SJSU | SCHOOL OF INFORMATION

School of Information Student Research Journal

Volume 10 | Issue 1 Article 8

May 2020

Library application of Deep Web and Dark Web technologies

Mollie L. Coffey
San Jose State University, mollie.coffey@yahoo.com

Acknowledgements

I would like to thank Dr. Anthony Bernier, San Jose State University, for his encouragement and support in submitting this article. I would also like to acknowledge Eastern Washington University JFK Library for providing access to research materials.

Library application of Deep Web and Dark Web technologies

Abstract

The Deep Web and Dark Web are legitimate tools for use in the field of information science, adding to the discussion of patron privacy. The American Library Association policies on privacy and confidentiality combined with the advancement of internet technology necessitate that library professionals become fluent in Dark Web usability in libraries.

Keywords

Dark Web, Deep Web, privacy, Tor browser, internet, libraries

Acknowledgements

I would like to thank Dr. Anthony Bernier, San Jose State University, for his encouragement and support in submitting this article. I would also like to acknowledge Eastern Washington University JFK Library for providing access to research materials.

About Author

Mollie Coffey has a Bachelor of Arts in psychology from Eastern Washington University and a Master of Library and Information Science from San Jose State University. She currently works as the Interlibrary Loan Specialist at Spokane Public Library. Mollie completed the original research for this article as a part of an Online Searching course at San Jose State University's School of Library and Information Science.

Introduction

Intellectual freedom, privacy, and unrestricted access to information remain the core values of information professionals. Libraries function as community education centers serving diverse populations with differing information needs. Information seeking using digital platforms requires specialized skills and tools to extract applicable results. Librarians remain a vital part of this process (Prasad, 2018). As online activity grows, the proliferation of surveillance and censorship becomes an undeniable challenge to intellectual freedom. Third-party data collection and online tracking creates difficulties for those entrusted with protecting patron privacy (Pekala, 2017). Increased awareness of these practices leads some libraries to examine the layers of the internet through a lens of privacy and security. The Dark Web promises anonymity and privacy through platforms such as the Tor browser, although there are potential risks involved. While information found on the Dark Web should not be considered reliable for academic research, justification exists for valid uses of the Dark Web in libraries to ensure patron privacy.

Internet Layers

There is more to the internet than meets the eye, with its three distinct layers of depth. The Surface Web, occupying 10% of the internet, contains those websites with visible contents resulting from search engine indexing (Beckstrom & Lund, 2019). These searchable, publicly available pages can be accessed from a standard web browser and connect to other pages using hyperlinks. However, information is being overlooked that was never intended to be hidden (Devine, Egger-Sider, & Rojas, 2015). This information, invisible to regular search engines, requires persistence and specialized search tools to locate. Beyond the Surface Web exist the Deep Web and the Dark Web.

Several distinct characteristics identify the differing layers of the internet (Table 1). Inaccurate understanding of the difference between the Surface Web and the Deep Web can be remedied. The Surface Web contains several billion websites and documents with diverse subsets which can be indexed by most search engines. The next layer, the Deep Web, includes millions of databases and dynamic web pages that often reside behind paywalls or require passwords. This layer contains higher quality information than is usually found on the Surface Web (Prasad, 2018).

Table 1: Characteristics of internet layers

Surface Web	Deep Web	Dark Web
• Freely accessible • Indexed by standard search engines • Mostly HTML files • Fixed content	• Not indexed • Proprietary databases • Dynamically generated content • Login authorization	• Specialized software/tools • Intentionally hidden data • Encrypted and anonymous • Difficult to track

The Deep Web

The Deep Web is often referred to as the Invisible Web or the Hidden Web because the web pages found here are beyond the reach of standard search engines. Deep Web sites have more focused and deeper content materials than Surface Web sites. Accessing material on the Deep Web often calls for skill, effort, and perseverance on the part of the searcher. Resources on the Deep Web frequently go undiscovered because of the user's inexperience in searching skills, otherwise known as "cognitive invisibility" (Devine & Egger-Sider, 2014). Exploring the Deep Web is significant for libraries, to expose the hidden information so that users can access resources largely ignored by popular search engines (Blandford, 2015). The need exists for information professionals to serve as digital guides in navigating information resources such as special collections and proprietary databases.

Despite its reputation, the unindexed material on the Deep Web can usually be found in ordinary databases. PubMed, LexisNexis, and Web of Science are all housed on the Deep Web. Users of these databases interact with the Deep Web regularly but may not be aware of it. 90% of traffic on the internet comes from the Deep Web (Chertoff, 2017). Current academic research relies heavily on the Deep Web, which provides an active element in improved higher education outcomes (Alyami & Assiri, 2018). Deep Web sites often contain dynamically generated pages, data intensive pages, and time-sensitive or short-lived pages.

