and freedom are corresponding concepts. Factually, though not chronologically, responsibility presupposes freedom and freedom can consist only in responsibility.”

Una corriente mayoritaria, un “bien hacer” normalmente es una rueda que todos empujamos en diferente medida, quizás sea tiempo de detenerla y de reconsiderar nuestro rumbo. En este caso la responsabilidad no es elegir un bando, sino abrir el espectro de posibilidades: presentar alternativas y visiones, y porque no, decidirnos a traspasar fronteras.

A lo largo de este trabajo hemos visto cómo las tecnologías estudiadas contribuyen tanto para asegurar un acceso universal como para preservar información de la censura y la manipulación, estas herramientas han sido, no solamente han servido de ayuda para instituciones y personas, sino que pueden llegar a ser el germen de una nueva cultura.

“You claim there are problems among us that you need to solve. You use this claim as an excuse to invade our precincts. Many of these problems don't exist. Where there are real conflicts, where there are wrongs, we will identify them and address them by our means. We are forming our own Social Contract. This governance will arise according to the conditions of our world, not yours. Our world is different.” (Barlow, 2019)

Barlow en su declaración de la independencia del ciberespacio plantea: “Ustedes afirman que existen problemas que resolver, y utilizan esa excusa para invadir nuestros precinctos, nosotros somos el conflicto, nosotros somos lo que está mal, nosotros estamos formando nuestro propio contrato social, esta gobernanza se hará bajos las condiciones de nuestro mundo, no del suyo. Nuestro mundo es diferente”.

Por su lado, Castells plantea que subvirtiendo el orden existente y desafiando las relaciones de poder podemos generar ese cambio y ese nuevo contrato.

Si sentimos o pensamos de forma distinta, adquiriendo nuevos significados y reglas para entender este significado, actuaremos de forma diferente, y terminaremos cambiando la forma en que funciona la sociedad, bien subvirtiendo el orden existente, bien alcanzando un nuevo contrato social que reconozca las nuevas relaciones de poder que resultan de los cambios en la mente pública. Por tanto, la tecnología de la comunicación que moldea un entorno comunicativo determinado tiene consecuencias importantes en el proceso de cambio social. (Castells, 2009)

Ese tipo de libertad de la que habla Bonhoeffer, el pararse de un lado u otro, del lado de la propiedad intelectual y de la vigilancia, o elegir ir más allá de los adjetivos y de los discursos (como dice Valery al comienzo de este capítulo) para tomar un rumbo de responsabilidad es una oportunidad que tenemos para equilibrar las cosas.

La Deep Web en conclusión no es ni más ni menos que una herramienta de libertad, darle su uso apropiado, potenciarla, abrazarla o cuestionarla, pero desde el conocimiento, el análisis y la crítica, nos enriquece, puede hacernos más libres y poderosos, puede ser germen de nuevos escenarios tecnológicos y por qué no, de una sociedad distinta, donde las estructuras de poder y las fronteras estén subvertidas.

Las redes de anonimato y en especial la Deep web con su fama de lugar recóndito y prohibido han demostrado ser refugio y bastión de ideales, lugar de interés académico por su contenido, por darnos la oportunidad de ser más libres. Lugar más allá de la máscara y la ocultación, que genera un espacio nuevo en el que podemos ser más nosotros y menos lo que se supone que deberíamos ser.

“Man is least himself when he talks in his own person. Give him a mask, and he will tell you the truth.” Oscar Wilde

