

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

May 2010

## Understanding the Deep Web

Rabia Iffat

Government Degree College Mahagaon Cross, Gulbarga District, India, Fiza\_kainat@yahoo.com

Lalitha K. Sami

Gulbarga University, lalita\_sami@rediffmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

---

Iffat, Rabia and Sami, Lalitha K., "Understanding the Deep Web" (2010). *Library Philosophy and Practice (e-journal)*. 364.

<https://digitalcommons.unl.edu/libphilprac/364>

## Understanding the Deep Web

Rabia Iffat

Librarian

Government Degree College  
Mahagaon Cross, Gulbarga District, India

Lalitha K. Sami

Professor, Department of Library & Information Science  
Gulbarga University  
Gulbarga, Karnataka, India

### Introduction

The most coveted commodity of the information age is indeed information. Information has become a basic need after food, shelter, and clothing. Due to technological advancements, a large amount of information is available on the Web, which has become a complex entity containing information from a variety of sources. Information is found using search engines. A searcher has access to a large amount of information, but it still far from the huge treasury of information lying beneath the Web, a vast store of information beyond the reach of conventional search engines: the "Deep Web" or "Invisible Web."

The contents of the Deep Web are not included up in the search results of conventional search engines. The crawlers of conventional search engines identify only static pages and cannot access the dynamic Web pages of Deep Web databases. Hence, the Deep Web is alternatively termed the "Hidden" or "Invisible Web." The term Invisible Web was coined by Dr. Jill Ellsworth to refer to information inaccessible to conventional search engines. But using the term Invisible Web to describe recorded information that is available but not easily accessible, is not accurate.

### Deep vs Surface Web

The Web can be divided into Surface Web and Deep Web. Surface Web is made up of static and fixed pages, whereas Deep Web is made up of dynamic pages. Static pages do not depend on a database for their content. They reside on a server waiting to be retrieved, and are basically html files whose content never changes. Any changes are made directly to the html code and the new version of the page is uploaded to the server. In static publishing, html text is pre-generated and stored as flat files on Web servers. These pages are less flexible than dynamic pages. Dynamic pages are created as a result of a database search. They are also called database-driven Web pages, wherein the content and the design are housed separately. The content is put in a database and is provided only when requested by the user.

The following are the differences between Surface Web and Deep Web (Bergman, 2001)

- Public information on the Deep Web is currently 400-550 times larger than the commonly-defined World Wide Web.
- The Deep Web contains 7,500 terabytes of information, compared to 19 on the Surface Web.

- Deep Web contains nearly 550 billion individual documents compared to one billion on the Surface Web.
- More than 200,000 Deep Web sites presently exist.
- Total quality 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.
- More than half of Deep Web contents reside in topic-specific database.
- Content of the Deep Web is evaluated by experts
- Information on Deep Web is more comprehensive than that of the Surface Web.

### **Concept of Deep Web**

The content of the Deep Web is rarely shown in a search engine result, since the search engine spiders do not crawl into databases and extract the data. These spiders can neither think nor type, but jump from link to link. As such, a spider cannot enter pages that are password protected. Web page creators who do not want their page shown in search results can insert special meta tags to keep the page from being indexed. Spiders are also unable to pages created without the use of html, and also links that include a question mark. But now parts of the Deep Web with non-html pages and databases with a question mark in a stable URL are being indexed by search engines, with non-html pages converted to html. Still, it is estimated that even the best search engines can access only 16 percent of information available on the Web. There are other Web search techniques and technologies that can be used to access databases and extract the content, including Librarian's Index to the Internet, which indexes access points to the content of the Deep Web.

### **Tips for Searching the Deep Web**

- To search the Deep Web requires time and patience
- With a specific URL, users can access a specific database
- Search for the specific database directory, i.e., while entering search terms, keywords should be entered along with the word "database." For example, for information on sports, use "sports database"

### **Tools for Searching the Deep Web**

In the past technological barriers made it difficult to access the Deep Web, but it is now possible to overcome these barriers. Many individuals and institutions have compiled a list of invisible Web directories. They include:

Complete Planet

Complete Planet searches more than 70,000 databases and specialty search engines.

