Burscher et al. 2016

Cleaning data

```python
for each document in documents:
    set cleaned text as "
    for each word in document.text:
        if word not noun, adjective, adverb or and word is not a name:
            add word to cleaned text
```

K-means data analysis

Each document is represented as a data point in a high-dimensional space.

```python
randomly select 8 points from the space
several times
for each point in points:
    find the documents closets to the point
    calculate the average position based on documents
    set the average position as new point
```