10 Bibliografía

- Abela, J. (2002). Las técnicas de análisis de contenido: una revisión actualizada. Fundación Centro de Estudios Andaluces.
- AccesNow. (2021) Shattered dreams and los opportunities: A year in the fight to #KeepItOn. https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_3.pdf
- ALA. (2013). Posthumously Honors Aaron Swartz with Madison Award. <https://www.districtdispatch.org/2013/03/american-library-association-honors-aaron-swartz-with-madison-award/>
- Aluffi, G. (2017). Es cierto: robo a los editores para dárselo a los científicos. https://elpais.com/elpais/2017/02/06/ciencia/1486399819_243966.html
- Araya-Moreno, C. (2016). Escritura binaria, felicidad y juego en la filosofía de Leibniz. El origen teológico del código binario de Leibniz. *LifePlay: Revista académica internacional sobre videojuegos*, (5), 17-41.
- Asociación de Bibliotecólogos del Uruguay. (2021). Comunicado de la ABU sobre recorte de fondos destinados a la Agencia Nacional de Investigación e Innovación (ANII). <https://www.abu.net.uy/2021/03/comunicado-de-la-abu-sobre-recorte-de-fondos-destinados-a-la-agencia-nacional-de-investigacion-e-innovacion-anii/>
- Babbage, C. (2010) Babbage's Calculating Engines: Being a Collection of Papers Relating to them; their History and Construction. Cambridge University Press.
- Bacon, F. (1605). Bacon; the Advancement of Learning. https://oll-resources.s3.us-east-2.amazonaws.com/oll3/store/titles/1433/Bacon_AdvancementLearning0414.pdf
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval (Vol. 463). New York: ACM press.
- Banks, M. (2008). On the way to the web: The secret history of the internet and its founders. Apress.
- Baran, P. (1964). On distributed communications: I. Introduction to distributed communications networks. RAND CORP
- Bardin, L. (1991). Análisis de contenido. Akal.
- Barlow, J. P. (2019). A Declaration of the Independence of Cyberspace. *Duke Law & Technology Review*, 18(1), 5-7.

- Bazli, B., Wilson, M., & Hurst, W. (2017). The dark side of I2P, a forensic analysis case study. *Systems Science & Control Engineering*, 5(1), 278-286.
- Belkin, N (1978), Information concepts for information science, *Journal of Documentation*, Vol. 34 Iss 1 pp. 55 - 85
- Bergman, M. (2001). White Paper: The Deep Web: Surfacing hidden value. *The journal of electronic publishing* 7(1). [Aug 2016].
- Berners-Lee, T. (2000). Tejiendo la red: el inventor del World Wide Web nos descubre su origen. *Siglo 21*.
- Bodó, B. (2018). The genesis of library genesis: The birth of a global scholarly shadow library. *Shadow libraries: access to knowledge in global higher education*.
- Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science*, 352(6285), 508-512.
- Bonhoeffer, D. (2008). *Ethics* (Vol. 6). Fortress Press.
- Boole, G. (1847). *The mathematical analysis of logic*. Philosophical Library.
- Brownlee, K. (2016). The civil disobedience of Edward Snowden: A reply to William Scheuerman. *Philosophy & Social Criticism*, 42(10), 965-970.
- Bush, V. (1945). As we may think. *The Atlantic monthly*, 176(1), 101-108.
- Bykov, I. A., Medvedeva, M. V., & Hradziushka, A. A. (2021, April). Anonymous Communication Strategy in Telegram: Toward Comparative Analysis of Russia and Belarus. In *2021 Communication Strategies in Digital Society Seminar (ComSDS)* (pp. 14-17). IEEE.
- Cusick, T. W., & Stanica, P. (2017). *Cryptographic Boolean functions and applications*. Academic Press.
- Capurro, R. (2016) Impactos de las tecnologías digitales de la información y comunicación en la filosofía, las artes y las ciencias. <http://www.capurro.de/campinas.html>
- Carracedo, J. (2004). *Seguridad en redes telemáticas*. McGraw Hill
- Castells, M. (2009). *Comunicación y Poder*. Alianza
- Castells, M. (2013). *Piracy cultures: How a growing portion of the global population is building media relationships through alternate channels of obtaining content*. Xlibris Corporation.

