**SE 311 – Software Architecture II**

Winter 2016

**Assignment #1**

The KWIC index system accepts an ordered set of lines, each line containing an ordered set of words, and each word consisting of an ordered set of characters. A KWIC index is formed by sorting and aligning the words within a line to allow each word (except the stop words) to be searchable alphabetically.

For example, given the sentence “KWIC is an acronym for Keyword In Context”, the index will contain the following entries:

**Initial Output**

**Key Value**

**KWIC** is an acronym for Keyword ...

KWIC is an **acronym** for Keyword In Context.

KWIC is an acronym for **Keyword** In Context.

... is an acronym for Keyword In   **Context**.

**After Sorting in Alphabetical Order**

**Key Value**

KWIC is an **acronym** for Keyword In Context.

... is an acronym for Keyword In   **Context**.

KWIC is an acronym for **Keyword** In Context.

**KWIC** is an acronym for Keyword ...

[Stop words are words which are filtered out before or after text processing. There is no single universal list of stop words; any group of words can be chosen as the stop words for a specific application.]

The KWIC problem was first used by David L. Parnas in his classic paper “*On the Criteria to be Used in Decomposing Systems into Modules*, Communications of the ACM, 15(12):1053-1058, December 1972”, to contrast different criteria for decomposing a system into modules.

Parnas proposed possible solutions for the KWIC index problem around four basic functions: *input*, *shift*, *alphabetize*, and *output*.

Since its introduction, the problem has received significant attention by the software engineering community and a lot has been published about it.

**Your assignment is to implement an object-oriented program that generates a KWIC index for a given set of input lines.**

Your program should be implemented either in Java or C++. It should read input lines from a file and write its output to a (different) file. It should also accept, as an option, a file containing stop words.

This is an individual assignment. You should submit your work **via Blackboard**, by Sunday, January 31, 2016, **11:59pm**.

There will be a 25% (absolute value) deduction for each week of lateness. Missing work will earn a zero grade.

Mr. Sheik Hassan (the TA for this course) will send out additional information regarding the submission logistics for this assignment.