A Review on Levels of Surfing Web

Yogesh R. Shelokar[1] , Prof P. Mundhare[2]
Department of Computer Science and Information Technology
[1] Student, M.E. First Year, Sanmati Engineering College, Washim
[2] Sanmati Engineering College, Washim

Abstract- *Surfing is an activity that started with the appearance of the planet wide net. With machine-readable text links, users will go not solely from one a part of a document to a different, however additionally from one document to a different, as well as those settled in remote sites. Surfing could be a favorite recreation for various individuals round the world UN agency has access to the net. Many users are hooked thereon, spending countless hours doing casual searches or other online activities. Searching on web nowadays will be compared to dragging a net across the surface of the ocean. While a good deal could also be caught within the internet, there is still a wealth of information that is deep, and therefore, missed. The reason is simple: Most of the Web's info on dynamically generated sites, and standard search engines never find it. Traditional search engines produce their indices by spidering on locomotion surface sites. To be discovered, the page must be static and linked to other pages. Traditional search engines cannot "see" or retrieve content within the deep net - those pages don't exist till they're created dynamically because the results of a particular search. Because traditional search engine crawlers cannot probe beneath the surface, the deep web has heretofore been hidden. Deep web is the name given to the technology of surfacing the hidden price that can't be simply detected by different search engines. The deep net is that the content that can't be indexed and searched by search engines. For this reason the deep net is additionally known as invisible net.*

Keywords- Surfing, Deep Web, Crawlers.

I. INTRODUCTION

The World Wide Web (WWW), additionally known as the online data house wherever documents and alternative net resources area unit known by Uniform Resource Locators (URLs), interlinked by machine-readable text links, and accessible via the Internet. English human Tim Berners-Lee unreal the globe Wide net in 1989.

Surfing the Web-to navigate through the globe wide net or web, sometimes by clicking with a mouse. The term additionally features a generic which means of paying time on the web.

Level 1 - Surface Web or Common Web: This is the level that is browsed everyday by netters; like YouTube, Facebook, Wikipedia and other famous or easily accessible websites can be found here such as google.

Level 2 - Bergie Web: This level is the last one normally accessible: all levels that follow this one have to be accessed with a proxy, or modifying your hardware. In this level you can find some "underground" but still indexed websites, such as 4chan. Freehive, Black Hat World or FTP servers and also the blocked Google search results

Level 3 - Deep Web: The Deep Web also called the Deep-net, the invisible web, the Undernet or the hidden web is World Wide Web content that is not part of the Surface Web, which is indexable by standard search engines. It should not be confused with the dark Internet, the computers that can no longer be reached via Internet, or with the distributed filesharing network Darknet, which could be classified as a smaller part of the Deep Web.

Level 4 - Charter Web: That is granted or authorized Government pages and banks.

Level 5 - Marianas Web: Practically nobody has reached here. For many, the "inaccessible level" is impossible to reach, however it contains the most valuable information. Very few have reached here, only those that have achieved computing and hacking mastery, among themselves Julian Assange. Only reached level 4 on browser and using a vpn.

II. LEVELS OF THE WEB

Many may consider the Internet and World Wide Web (web) to be synonymous; they are not. Rather, the web is one portion of the Internet, and a medium through which information may be accessed. In conceptualizing the web, some may view it as consisting solely of the websites accessible through a traditional search engine such as Google. However, this content—known as the “Surface Web”—is only one portion of the web. The Deep Web refers to “a class of content on the Internet that, for various technical reasons, is not indexed by search engines,” and thus would not be accessible through a traditional search engine.

