

Information Technology Essentials — Lecture 12

Dr. Karim Lounis

Fall 2023



Web Technology

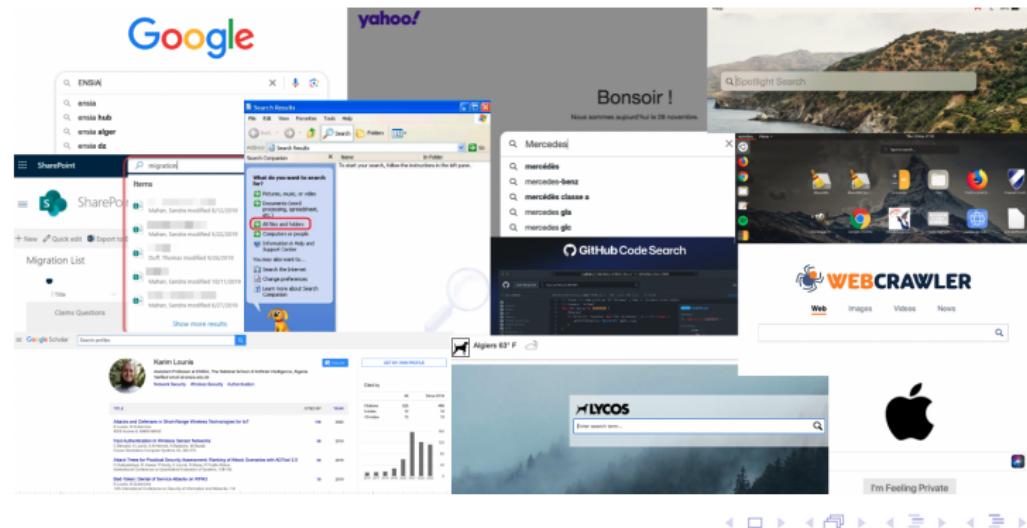
Search Engines

?

Search Engines

Definition

Search Engines. Are specialized software systems designed to search and retrieve information from a data sources, typically by indexing and organizing content to provide relevant results in response to various user queries.



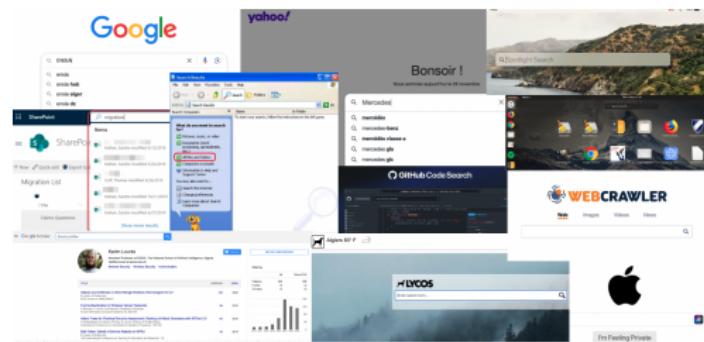
Search Engines

Definition

Search Engines. Are specialized software systems designed to search and retrieve information from a data sources, typically by indexing and organizing content to provide relevant results in response to various user queries.

What do Search Engines do?

The primary function of search engines is to allow users to input queries or keywords and then return a list of results that match the user's search criteria.



These results typically include links to documents, images, videos, applications, websites, etc.

Type of Search Engines

Depending on their application, there exist various types of search engines:

- **Desktop Search Engines.**
- **Enterprise Search Engines.**
- **Code Search Engines.**
- **Mobile Search Engines.**
- **Academic Search Engines.**
- **Web Search Engines.**

