This lab forms a simple demonstration of the very topical issues of text and sentiment mining.

- Text mining is described by:
  
  http://searchbusinessanalytics.techtarget.com/definition/text-mining as:

  'the analysis of data contained in natural language text. The application of text mining techniques to solve business problems is called text analytics.'

  Text mining can help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Facebook, Twitter.'

- Sentiment, or opinion, analysis / mining defined by:
  
  http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining

  As ‘Opinion mining, which is also called sentiment analysis, involves building a system to collect and categorize opinions about a product. Automated opinion mining often uses machine learning, a type of artificial intelligence (AI), to mine text for sentiment.’

  ‘Opinion mining can be useful in several ways. It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features.'

Therefore, for this lab, we would like you to use the two data sources we asked you to syndicate dynamic content from in part (8) of lab 1, or two entirely new data sources, to record and visualise keyword frequencies, where this data may lead to potentially useful ‘business insights’ by end users in roles such as:

1. **Business Intelligence**: As a source of data for analysis where raw data can be used in data mining fact and decision tables.

2. **Tracking and Trending**: Where an organisation might place mouse-over adverts in web pages based upon popular keywords, or to track frequencies of keywords to determine market share.

We require you to choose three or more phrases, each of which includes specific keywords, i.e. names of people, places or events, from your two data sources from lab 1, and to monitor the phrases for a total of seven consecutive days within the duration of the assignment.

For example, you could use RSS or Twitter feeds concerning international news stories, and your phrases might be related to these, or perhaps dramatis personae in two data sources concerning the newly started football season, or the recent Scottish independence referendum et al.

If you are unable to decide which data sources and phrases you would like to use, please refer to Mark Levene or Martin O’Shea.

Just to remind you, lab 1 allowed you to use RSS feeds, Twitter, Google+, Facebook or other web services as data sources.

So at the end your seven day period, what do we want you to submit?

1. An explanation of why you chose your particular phrases and the data sources you used to derive them.
2. A simple visualisation(s) displaying the frequency of each phrase, and how they fluctuate over the seven day period of monitoring.

3. A statement of why you think your phrases are relevant and to whom.

4. What can you say about sentiment concerning your keywords, the items of data containing them, fluctuations in their frequencies, and the data sources you have used?

In order for you to do this, we have provided SentiSEWN which is available at:

http://qzone.dcs.bbk.ac.uk:8080/SentiSEWN/

SentiSEWN is a simple web application written for this assignment. **Before using SentiSEWN, you must read the application’s About page.**

The front page of SentiSEWN is illustrated below:

![SentiSEWN Front Page](image)

This page requires you to enter the text you want to analyse for sentiment, keywords you want the sentiment to relate to, together with a rating for your sentiment and also a brief description concerning your rating of the sentiment.

The results of the sentiment analysis will be displayed on a second page:
Allowing you to compare both sets of results.

For each piece of text analysed, we require you to list the text, keywords, your rating and sentiment, together with the outputs from SentiSEWN, and any further comments you may have comparing your sentiment to the automated sentiment.

If you encounter any difficulties with SentiSEWN or if it is unavailable, please advise Martin O’Shea immediately.

In such a case, please continue with Labsheet 2 by visiting SentiStrength and use it as per the following screen dumps:

1. Text for sentiment analysis should be entered in the area shown in the red box below and click Detect Sentiment Around Keywords:

2. The results of the sentiment analysis will be displayed on the top of the page shown below:
The results provided by SentiStrength on the second page above should then be recorded as described above, together with your own ratings and explanations of your text.

**What to hand in:** A short (at most four A4 pages) in report in .doc(x) or .pdf format containing the deliverables above, to be submitted via the Lab 2: Tracking and visualising keyword frequencies for sentiment analysis drop box in Moodle.

**Submission Deadline:** 23 10 2014.

**Late assignments:** No extensions are available as for this lab and any late submissions will be graded as per the guidelines of the relevant course being studied.