Information on the Deep Web includes content on private intranets (internal networks such as those at corporations, government agencies, or universities), commercial databases like Lexis Nexis or Westlaw, or sites that produce content via search queries or forms. Going even further into the web, the Dark Web is the segment of the Deep Web that has been intentionally hidden. The Dark Web is a general term that describes hidden Internet sites that users cannot access without using special software. While the content of these sites may be accessed, the publishers of these sites are concealed. Users access the Dark Web with the expectation of being able to share information and/or files with little risk of detection. In 2005, the number of Internet users reached 1 billion worldwide. This number surpassed 2 billion in 2010 and crested over 3 billion in 2014. As of July 2016, more than 46% of the world population was connected to the Internet.9 While data exist on the number of Internet users, data on the number of users accessing the various layers of the web and on the breadth of these layers are less clear [4]. The size of the World Wide Web (The Internet) The Indexed Web contains at least 4.43 billion pages. The Dutch Indexed Web contains at least 153.8 million pages [1].

Level one: Common Web or Surface Web

Also know as crawlable Web" and "public Web If we imagine web as an ocean, the surface web is the top of the ocean which appears to spread for miles around, and which can be seen easily or "accessible" Example Reddit , dig, Temp Email Services, Newgrounds, Vampire Freaks, Foreign Social Networks ,human intel tasks , web hosting ,MYSQL Databases,college Campuses The Surface Web (also called the Visible Web, Indexed Web, Indexable Web or Lightnet) is the portion of the World Wide Web that is readily available to the general public and searchable with standard web search engines. It is the opposite of the deep web. The Indexed or Surface Web contains at least 4.45 billion pages (Thursday, 23 August, 2018).[1].

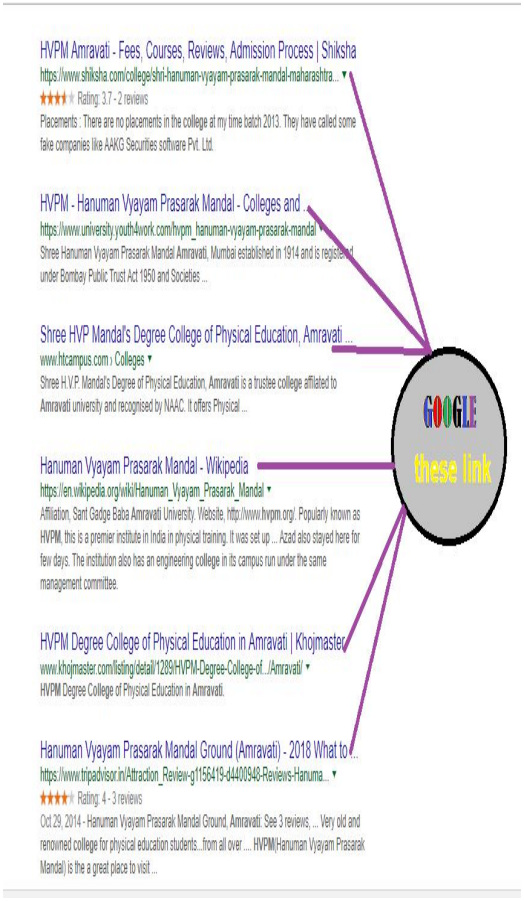

FIG 1. EXAMPLE OF SURFACE WEB

Level two: Bergie Web

This level is the last one normally accessible: all levels that follow this one have to be accessed with a proxy, Tor or by modifying your hardware. In this level you can find some "underground" but still indexed websites, such as 4chan. Freehive, Black Hat World or FTP servers and also the blocked Google search results.

Level 2 Web – Bergie Web	
-FTP Servers	-4chain
-Google Locked results	-RSC
-Honeypots	-Freehive
-Loaded Web Servers	-Let Me Watch This

TABLE 1. BERGIE WEB

Level Three: Deep Web or Invisible Web or Hidden Web

The third level onwards Deep Web starts, no search-engines are able to index these sites and they need some sort of proxy network like Tor, I2P, freenet or JonDo to become accessible. Although this is the Deep Web, most content on

level 3 is publicly accessible with proxy, without any sort of restrictions. Which is not directly accessible, we need some special software or proxy to connect or access deep web E.g: Government/Business Research, Hackers, ScriptKiddies, Virus Information, Illegal and Obscene Content (CP, Gore, Suicides, etc…)

WHAT IS DEEP WEB?