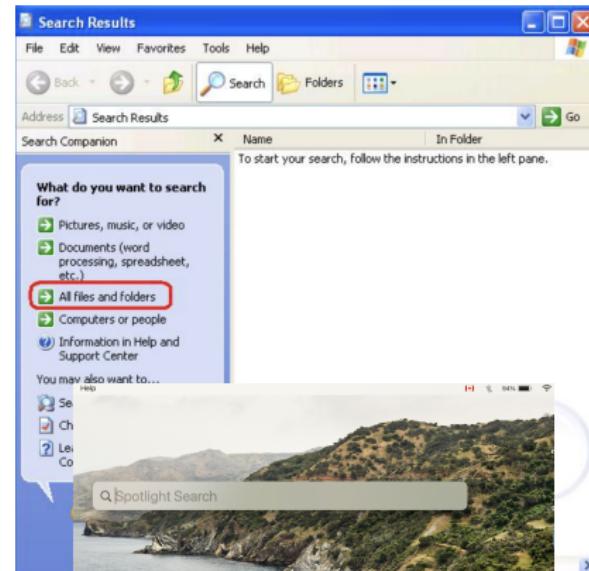
Type of Search Engines

Depending on their application, there exist various types of search engines:

- **Desktop Search Engines.**
- **Enterprise Search Engines.**
- **Code Search Engines.**
- **Mobile Search Engines.**
- **Academic Search Engines.**
- **Web Search Engines.**

E.g., **Microsoft Windows Search** and
Apple's macOS Spotlight (and others)

These are built-in application search engines allowing users to search for files and applications on their desktops.



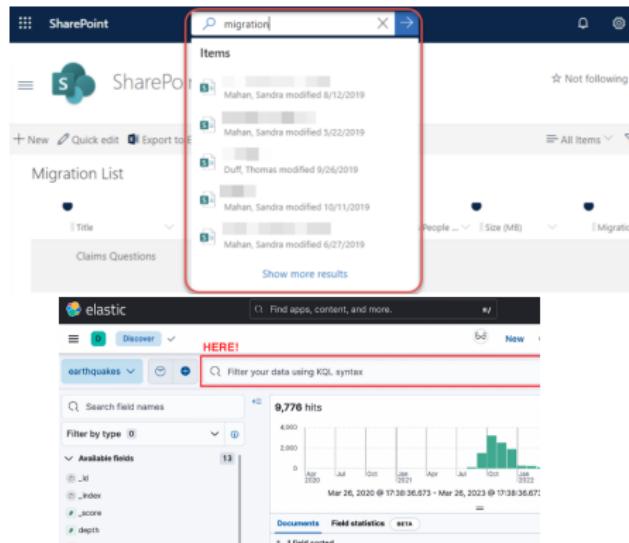
Type of Search Engines

Depending on their application, there exist various types of search engines:

- Desktop Search Engines.
- Enterprise Search Engines.
- Code Search Engines.
- Mobile Search Engines.
- Academic Search Engines.
- Web Search Engines.

E.g., Microsoft SharePoint Search and
Elastic Search (and others)

They helps users and employees to find public information (e.g., files, sites, news, and images) within a company.



Type of Search Engines

Depending on their application, there exist various types of search engines:

- Desktop Search Engines.
- Enterprise Search Engines.
- **Code Search Engines.**
- Mobile Search Engines.
- Academic Search Engines.
- Web Search Engines.

E.g., GitHub Code Search, source-graph, and OpenGrok (and others)

Used to search through source code repositories. They help developers find specific code snippets, functions, or files.

The image shows two side-by-side screenshots of search interfaces for source code. On the left is GitHub Code Search, displaying a search results page for 'rust-lang/rust' with a query 'or:rust-lang'. It shows a snippet of Rust code from the file 'src/lib.rs'. On the right is the OpenGrok interface, showing a search results page for 'full:pdf' with a relevance sort. It includes search filters for 'Definition', 'Symbol', 'File Path', 'History', and 'Type'.

