Instructions to run the indexing pipeline

Option 1 - Original project files
Assignment1 folder is the Eclipse project just as the program was developed, containing all original files.
Use Eclipse IDE, open the folder as a project (by adding it to your local eclipse workspace) and run the class Assignment1.java.

Option 2 - Standalone version
Standalone version folder contains the whole project exported as a runnable .jar. You will find Run.bat, which will automatically run the Assignment1Standalone.jar file.
The folder named “taggers” containing the trained POS tagger is needed by the .jar file to work.

Further instructions:
When prompted by the terminal: “Enter input website URL:”, a URL with the format *http://csee.essex.ac.uk/staff/udo/index.html* must be entered.
Error will occur if the user inputs something like: *csee.essex.ac.uk/staff/udo/index.html* or *www.csee.essex.ac.uk/staff/udo/index.html*.

Sample outputs folder contains outputs obtained for each step of the pipeline, as explained in the report, for the websites http://csee.essex.ac.uk/staff/udo/index.html and http://irsg.bcs.org/ksjaward.php