International Journal of Latest Trends in Engineering and Technology
Vol.(9)Issue(2), pp.167-174
DOI: http://dx.doi.org/10.21172/1.92.28
e-ISSN:2278-621X

UNVEILING DEEP WEB, A HIGH-QUALITY, QUANTITATIVE INFORMATION RESOURCE

S. Suneetha[1] & Dr. M. Usha Rani[2]

Abstract: Data is the fuel of today. Web is an information space from where data can be accessed through the internet. Search is one of the keys to Web's success. Accessing Web for desired information is like a perennial stream of knowledge. The more we dig the more thirst can be quenched. Hidden and non-indexable content cannot be accessed by traditional search engines such as, Google, Yahoo and Bing and thus the Deep Web is overlooked. Deep Web is the portion of the Web that is hidden from the view of search engines. Looking and slithering of the Deep Web without appropriate technology would be an inconceivable task. An extensive study about Deep Web is carried out to explore various means for unveiling the Deep Web. A Web Search should produce large quantity of high-quality results from the most relevant Web resources. Search engines are a good starting point, but they cannot probe beneath the Surface Web. Searching must incorporate the complete Web, both Surface Web and Deep Web. The primary contribution of this paper is to improvise the possibilities of unveiling the Deep Web resources in an effective and efficient manner. Our goal is to design, develop and implement an application that encompasses the complete Web – an integration of Surface Web and Deep Web, in Web Search. We hope that this overview provides an initiative for fruitful research in this stimulating area of Deep Web Exploration. We are attempting to integrate Deep and Surface Web information to cover the entire Web via query-based application.
Keywords: Surface Web, Deep Web, Surfacing Deep Web, TOR (The Onion Router), Deep Web Exploration Framework

1. MOTIVATION

Web is segmented into three layers: Surface Web, Deep Web, and Dark Web, in terms of content accessibility. Surface Web is the top layer and comprises of Web content that is readily accessible by using search engines. Standard search engines such as, Google, Yahoo and Bing index this Web content for making it easier to find. This indexing process is based upon the links on each page and the connections between them. Deep Web is a layer beneath the Surface Web that is hidden from the view of standard search engines. It is difficult to keep track of Deep Web. Dark Web is a small portion of the Deep Web that can be accessible only with special software to ensure anonymity. We are concerned with the Deep Web: a valuable trove of content that lies between the Surface Web and the Dark Web.

Deep Web or Hidden Web or Invisible Web or Encrypted Web refers to 'the portion of the Web that can be neither indexed nor searched by regular search engines and that require layers of encryption to be visited'. Only 4% of Web Content is available via Search Engines and the rest 96% is the Deep Web. Deep Web comprises of all information that resides in autonomous databases behind portals and information providers' Web front-ends. Content in the Deep Web is dynamically-generated in response to a query through a Web site's search form. Thus, Deep Web cannot be accessible by regular search engines. Unlocking/Unveiling this vast Deep Web content presents a major research challenge. There is a pre-requisite for designing, developing and implementing specific application for encouraging Deep Web Exploration/Search.

Standard search engines cannot meet all the data demands of the Web user. Deep Web is to be explored in order to gain competitive advantage. More timely, relevant and reliable information can be obtained from the Deep Web that is almost freely accessible. Exploring Deep Web physically without appropriate technology would be an inconceivable task. It is an expensive and time-consuming task. Other pitfalls include: information overload, dynamism and lack of indexing that prevent traditional search engines to access contents of the Deep Web.

There has been increasing interest in applying Web Mining techniques in building Deep Web crawlers. By virtue of this, the objective of the current research paper to make the Web user aware of the presence of the Deep Web, and its importance, as well as to initiate research in this area.

In this paper, we provided some information about (ABCs-About Basic Concepts of) Deep Web. The primary goal is to design and develop a contemporary Deep Web Exploration framework for indexing and ranking Deep Web contents based on the key word provided by the user through Tor interface. Providing better indexing and ranking improves search results. Further, this application based Deep Web Exploration framework is to be implemented both on Desktop and Android Mobiles.

[1] Research Scholar, Dept of Computer Science, Sri Padmavati Mahila VisvaVidyalayam, Tirupati, India
[2] Professor, Dept of Computer Science, Sri Padmavati Mahila VisvaVidyalayam, Tirupati, Andhra Pradesh, India

2. **ABOUT BASIC CONCEPTS OF DEEP WEB**

The term 'Deep Web' is coined by Mike Bergman in 1990 because of its massive size. [11] Deep Web is all the information stored online that isn't indexed by search engines. Accessing it requires filling out forms to secure specific content, searching a database, or logging into a specific set of content.

