Project #2 – KOOGLE Search Engine

Story:

The project is to create a prioritized collection of web site information that can be searched by keywords input by the user. The goal is for a user to enter search terms (words) and have a prioritized list of matching sites (name and URL) be displayed to him or her. Thus the project is divided into two main compartments: Database building & Queries' responses.

The database compartment is where the program creates and arrange its database of websites. To do that, it gets as an input a text file including all the names of the websites that should exist in the database. Then it goes through all the files of websites to build a collection of them including all the information they contain (based on description ~keywords~ & priority.) Every website will have a priority rating associated with it: low, medium, and high that would reflect how trusted the website is. Websites with a higher priority would be displayed first in search results.

After the database has been built, the program moves to the second phase (compartment), where it mainly gets queries from the user in a specific form (combination of <Keywords> and <Linkers> such that there's a <Linker> between every two <Keywords>). The user can add negation sign to the keywords for terms that will be excluded in the search. The program then would display the top 5 websites that match the search query. The results are ordered by the following (in priority order):

1. Number of unique words that match the file/URL

2. Priority of website IV.

3. alphabetical (by name of the website)

The purpose of this program is to answer with the help of the program are related to the complexity of the searches and their time consumptions. It would try to account for several common search scenarios and calculate their time consumption.