**Search Engine User Manual:**

**Interactive Mode:**

Purpose: Allows to enter search queries with a combination of AND/OR, NOT, USERNAME, DATEGT, DATELT as defined in the project handout and then to select from a list of maximum 15 pages from the current Index that contain all the specified words and pass the required filters (will return 0 pages if no page in the Index contains a page that meets all criteria):

Set up before search:

1. All XML files must be in local folder
2. Index and pageRange files (used for returning page text form XML files) must be in local folder

Rules for User Query

1. Must begin Query with AND or OR (even if doing 1 word query) as first keyword
2. Directly after the first keyword enter pages to be ANDed or ORed
3. Next can enter other keywords (NOT, USERNAME, DATELT, DATEGT) followed by the arguments with the following rules
4. NOT: arguments (word) after specify to remove pages from list if they contain any of the words in the argument
5. USERNAME: specify exact username to only return pages with that username
6. DATELT: Enter date as YYYY-MM-DD (will only apply filter for the first argument entered)
7. DATEGT: Enter date as YYYY-MM-DD (will only apply filter for the first argument entered)

Running Interactive Mode:

1. Put “Interactive” as command line argument and run program

2. Wil be prompted for data structure you wish to load stored off index int

a. Enter 0 for HashTable

b. Enter 1 for AVL tree

- After which it will display the file from which the index is being loaded from

3. Enter the desired search query when prompted (following the user query rules)

4. The program will display a maximum of 15 pages the meet requirements of search

- If no pages satisfied criteria, program will output no pages could be found

.5. Select the ranking (1 – 15) of the page that you want to view. The page will then be printed out on the screen

6. Prompt to continue viewing pages from last search or go to new search

a. Enter 1 to continue viewing

b. Enter 0 ( any other character) to go to new search

7. Continue entering search queries and following same process or enter “END” to quit program

**Maintenance Mode**

Purpose: Maintenance mode allows the user to add files into the index in hard memory, or delete it from memory.

Set-up before start:

1. XML Dump files, Index, and Page range documents must be stored in local folder.

Upon starting maintenance mode, the index will be loaded from hard memory into RAM.  The user will then be prompted to enter a command.  The options are:

1. Clear Index
2. Add New File

Option 1 will clear the index in RAM.  Option 2 will prompt the user to enter the name of the file he/she wants to index.  Files are added one at a time.

The user can then press “1” to continue or “0” to exit maintenance mode.  Upon exiting the program, the updated index will be written out to hard memory, as well as the new page range.

Rules:

1. The format of the new file should be in the form “WikiDumpPart180.xml”.
2. The page ranges in the new file must not match existing ones
3. The first new file should always be number 180, and successive files should be one higher.

This format ensures correct retrieval of pages from the new files being indexed.

**Stress Test Mode:**

Purpose: To automatic interactive and maintenance mode with a text file as the script

Rules before Running Stress Test:

1. All rules for Maintenance and Interactive
2. Stress test text file used must be in current folder

Commands:

|  |  |  |
| --- | --- | --- |
| Keyword | Description | Arguments Number and Type |
| CI (Clear Index) | Clears the contents of current index loaded in the interactive mode that is being searched | None |
| BI (Build Index from Memory) | Loads the index stored off in a file in the folder into index able to be searched as data structure specified | Number specifying type of Data Structure:  1 = Avl Tree  0 = Hash Table |
| IF (Index New File) | Adds specified file to current index | Name of file to add |
| SE (search) | Specifies search specified by user command (same as input for search in interactive) | Search criteria (following rules of Interactive Mode) of variable length |
| NS (new search) | Clears all data members of Interactive Mode before a new search is performed | None |
| PP (Pick Page) | Selects Page to print out of the last search performed | Number of page rank to print out |
| ST (start time) | Stores of current time as start time of procedure | None |
| ET (end time) | Stores current time as stop time of procedure | None |
| PT (print time) | Outputs (cout and to file) the time difference between the stored stop and start time | None |

Rules for Order:

1. BI must be specified before IF, SE, PP
2. PT can only occur after at least one occurance of ST and ET
3. Before every new search use NS command

Procedure:

1. Create Stress test file according to rules and save in local directory
2. Run program using StressTest as first command line argument
3. When prompted enter name of stress test mode script
4. All output “Print times” will be stored in StressTestOutput.txt in local director