The Deep Web is the content that resides in searchable databases, the results from which can only be discovered by a direct query. Without the directed query, the database does not publish the result. When queried, Deep Web sites post their results as dynamic Web pages in real-time. Though these dynamic pages have a unique URL address that allows them to be retrieved again later, they are not persistent. The invisible web consists of files, images and web sites that, for a variety of reasons, cannot be indexed by popular search engines. The deep web is qualitatively different from the surface web.

Deep web sources store their content in searchable databases that only produce results dynamically in response to a direct request. But a direct query is a "one at a time" laborious way to search. Deep web's search technology automates the process of making dozens of direct queries simultaneously using multiple-thread technology. The Deep Web is made up of hundreds of thousands of publicly accessible databases and is approximately 500 times bigger than the surface Web.[3]

Fig. 2 displays the distribution of deep Web sites by type of content [6]

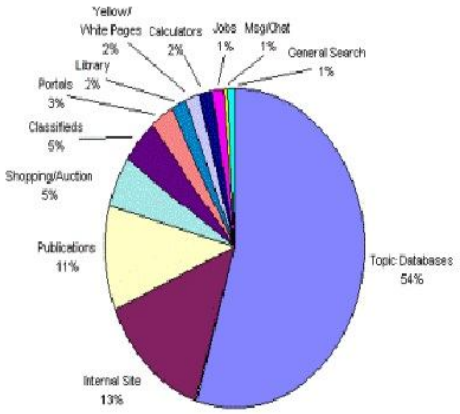

FIG. 2 DISTRIBUTION OF DEEP WEB SITES BY CONTENT

IMPORTANCE OF DEEP WEB

- ✓ Public information on the deep Web is currently 400 to 550 times larger than the commonly defined World Wide Web.
- ✓ The deep Web contains 7,500 terabytes of information compared to nineteen terabytes of information in the surface Web.
- ✓ The deep Web contains nearly 550 billion individual documents compared to the one billion of the surface Web.
- ✓ More than 200,000 deep Web sites presently exist.
- ✓ Sixty of the largest deep-Web sites collectively contain about 750 terabytes of information -- sufficient by themselves to exceed the size of the surface Web forty times.
- ✓ 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.
- ✓ Total quality content of the deep Web is 1,000 to 2,000 times greater than that of the surface Web [3]

There are two fundamentally deferent approaches to incorporating the deep web into search or topic exploration engines [7].

Deep Web Crawl. Crawl as much of the deep web as possible and incorporate it into a conventional search engine index.

Federated Search. Use APIs to access deep web sources at query-time and construct results pages based on their responses.

Wright [8], using a fishing analogy, calls these approaches trawling and angling respectively, while Madhavan et al. [9] calls them surfacing and virtual integration. These approaches are also analogous to the warehousing and mediation approaches in data integration.