Illustrations include online forums and scanned documents. To get some information like accessing a full length research paper from the Web, one must go to the concerned site and then search through the records, and this requires credentials or a document number to access. In addition to the restricted content, some web pages are either not linked to other or have specific coding to prevent search engine crawls from indexing them. Since current search engines search the Web by going from link to link, they are unable to get to this content. The Deep Web, though less accessible, is potentially more valuable for businesses and organizations, since it is larger than the Surface Web and the content it serves up is a more authoritative source of information.

Deep Web includes dynamic web pages, blocked sites (like those that ask you to answer a CAPTCHA to access), private sites (like those that require login credentials), unlinked sites, non HTML/contextual/scripted content, and limited access networks. The following hidden resources are primarily a part of the Deep Web [11]:

- Dynamic content includes the content that is dynamically in response to a submitted query or accessed only through a form.
- Unlinked content includes Web pages that are not linked to others.
- Limited access content comprises of sites that limit access to their pages in a technical way by using special HTTP headers that prohibit search engines from browsing them. Engines such as, Google do not index pages outside of HTTP or HTTPS.
- Non HTML/Text content comprises of the content in multimedia files or specific file formats not supported by regular search engines.
- Limited Access Networks include the resources and services that are not normally accessible with a standard network configuration.
- Most search engines do not index text content using the Gopher protocol and files hosted on FTP and so they are a part of the Deep Web.
- Some Web content is hidden intentionally from the regular internet and can be accessible only with special software, such as Tor, I2P, or other software is the part of Deep Web.
- Web Archives are also a part of the Deep Web.

A few Key Points about Deep Web include [3]:

- Deep Web is the largest budding segment of fresh information on the internet.
- The quantity (content) of Deep Web is immense than that of the Surface Web.
- Deep Web sites are narrower, with deeper content, than conventional surface sites.
- More than half of the Deep Web content resides in topic-specific searchable databases.
- The quality of content on Deep Web, even, is much better than that of the Surface Web.
- 95% of the Deep Web content is publicly accessible information, absolutely free.
- Deep Web content is highly relevant to every information need, market, and domain.
- Professionals, especially researchers have several advantages in surfing the Deep Web.
- Deep Web is a treasure trove of incredible and interesting information that is hard to access efficiently in bulk.

2.1 Deep Web vs. Dark Web

Deep Web is often confused with Dark Web. Deep Web and Dark Web share few traits in common: like being outside the reach of search engines, Website URLs appended with '.onion', scattered among servers around the World rather than being at a single, centralized place. Yet, there is an important distinction. Deep Web contents are not hidden intentionally. They can be easily accessible through normal channels and doesn't require special software. It is necessary to know where to look in addition to what to look for. Specialized search engines (such as, DuckDuckGo, Torch and IceRocket), directories, and Wikis can help users locate the data being searched.

Dark Web is a small part of the Deep Web whose contents are hidden intentionally for some specific purpose. Dark Web demands special tools to access its contents and requires the highest level of encryption. Dark Web harbours criminal activity and often used by journalists and others to exchange sensitive information. Grams, Ahmia and Reddit are search engines available for browsing Dark Web. The beauty of the Dark Web is its anonymity. But, no one knows who is on the other side. So, necessary precautions need be followed while browsing the Dark Web. [5]

Domains on the Deep Web end with .onion and are reachable only via Tor. Tor (The Onion Router), a special encryption browser allows searching the Deep Web anonymously. It is the most widely used, popular distributed and anonymous software for browsing Deep Web, because of the reasons of speed and security. The names of Deep Websites are a random mix of letters and numbers such as, asdf579lkj13.onion'. These domains change regularly and a Website under one domain may not be under the same later.

2.2 About .onion

.onion obtained its name as the data that is sent passes through various layers of encryption and decryption (like the various layers of an onion) when it passes from source to destination. The purpose of using such a system is to make both the information provider and the person accessing the information more difficult to trace, by an intermediate network host, or by an outsider. Actually, .onion domains will not be available in the Internet DNS root. But, Web browsers such as, Tor can access sites with .onion addresses by sending the request through the network of Tor servers with the installation of appropriate proxy software. Deep Web search engines connect to the onion service via Tor and relays and then deliver the final search result to the browser on the Surface Web.

