

# Search techniques for efficient information retrieval on web (Part-1)

**Ms. Sapna**

**Assistant Professor**

**Department of Library & Information Science**

**Mahatma Gandhi Central University**

**PHONE NO.:- +91-8988167975**

**Email-Id: [sapnasna121@gmail.com](mailto:sapnasna121@gmail.com)**

# Learning Objectives

- Be introduced to historical evolution of search engines
- Be able to make difference between different types of search engines
- Identification of examples of search engines
- Know about faceted and federated search techniques

- With the rapid advancement of technology, a large amount of information is being made available on the web however, retrieving relevant results from the web search engine is quite difficult. The reasons for this are:
  - (a) abundance of information on the web and
  - (b) lack of proper description and organization of information on the web
- Due to the massive amount of information on the web, it has become very difficult and overwhelming for a user to retrieve relevant information from the web

- To overcome the issues in information retrieval, various search tools and techniques are being constantly developed
- Search engines are one among such discovery tools

# Web Search Engine

- A search engine acts as a practical application of information retrieval techniques to large-scale information collections (Croft et al, 2015)
- Web Search Engine is a type of program designed to help find and access information stored on the World Wide Web

## S.No. Search engine

1 Crawler-Based Search Engines

### Features

- Searching is divided into two phases: The back-end phase and front-end phase.
- For a specific query, the crawler-based search engines are quite efficient in finding relevant information

*Example: **Google***

2 Meta-search Engines

- Don't have their own database of data. They search the databases of other search engines.
- It allows the user to search several search engines simultaneously i.e. main advantages of meta search engine
- Examples of meta search engine is **Dogpile**.

3 Vertical Search Engines

- It focuses on a particular domain of search.
- Facilitate more accurate, relevant and faster search by indexing in specific domains

*Example: **Answers.com***

S.No.	Search engine	Features
4	Hybrid Search Engines	Hybrid search engines is a combination of both crawler-based results and human-powered directories <i>Example: <b>DMOZ</b></i>
5	Specialized Search Engines	A search engine which is specialized in a particular topic usually produces a better quality of results and most relevant documents as compare to general search engine <i>Example: <b>Scirus</b></i>

# Historical Evolution of Search Engines: From Archie to Google Today

- Archie was first search engine created by Alan Emtage in 1990, a student at McGill University in Montreal. Its original name was "archives" but it was shortened to Archie
- Beginning with Archie in 1990, considered the first search engine, moving on to Excite and Lycos and Infoseek
- AltaVista and Yahoo were launched in 1995, and brought many important features to the web searching
- Google, developed by Larry Page and Sergey Brin, Stanford University, Stanford in 1998
- After that DogPile, Duck Duck Go, Live Search, Bing , Scinet came into existence



# Open Source Search Engines

S.no.	Search Engine	Features
1.	<b>Lucene</b>	<ul style="list-style-type: none"><li>· Fielded searching (e.g. title, author, contents) sorting by any field</li><li>· Powerful, Accurate and Efficient Search Algorithms</li></ul>
2.	<b>Apache Solr</b>	<ul style="list-style-type: none"><li>· Near Real-Time Indexing</li><li>· Faceted Search and Filtering</li><li>· Query Suggestions, auto correction of Spelling errors</li></ul>
3.	<b>Sphinx</b>	<ul style="list-style-type: none"><li>· SQL database indexing</li><li>· Advanced full-text searching syntax</li><li>· Better relevance ranking</li></ul>
4.	<b>Xapian</b>	<ul style="list-style-type: none"><li>· It allows developers to easily add advanced indexing and search facilities to their own applications</li></ul>
5.	<b>Nutch</b>	<ul style="list-style-type: none"><li>· Create plug-ins for media-type parsing, data retrieval, querying and clustering</li><li>· Highly robust and scalable</li></ul>

# Search Techniques

## Federated Search

- Search for information in multiple databases
- Single search interface
- Search multiple library catalogs (OPACs), Web sites and citation databases all at once
- Integrated search result technology

