**”PyTLDR1 is an open-source Python module used for ex- tractive document summarization. It includes three implemen- tations of ATS: a TextRank based, a Latent Semantic Analysis based and a sentence relevance score based summarizer. In this work we evaluate the summarization algorithms provided by the PyTLDR module. ”Initially, we consider the CNN/DailyMail data set presented to us. However, we decide to go with DailyMail articles directly from the dailymail website. This decision enables us to better assess the performance of our text summarization tool through the standard ROGUE-N evaluation metric.**

**B. Data sets**

We decided that the data set used in this project would be dailymail articles retrieved from the dailymail site. This was done as the dailymail articles commonly start with a small summary in the form of bullet points done by the journalist writing the article. We could then easily use this bullet point summary to compare differences between our Python script and the summary done by the journalist in question through the ROGUE evaluation. The dataset used was extracted from the dailymail site manually as we needed to ensure that all of the articles had the bullet point summaries.

We tested 73 different documents fetched from the dailymail.co.uk website. The links to the documents were stored in a text document and automatically processed with a Python script.

For the automatic processing of the data set, we developed a Python script that creates the summaries using both PyTLDR and the newly developed named-entity summarizer. This process bypassed the use of the GUI for faster processing of large amount of data. The script also measured the time each algorithm took to produce the summaries.

When testing the data, we created summaries that were of equal length to those of the human written reference texts, measured in the amount of sentences.