2.3 About Tor (The Onion Router)

Tor (The Onion Router) is free open source browser software used to surf the internet, especially the Deep Web anonymously. Tor browser is built as a modified version of the Mozilla Foundation's Firefox. It has its own privacy and security settings, reachable via a green onion icon in the toolbar. It makes use of use of onion routing process and can access .onion sites that cannot be accessed by standard Web browsers and search engines. Tor is popular for its speed and security.

Tor stands for The Onion Router that refers to the onion in technical terms. Two key aspects are related to the onion routing. First, the Tor network is composed of volunteers where computers are used as 'nodes'. During normal browsing, information travels across the Internet in packets. When a Tor user visits a Website, Tor creates a random path of nodes for the packet to reach the server. Another important aspect of onion routing is about the construction of packets. In general, a packet will include the sender's address and the destination. When using Tor, the packet is wrapped in successive layers of packets. When the user sends the packet, the top layer consists of information about Router A, the first stop on the circuit. When the packet reaches Router A, it takes off the first layer. The next layer informs Router A to forward the packet to Router B. Router A does not know the ultimate destination, only that the packet came from the user and went to B. Router B peels off the next layer, seeing that the next stop is Router C. The process continues until the message reaches its destination. No router is aware of the destination. At each stop, the node only knows the available information: the last place the packet was, and the next place it will be. No node knows the complete path, and neither would anyone who observes the message being sent from a node.

Tor is free open source software that is easy to download, install and set up. It can be used on Windows, Mac OS X, or Linux and works on all types of internet protocols such as, HTTP, FTP, gopher etc., Tor provides advanced browsing capabilities and is referred to as 'hidden service' for supporting anonymous browsing. Though browsing with Tor provides some advantages of accessing the desired information from the Deep Web anonymously reducing exposure, still it has risks. Bandwidths are reduced when using Tor and Tor doesn't encrypt the traffic. There is no official Tor browser for the iPhone and iPad yet. VPN is to be used on the top of Tor for additional security.

3. ACCESSING THE DEEP WEB

The methodology for accessing Deep Web is different from that of accessing Surface Web because of the hidden nature of the contents of the Deep Web. Some parts of the Deep Web are accessible through normal browser. Most of the Deep Web content is accessible, provided that the user knows where it is. Deep Web can be accessed like a private website, once the URL is known. But, this is not desired because of performance and security reasons [3].

There are two primary issues in accessing the Deep Web: Proper understanding of the interfaces (modelling of the query interfaces and perform schema matching) and efficient exploration of the data (determining which queries/browsing requests to issue, given the extremely restrictive input and output interfaces of a hidden web repository) [6].

3.1 Taxonomy for Surfacing Deep Web

Basically, there are two approaches for accessing the Deep Web: Data Integration and Surfacing. Data integration is about creating vertical search engines for specific domains. In this approach, mediator form is created for each domain and semantic mappings between individual data sources and the mediator form. This approach is expensive in terms of building and maintaining mediator forms and semantic mappings. Also, it is difficult to identify the domain and the forms within the domain that are relevant to a keyword query. Domain boundaries are not clearly defined on the Web. Surfacing the Deep Web includes pre-computing the most relevant form submissions for all interesting HTML forms followed by indexing the URLs resulting from these submissions. This approach leverages the existing search engine infrastructure and hence allows the seamless inclusion of Deep Web pages into Web search results; and so this surfacing approach is preferred [4].

Deep Web contents can be accessed by using any one of the following techniques. A brief taxonomy for Surfacing Deep Web [3]:

Using Deep Web Crawlers, the user accesses contents of the hidden Web by issuing a query through the search form provided by the Web site, which in turn gives a list of links to relevant pages on the Web. The user then looks at the obtained list and follows the associated links to find interesting pages. Resource discovery and Data extraction are the main tasks of Deep Web Crawler. Deep Web crawlers enable indexing, analysis and mining of hidden Web content. The extracted content can be used to categorize and classify the hidden databases. The results so obtained can then be used along with that of the existing search engines. Deep Web Crawlers may be generic (perform Breadth search) or focused (perform Depth based search) on the Deep Web. Focused Crawlers can be classic, semantic or learning Web crawlers. . But, crawling Deep Web requires huge amount of computing resources.