<http://www.completeplanet.com>

Worldwide science.org

This is a gateway to national and international scientific databases. It provides public access to more than 200 million pages of international research information. It was developed by the US Department of Energy, the British Library, and eight countries ranging from Australia to Japan. It uses federated search technology, searching a variety of databases, aggregating the results, and ranking them.

<http://scilibupdate.blogspot.com/2007/07/worldwidescienceorg.html> <http://worldwidescience.org/>

## INFOMINE

INFOMINE is a virtual library of Internet resources relevant to university faculty, students, and research staff. It contains resources such as databases, electronic journals, electronic books, bulletin boards, mailing lists, library catalogs, articles, directories of researchers, and many other types of information. Librarians from the University of California, Wake Forest University, California State University, the University of Detroit - Mercy, and other universities and colleges have contributed to building INFOMINE.

<http://infomine.ucr.edu/>

## Techweb.com

This site is a path finder for locating IT information by describing useful search tools, portals, and websites.

<http://www.techdeepweb.com>

## Directory of Open Access Journals

DOAJ is maintained by the Lund University, and a collection of searchable scientific and scholarly journals on the Web

<http://www.doaj.org/doaj?func=subject&cpid=20>

## World Fact Book

This is a searchable directory of flags of the world, reference information, maps, country profiles and more

<http://www.worldfactbook.com>

## Infoplease.com

Infoplease is a searchable Deep Web database. Results are from encyclopedias, almanacs, dictionaries, and other resources.

<http://www.infoplease.com/>

## The Library Spot

This is a collection of databases, online libraries, references, and other information.

<http://www.libraryspot.com>

## SCIRUS

SCIRUS is a comprehensive scientific research tool. It has 450 million items indexed, and allows searches for journal content, scientists' homepages, courseware, preprints, patents, institutional repositories, and website information.

<http://www.scirus.com>

## PUBMED

Pubmed is a service of the US National Library of Medicine, with more than 18 million citations from Medline and other life science journals, going back to 1948.

<http://www.ncbi.nlm.nih.gov/pubmed/>

## OAIster

OAIster is a multidisciplinary database, largely based in university and research institution digital libraries. Typical collections include theses, technical reports, research papers, and image collections, which are often difficult to access. OAIster harvests from Open Archives Initiative (OAI) compliant institutional repositories.

<http://oaister.worldcat.org/>

## Research and Mine the Invisible Web.

This is a site that links to 99 ways of accessing the Deep Web using search engines, Deep Web databases, and directories.

<http://www.collegedegree.com/library/college-life/99-resources-to>.

## Librarians' Index to the Internet.

This is a Web directory that offers a searchable and browsable collection of websites maintained by librarians. The Web Directory is organized into 14 main topics and about 300 related topics. It is possible to access Deep Web databases through LII, by typing a broad topic and adding the words "and databases," e.g., "chemistry and databases."

<http://lii.org/>

## Turbo10

Turbo10 is a metasearch engine that allows a user to search the Deep Web through a collection of databases and search engines, which can be customized.

<http://www.turbo10.com>

## LexiBot

This is a specialized search tool developed by BrightPlanet for searching the Deep Web. LexiBot enables users to conduct searches using simple text, natural language, or Boolean queries on hundreds of databases simultaneously, to filter and analyze data, and to publish the results as Web pages.

<http://www.LexiBot.com>

## Nuclear Explosions Database Geoscience

This is Australia's database of nuclear explosions that provides location, time, and size of explosions worldwide since 1945.

<http://www.ga.gov.au/oracle/nuclear-explosion.jsp>

## **Conclusion**

The advent of Internet and access to global information was a great benefit, even though information managers had the difficult task of organizing, retrieving, and providing access to precise information. Users depend on the popular search engines and portals, which cannot provide access to the hidden store of valuable information available in the Deep Web. To access the information available on these databases, users will have to become familiar with the structure of the Deep Web. Any information created should be shared and used, since that alone leads to the creation of more information. When a specific database is created, information regarding its existence should be published so that users will be aware and make maximum use of available information.

## **References**

Bergman, M.K. (2001). White Paper: The Deep Web: Surfacing Hidden Value. Ann Arbor, MI: Scholarly Publishing Office, University of Michigan, University Library 7(1). DOI: <http://dx.doi.org/10.3998/3336451.0007.104>