User Manual

CS 576

Avichal Chum

Import all files into the kwic project. The main file is kwic.java

Upon running it, either pass the input text file as a parameter of run it without any parameters and the system will ask you for an input file name. Enter the file name and press enter.

The program will index the file and display the index successful message. Press enter to display the results. (Note: The results will be displayed on the console and not in an output file).

Since there is no way to give the index word in **bold** in the console, the index word is always displayed in the middle with 30 characters preceding and following it. Also, to make the index word clear, I have chosen to display the index word before each context line, followed by the context line. Even though this wasn't required, I made it appear so that the index word could be interpreted clearly. All index words have been arranged alphabetically and all special characters and stop words removed. (Can be change in the string[] ignorelist). Instead of repeating index word multiple times, the program displays all found contexts of the index words with their corresponding line number/page number (which an actual KWIC program does).

The program stops execution automatically after displaying the results. The program stops index after the first *blank line It finds.* Therefore the input document cannot have blank documents. It may have special characters though. The program reads a maximum of `10000000 lines` (can be changed in the program). It also stores a maximum of 10 instances of the same index word (Again can be changed in the program: KwicSearch.java).

If the input document name is not found, It will ask the user for a name. If it still not found, it recursively keeps on asking the user for a valid document name until either the document is found or the user enters the 'exit' command.

The Emma report has been submitted as a separate archive 'Emma Report.zip' where as the FindBugs document has also been submitted as a separate document.

The Junit test files are: KwicTest.java, KwicSearchtest.java, VariableValuesTest.java and AllTests.java . The readme for these files are given in the Junit Readme.docx

Acknowledgments: Most of the program code has been developed by myself. Some logic based help was taken from various websites and the codes for multiple other algorithms was studied in order to build the final KWIC algorithm. The algorithm has been tested using many test cases and so far has passed all the test cases (except some that Junit could not cover). The code and resource files have also been made available on github at https://github.com/avichum/KWIC

Some of the resources referred are:

http://en.wikipedia.org/wiki/Key_Word_in_Context
http://www.tutorialspoint.com/junit/junit_suite_test.htm

https://github.com/franklingu/KWIC-System
http://www.vogella.com/tutorials/JUnit/article.html
http://wuyangnju.googlecode.com/svn/trunk/java/kwic/src/kwic/ms/KWIC.java
https://code.google.com/p/t2framework/wiki/JUnitQuickTutorial
https://sites.google.com/site/drriggsnewsite/classlist/cop4020proglangfall12/calendarplfall12/exam-1-review/kwic-in-python-c-c-java

Sample Run:

For wiki.txt :

```
Enter file name and press enter
wiki.txt

You entered: wiki.txt

 Starting indexing of file
Indexing complete. Press enter to display generated KWIC index.


acronym
1         :                   KWIC is an acronym for Key Word In Context, the
an
1         :                      KWIC is an acronym for Key Word In Conte
common
1         :Key Word In Context, the most common format for concordance lines
concordance
1         :t, the most common format for concordance lines the free encyclopedia
context
1         :is an acronym for Key Word In Context, the most common format for c
encyclopedia
2         :or concordance lines the free encyclopedia
format
1         :d In Context, the most common format for concordance lines the fre
free
2         :mat for concordance lines the free encyclopedia
in
1         :IC is an acronym for Key Word In Context, the most common form
is
1         :                         KWIC is an acronym for Key Word In Co
key
1         :       KWIC is an acronym for Key Word In Context, the most com
kwic
1         :                              KWIC is an acronym for Key Word In
lines
1         :common format for concordance lines the free encyclopedia
most
1         : for Key Word In Context, the most common format for concordance
```