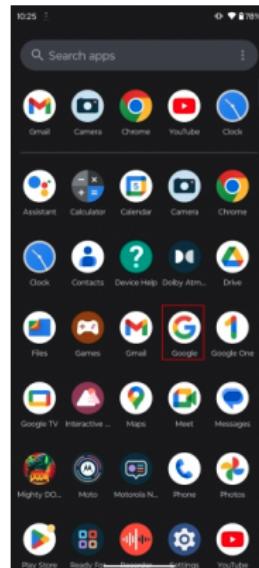
Type of Search Engines

Depending on their application, there exist various types of search engines:

- Desktop Search Engines.
- Enterprise Search Engines.
- Code Search Engines.
- Mobile Search Engines.
- Academic Search Engines.
- Web Search Engines.

E.g., Built-in search tools in Android, Windows mobile, and iOS (and others).

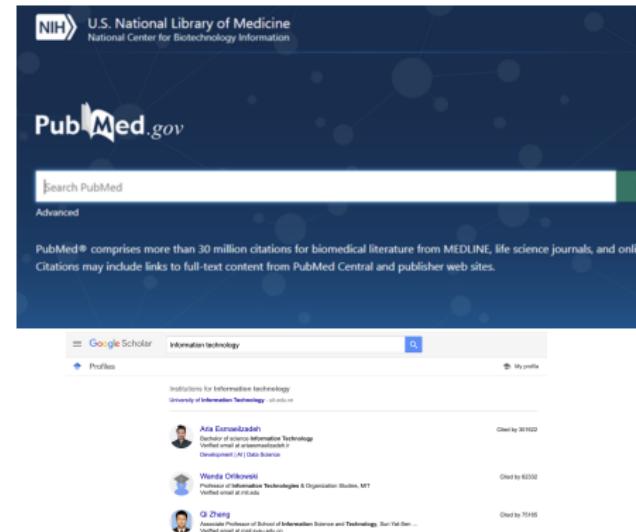
They provide search functionalities for apps, contacts, and local content on smartphones.



Type of Search Engines

Depending on their application, there exist various types of search engines:

- Desktop Search Engines.
- Enterprise Search Engines.
- Code Search Engines.
- Mobile Search Engines.
- Academic Search Engines.
- Web Search Engines.



E.g., Google Scholar and PubMed (and others)

Used to find scholarly content, including research papers, articles, and publications, about people in academia and research.

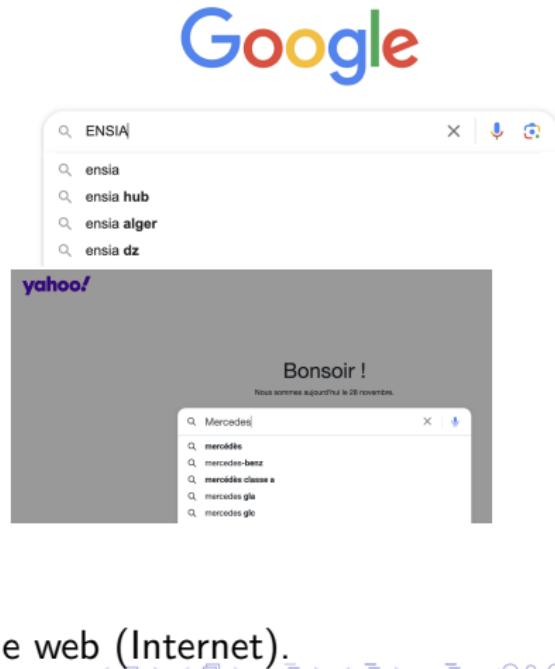
Type of Search Engines

Depending on their application, there exist various types of search engines:

- Desktop Search Engines.
- Enterprise Search Engines.
- Code Search Engines.
- Mobile Search Engines.
- Academic Search Engines.
- Web Search Engines.

E.g., Goolge.com, Yahoo, Yandex, etc

They are used to find relevant content
(websites, images, videos, files, etc) over the web (Internet).



Web

?

Web (WWW).

Definition

Web. A.k.a., WWW (World Wide Web), is a networked system of interlinked hypertext documents and multimedia content that is publicly accessible by users worldwide using web-browsers.

