# **Exercise 1- Explore the BBC news archive**

For this exercise you’ll get the [BBC text archive](http://mlg.ucd.ie/datasets/bbc.html). Your job will be to tokenize the dataset, removing common stopwords. A great source of these stop words can be found [here](https://github.com/Yoast/YoastSEO.js/blob/develop/src/config/stopwords.js).