- Castells, M. (2014). El impacto de internet en la sociedad: una perspectiva global. *C@mbio*, 19, 127-149.
- Castillo, T (2018). Multa histórica de la UE a Google por Android: 4.340 millones de euros por abuso de posición dominante.
<https://www.xataka.com/moviles/multa-historica-ue-google-android-4-340-millones-euros-abuso-posicion-dominante>
- Centro de Archivo y Acceso a la Información Pública. (2022). Cainfo registró 69 amenazas y restricciones a la libertad de expresión en el último año.
<http://cainfo.org.uy/sitio/cainfo-registro-69-amenazas-y-restricciones-a-la-libertad-de-expresion-en-el-ultimo-ano/>
- Cerf, V., Dalal, Y., & Sunshine, C. (1974). *Specification of internet transmission control program* (pp. 1-70). RFC 675, December.
- Chitika. (2018) The Value of Google Result positioning. disponible en:
<https://chitika.com/2013/06/07/the-value-of-google-result-positioning-2/>
- Codina, L. (2000). Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista española de documentación científica*, 23(1), 9-44.
- Codina, L., Pedraza, R., Noci, J. D., Rodríguez-Martínez, R., Pérez-Montoro, M., & Cavaller-Reyes, V. (2014). Sistema Articulado de Análisis de Cibermedios (SAAC): Una propuesta sobre el qué y el cómo para estudiar medios de comunicación digitales. *Hipertext. Revista Académica sobre Documentación Digital y Comunicación Interactiva*, (12).
- Copeland, D. A., & Copeland, D. (2010). *The media's role in defining the nation: the active voice* (Vol. 5). Peter Lang.
- Córdoba, S. (2017) Nick, cofundador de TOR browser explica cómo navegar con seguridad.
<https://cacharrerodelaweb.com/2017/06/cofundador-tor-explica-navegar-seguridad.html>
- da Cruz, F. (19XX), *Columbia University Computing History :The Jacquard Loom*.
<http://www.columbia.edu/cu/computinghistory/#eckert>
- Davies, D. W., Bartlett, K. A., Scantlebury, R. A., & Wilkinson, P. T. (1967). A digital communication network for computers giving rapid response at remote terminals. *Proceedings of the ACM Symposium on Operating System Principles*
- Davis, M. (2014). Logic and the development of the computer. In *Computational Logic* (pp. 31-38).
- Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.

- Devine, J., & Egger-Sider, F. (2004). Beyond Google: the invisible web in the academic library. *The Journal of Academic Librarianship*, 30(4), 265-269.
- EFE (2013). Cronología del 'caso Snowden', el joven que reveló el espionaje masivo de Estados Unidos.
<https://www.20minutos.es/noticia/1850380/0/caso-snowden/cronologia/espionaje-ee-uu/>
- El Observador. (2021) El impacto del ajuste en la ANII en el acceso a la literatura científica internacional.
<https://www.elobservador.com.uy/nota/el-impacto-del-ajuste-en-la-anii-en-el-acceso-a-la-literatura-cientifica-internacional-20213515251>
- Ermoshina, K., & Musiani, F. (2021). The Telegram ban: How censorship “made in Russia” faces a global Internet. *First Monday*, 26(5).
- Essien, E. E. (2019). Relevance of the Deep Web to Academic Research. *International Journal of Natural and Applied Sciences (IJNAS)*, VOL. 12, Special Edition
- Facultad de Información y Comunicación. (2021). Reflexiones sobre Timbó: su relevancia para la investigación científica.
<https://fic.edu.uy/noticia/reflexiones-sobre-timbo-su-relevancia-para-la-investigacion-cientifica>
- Fall, K. R., & Stevens, W. R. (2011). *TCP/IP illustrated, volume 1: The protocols*. Addison Wesley.
- Fernández, L (2015). Ciencia-espectáculo: Impresionando al Rey.
<http://www.detectivesdelahistoria.es/ciencia-espectaculo-impresionando-al-rey/>
- Foucault, M. (1967). *Des espaces autres*.
<https://www.cairn.info/revue-empan-2004-2-page-12.htm>
- Fawkes, G. (2021). VPN Use and Data Privacy Stats for 2021.
<https://www.vpnmentor.com/blog/vpn-use-data-privacy-stats/>
- Fundeu (2018). Internet profunda, mejor que Deep Web.
<https://www.fundeu.es/recomendacion/internet-profunda-mejor-que-deep-web/>.
- Freedom on the Net 2016, Silencing the Messenger: Communication Apps Under Pressure.
https://freedomhouse.org/sites/default/files/2020-02/FOTN_2016_BOOKLET_FINAL.pdf
- Garfield, E. (1964). " Science Citation Index"-a new dimension in indexing. *Science*, 144(3619), 649-654.