Schema matching can be done by Web source virtual integration system. The vertical integration system discovers relevant Web sources to user's query and avoids the user to search each Web source separately. This technique greatly reduces the cost of extraction of Web pages and then processing them. There are two types of virtual integration approach: vertical search engine that integrates the same kind of information from multiple Web sources and the second one is the topical search portal that integrates all information about a particular domain. Various other approaches of schema matching include: schema-only based, content based, hybrid and composite matchers.

Data mining on the Deep Web can produce important insights. Clustering can be performed on Web sources to process only domain specific content. Deep web has huge amount of information, hence the number of relevant Web pages returned might be very less. Ontologies can assist the web search, to reduce the number of irrelevant web pages returned by the search engine. Vision-based approach utilizes the visual features to extract data from Deep Web pages and is programming language independent.

3.2 Accessing Deep Web with Tor browser

In general, the following steps are to be followed for accessing the Deep Web with Tor browser [10]:

Step-1: Install Tor Browser Bundle.

Tor can be downloaded from torproject.com free of cost, installed easily and finally a secure connection is to be established by connecting our internet to Tor via a relay.

Step-2: Configure Tor Browser for better security.

To accomplish this, the security level is to be changed from Low to High in 'Privacy and Security Settings' menu item, for anonymity.

Step-3: Find the working URLs of Deep Web sites.

Deep Web links are dynamic as .onion of the sites change regularly to keep them untraceable. A website that was visited yesterday may not be under the same domain today. So, finding a working link of the Deep Web site is difficult. .onion directories and search engines are available on the internet to easily find the working Deep Web addresses.

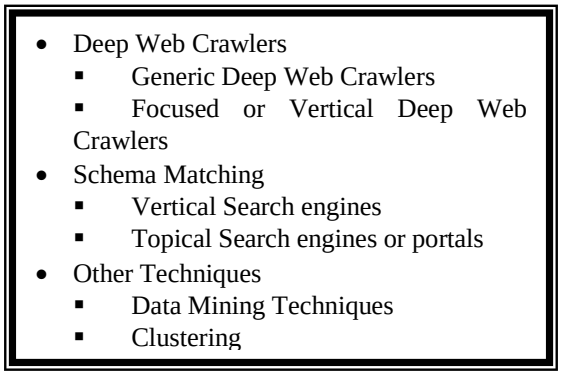
Popular .onion link directories include: http://thehiddenwiki.org and http://hwikis25cffertqe.onion, the homepage of Hiddenwiki that displays hidden Deep Web site links category wise. DuckDuckGo, Torch and Ice Rocket are few Deep Web search engines used to search the whole Deep Web like standard search engines like Google.

Eg:- If we want to chat with someone without revealing original identity and looking for user friendly sites like facebook, then, visit 'hwikis25cffertqe.onion' site, scroll down a little bit to find 'Social Networks section' and finally click on any one of the links on that section to access social media websites. If facebook link is selected from that section then it will take to the Deep Website version of facebook and its URL is 'www.facebookcorewwwi.onion'.

3.3 Deep Web Browsers for Windows, Linux, Android, iPhone

Some of the famous Deep Web browsers for Windows, Android, iPhone & Linux follow [7]:

- Tor Browser is the best and the most widely used browsers for accessing the Deep Web by netizens. It is available free of cost, easy to download, install, configure and use, and requires no other additional software.
- Orbot is the only mobile application available to access Deep Web on Android mobiles. Orbot developed by the Tor project team acts as a proxy server and protects the user's browsing activity as well as internet activity. It is free to download and easy to use. Internet speeds will be reduced while browsing via Tor network, using this application.



- VPN Browser developed by Art Fusion Labs, is used for accessing Deep Web on iPhones.
- Operating Systems with inbuilt Deep Web browser
  Operating systems such as, Tails OS, Kali Linux and Qubes have inbuilt Tor browser and a complete set of tools for browsing the Deep Web.
- Browser Plug-ins for browsing Deep Web
  To convert the standard browser like Chrome into a Deep Web browser, plug-ins/extensions are available.