While it is technically impossible to accurately measure the size of the Deep Web, some estimates put it at 500 times the size of the Surface Web (Weimann, 2016), while other researchers consider the Deep Web to be 5000 times larger (Chertoff, 2017). Regular search engines index less than 16% of the Surface Web and 0.03% of all the information that exists online (Weimann, 2016). The Deep Web can be accessed through databases and directories, using specialized search engines that provide more precise search results on more specific topics. These advanced search tools allow the user to modify and customize searches for improved results (Prasad, 2017). As most libraries offer access to hundreds of different databases to their users it befits the information professional to understand the Deep Web and how to utilize its features to locate accurate information. Library and information professionals are trained to find relevant resources more quickly and efficiently on the Deep Web than casual information seekers (Crockett, 2018).

The Dark Web

A very small, hard to access allocation (0.01%) of the Deep Web is called the Dark Web (Chertoff, 2017). The Dark Web operates on a deeper layer of the Internet that thrives on anonymity. These sites are not freely available to the public, with Internet Protocol addresses being hidden to ensure confidentiality and anonymity. Internet Protocol (IP) addresses are numeric strings that identify devices on networks to enable them to be recognized by other systems. Dark Web pages tend to be unreliable, coming and going regularly, leaving directories peppered with dead links as websites disappear or change locations. The relative hardship in merely

finding hidden websites, as well as the strict anonymity, bolsters the Dark Web's air of mystery.

The phrase Dark Web itself gives an impression of illegality, although legitimate purposes for its use do exist. Residents of countries with high censorship, transgender individuals in repressive regions, and undocumented immigrants possess the right to information access and need privacy protection to exercise that right. The Dark Web provides all a method of protecting their personal information and privacy amid the incursion of data collection practices (Monk, Mitchell, Frank, & Davies, 2018). While the Dark Web is a place that has a somewhat dubious reputation, and indeed has gained some notoriety for actionable activities, it has uses beyond the nefarious. The evolution of the Dark Web came from a need for a secure avenue for military communications and expanded to be known as a guarantor of online privacy (U & Thampi, 2019). The same anonymity that allows blatantly illegal sites to remain viable also to remain viable also enables people in countries with oppressive regimes to communicate with the outside world, conducting political dissent without fear of retribution. Its significance lies not in its efficacy, but in its existence. Some social media sites, such as Facebook, have set up Dark Web versions of their websites in order to protect dissidents and others needing anonymity to connect with people. Jardine (2016) propounds the "dark web dilemma": shutting down the network will not actually stop illegal activity but will hurt people who rely on it to exercise their free expression and access to information. While some may experiment with Dark Web access for curiosity, the anonymous nature and freedom of expression become the sustaining factor in its continued use (Mirea, Wang, & Jung, 2019).

At the core of the Dark Web is an intricate array of routing that provides anonymity for users accessing it. Unlike the Surface Web and most parts of the Deep Web, the Dark Web involves special technology to access the websites hosted there. To gain admission to these anonymous sites a Tor browser must be employed. Tor is an acronym for The Onion Router. Dark Web sites run on a special server which delivers content to Tor browsers. Tor, a modified version of the Firefox browser, remains the most prominent tool of Dark Web users (Monk, et.al., 2018). It uses a set of encryption tools, services, and nodes that hide and change IP addresses and encrypt data to and from computers, protecting both the visitor and the website operator. Tor routes connections through a series of relay machines running open source software that is encrypted as it is routed. These relays form the infrastructure of the Tor network, passing information through layers of anonymizing encryption (Macrina, 2015). The United States Naval Research Laboratory developed the Tor browser as a means of protecting government communications using a secure method of routing (Weimann, 2016), and the system soon developed into a non-military project (Macrina, 2015). In 2004 the software became public and was offered as a free service to advance unconstrained access to the Internet for those who face persecution for online communication. Like many innovations, it became subverted and began to connect illegal goods and services to willing customers.

The transitory nature of most of the Dark Web content does not qualify it as a reliable information source for scholarly research. While the Dark Web is not

a sound authority for academic needs, Tor accommodates many beneficial functions. The Tor browser's ability to anonymize its users is unique in the way it protects the identities of people at risk (Bayle, Compoe, Ehrick, Hubbell, Lowe, & Ridge, 2017).Tor works with victims of domestic violence, showing them how to access the Internet safely without revealing their location and activities to their abusers (Devine, et al, 2015) and has been involved in protests in the Middle East and North Africa (Goldsborough, 2016). For these and other reasons, Tor won the Free Software Foundation's award for Projects of Social Benefit in 2010. Beginning as a military project, Tor continues to receive a large portion of its funding from the Department of Defense and the State Department (Jardine, 2018).

