Java Application: SimilarWords

Written by: Keith Fenske, http://www.psc-consulting.ca/fenske/
First version: Thursday, 5 January 2006
Document revised: Saturday, 8 March 2008
Copyright © 2006 by Keith Fenske. Released under the GNU General Public License (GPL).

Description

SimilarWords is a graphical Java 1.4 Swing (GUI) application to help find unique words or identifiers, by avoiding words that are similar to each other. Two words are considered similar by this program if you can convert one word into the other by:

(1) Changing one letter: bat / cat;
(2) Adding one letter: bat / boat;
(3) Deleting one letter: frog / fog; or
(4) Reversing two consecutive letters: two / tow.

All of these changes are common typographical errors, and avoiding words that are too similar provides a degree of error checking. Not shown is a fifth common error: "off by one" where fingers are shifted one key in some direction, resulting in numbers like 1342 instead of 2453, or "hal" instead of "ibm" (a reference to the movie, 2001: A Space Odyssey). This is more of a problem with numeric keypads than with alphanumeric keyboards. Words in this program are defined to be any combination of letters or digits (that is, an alphanumeric string), and are converted to lowercase for processing. This range of words is what you might expect for identifiers such as login accounts, program module and variable names, or abbreviations.

For convenience, there are two lists of words: one of input words and another of "known" words. The known words can be loaded from a file, or added to from the input list. The intention is to maintain a list of words that have been previously checked, stored in a file, and to test new words against this list before they are added.

GNU General Public License (GPL)

SimilarWords is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General

Public License for more details. You should have received a copy of the GNU General Public License along with this program. If not, see the http://www.gnu.org/licenses/ web page.

Additional Comments

People make mistakes when they type numbers, words, almost anything. The most common mistakes are the four categories listed above. Most mistakes with letters don't result in proper words. Most mistakes with numbers do. Consider telephone numbers. If you change one digit or reverse two digits, you usually get a valid telephone number, which is a "wrong" number for you, but still someone answers and you have to excuse yourself. This happens because telephone numbers use almost all possible combinations of digits. It doesn't happen as often with words because there are many more combinations of letters than there are words, and words that we don't recognize become nonsense.

To avoid this problem with telephone numbers, don't assign every number. How many to reserve depends upon how careful you want to be. Hint: prime numbers are very helpful here, especially numbers that are relatively prime to ten. Assign only those telephone numbers that have a certain remainder when divided by a chosen prime number. The prime number 3 is too small, five won't work at all, seven is better, eleven beats seven, etc. The bigger the prime number, the more error checking you have. What's the smallest prime number that will catch each of the four categories of mistakes listed above (individually)? What's the smallest prime number that will catch all four categories (collectively)?

You can't use prime numbers on words. You can compare words to each other and choose only those words that can't be confused by the four categories listed above.

Installation

You must have the Java run-time environment (JRE) installed on your computer. SimilarWords was developed with Java 1.4 and should run on later versions. It may also run on earlier versions, but this has not been tested. For Macintosh computers, the version of Java is determined by your version of MacOS. For Windows, Linux, and Solaris, you can download the JRE from Sun Microsystems:

Sun Java
- JRE for end users: http://www.java.com/getjava/
- SDK for programmers: http://developers.sun.com/downloads/
- IDE for programmers: http://www.netbeans.org/

Once Java is installed, you need to put the program files for SimilarWords into a folder (directory) on your hard drive. The name of the folder and the location are your choice, except it is easier if the name does not include spaces. Assume that files will go into a C:\JAVA folder.

Then create the folder and unpack the Java *.class files into this folder (if you received the program as a ZIP file). The files look something like this:

```
GnuPublicLicense3.txt  (36 KB, legal notice)
RunJavaPrograms.pdf  (85 KB, more notes about running Java)
SimilarWords1.class  (14 KB, executable program)
SimilarWords1.doc  (33 KB, this documentation in Microsoft Word format)
SimilarWords1.gif  (15 KB, sample program image)
SimilarWords1.java  (39 KB, source code)
SimilarWords1.pdf  (72 KB, this documentation in Adobe Acrobat format)
SimilarWords1User.class  (1 KB, helper class for main program)
```

To run the program on Windows, start a DOS command prompt, which is Start button, Programs, Accessories, Command Prompt on Windows 2000/XP. Change to the folder with the program files and run the program with a "java" command:

```
c:
cd  \java
java  SimilarWords1
```

The program name "SimilarWords1" must appear exactly as shown; uppercase and lowercase letters are different in Java names. Should you find this program to be popular, you can create a Start menu item or desktop shortcut on Windows 2000/XP with a target of "java SimilarWords1" starting in the "c:\java" folder. One complication may arise when trying to run this program. Java looks for an environment variable called CLASSPATH. If it finds this variable, then that is a list of folders where it looks for *.class files. It won't look anywhere else, not even in the current directory, unless the path contains "." as one of the choices. The symptom is an error message that says:

```
Exception in thread "main" java.lang.NoClassDefFoundError: SimilarWords1
```

To find out if your system has a CLASSPATH variable defined, type the following command in a DOS window:

```
set  CLASSPATH
```

To temporarily change the CLASSPATH variable to the current directory, use the following command line:

```
java  -cp  .  SimilarWords1
```

To permanently change the CLASSPATH, you must find where it is being set. This may be in an old AUTOEXEC.* file in the root directory of your system disk (usually the C:\ folder), or it may be in Control Panel, System, Advanced, Environment Variables on Windows 2000/XP.

Removal or Uninstall

To remove this program from your computer, delete the installation files listed above. If the folder that contained the files is now empty, you may also delete the folder ... if you created the folder, of course, not the system. If you created desktop shortcuts or Start menu items, then delete those too. There are no configuration or preference files, and no information is stored in the Windows system registry. You don't need an "uninstall" program.

---

file: SimilarWords1.doc 2008-03-08