

# سازمان سما

وابسته دانشگاه آزاد اسلامی

دانشگاه سما واحد حاجی آباد



# زبان تخصصی مهندسی کامپیوتر

---

منبع : زبان تخصصی دکتر حسن پور

---

حمیدرضا رضاپور

[WWW.HREZAPOUR.IR](http://WWW.HREZAPOUR.IR)

# زبان تخصصی مهندسی کامپیوتر

درس هشتم:

## Hunting Spiders through Search Engines

# Part I- Writing Development: Using Comparison

یکی از راه های توسعه پاراگراف استفاده از مفهوم مقایسه میباشد، یعنی مقایسه بین دو چیز یا دو جنبه مختلف از یک چیز. چند ساختار معمول از بکارگیری مفهوم مقایسه در یک جمله در زیر آمده است:

- **Adjective/preposition** (the same/as; similar/to; like)

## Examples:

A meta-search engine employs the search results of other search engines **the same way as** a searching tool.

A meta-search engine employs the search results of other search engines **similar to** a searching tool.

**Like** meta-search, a searching tool employs the search results of other search engines.

- **Attached statements** (and.....too; as so; and .....either; and neither)

**Examples:**

A meta-search engine employs the search results of other search engines **and** a searching tool does **too**.

A meta-search engine employs the search results of other search engines **and so** does a searching tool.

A meta-search engine does not employ the search results of other search engines **and** a searching tool does not **either**.

A meta-search engine does not employ the search results of other search engines **and neither** does a searching tool.

# Part II-Vocabulary

## **Hierarchical (adj)**

سلسله مراتبی

- any system of persons or things ranked one above another .

## **Venture (n)**

جرات، سرمایه گذاری

- A risky or daring journey or undertaking

## **Premise (n)**

فرض مقدم،

- A previous statement or proposition from which another is inferred or follows as a conclusion: *If the premise is true, then the conclusion must be true.*

## **Crawler (n)**

خزنده

- A thing that crawls or moves at a slow pace.

**Retrieve (v)**

بازیابی کردن

- Find or extract (information stored in a computer).

**Exclusion (n)**

استثناء، محرومیت

- The process or state of excluding or being excluded: drug users are subject to exclusion from the military.

**Proximity (n)**

نزدیکی، مجاورت

- Nearness in space, time, or relationship: do not operate microphones in close proximity to television sets.

**Customize (v)**

سفارشی کردن، بهینه کردن

- Modify (something) to suit a particular individual or task.

## **Part III- Reading**

### **FUNCTIONS OF A WEB SEARCH ENGINE**



# Section 1: Pre-reading Questions

- What is a web search engine?
- What are the features of a web search engine?
- Who uses web search engines and why does s/he use them?

## Section 2: Reading Passage

### 1. Definition and types of web search engines

A web search engine is defined as a software system that is designed to search for information on the World Wide Web. There are three main types of searching tools that have evolved. First, human beings have extensively programmed a system of predefined and **hierarchically** ordered keywords which is defined as web directory. A web directory does not display lists of web pages based on keywords; instead, it lists web sites by category and subcategory. An example of a well-known general web directory is DMOZ (<http://www.dmoz.org>). Second, a system generates an "inverted index" by analyzing the texts it locates on the network of a search engine. This type is a commercial **venture** supported by advertising **revenue** and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. For example, search engines, such as Google, that do not accept money for their search results make money by running search related **ads** along with the regular search engine results. The search engines make money every time someone clicks on one of these advertisements.

منافع، درآمد - آگهی

Finally, a **meta-search engine** is a searching tool that employs the search results of other search engines and **aggregates** the results into a single list or displays them according to their source. The meta-search engine enables users to enter search criteria once and access several search engines simultaneously. Meta-search engines operate on the **premise** that the Web is too large for any one-search engine to index it and more **comprehensive** search results can be obtained by combining the results from several search engines. This meta-search engine, for instance, may save the user from having to use multiple search engines separately. Examples of meta-search engine are Dogpile ( <http://dogpile.com>) and Yeppi (<http://yepi.com>).

فرا موتور جستجو - تجميع (جمع) کردن - بسیط، روان

The search results are generally presented on lines of results and often referred to as Search Engine Results Pages (SERPs). The information from search results may be used for specialists on web pages. Some search engines also **retrieve** available data from databases or open directories. Unlike web directories, which are maintained only by human editors, search engines maintain real-time information by running an algorithm on a web crawler. Web search engines work by storing information about many web pages, which are retrieved from the Hypertext Markup Language (HTML) pages. HTML is a standardized **annotation** system for **tagging** text files to achieve font, color, graphic, and hyperlink effects on World Wide Web displayed pages. A Web **crawler** as shown in Figure 1, also known as a spider, automatically follows every link on the site and retrieves the pages.

