**Abstract**

**An analysis of the historical newspapers at the National Library**

Diana Silvia Teodorescu, dst1@aber.ac.uk

The project will be an analysis of the data collection provided from the historical newspapers found at the National Library. The newspapers are digitised and have a basic API that will be used to search, request and extract specific data. The point of the project is to discover interesting themes and concepts (research studies, advertised services, positive/negative) within the articles and using data mining processes and statistics to present them in an interesting way to the user.

I am considering a few options regarding the data processing that would be made available through a web application. I will look into connecting the data resulted through analysis in the newspapers with external data to attract interest into the past and the changes that occurred over the years.

Some of the data analysis that could be made are: Sentiment and Mood Analysis; domains (politics, sport, science); statistical representations through graphs for word frequency or subject frequency;

During this project I am looking forward to learn more about statistical techniques and ways to represent them into a user friendly way using charts and beautiful graphical representations.   
  
The parts of the project that most worry me are actually the ones I want to learn more about like: machine learning and statistical techniques. I feel that at the moment I don't have enough knowledge about these areas and I am eager to do some research and learn more about what are the best ways to implement statistics and machine learning into code.