Course 5. Information searching tools

Outline

1. Building a search strategy
2. Scientific information sources
3. Scientific information on the web

Introduction

Various search tools can offer a range of features, including private browsing, personalized results and filtering options. Whether you're writing an essay or conducting scientific research, learning about the various search tools available can help you determine which one might best fit your needs.

You can find different types of information by using various search tools:

- Library catalogue: a great place to begin your search for academic sources as you can search by title or keywords
- Databases: perform targeted searches using advanced search functionalities
- Google: search for specific types of information on credible websites and in different formats.

1. Building a search strategy

A search strategy is a logical series of steps for planning and preparing an efficient way to collect information on a given topic. A search strategy shows how these terms combine in order to retrieve the best results.

Different databases work in different ways, so you need to adapt your search strategy for each of the databases you use.

Databases have several features that will allow you to build searches that will focus your search to relevant information. These features include search techniques that will allow you to search more efficiently and effectively.

2. Sources of Information

2.1. Search engines

Search engine, computer program to find answers to queries in a collection of information, which might be a library catalog or a database but is most commonly the World Wide Web.

- **Refseek (http://www.refseek.com)**
- **WorldCat (http://www.worrldcat.org)**
- **Springer (http://link.springer.com)**
- **Microsoft_Academic (http://academic.microsoft.com)**
- **Bioline (http://www.bioline.org.br)**
- **EThOS (http://ethos.bl.uk)**
- **Science.gove (http://www.science.gov)**
- **ISEEK Education (http://education.iseek.com)**
- ***Base_search* (http://www.base-search.net)**
- ***pdfdrive* (http://www.pdfdrive.com)**
- **Sci-hub (https://sci-hub.se)**

2.2. Metasearch engine

Metasearch is to utilize **multiple other search systems** (called component search systems) to perform simultaneous search. A metasearch engine is a search system that enables metasearch.

To perform a basic metasearch, a user **query is sent to multiple existing search engines** by the metasearch engine; when the search results returned from the search engines are received by the metasearch engine, they are merged into a single ranked list and the merged list is presented to the user.

Key issues include how to pass user queries to other search engines, how to identify correct search results from the result pages returned from search engines, and how to merge the results from different search sources. More sophisticated metasearch engines also perform search engine selection (also referred to as database selection), i.e., identify the search engines that are most appropriate for a query and send the query to only these search engines. To identify appro- priate search engines to use for a query requires to estimate the usefulness of each search engine with re- spect to the query based on some usefulness measure.

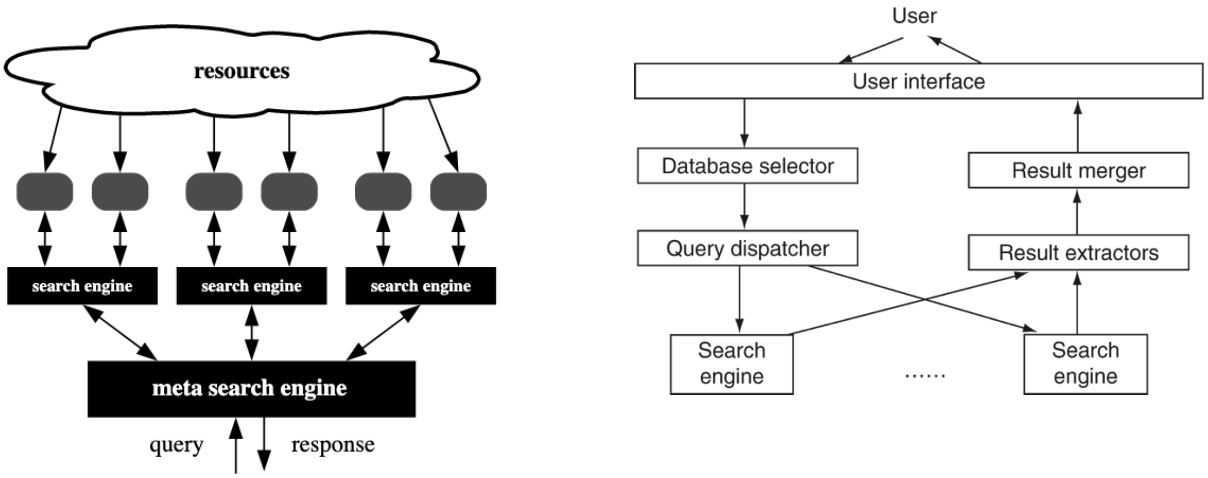

Figure 1. Metasearch engine component architecture.

- www.search.com,
- www.dogpile.com,
- www.yippy.com

3.3. Multi-search engine

Multisearch is a multitasking search engine which includes both search engine and metasearch engine characteristics with additional capability of retrieval of search result sets that were previously classified by users.

Example :

- Creative Commons Search,
- Ecosia,
- Disconnect,
- le moteur de recherche de Maxthon,
- HooSeek .

3.4. Database

Data **is information** that can be **linked to any object**.

It can take many different forms: numbers, bytes, images, and text files…

A database is an **organized collection of structured information**, or data, typically stored electronically in a computer system.

A database is usually controlled by a database management system (DBMS).

Most databases use structured query language (SQL) for writing and querying data.

A database consists of **five main components**:

- Hardware (includes all the physical devices needed to install and run a database),
- software,
- data,
- procedure,
- and database access language.

The Biological data can be broadly classified as:

Biological Databases	**Information they contain**
1. Bibliographic databases	Literature
2. Taxonomic databases	Classification
3. Nucleic acid databases	DNA information
4. Genomic databases	Gene level information
5. Protein databases	Protein information
6. Protein families, domains and functional sites	Classification of proteins and identifying domains
7. Enzymes/ metabolic pathways	Metabolic pathways

Some databases accessible via the server of the University of Bejaia :

- Science direct : (www.sciencedirect.com)
- Springer link: (www.springerlink.com)
- Jstor: (http://www.jstor.org)
- Global plant on Jstor: (http://www.plants.jstor.org)
- NCBI (www.ncbi.nlm.nih.gov)
- Geo science world: (http://www.geoscienceworld.org)
- SNDL : Système National de Documentation en Ligne (https://www.sndl.cerist.dz) qui offre l'accès à plusieurs bases de données.
- Web of science : (https://webofknowledge.com) et Scopus (www.scopus.com), qui sont supports d'analyses bibliométriques.