توضیح و تفسیر، توشیح، زیرنویس - بر چسب زدن، عنوان دادن

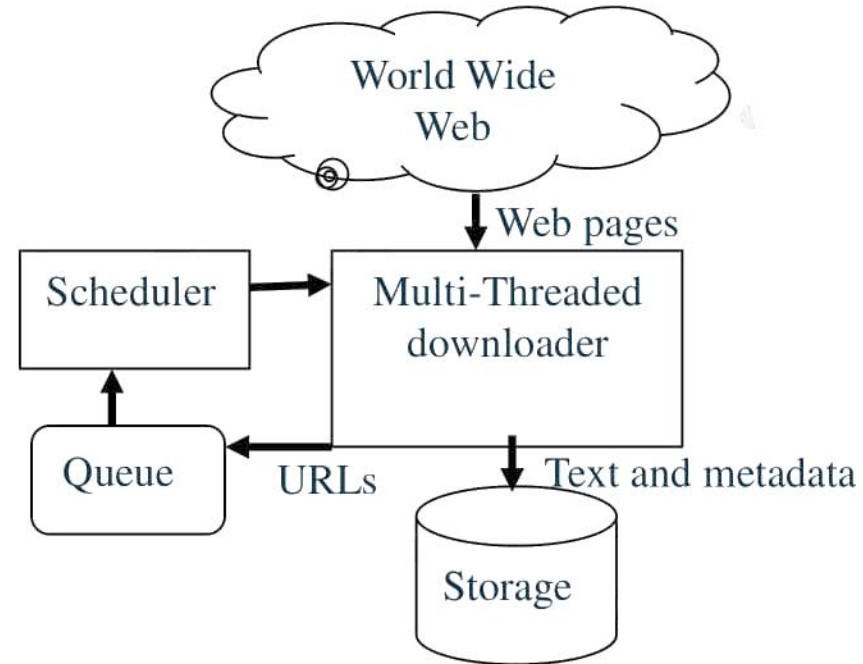


Figure 1: High-level architecture of a standard Web crawler.



The site owner can make **exclusions** by using robot.txt. The contents of each page are then analyzed to determine how it should be indexed (for example, words can be extracted from the titles, page content, headings, or special fields called meta tags). Data about web pages are stored in an index database for use in later queries. A query from a user can be a single or sequence of keywords. The index helps find information relating to the query as quickly as possible. Some search engines, such as Google, store all or part of the source page (referred to as a Cache) as well as information about the web pages, whereas others, such as AltaVista, store the keywords of pages they find. The cached page always holds the actual search text since it is the one that was actually indexed, so it can be very useful when the content of the current page has been updated and the search terms are no longer in it. Increased search **relevance** makes these cached pages very useful, but not just because they may contain data that may no longer be available elsewhere.

وابستگی، ارتباط

When a user enters a query into a search engine (typically by using keywords), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document title and sometimes parts of the text. The index is built from the information stored and the method by which the information is indexed. From 2007, Google.com search engine has allowed users to search using dates by clicking on *Tools* tab in the initial search results page, and then selecting the desired date ranges. Most search engines support the use of the Boolean operators AND, OR, and NOT to further specify the search query. Boolean operators are for **literal** searches that allow the user to **refine** and extend the terms of the search. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called **proximity** search, which allows users to define the distance between keywords in documents. In Google, for example, if you were searching for "Memory AND CPU" before, you could now use "Memory AROUND(1) CPU" instead. This will tell Google that you're looking for Memory and CPU to appear in close proximity to each other. If you want to extend the range a bit, increase the number (e.g. AROUND(2), AROUND(3), etc.). There are also concept-based searches where the search involves using statistical analysis on pages containing the words or phrases you search for. Also, natural language queries allow the user to type a question in the same way one would ask it to a human being. A site with this feature would be *ask.com*.

لغوی (دقیق) - پالایش یا تلخیص کردن

The usefulness of a search engine depends on the relevance of the results set it gives back. For instance, there may be millions of web pages that include a particular word or phrase, but some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to list the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be presented, varies widely from one engine to another. The methods also change over time as Internet usage changes and new techniques evolve.



## *2. Search engine bias*

Although search engines are programmed to rank websites based on their popularity and relevancy, **empirical** studies indicate various political, economic, and social biases in the information they provide. These biases can be a direct result of economic and commercial processes (e.g., companies that advertise with a search engine can also become more popular in its organic search results), and political processes (such as the removal of search results to **comply** with local laws).

Biases can also be the result of social processes, as search engine algorithms are frequently designed to exclude **non-normative** viewpoints **in favor of** more "popular" results. Indexing algorithms of major search engines **skew** towards the coverage of US-based sites, rather than websites from non-US-based sites.

تجربی - پیروی کردن - نا هنجار - موافق، جهت حمایت از - تمایل (انحراف) داشتن

### ***3. Customized results and filter bubbles***

Many search engines, such as Google and Bing provide **customized** results based on the user's activity history. This leads to an effect that has been called a filter bubble. The term describes a **phenomenon** in which websites use algorithms to selectively guess what information a user would like to see, based on information about the user (such as location, past click behavior and search history). As a result, websites tend to show only information that agrees with the user's past viewpoint, effectively isolating the user in a bubble that tends to exclude contrary information. Prime examples are Google's personalized search results and Facebook's personalized news stream. According to Eli Pariser, who also coined the term Pariser, users get less **exposed** to conflicting viewpoints and are isolated intellectually in their own informational bubble. Pariser **related** an example in which one user searched Google for "BP" and got investment news about British Petroleum while another searcher got information about the Deepwater Horizon oil spill and that the two search results pages were "**strikingly different**". The bubble effect may have negative implications for **civic discourse**, according to Pariser. Since this problem has been identified, competing search engines have emerged and seek to avoid this problem by not tracking or "bubbling" users.

پدیده - در معرض قرار گرفتن - نقل کردن - اختلاف خیلی واضح - گفتگوی مدنی، مباحثه اجتماعی



## Part IV- Reading comprehension

**Mark each statement as T (True), F (False), or NG (Not Given) to the information in the reading comprehension passage.**

1. Three kinds of search engines are discussed in the passage.
2. Meta-search engine helps users to have access to several search engines at the same time.
3. Some search engines receive data from an unavailable database.
4. HTML is a writing style used by some researchers.
5. The cached page keeps information as it is normally indexed.
6. Search results in a search engine are listed based on the amount of text.
7. Search engines are programmed based on only political biases.
8. The bubble effect has a positive effect on civic discourse.