Discussion

Information professionals serve as exponents of information literacy and are in the forefront of a paradigm shift into information fluency. In the digital age, the ability to intuitively interpret all forms and formats of information is intrinsically linked to the knowledge of how to apply this fluency to real-world tasks and problems (Crockett, 2018). Librarians traditionally guide and provide assistance in learning and discovery to disparate, and at times marginalized, users but must adopt a new role to remain relevant amidst digital revolution. Patron privacy and user experience compete for importance to library professionals (Pekala, 2017). While intellectual freedom is codified in the American Library Association Code of Ethics and the Library Bill of Rights, information accessed online leaves users vulnerable to surveillance and exposure (Childs, 2017). Libraries unintentionally undermine user privacy with unsecured connections to their websites. Long the champions of user privacy, libraries still employ website practices at odds with their stated privacy values (O'Brien, Young, Arlitsch, & Benedict, 2018).

The ubiquity of the internet necessitates a revision of basic assumptions concerning anonymity and the web (Sarda, Natale, Sotirakopoulos, & Monaghan, 2019). Privacy has become an increasingly complex issue for library professionals. All online activity can be tracked, traced, compiled, crunched, bought and sold without authorization unless privacy tools are utilized (Coleman, 2019).

Librarians have always educated patrons in information literacy and must now also educate them on privacy literacy by focusing on understanding the risks and obligations inherent in sharing information online (Wissinger, 2017). Bridging the knowledge gaps that prevent comprehension of the usage of privacy-protection technologies are the next steps taken toward more secure data (Maceli, 2018). To further this end, the Library Freedom Project (LFP) was born (https://libraryfreedomproject.org/). LFP instructs library professionals on surveillance threats, confidentiality rights, and privacy tools, shifting the focus to a more privacy-centric standard. LFP offers resources for teaching library users how to safeguard their information and protect their online presence (Hennig, 2018).

Dark web browsers such as Tor are easily installed on public computers, but may run afoul of local, state, and federal regulations (Beckstrom & Lund, 2019). Bayle et.al. (2017) believe Tor to be an ideal browser for public library computers to ensure patron privacy. A risk/benefit analysis shows an even split: criminal

activity and potential malware (risks) versus protection of personal data and free access to information (benefits). Library administrators are obliged to consider providing access to Tor and other anonymous platforms as a way of broadening organizational missions. These platforms exist as a powerful tool for libraries to offer privacy protection to their users, provided that the potential risks are understood, and precautions are taken to mitigate them (Bayle, et.al., 2017). O'Brien, et.al, (2018) offer five interrelated areas for taking action:

- Implementation of secure web protocols
- User education
- Privacy policies
- Informed consent
- Risk/benefit analysis

Conclusion

Librarians play crucial roles in the global information infrastructure. The Deep and Dark Webs as tools for patron privacy and confidentiality have become a valuable resource in the digital age (Maceli, 2018). Libraries' ability to safeguard intellectual freedom has evolved through new and constantly changing information technologies. Privacy threats to patron information remain a concern in the library field. This right to privacy as advocated by the American Library Association highlights the necessity for library professionals to understand and demonstrate Dark Web technologies. Librarians can educate patrons about privacy threats and the protective measures available to mitigate them. Effectively communicating the privacy risks associated with online information discovery is key.

Exploring the Deep Web, and by association the Dark Web, requires special tools and techniques, with different means of reaching various layers of depth. Anonymity networks are also used constructively. The anonymity and security protection afforded to criminals can be used to shield those living in countries with tyrannical governments where there is a legitimate risk of prosecution, imprisonment, or death for leaking information or having dissenting opinions. The potential exists for further research into how libraries are using Dark Web technologies to protect patron privacy.

Dark web platforms share the core library values of privacy, intellectual freedom, and information access. Using and teaching the use of the Tor browser upholds these values. Consequently, privacy literacy instruction in Library and Information Science programs will be enhanced by the addition of privacy-protection technologies. Producing digital-savvy librarians creates the framework for shaping digital-savvy patrons.

References

Alyami, H.Y., & Assiri, E.A. (2018). Invisible web and academic research: A partnership for quality. *International Education Studies, 11*, 84-91. doi:10/5539/ies.v11.n4.p84

Bayle, E., Compoe, S., Ehrick, R. Hubbell, D., Lowe, B., & Ridge, J. (2017).