In the web:

- There are trillions of **webpages**.
- A **webpage** is a document, typically written in HTML (Hypertext Markup Language), and that is interpreted by a **web browser**.
- It contains HTML source code, text, and **multimedia content** (i.e., references to images, videos, audio, animation, interactive maps, WebRTC, 3D Models, etc).



الشبكة العنكبوتية العالمية

Web (WWW).

Definition

Web. A.k.a., WWW (World Wide Web), is a networked system of interlinked hypertext documents and multimedia content that is publicly accessible by users worldwide using web-browsers.

In the web:

- When a collection of **webpages** are linked together, they constitute a **website** (webpages in a folder).
- **Webpages** are linked together using **hyperlinks**. A **hyperlink** could be a text, an image, or any type of **clickable object**, that takes you from one webpage to another one (or same page), once clicked.



الشبكة العنكبوتية العالمية

Web (WWW).

Definition

Web. A.k.a., WWW (World Wide Web), is a networked system of interlinked hypertext documents and multimedia content that is publicly accessible by users worldwide using web-browsers.

In the web:

- A website is identified by its **URL** (Uniform Resource Locator), which is its address. E.g., <https://www.ensia.edu.dz> is the URL of the school's website.
- When you type a URL on your browser, your computer request a DNS server to resolve the URL into an IPv4 address, so that communication occurs (webpages retrieval).



الشبكة العنكبوتية العالمية

Web (WWW).

Definition

Web. A.k.a., WWW (World Wide Web), is a networked system of interlinked hypertext documents and multimedia content that is publicly accessible by users worldwide using web-browsers.

In the web:

- **Websites** are hosted in a computer, called **webserver**.
- Modern **webpages** do not only contain HTML source code, but also code written using other web-programming languages, such as **CSS, javascript, and php**.
- HTML & CSS r used 2 create **static webpages**. Javascript & php r used for **dynamic webpages**.



الشبكة العنكبوتية العالمية

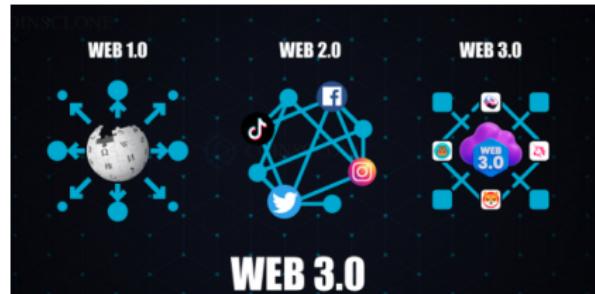
Web (WWW).

Definition

Web. A.k.a., WWW (World Wide Web), is a networked system of interlinked hypertext documents and multimedia content that is publicly accessible by users worldwide using web-browsers.

In the web:

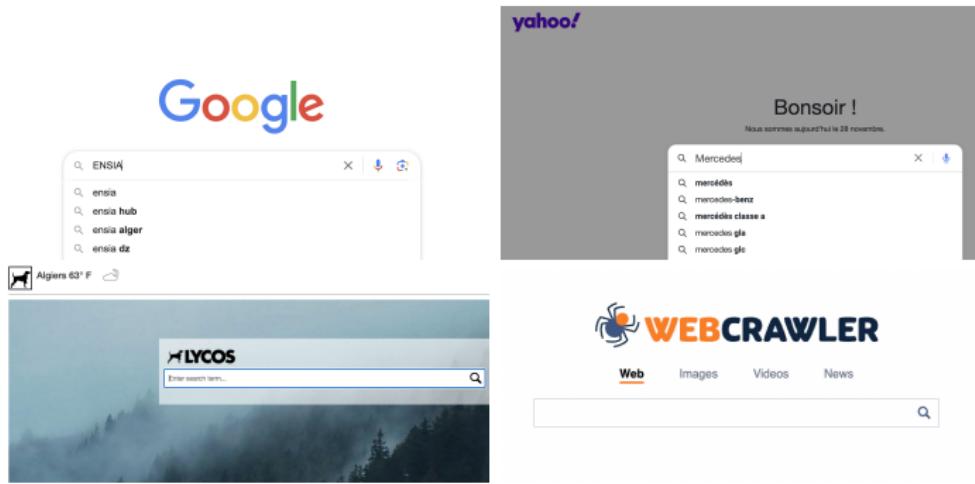
- **Websites** have undergone various changes and improvements:
 - **Web 1.0.** Basic informational webpages & limited interactivity.
 - **Web 2.0.** More dynamic and interactive web content.
 - **Web 3.0.** Incorporating artificial intelligence (AI, VR, AR).
- Note that the terms Internet and the World Wide Web cannot be used interchangeably.