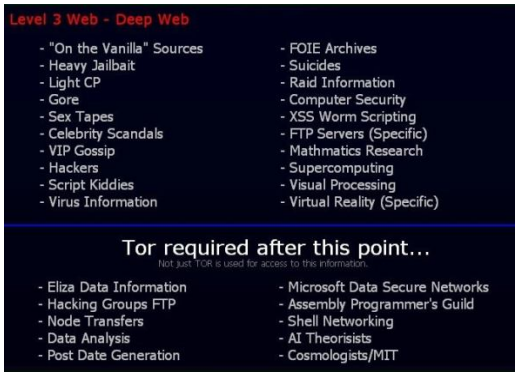

FIG 3.1 DEEP WEB AND DARK WEB

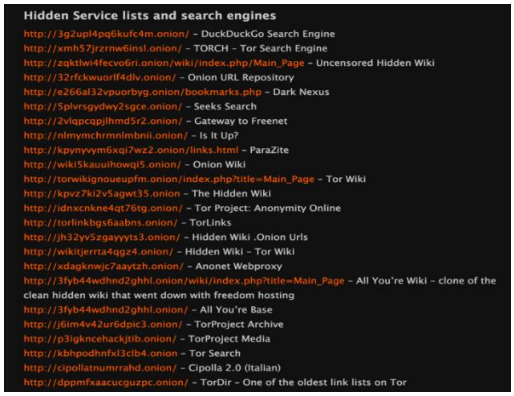

FIG 3.2 DARK WEB

Level Four: Charter Web

This level is also divided into two parts, the first can be accessed through the Tor network, it contains such things as drugs and human trafficking, banned films and books, black markets. Includes the Hidden Wiki (usually the first website you will access when trying to get into the deep web), which is like the deep web website that contains the link for many other charter web websites. Websites become more restrictive and begin using stronger security measures like registration & login, invite only memberships, open only for a specific time and/or dates, restricted to certain IP addresses, or a combination of the above, etc. Unlike most websites, they are not interested in maximizing traffic hits and keep a very low profile even in the Deep Web The second part is accessed by a hardware modification: a "CSS" "Shell closed system" and contains over 80% of the web, not in volume but in concentration of information, this part of the web charter contains unconditional PC, information on the experimental material and also dark information, such as the "Law 13", the experiences of World War II, and even the location of Atlantis . These comprise of a single computer or a network of systems that are not connected to any external network at all. They can only be accessed from within the network. It is not possible to connect to these systems unless the attacker can physically access these systems. Many companies have sensitive internal networks that are behind a firewall (green zone), that is a different scenario and is still considered insecure in this context. The CSS networks have no physical (wired or wireless) connections to any other network. What these types of networks contain is left to the reader's imagination. [2].

FIG 4. CHARTER WEB

Level Five: Marianas' Web

Having stolen it's name from the deepest part of the ocean the Mariana Trench (Mariana Trench - Wikipedia), Mariana's Web is said to be a World Wide Network that connects the government computers to one another making it easy to secretly transfer data over the internet. The Data is said to include top secret government information about things the public is not yet aware of (such as extraterrestrial life etc). It is assumed to be using secure internet protocols that encrypt all the information to stay anonymous and untraceable. While Mariana's Web is considered world wide by most conspiracy theories it is mostly associated with the American government (since they were the ones to launch the TOR network anyways) The closest verified version we have of the Mariana's web are private networks used by the government, such as SIPRnet, NSAnet, OPENnet/OPENnet+, RIPR, NIPRnet, JWICS, GWAN (Global wide areal network etc) [5]

Examples:-

- Videos snuff Choose the victim online :- ydbcnqhopqmeltucyeik3.clos/
- Classified Information :- duanjfheyu5clq9ia7yt6.clos/
- trafficking and organs :- g5h6j8t66t4jhbn9iksh6.clos/
- erlast Project :- jquydhr7nqb59j16jh6d0.clos/

Note: - also the domain name like .bit, .lib, .emc, .coin, .bazar but not this entire site are active.

III. CONCLUSION

Searching on web nowadays is often compared to dragging a net across the surface of the ocean. While an excellent deal is also caught within the web, there's still a wealth of data that's deep, and thus, missed. The reason is simple: Most of the Web's info on dynamically generated sites, and normal search engines never realize it. Specific vertical market services are already evolving to partially address this challenge. These probably ought to be supplemented with a persistent question system customizable by user that might set the queries, search sites, filters and schedules for recurrent queries.

REFERENCES

[1] http://www.worldwidewebsize.com/
[2] https://infoseceye.wordpress.com/tag/bergie-web/
[3] http://www.completeplanet.com.
[4] Congressional Research Service 7-5700 www.crs.gov R44101
[5] https://www.quora.com/What-is-the-Marianas-Web
[6] Sumit Bindal, Himank Singh Muktawat; "Deep Web", https://www.researchgate.net/publication/261773660; March 2010.
[7] Anand Rajaraman. "Kosmix: High Performance Topic Exploration using the Deep Web". VLDB 2009.
[8] A. Wright. Searching the Deep Web. In CACM, 51(10):14-15, October 2008.
[9] J. Madhavan, D. Ko, L. Kot, V. Ganapathy, A. Rasmussen, A. Y. Halevy. Google's Deep Web crawl. PVLDB 1(2): 1241-1252, 2008.