3.4 Top 10 Best Deep Web Search Engines of 2017

The following is the list of top 10 Best Deep Web search engines that are used to search Deep Web easily [12]:

1. Torch is the most popular and widely used Deep Web search engine providing about 90,000 search results everyday for Tor users.
2. not Evil will provide the most appropriate content as a part of its search results without popup or text ads.
3. DuckDuckGo is the only search engine which will not track the things that we are doing online. It also allows customizing the look and feature of the search easily.
4. Grams is the dark net search engine widely used to search multiple marketplaces for specific products listing along with their features.
5. Ahmia is an open source dark net search engine for Tor hidden services that allows adding new sites to its database.
6. Archive is used to access either Surface or Deep Web. It takes more time to find a Deep Web site on archive and hence it is difficult to access this non-profit library of free books, music, movies and more.
7. Google requires some special keyword on the search box to show the results from Deep Web.
8. onion.link accesses Deep Web without using Tor browser.
9. LookAhead is used to find journals, political news and research papers easily in a completely traceable way.
10. Live Leak is one of the best available search engines for finding Deep Web videos easily. It also provides an option to upload a video related to the Deep Web.

3.5 Accessing Deep Web on Android

Deep Web can be accessed not only from the desktop but also from hand-held gadgets and Android-powered Televisions. Deep Web can be accessed on Android mobiles by using anyone of the below methods [9]:

- Orbot is an Android application used to browse Deep Web on Android mobiles. It is available for free in the play store and easy to install. Once it is installed and created a connection to the Deep Web successfully, just the Deep Web address is to be entered in the browser and enter key is to be pressed. It is one of the easiest methods available.
- Orfox is the official Tor project Web browser for Android, a modified version of Mozilla Firefox. It is highly recommended method for surfing the Deep Web with Tor project's Orbot application, on Android mobiles. This browser software provides highest security on Android. The drawback with this browser is that it won't support all the Mozilla plug-ins because of version issues.
- Deep Web Search Engines: One can visit the site of Deep Web search engine from the Android browser and explore the Deep Web just like a normal browser. The search results so obtained will be only from certain part of the Deep Web. But, this method is not recommended for accessing Deep Web pages.
- Using Google: Search engine giant Google can be used to access Deep Web as well. On the Android phone browser visit Google.com. Then, special key words should be entered in the Google Search box to get the results from the Deep Web. Copyrighted materials should be downloaded using this method.

For example, to search a collection of pdf files related to technology one has to type 'indexof: .pdf Technology' in the Google's search field and hit Search. By doing so, the user will get almost all digital files for free. The search results will be displayed and selecting one of the links will direct to the server that has the concerned file.

Though Orbot and Orfox can be used together to browse the Deep Web, Orwall should also be used along, to enable all the applications on the Android device to use the Tor network for their data transmission. Orwall is a firewall for Android that makes use of iptables and special set of rules to redirect the traffic for Android applications through Tor [8]. All the above are not cent percent foolproof/offer silver bullet for total privacy.

4. DEEP WEB EXPLORATION/SEARCH

Search is one of the keys to Web success. A lot of significant information resides in the Deep Web especially in accessible databases amazingly valuable for an engaged hunt. Traditional search engines cannot probe underneath the Surface Web into searchable databases. This inability of surface Web crawlers motivated us towards Deep Web Exploration/Search.

4.1 Working of a Typical Search Engine

The user can search for the desired Web pages, with a Search engine. He/she can evaluate the obtained search results and make use of the information retrieved by the search engine that is most promising. The broad design of a search engine generally consists of three activities: Crawler, Indexer and Search modules. Cyclic architecture of a search engine is shown in the following figure [2]:

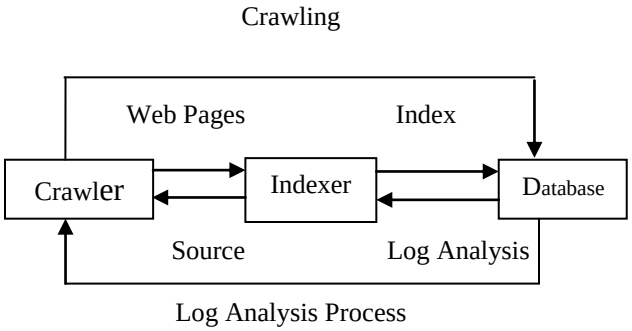


Figure 2. Working of a Typical Search Engine

The Web Crawler is the first stage that downloads Web documents which are then indexed by the indexer for later use by the search module, with a feedback from other stages. Besides simple indexing, the indexer module helps the Web crawler in arranging the Web pages based on the ranking, to make crawler more selective in downloading of high ranking Web pages. The search process, through log file analysis process, optimizes both indexer and the crawler by providing information about 'active set' of Web pages seen and sought by the users. At last, the Web crawler provides on-demand crawling services for search engines.

