

Exercise: Concordancing Java Source Code

This exercise will give you a chance to use the `java.util.Vector`, `java.util.Hashtable` and `java.io.StreamTokenizer` classes. You will also gain a little experience in using some of the other Java streams.

1. Background

Concordancing is an application traditionally used in Computer Science to demonstrate the power and flexibility of a programming language.

A concordance is simply a list of words contained in a text, together with a list of all the numbers of the lines on which the word occurs. For example:

And	1 5 9 12 13 17
Andrew	1 2 6
Ball	3
Chocolate	17 18 19 20

2. The Exercise

You should write a Java application that will create a concordance of the identifiers in a Java program source code file (specified as a parameter to the application).

Note:

The concordance should not contain:

- String contents.
- The contents of comments
- Numbers and the various pieces of punctuation used by Java (operators, etc.)

Optional:

- Print the contents of the concordance in alphabetical order
- Create a “stop list” of Java keywords and ensure that these are not added to the concordance