Programming Task

Data File

The file test.txt is a space-separated Information Retrieval data file with some Query/Document data. Each row has features associated with a document retrieved in response to a query (these features have mainly been constructed from both document and query content, though some are purely query or document based).

Each row includes the following in order:

- A '0' or '1' which represents the judged relevance of the document to the query.
- The query ID (a string preceded by the text "qid:").
- A series of numeric feature values for a document returned in response to the query. Each feature value is prefixed with its feature number and then a ":".
- The document ID (a string preceded by the text "#docid = ")

All the documents retrieved in response to a particular query are in a contiguous chunk of rows – you can make this assumption when reading the file.

Your task

- Read the file into some classes defined by you, and designed in a way which will facilitate analysis.
- Do some analysis of the data – you can choose what analysis to do, but it should be non-trivial and should make good use of your classes. If your analysis includes some modelling, there is no requirement or expectation that you should use Infer.NET.
- Present the results of the analysis graphically in a manner of your choosing.

Your constraints

- You must use Visual Studio.
- You must use a high level language, though a .NET language (C#, F#, C++/CLI, IronPython, Visual Basic, etc.) is preferred.
- You can use low-level libraries of your choice.
- Your program must present a simple user interface which will allow loading and analysis of other files with the same format.
- You can ask questions by e-mail at any time – we will try to be responsive.
- You have a week from the date your receive this to complete.
- Feel free to use any reference resources (i.e. books, internet, etc.).

You will be judged on

- The quality of your software design.
- Your attention to detail.
- The depth of your analysis.