Search Engine User's Manual

Regular Operation:

Our search engine in regular operation mode features four possible options the user may select, which encompass search, parsing, and exiting the search engine while also allowing the user to get into advanced operation mode. Additionally, had we been able to completely implement the Hash Table Data Structure in this project, the user would have been able to select upon program startup, to use either an AVL Tree or a Hash Table implementation in which to store the data.

Advanced Operation:

In advanced operation mode, the following functionality is also accessible, however, due to time constraints, not all of it has been realized. Advanced operation mode adds to the regular operations menu the option to switch between Hash Table and AVL Tree Data Structures, a feature which we failed to implement, it also gives the option to enter into stress test mode, a feature which we felt was less important than meeting the given time constraints, finally, we also have the option to clear out the index of documents, which completely clears all documents which the search engine can iterate over.

Search Mode:

Note: In order to use this mode, you must first parse some XML documents.

Search mode presents the user some instructions on how to format searches, i.e. to use AND and OR as keywords to specify either a search of the intersection of the corpus of documents in which the keywords exist or the union of the documents in which the keywords exist, respectively. It also allows the user to exclude certain keywords so as to remove any documents in which these keywords appear from the returned results. The user is also asked to make all keywords lowercase. The following outputs are possible: If a certain search term does not exist in a compound search, the search engine will return output saying so, and will continue to search. The final output is then either a list of documents in which the search terms appear, with the relevant rankings, or a statement of "No Results Found." In the latter case, the user will need to press enter to return to the menu, in the former case, the user will then be asked to select a document and then the document will be displayed to the screen. (In the case that the document is not properly stored on disk a message of "Unable to open document will be displayed.) The user will then be returned to the menu.

Parse Document:

Note: Undefined behavior may occur if subdirectories to be parsed contain non XML documents.

In order to use the search engine, you must first parse some XML documents so that the search engine can know of their existence. Firstly, you must have a directory full of (many or just a single) XML documents in the directory containing the executable. Upon selecting parse document, the parser will ask the user to select either the default directory "WikiDump" or another directory (either should exist in the directory in which the executable resides). The parser will then index and split the XML documents into useable forms. Once complete the parser will ask whether you want to parse more directories, if so, select "y" and the process will restart.

Toggle Advanced Menu:

Selecting this option will switch to advanced operation mode.

Toggle AVL/Hash Table Mode:

Would switch between AVL and Hash Table Implementations had the Hash Table Implementation been completed.

Stress Test Mode:

Not implemented due to time constraints.

Clear All Documents:

Clears all documents from the index upon completing proper prompts.

Exit:

Exits search engine upon completing proper prompts.