- Geselowitz, M (2019) The Jacquard Loom: A Driver of the Industrial Revolution It automated weaving and inspired punch-card computers. <https://spectrum.ieee.org/the-jacquard-loom-a-driver-of-the-industrial-revolution>
- Google. (2018). Reporte de transparencia. https://transparencyreport.google.com/traffic/overview?group=REGION&disruption_history=start:1235865600000;end:1238543999999;product:21;region:CN&lu=disruption_history
- Google. (2021). Reporte de transparencia. <https://transparencyreport.google.com/user-data/overview>
- Gordy, C. (2016). Why ProPublica Joined the Dark Web. <https://www.propublica.org/podcast/why-propublica-joined-the-dark-web>.
- Grosskurth, A., & Godfrey, M. Architecture and Evolution of the Modern Web Browser, June 2006.
- Himanen, P. (2002). *La ética del hacker y el espíritu de la era de la información*. Destino.
- Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Tzovaras, B. G., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife*, 7, e32822.
- Home, R. W. (1979). Nollet and Boerhaave: a note on eighteenth-century ideas about electricity and fire. *Annals of science*, 36(2), 171-175.
- Internet Hall of Fame. (2013). Global connector: Ida Holz. <https://www.internethalloffame.org/inductees/ida-holz>
- Isaacson, W. (2014). *The innovators: How a group of inventors, hackers, geniuses and geeks created the digital revolution*. Simon and Schuster.
- Issa, D. Cummings, E. (2013). Ask for Briefing on Swartz Prosecution. <https://upload.wikimedia.org/wikipedia/commons/3/39/2013-01-28-DEI-EEC-to-Holder-re-Aaron-Schwartz-prosecution.pdf>
- Inger, S., & Gardner, T. (2013). Library technology in content discovery—evidence from a large-scale reader survey. *Insights*, 26(2).
- Free Software Foundation. (2011). 2010 Free Software Awards announced. <https://www.fsf.org/news/2010-free-software-awards-announced>
- Jeffries A., Leon, Y. (2020). Google's Top Search Result? Surprise! It's Google. <https://themarkup.org/google-the-giant/2020/07/28/google-search-results-prioritize-google-products-over-competitors>

- Jiménez Carreira, M., & Álvarez Marcos, J. (2018). Metodología de análisis y evaluación de sitios web 2.0. El caso de las Diócesis Andaluzas. *Revista de Comunicación*, 17(2).
- Kandari, J., Jones, E. C., Nah, F. F. H., & Bishu, R. R. (2011). Information quality on the world wide web: development of a framework. *International Journal of Information Quality*, 2(4), 324-343.
- Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys (CSUR)*, 32(2), 144-173.
- Krippendorff, K. (1990). *Metodología de análisis de contenido: teoría y práctica*. Grupo Planeta .
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Kuckartz, A. M., Kuckartz, U., & Andaluces, F. N. C. D. E. (2002). Qualitative text analysis with MAXQDA.
- Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. Sage.
- Labarca, F. (2016). Sci-Hub: La página universitaria "pirata" más buscada en la "Deep Web" disponible en: <https://www.latercera.com/noticia/sci-hub-la-pagina-universitaria-pirata-mas-buscada-en-la-deep-web/>
- Leal Carretero, F. (2017). ¿Qué función cumple la argumentación en la metodología de la investigación en ciencias sociales?. *Espiral (Guadalajara)*, 24(70), 9-49.
- Leibniz, G. W. (1703). Explication de l'arithmétique binaire, qui se sert des seuls caracteres 0 et 1, avec des remarques sur son utilité, et sur ce qu'elle donne des anciennes figures chinoises de fohy. <https://hal.archives-ouvertes.fr/ads-00104781/document>
- Lepper, G. (2000). *Categories in text and talk: A practical introduction to categorization analysis*. Sage.
- Leray, C. (2008). *L'analyse de contenu: De la théorie à la pratique, la méthode Morin-Chartier*. PUQ.
- Levy, S. (1984). *Hackers: Heroes of the computer revolution (Vol. 14)*. Anchor Press/Doubleday.
- Lewandowski, D., & Mayr, P. (2006). Exploring the academic invisible web. *Library Hi Tech*, 24(4), 529-539.