4.2 Surface Search Engines vs. Deep Web Search Engines

Traditional search engines such as, Google and Yahoo have limitations in indexing the Deep Web pages and thus in accessing the Deep Web contents. Moreover, the working of Deep Web is entirely different from that of the traditional Web. In Deep Web, there are no tags that link the contents and hence no chain of links to be followed. On the other hand, the hidden Web pages are dynamically generated and therefore in Deep Web retrieval process, quality of a page cannot be predicted with its reference.

It is challenging to locate the Deep Web databases, because they are not registered with any traditional search engines, sparsely distributed and keep constantly changing. Besides efficiency, quality and coverage on relevant Deep Web sources are also challenging. A Deep Web Crawler must produce a large quantity of high-quality results from the most relevant content sources.

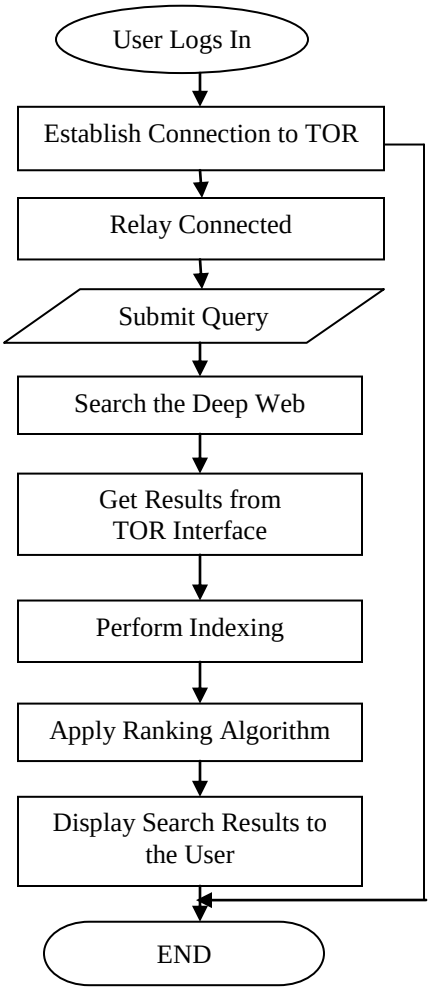
Two major advantages of Deep Web Search engines include: Analytic capabilities and Versioning of Web pages. Instead of storing metadata and only top keywords as in regular search engines, Deep Web Harvest engine retrieves the actual raw text from Web pages. This can be integrated directly into nearly any analytic technology. By combining scalable Deep Web harvesting capabilities with custom analytic technology helps customers visualize, analyze, and ultimately create intelligence from large data sets.

Thus, there is a need for developing a contemporary framework to access Deep Web contents.

5. DEEP WEB EXPLORATION FRAMEWORK

'Deep Web Exploration/Search' is a latest technique that extracts information by probing underneath the visible Web. It is an imperative subject of Data Mining. Providing better indexing and ranking improves search results. A Deep Web information retrieval framework was proposed by [2] after survey of technical literature available on Deep Web search. The taxonomical steps of the framework are classified into the four major parts: Query Interface Analysis, Values Allotment, Response Analysis and Navigation, Relevance Ranking. In this framework, different tasks in Deep Web crawling are identified, arranged and aggregated in sequential manner. This Deep Web Extraction framework [2] serves as a foundation for understanding entire working of Deep Web crawling mechanisms as well as it facilitates to identify the limitations of present Web crawling mechanisms in searching the Deep Web. We extracted the qualified steps from this framework and developed a Deep Web Exploration framework for indexing and ranking Deep Web pages through Tor interface. Using this framework as basis, we customized the existing Data Mining algorithms for classification, indexing and ranking so that they can be applied to the large scale crawling of the Deep Web by overcoming the limitations of existing ones. Indexing plays a vital role in Deep Web whose contents are not indexed. Ranking is the heart of search. Ranking refers to the order in which search engine should return the results (URLs) produced in response to a user's query. Before presenting results for the user's search query, they are ordered according to relevance.

The flow of workand algorithm of our proposed Deep Web Exploration framework is presented in figure 3.



The algorithm for extraction of data using our proposed Deep Web Exploration framework:

- Initially, the user submits the query in the search box.
- Internally, the connection to TOR is established and Deep Web contents that match with the user's query are retrieved through TOR interface.
- The retrieved contents are indexed by using our customized indexing algorithm.
- Then, the indexed contents are ranked by using customized ranking algorithm.
- Finally, the search results (links) are displayed to the user.

