

Kwirē
Search
Engine

User Manual

This document is the User Manual for the KWIRE SEARCH ENGINE; provides instructions for using program. The project is part of the Data Structures course (CSE 2341) at SMU.
December, 2014 Edition, Ver1.0

Team ZombieCoders:
Aurora Havens
Emely Villeda-Principe

Contents

1. Introduction	2
2. Maintenance Mode	3
2.1. Adding Documents	3
2.2. Clearing Index	4
2.3. Error Messages	4
3. Interactive Mode	5
3.1. Loading Index	5
3.2. Searching	6
3.3. Error Messages	8
4. Stress Test Mode	9
4.1. Supplying Commands	9
4.2. Error Messages	10

Introduction

Kwirē Search Engine

Project Summary:

The Kwirē Search Engine is designed to allow users to quickly create a persistent index using data from Wiki documents in order to be able to later locate data efficiently. Users can make search queries that contain multiple conditions, including certain terms but excluding others. The results of each search will be ranked by relevancy, which will help users identify which documents are most likely to satisfy their searching needs.

Program Functionality:

- Add documents to an index by supplying a path to the files
- Clear both the index file on disk and data structure
- Load index data into one of two data structures
- Enter a search query to find documents containing the query
- Supply a command file to test program commands

Maintenance Mode

In Maintenance Mode, users can create an index file for later use in Interactive Mode. The file will be written on disk and persistent – even when the program ends, the file will continue to be available for use. Maintenance Mode includes two main commands, ‘add’ and ‘clear’.

2.1 Adding Documents

Users can add documents to the index by supplying a path to the file(s) containing the properly marked-up XML documents. There are two ways to add to an index – building on an already existing file or creating a new file.

```
***** Would you like to add to an already existing index file or create a new index? *****
      'old' - Add to existing index file
      'new' - Add to new index file
:: █
```

Adding to an Already Existing Index

Adding to a previously created index allow users to include new documents without having to start all over and add the previous documents again. By entering the ‘old’ command, users choose to open the already existing index on disk and append data to the file.

Users will then be prompted to supply the path to the new files and the type of data structure they wish to utilize. Once the new info is added to the file, the user can choose to continue in Maintenance Mode or return to the main menu.

Creating a New Index

Creating a new index involves deleting the previously created index on file and, if applicable, the data structure created using that data. This option allows users to remove data from the file if it is no longer needed/relevant and start over with the new files from the supplied path.

Once the info is written to the file, the user can choose to continue in Maintenance Mode or return to the main menu.

2.2 Clearing Index

This option allows users to clear the data contained in the index file and, if applicable, data structure without having to add new data.

```
***** Entering Maintenance Mode...input one of the following commands *****
    'add' - Add document(s) to index
    'clear' - Clear index file/structure
    'return' - Return to mode selection
:: █
```

Note: Clearing the index deletes the index file on disk and empties the index data structure (if it exists). The user must parse new documents in order to utilize search engine functions.

2.3 Error Messages

The following is a list of potentials that may arise when using Maintenance Mode, along with an explanation/solution for each.

- **!!ERROR: Incorrect command input**
This error occurs when the user doesn't input one of the two main commands for Maintenance Mode. To resolve, the user must input 'add' or 'clear' to utilized MM or input 'return' to exit the mode.
- **!!ERROR: Could not open ... - Supply a different path**
This error occurs when the user supplies a non-existing path for the Wiki XML documents. To resolve, the user must supply an existing path.
- **!!ERROR: Incorrect data structure request**
This error occurs when the user inputs a command to create a data structure other than an AVL Tree or Hash Table. To resolve, the user must input 'avl' or 'hash'.
- **!!ERROR: Could not open index file on disk**
This error occurs when the index file is being written in but either doesn't exist or is corrupted. To resolve, the user may have to manually delete and create an empty index file.

Interactive Mode

In Interactive Mode, users load data from an existing index file on disk to quickly and locate documents. The data structure can be recreated whenever the user changes the index file in Maintenance Mode. Interactive Mode includes two main commands, 'load' and 'search'.

3.1 Loading Index

Users can load data from an existing index file on disk into a data structures for use in searching. There are two data structure types available for the user to load into – AVL Tree or Hash Table.

```
**** Storing to data structure...input which of the following structures you would like to use ****
    'avl' - Load index to AVL Tree
    'hash' - Load index to Hash Table
:: █
```

Loading to AVL Tree

An AVL Tree is a self-balancing binary tree type data structure. Users may choose to use this structure to store the index data.

Note: The AVL Tree currently implemented in the program has found to be slow and less efficient at storing large sets of data. (See our “Data Structures – Case Study” document for more details)

Loading to Hash Table

A Hash Table is a data structure used to implement an associative array, a structure that can map keys to data. Users may choose to use this structure to store the index data.

Note: This is the recommended data structure to use, as tests using different cases have shown that it performs better than the AVL Tree. (See our “Data Structures Report” document for more details)

3.2 Searching

This option allows users to search for documents containing certain terms by entering a properly formatted Boolean query. There are three Boolean operators that may be used in a query search – ‘AND’, ‘OR’, and ‘NOT’.

```
***** Enter your query - You may use the following boolean expressions in your search *****
'AND' - (Prefix) Searches must include both terms
'OR' - (Prefix) Searches may include one or both terms
'NOT' - Searches must exclude the term
:: █
```

‘AND’ operator

The ‘AND’ operator is a prefix that sets the condition that all terms trailing the operator must be included in a document.

