

## NLP for News Clustering Overview

Natural Language Processing clustering algorithm for multi-media news content.

**Project Description** When a news story breaks it's important to see a diversity of opinion from around the globe on that headline. Currently that requires the manual checking of individual news provider sites and social media channels.

Therefore, it is desirable to be able to offer a headline grouping, with all the content about that specific headline whether it be a video, tweet or text publication all in one place.

Currently a data gathering program has been written and a rudiment NLP clustering algorithm implemented (see website link below for a demo)

The aim of this project is to improve on the existing algorithm by using more advanced clustering techniques.

**What you will work on** The project can be divided into smaller constituent parts:

1. Mapping an article (short text summary) to a single vector through a learned embedding matrix.
2. Identifying clusters in the embedding matrix
3. Updating clusters live based a new articles being added over time (not so trivial)
4. Clustering of clusters to identify headline groups belonging to a particular theme.
5. Extracting meaningful video transcripts/summaries to be used for clustering
6. Extracting meaningful podcast transcripts/summaries to be used for clustering

**Who we are looking for**

1. Competent in Python and optionally experience with toolkits: SkLearn, Pandas, TensorFlow.
2. Linear Algebra
3. Problem solving skills

**Supervisor Profile** Name: Brad Fowler Current Position: CEO Website: <https://scopenews.co.uk/get> Linkedin: <https://www.linkedin.com/in/bradley-fowler/>