Errata and changes and updates:

1. grading: the speed of your application will be taken into account compared to other applications WRITTEN IN THE SAME OR SIMILAR LANGUAGES. scripts get compared to scripts, compiled get compared to compiled.

2. Correction will also be performed in a Windows system if that is the requirement for your script. You MUST state this requirement in your readme! otherwise I will default to linux.

4. Input for script 3 cannot be an inverted index as given in the assignment. As some students have noted, you need the frequencies of terms in documents to be able to calculate the TF.IDF score. The format will be the same however,

	D1	D2	D3
term1	1	1	2
term2	1	2	1
term3	2	4	1

etc, terms down the left, a header row of documents across the top, tab delimited, each line indicated by a carriage return /n (newline character) , not necessarily a /r/n as used in windows).

5. You can use libraries that come installed with the language you use. You cannot use any languages that require an additional download. Eg: I must be able to use you script WITHOUT an internet connection.

6. If necessary, I will make an exception for tartarus.org so the porter stemmer algorithm can be downloaded but, a better option would be for you to download it and include the source code in your own code, not as a separate executable.

7. This is an assignment for you to figure out. I am not going to correct your code as you go along! (so please stop sending me code to review for you). I will answer questions about information retrieval and about the assignment but NOT about how to program in a particular coding language.

8. the input files given are for example only so you can see the format. The actual files used for testing will be larger and more complex. Your code must be written to run separate scripts and I will use MY FILES and not the output of your code as input to each section. If my files don't work I will try it with your output but you will lose marks for not correctly completing the assignment. Remember this, especially if you are creating additional files to store between scripts.