Web Search Engines

Definition

Web Search Engine. Is a specialized software system, presented as an online service, and designed to search and retrieve information from the web (www) following users' requests through their web-browser.



Web Search Engines

Definition

Web Search Engine. Is a specialized software system, presented as an online service, and designed to search and retrieve information from the web (www) following users' requests through their web-browser.

There are many web search engines out there:

- **Open-source web search engines.** E.g., DataparkSearch Engine (2000), Sphinx Search (2001), Xapian (2002), Nutch (2002), Apache Solr (2004), YaCy (2004), Elasticsearch (2010), and OpenSearch (2021).

and more ...

- **Closed-source web search engines.** E.g., Yahoo Search (1995), Google search engine (1997), Yandex (1997), Microsoft Bing (2009), Baidu (2000), and DuckDuckGo (2008), . . . , Yooz (2015), preasearch (2017), Brave (2021), and You (2021).

and more ...

Web Search Engines

Definition

Web Search Engine. Is a specialized software system, presented as an online service, and designed to search and retrieve information from the web (www) following users' requests through their web-browser.

Very early web search engines: Archie (1990 — indexing and searching FTP archives), W3Catalog (1993 — indexed webpages), Aliweb (1993), JumpStation (1993), WWW Worm (1993), WebCrawler (1994 — full-text web search engines), and Lycos (1994). Some of them are still working:

- WebCrawler (www.webcrawler.com),
- Lycos (www.lycos.com),
- MetaCrawler (www.metacrawler.com),
- Dogpile (www.dogpile.com),
- Yahoo search (www.search.yahoo.com)
- Yandex (www.yandex.com)
- And others...

Web Search Engines

Definition

Web Search Engine. Is a specialized software system, presented as an online service, and designed to search and retrieve information from the web (www) following users' requests through their web-browser.



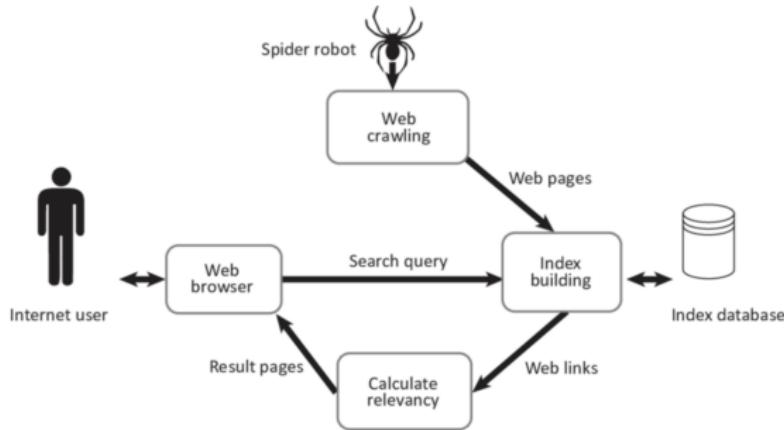
Certain web search engines do not crawl the web. They combine web search engines results and compile the result into one listing. This is the case of MetaCrawler and DogPile, for instance.

Web Search Engines

Definition

Web Search Engine. Is a specialized software system, presented as an online service, and designed to search and retrieve information from the web (www) following users' requests through their web-browser.

