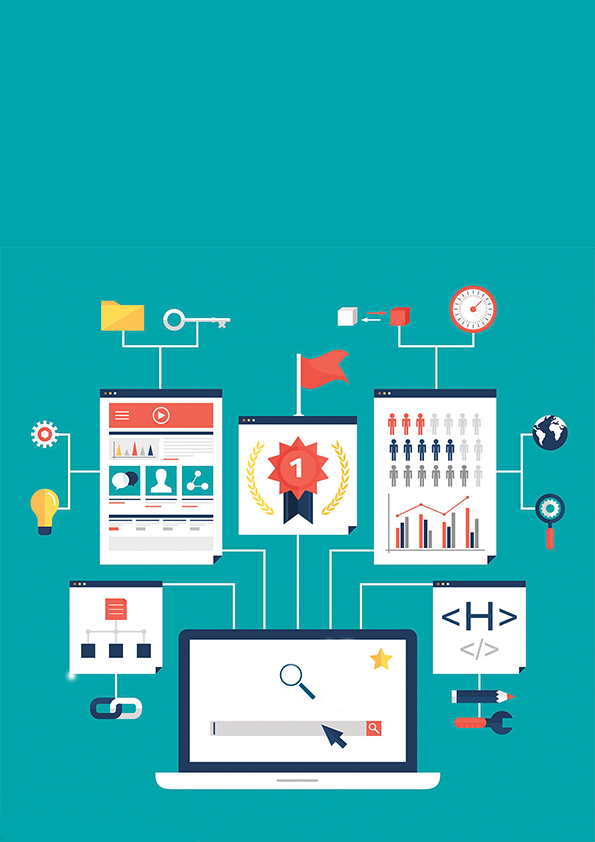
****

CS3343-

**Introduction**

Search Engine is a software system designed to search for information on the World Wide Web. In the century of information, people’s demand for information acquisition is getting higher and higher. Various search engines have emerged in recent years, based on the principle it seems extremely available to build a small size search engine. There is no need to connect the Internet all the time, and the dataset provided by ourselves is more inclined to users own orientations.

The process of building search engine software can be roughly divided into two parts, preprocessing and querying. In the part of preprocessing, firstly obtaining an original webpages dataset in a certain formal saved locally. However, the pages information is not convenient to access randomly, so recording the index of webpage in the dataset by extracting keywords of the page. In order to inquire easier, reversing the index obtaining a hash map from keywords to URLs. Furthermore, the result of the obtained URLs should rank in a logical order, the webpage tends be more in line with users requirements is ranked in front based on TF-TDF.

Considering the input users given freely, tokenizer is also included in the software.

Commonly saying, user input one or several words, a list of corresponding webpages’ URLs is returned. Then user can obtain the information from the obtained URLs.

**Use Cases and Requirements Specifications**

This Search Engine Software tends to design for web engineer to implement site search and users have demand for find information based on their own dataset.

First, the program will ask users whether to add new webpages or not, if yes, input the webpage information.

Then, users will be asked to input keywords, the number of keywords is unlimited, also there is no case sensitive.

Finally, the relevant URLs will be output in a correlation priority order. Users can look up the webpage from the given list to obtain the requested information.