Example: AND programming computer

This query should return all documents that contain both the words programming **and** computer.

‘OR’ operator

The ‘OR’ operator is a prefix that sets the condition that at least one of the terms trailing the operator must be included in a document.

Example: OR programming computer

This query should return all documents that contain either programming **or** computer **or** both.

‘NOT’ operator

The ‘NOT’ operator sets the condition that all the terms trailing the operator must be excluded from the results.

Example: programming NOT computer

This query should return all documents that contain programming **but not** computer.

Note: ‘NOT’ cannot be used as a prefix, but it can be combined with either of the prefix operators to include multiple terms. Single word queries that don’t use any of the operators are allowed as well.

After a list of matching documents is returned, users have the option of viewing any of the documents by inputting the document number. To finish the search session, the user must input a '-1'.

Example of Viewing a Document:

```
***** Input the number of the document you wish to view or input '-1' to end search session *****
:: 1

Document 1, 'Main Page'
=====
<div style="width:100%; margin:0px; padding:0px 5px 5px 5px;">
<div style="clear:both; text-align:center; font-size:95%; margin:0em 1em 0em 1em;">
<div id="mf-overview">[[Wikibooks:Welcome|Overview]] *</div>
[[Using Wikibooks]] *
[[Wikibooks:Reading room|Questions]] *
[[Help:Contents|Help]] *
[[Wikibooks:Card Catalog Office|Browse]]
</div>
<div style="width:50%; float:right; font-size:95%;">
<div style="float:right; width:30%;">
* [[Subject:Science|Science]]
* [[Subject:Social sciences|Social sciences]]
* ''[[Subject:Books by subject|All subjects]]''
</div>
<div style="float:right; width:30%;">
* [[Subject:Languages|Languages]]
* [[Subject:Mathematics|Mathematics]]
* [[Subject:Miscellaneous|Miscellaneous]]
</div>
<div style="float:right; width:30%;">
* [[Subject:Computing|Computing]]
* [[Subject:Engineering|Engineering]]
* [[Subject:Humanities|Humanities]]
</div>
</div>
<div style="width:40%; text-align:center;">
<div style="font-size:162%; padding:0.1em;">[[Wikibooks:Welcome|Welcome]] to [[Wikibooks:What is Wikibooks|Wikibooks]],</div>
<div style="font-size:95%; margin-top:0.2em;">the open-content textbooks collection that [[Help:Contributing|anyone can edit]].</div>
<div id="pagecount" style="font-size:85%;">[[Subject:Books by subject|{{formatnum:{{NUMBEROFBOOKS}}} books]] with [[Special:Allpages|{{NUMBEROFARTICLES}} pages]]</div>
</div>
<div style="clear:both; margin:25px 0px 0px 0px; padding:0px;">
{ id="mp-content" style="width:100%; border-spacing:10px 0px;"
| style="width:33%; padding:10px; vertical-align:top; background-color:#dddfdf;" | {{Main Page/Featured}}
| style="width:33%; padding:10px; vertical-align:top; background-color:#dddfdf;" | {{Main Page/Wikijunior}}
| style="width:33%; padding:10px; vertical-align:top; background-color:#dddfdf;" | {{Main Page/Recipe}}
}
</div>
</div>

{{:Main Page/Sisters}}

[[Category:Main page| ]]
=====
```

3.3 Error Messages

The following is a list of potentials that may arise when using Interactive Mode, along with an explanation/solution for each.

- **!!ERROR: Incorrect command input**

This error occurs when the user doesn't input one of the two main commands for Interactive Mode. To resolve, the user must input 'load' or 'search' to utilized IM or input 'return' to exit the mode.

- **!!ERROR: Incorrect data structure request**

This error occurs when the user inputs a command to create a data structure other than an AVL Tree or Hash Table. To resolve, the user must input 'avl' or 'hash'.

- **!!ERROR: Could not open index file on disk**

This error occurs when the index file is being read from but either doesn't exist or is corrupted. To resolve, the user must return to Maintenance Mode and create a new index file. In the worst-case scenario, the user may have to manually delete and create an empty index file.

- **!!ERROR: Unable to complete request - Must load data structure before you can search**

This error occurs when the user attempts to search through the index data without having first loaded it into one of the two available index structures. To resolve, the user must first load the data into either one of the data structures in Interactive Mode.

- **!!ERROR: Requested pages number is not in the query list of document results**

This error occurs when the user attempts to view the contents of a document that wasn't on the query search results. To resolve, the user must either input a page number of one of the results or input '-1' to end the search session.

Stress Test Mode

In Stress Test Mode, users can supply a 'command file' containing a set of commands used in the other two search engine modes. The output of this mode will be the amount of time it takes to complete each of the commands in the file.

4.1 Supplying Commands

In order to run this mode, the user must enter a command text file that exists on disk.

The following is the list of commands in the default command file:

```
mmode
add
new
default
avl
return
imode
load
avl
search
main
return
mmode
clear
add
new
default
hash
return
imode
```

load
hash
search
main
return
exit

4.2 Error Messages

The following is a list of potentials that may arise when using Stress Test Mode, along with an explanation/solution for each.

- **!!ERROR: Could not read from commands file**
This error occurs when the command file is being read from but either doesn't exist or is corrupted. To resolve, the user must supply an existing file/path to the file. In the worst-case scenario, the user may have to manually delete and create an empty index file.