Technology: XML & Z39.50 protocol

*Example* : Mednar.com



**i** Full Record

**i** Title

**i** Author

**i** Date Range

**i** Match

From  To

Limit to Available Full-Text **i**

Clear

Search

**+**  All Categories

**+**  Commercial Databases

**+**  Medical Societies

**+**  NIH Resources

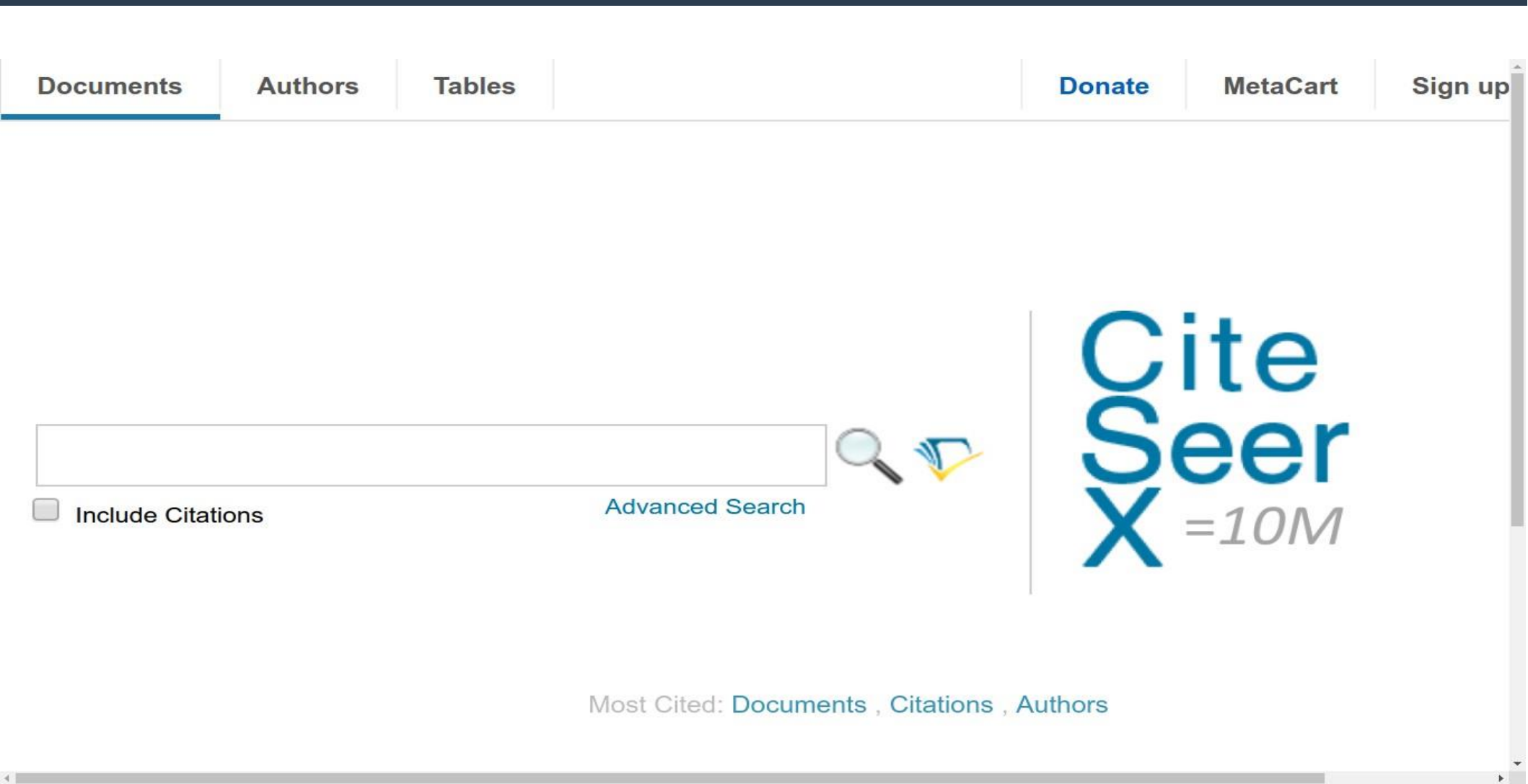
**+**  Other Government Resources

**+**  Patents

# Faceted Search

- It is a technique for accessing information organized according to a faceted classification system
- Also known as faceted navigation or faceted browsing
- Many online library catalogs are using faceted search interfaces
- Faceted search provides a platform for interactive information retrieval

Example: The CiteSeerX project at the Pennsylvania State University permits faceted search for academic documents and Furthermore proceeds to extend under different facets, for example, table search



# References:

1. Apache Lucene core. Retrieved from <http://lucene.apache.org/core/>
2. Arrigo, M., Gentile, M., Taibi, D., & Di Giuseppe, O. (2005). Specialized search engines for E-learning. *Recent Research Developments in Learning Technologies*.
3. Croft, W. B., Metzler, D., & Strohmann, T. (2015). Search engines. *Pearson Education*.
4. Shade O, Kuyoro, Okolie Samuel, O, & Kanu Richmond, U. (2012). Trends in web- based search engine. *Journal of Emerging Trends in Computing and Information Sciences*,3(6).

**Thank you**

*Stay Home, Stay Safe & Save Lives*