Steps in proposed Deep Web Exploration Framework

Figure 3. Flow of Work and Algorithm of Proposed Deep Web Exploration Framework

Deep Web Exploration framework takes user's query as input makes use of customized indexing and ranking algorithms to extract implicit and potentially useful links from the Deep Web, which are then presented to the user for his/her use. The elegant way in which the search results are returned has been well studied and found be remarkably effective. The proposed

Deep Web Exploration framework is implemented on Desktops. An overall better performance is visualized than the currently existing works with this application-based framework.

6. CONCLUSION

WWW is a large source of information that contains data in either Surface Web or Deep Web. Standard search engines such as, Google and Yahoo cannot probe beneath the Surface Web. Nevertheless, Deep Web is a gold mine of high quality quantitative information that is difficult to access directly. Deep Web is the set of information resources on WWW, unindexed, and thus cannot be accessed by using standard search engines. A great deal of work is available on Deep Web Data Extraction but each one has its own limitations. Much of the work is vision-based. It is a highly unexplored area of research and in great need of full automation. Even, accessing Deep Web through Mobiles is still in its infancy.

Deep Web Exploration fosters innovation in the search industry. For researchers in the area of Data Mining, Deep Web offers tremendous opportunities. Through this paper, we are providing the importance of Deep Web and related information for accessing the Deep Web to initiate research in this area. We presented a Deep Web Information Exploration framework for indexing and ranking, thus, accessing Deep Web. The contemporary framework provides access to Web's investments in crawling, indexing and ranking and thus serves as a platform for the next generation of search innovations serving several Web users. Indexing plays a vital role in Deep Web whose contents are not indexed. Ranking is the heart of search. 'Ranking' refers to the order in which search engine should return the results (URLs), produced in response to a user's query i.e. to display more relevant pages to the user on priority basis. It personalizes the looking as requested by the user so that the outcome appears simple and precise to the user.

7. FUTURE SCOPE

In continuation to the design and development, we are implementing Deep Web Exploration framework for indexing and ranking with the customized algorithms for indexing and ranking. The application-based framework not only works on desktops but we are also planning to access Deep Web resources on Android mobiles in a fully automated manner through Tor. Search engines are a good starting point, but they only skim the surface of available content. Deep Web provides high quality quantitative information for creating actionable intelligence. However, searching must evolve to cover the whole Web. Both Surface Web and Deep Web information can be integrated through Directed query technology and our research is proceeding towards this direction.

8. REFERENCES

[1] Carlos Castillo, "Effective Web Crawling", Ph.D. Thesis, Department of Computer Science, University of Chile, November 2004.

[2] Dilip Kumar Sharma and A. K. Sharma, "Deep Web Information Retrieval Process: A Technical Survey", Source Title: Models for Capitalizing on Web Engineering Advancements: Trends and Discoveries DOI: 10.4018/978-1-4666-0023-2.ch005.

[3] Khushboo Khurana and Dr. M. B. Chandak, "Survey of Techniques for Deep Web Source Selection and Surfacing the Hidden Web Content", IJACSA, Vol. 7, No. 5, 2016 pp.409-418 www.ijacsa.thesai.org.

[4] Michael J. Cafarella, Jayant Madhavan and Alon Halevy, "Web-Scale Extraction of Structured Data", Available at: http://web.eecs.umich.edu/~michjc/papers/cafarella-webscale-sigmodrecord08.pdf.

[5] Michael K. Bergman "White Paper: The Deep Web: Surfacing Hidden Value", Bright Planet - Deep Content, September 24, 2001.

[6] Nan Zhang and Gautam Das, "Exploration of Deep Web Repositories", 37th International Conference on Very Large Data Bases, Proceedings of the VLDB Endowment, Vol. 4, No. 12 VLDB Endowment 2150-8097/11/08.

[7] http://www.pcsutra.com/deep-web-browser/

[8] http://www.pcsutra.com/access-deep-web-android-phone/

[9] https://www.bonkersabouttech.com/security/how-to-access-the-deep-web-on-your-android-phone/422

[10] http://www.pcsutra.com/access-the-deep-web-with-tor/

[11] https://en.wikipedia.org/wiki/Deep_web

[12] http://www.pcsutra.com/deep-web-search-engines/