Patron privacy: Is the Tor browser right for library use? *Computers in Libraries, 37,* 10-13. https://ezproxy.library.ewu.edu/login?url=https://search-proquest-com.ezproxy.library.ewu.edu/docview/1930069438?accountid=7305

Beckstrom, M., & Lund, B. (2019). *Casting light on the Dark Web: A guide for safe exploration.* Lanham, MD: Rowman & Littlefield.

Blandford, A. (2015). Google, public libraries, and the Deep Web. *Dalhousie Journal of Interdisciplinary Management, 11,* 1-18. doi:10.5931/djim.v11I0.5525

Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy, 2,* 26-38. doi:10.1080/23738871.2017.1298643

Childs, L. (2017). To uphold and resist: Protecting intellectual freedom through progressive librarianship. *The Serials Librarian, 73,* 58-67. doi: 10.1080/0361526X.2016.1270248

Coleman, G. (2019). How has the fight for anonymity and privacy advanced since Snowden's whistle-blowing? *Media, Culture & Society, 41,* 565-571. doi:10.1177/0163443719843867

Crockett, L.W. (2018). Librarians lead the growth of information literacy and global digital citizens. *Knowledge Quest, 46,* 28-33. https://ezproxy.library.ewu.edu/login?url=https://search-proquest-com.ezproxy.library.ewu.edu/docview/2034285237?accountid=7305

Devine, J., & Egger-Sider, F. (2014). *Going beyond Google again: Strategies for using and teaching the Invisible Web*. Chicago, IL: Neal-Schuman.

Devine, J., Egger-Sider, F., & Rojas, A. (2015). The evolving impact of the Invisible Web: Exploring economic and political ramifications. *Journal of Web Librarianship, 9,* 145-161. doi:10.1080/19322909.2015.1077183

Goldsborough, R. (2016). Stay clear of the Darknet. *Tech Directions, 75,* 12.

Hennig, N. (2018). Applying best practices. *Library Technology Reports, 54,* 29-33. doi:https://doi.org/10.5860/ltr.54n3

Jardine, E. (2016). The Dark Web dilemma. *Australasian Science, 37,* 12-14. https://ezproxy.library.ewu.edu/login?url=https://search-proquest-com.ezproxy.library.ewu.edu/docview/1798372691?accountid=7305

Jardine, E. (2018). Tor, what is it good for? Political repression and the use of online anonymity- granting technologies. *New Media & Society,20,* 435-452. https://doi org.ezproxy.library.ewu.edu/10.1177/1461444816639976

Maceli, M.G. (2018). Encouraging patron adoption of privacy-protection technologies: Challenges for public libraries. *IFLA Journal, 44,* 195-202. https://doi.org/10.1177/0340035218773786.

Macrina, A. (2015). The Tor browser and intellectual freedom in the digital age. *Reference & User Services Quarterly, 54,* 17-20. https://ezproxy.library.ewu.edu/login?url=https://search-proquest-com.ezproxy.library.ewu.edu/docview/1691315816?accountid=7305

Mirea, M., Wang, V., & Jung, J. (2019). The not so dark side of the darknet: A qualitative study. *Security Journal, 32,* 102-118. doi:10.1057/s41284-018-0150-5

Monk, B., Mitchell, J., Frank, R., & Davies, G. (2018). Uncovering Tor: An Examination of the network structure. *Security and Communications Networks, 2018,* 1-12. https://doi.org/10.1155/2018/4231326

O'Brien, P., Young, S.W.H., Arlitsch, K., & Benedict, K. (2018). Protecting privacy on the web. *Information Review, 42,* 734-751. doi:10.1108/OIR-02-2018-0056

Pekala, S. (2017). Privacy and user experience in 21st century library discovery. *Information Technology and Libraries, 36,* 48-58. https://doi.org/10.6017/ital.v36i2.9817

Prasad, M.R.M. (2017). Deep Web: Librarian's perspective. *Pearl: A Journal of Library and Information Science, 11,* 418-423. doi:10.5958/0975-6922.2017.00054.7

Sarda, T., Natale, S. Sotirakopoulos, N., & Monaghan, M. (2019). Understanding online anonymity. *Media, Culture, & Society, 41,* 557-564. doi:10.1177/0163443719842074

U., A., & Thampi, S. M. (2019). Dark Web and its research scopes. In A. Sari (Ed.), *Applying Methods of Scientific Inquiry into Intelligence, Security, and Counterterrorism* (240-268). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-8976-1.ch010

Weimann, G. (2016). Going dark: Terrorism on the Dark Web. *Studies in Conflict & Terrorism, 39,* 195-206 https://doi.org/10.1080/1057610X.2015.1119546.

Wissinger, C. (2017). Privacy literacy: From theory to practice. *Communications in Information Literacy, 11,* 378-389. doi:10.15760/comminfolit.2017.11.2.9