- Licklider, J., & Clark, W. E. (1962, May). On-line man-computer communication. In *Proceedings of the May 1-3, 1962, spring joint computer conference* (pp. 113-128).
- Licklider, J. (1963). Memorandum for: Members and affiliates of the intergalactic computer network; topics for discussion at the forthcoming meeting. *Washington, DC: Advanced Research Projects Agency. Retrieved April, 21.*
- Licklider, J., & Clapp, V. W. (1965). *Libraries of the Future* (Vol. 2). Mit Press.
- Licklider, J. C., & Taylor, R. W. (1968). The computer as a communication device. *Science and technology*, 76(2), 1-3.
- Lodder, J. M. (2009). Binary Arithmetic: From Leibniz to von Neumann. *Resources for Teaching Discrete Mathematics*, 169-178.
- Loughridge, C. (191X). Key to the Bi-literal Cipher of Francis Bacon. https://ia800302.us.archive.org/28/items/keytobiliteralci00loughrich/keytobiliteralci00loughrich_bw.pdf
- Madhavan, J. et al. (2009). Harnessing the deep web: Present and future. *CIDR Perspectives*
- Madigan, K (2017). Judgment Against Sci-Hub is a Win for Authors and Publishers. <https://cip2.gmu.edu/2017/06/27/judgment-against-sci-hub-is-a-win-for-authors-and-publishers/>
- Martín, A (2017) La Unión Europea impone a Google la mayor multa por abuso de posición dominante. <https://hipertextual.com/2017/06/multa-google-abuso-de-posicion-dominante>
- Marchiori, M. (1997). The quest for correct information on the Web: Hyper search engines. *Computer Networks and ISDN Systems*, 29(8-13), 1225-1235.
- Martín-Rodríguez, F., García, G. B., & Lires, M. Á. (2010, November). Technological archaeology: Technical description of the Gauss-Weber telegraph. In 2010 Second Region 8 IEEE Conference on the History of Communications (pp. 1-4). IEEE.
- Menabrea, L. F., & Lovelace, A. (1842). Sketch of the analytical engine invented by Charles Babbage.
- Meredith, S. (2018) Facebook-Cambridge Analytica: A timeline of the data hijacking scandal. <https://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html>
- Metcalfe, R. M., & Boggs, D. R. (1976). Ethernet: Distributed packet switching for local computer networks. *Communications of the ACM*, 19(7), 395-404.

- McCarthy, N (2016). Where Popular Apps Are Under Assault.
<https://www.statista.com/chart/6839/where-popular-apps-are-under-assault/>
- Morar, F. S. (2015). Reinventing machines: the transmission history of the Leibniz calculator. *The British Journal for the History of Science*, 123-146.
- Nature. (2016). Nature's 10.<https://www.nature.com/articles/540507a>
- Nelson, T. H. (1967). Getting it out of our system. *Information retrieval: A critical review*, 191-210.
- Nelson, T. (1987). Literary Machines 87.1, published by the author.
- Nelson, T.H. Transcopyright for the web: A clean legal and technical system for reusing on-line content. <http://transcopyright.org/>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin
- Pavithra, M., Vasanth, S., Rajmohan, R., & Jayakumar, D. (2018). ZERONET: AN OVERVIEW. *International Journal of Pure and Applied Mathematics*, 119(14), 1347-1351.
- Paywall: The Business of Scholarship (2018). <https://paywallthemovie.com/>
- Peirce, C. S. (1910). *The Rationale of Reasoning*. MS [R] 663.
- Petzold, C. (2000). *Code: The hidden language of computer hardware and software*. New York: Microsoft Press.
- Breton, T(2021). Pavel Durov, The live wire, Disrupter No.1 Russia
<https://www.politico.eu/list/politico-28-class-of-2021-ranking/pavel-durov/>
- Reporters sans Frontieres (2016). Reporteros sin Fronteras y Torservers.net, socios contra la vigilancia y la censura en línea.
<https://rsf.org/fr/actualites/reporters-sans-frontieres-et-torserversnet-partenaires-contre-la-surveillance-et-la-censure-en-ligne>
- Reporters Sans Frontiers (2018), Periodistas: Siete reflejos de seguridad digital para adoptar.<https://rsf.org/fr/actualites/journalistes-sept-reflexes-de-securite-numerique-adopter>
- Reporters Sans Frontiers (2021). Uruguay. <https://rsf.org/es/pais/uruguay>