How do web search engines generally operate?



Web Search Engine Operation

Web search engines generally work in the same way, they all perform three basic tasks:

- ① **Crawling.** Search engines use programs (crawlers or spiders) to permanently crawl (scan and follow links) the web to discover webpages.
- ② **Indexing.** After crawling the web, search engines collect and categories information about the discovered webpages and put the result into a database, called index. The search engine uses this index as a source of information to construct the search engine result pages (SERPs).
- ③ **Ranking.** Search engines use some dedicated algorithms to decide which content is most useful to the searchers (Ranking), and hence will be placed on top of the results. E.g., the Google's ranking algorithm is called PageRank. It was named as such after its inventor Larry Page.

Web Cookies

Definition

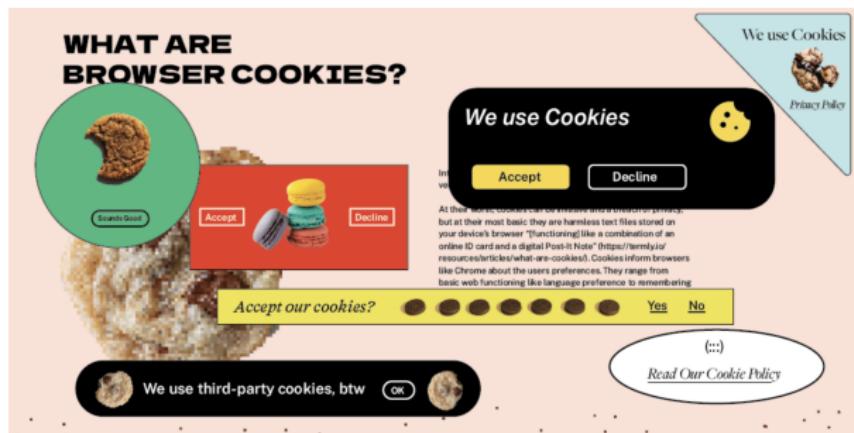
Web Cookie. Is a small block of data created by a webserver while a user is browsing a website and is sent and placed on the user's web-browser.

The web-browser includes the cookies whenever a subsequent request is made to the same website.

This allows the server to identify & recognize the user (session).

The web-cookie is placed in a Set-Cookie header within an HTTP message (name, expiry date, ...).

A.k.a., HTTP cookie, Internet cookie, browser cookies, or just cookie.



Web Cookies (Why do cookies exist?)

They exist because HTTP is a stateless protocol (i.e., forgets how you are).

Web cookies allow the following to be performed:

- Remain logged in within the services of a domain. E.g., within the google.com domain, you remain logged in www, mail, drive, docs, sites services, and even youtube.com (Terrible UX, otherwise).
- Tracking user behavior, collecting data for analytics, and delivering targeted advertisements. Some data is shared with others.
- Storing user preferences and settings (e.g., language, theme, etc).
- Remembering shopping cart content while browsing an EC site (orange/blue).
- Site analytics: time spent on website, click-paths, page views, ...
- Remembering user input in lengthy forms.

Web Cookies (Why do cookies exist?)

They exist because HTTP is a stateless protocol (i.e., forgets how you are).

Web cookies allow the following to be performed:

- Rema
googl
servic
 - Track
target
 - Storing user preferences and settings (e.g., language, theme, etc).
 - Remembering shopping cart content while browsing an EC site (OK).
 - Site analytics: time spent on website, click-paths, page views, ...
 - Remembering user input in lengthy forms.
- This website uses cookies
- We use cookies to personalise content and ads, to provide social media features and to analyse our traffic. We also share information about your use of our site with our partners who may combine it with other information that you've provided to them or that they've collected from your use of their services. Your preference can be changed by clicking 'Manage cookies'. By clicking 'Accept' you accept the use of all cookies as described in our [cookie-statement](#).
- [Manage cookies](#)
- [Accept](#)
- n the
sites
- re ring

- End.