Expand the compressed downloaded file to a directory.

For Running on Eclipse:

1. Open Eclipse and specify the workspace as the directory where you expanded the downloaded content.

2. Make a new project in Eclipse with the name "Language Identifier". It will catch my project.

3. Now there are two source files (.Java) i.e. LanguageCorpus.java & LanguageIdentifier.java. They both can be run as independent applications. But first of all you need to run LanguageCorpus.java atleast once so that it can build the hashed corpus and save it on the disk (which I have saved with the file extension .model).

4. LanguageIdentifier.java actually identifies the language based on the statistical analysis of the corpus which is saved in the .model files. There we first need to build the .model files using LanguageCorpus.java. But, if you have run the LanguageCorpus.java atleast once to build these .model files, you can run the LanguageIdentifier.java directly.

5. To run LanguageCorpus.java, set the 6 arguments of the program. First five of them are the names of the corpus text files with the complete path in the following order of the languages, English, German, French, Italian, Dutch.
   And the sixth argument is the directory with complete path where you want to save the .model files (which will save the hashed corpus).
   e.g. F:\corpus\English.txt F:\corpus\German.txt F:\corpus\French.txt F:\corpus\Italian.txt F:\corpus\Dutch.txt F:\model

6. To run LanguageIdentifier.java, set the arguments of the program as the names of the test text files with complete path. You can any number of test files by the arguments. And last argument Is the directory with complete path where you have saved your .model files.
   e.g. F:\test\text_en.txt F:\test\text_de.txt F:\test\text_nl.txt F:\model

For Running on command line:

1. To run LanguageCorpus.java, with the 6 arguments of the program. First five of them are the names of the corpus text files with the complete path in the following order of the languages, English, German, French, Italian, Dutch. And the sixth argument is the directory with complete path where you want to save the .model files (which will save the hashed corpus).
   e.g. F:\corpus\English.txt F:\corpus\German.txt F:\corpus\French.txt F:\corpus\Italian.txt F:\corpus\Dutch.txt F:\model

2. To run LanguageIdentifier.java, with the arguments of the program as the names of the test text files with complete path. You can give any number of test files by the arguments. And last argument Is the directory with complete path where you have saved your .model files.
   e.g. F:\test\text_en.txt F:\test\text_de.txt F:\test\text_nl.txt F:\model