- ReturnonNow (2015). 2015 Search Engine Market Share By Country.
<https://returnonnow.com/internet-marketing-resources/2015-search-engine-market-share-by-country/>
- Rossi, A. (2018). ¿Burbujas de filtro? Hacia una fenomenología algorítmica. *Mediaciones de la Comunicación*, 13(1), 263-281.
- Rushkoff, D. (2019). *Team Human*. Norton & Company.
- Rushkoff, D. (2019). We shouldn't blame Silicon Valley for technology's problems we should blame capitalism. <https://qz.com/1529476/we-shouldnt-blame-silicon-valley-for-technology-problems-we-should-blame-capitalism/>
- Sáez, H. (2018). El sesgo de confirmación: por qué todo lo que vemos parece confirmar lo que pensamos. https://medium.com/@hugo_saez/el-sesgo-de-confirmación-por-qué-todo-lo-que-vemos-parece-confirmar-lo-que-pensamos-i-4b2591cdf22e.
- Sar, A. (2016). Las pulsaciones de una expedición telegráfica A 150 años del tendido del primer cable submarino en el Río de la Plata. *Claves. Revista de Historia*, 2(3), 73-98.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.
- Scheuerman, W. E. (2014). Whistleblowing as civil disobedience: The case of Edward Snowden. *Philosophy & Social Criticism*, 40(7), 609-628.
- Seymour, T., Frantsvog, D., & Kumar, S. (2011). History of search engines. *International Journal of Management & Information Systems (IJMIS)*, 15(4), 47-58.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10-21.
- Shannon, C. (1953). Communication Theory-Exposition of fundamentals. *Transactions of the IRE Professional Group on Information Theory*, 1(1), 44-47.
- Singh, D. (2017). Publishers take ResearchGate to court, alleging massive copyright infringement. <https://www.science.org/news/2017/10/publishers-take-researchgate-court-alleging-massive-copyright-infringement>
- Swartz, A (2008). *Guerilla Open Access Manifesto*.
https://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.

[txt](#)

- Tate, M. A. (2009). Web wisdom: How to evaluate and create information quality on the Web. CRC Press.
- Taylor, N. K. (1992). Charles Babbage's Mini-Computer-Difference Engine No. 0. *Bulletin of the Institute of Mathematics and its Applications*, 28(6), 112-114.
- Techopedia (2018). Definition - What does Web mean? <https://www.techopedia.com/definition/5613/web>
- Terzin, R. (2021). Telegram Statistics. <https://cybercrew.uk/blog/telegram-statistics/>
- The Tor Project. (2018). Relay users. <https://metrics.torproject.org/userstats-relay-country.html?start=2012-01-03&end=2018-12-01&country=all&events=off>
- The Tor Project. (2019). Historia. <https://www.torproject.org/es/about/history/>
- The Tor Project. (2021). Top 10 countries by possible censorship events. <https://metrics.torproject.org/userstats-censorship-events.html?start=2000-07-03&end=2021-10-01>
- Thoreau, H. D. (1866). *Resistance To Civil Government [Civil Disobedience]* (pp. 145-171). Yale University Press.
- Travis, J. (2016). In survey, most give thumbs-up to pirated papers. <https://www.sciencemag.org/news/2016/05/survey-most-give-thumbs-pirated-paper>.
- TrendTic (2017). Entrevista a Isabela Bagueros project manager de Tor. <http://www.trendtic.cl/2017/01/entrevista-a-isabela-bagueros-project-manager-de-tor/>
- Valéry, P. (1895). *Introduction à la méthode de Léonard de Vinci*. Éditions de la Nouvelle revue française.
- Varde, A., Suchanek, F., Nayak, R., & Senellart, P. (2009, April). Knowledge discovery over the deep web, semantic web and XML. En: *International Conference on Database Systems for Advanced Applications* (pp. 784-788). Springer, Berlin, Heidelberg.
- Villada, V., & Jiménez, V. (2017). La Web Semántica y la Web Profunda como Sistemas de Información: Análisis a una realidad. *Revista Antioqueña de las Ciencias Computacionales*, 7(1).
- VPNmentor (2018). VPN Use and Data Privacy Stats for 2018. <https://www.vpnmentor.com/blog/vpn-use-data-privacy-stats/>

Wang, S., Gao, Y., Shi, J., Wang, X., Zhao, C., & Yin, Z. (2020, June). Look Deep into the New Deep Network: A Measurement Study on the ZeroNet. en: International Conference on Computational Science (pp. 595-608). Springer

W3 (2018). What is the difference between the Web and the Internet?
<https://www.w3.org/Help/#webinternet>

Worldwidewebsize (2018). <https://www.worldwidewebsize.com/>