

## Week 6 Research Notebook

### From the research topic “Universities”:

Cortés, Julián. (2018). Mission statements of universities worldwide - Text mining and visualization. *Intangible Capital*. 14. 584-603. 10.3926/ic.1258.

In the above paper, The author first creates a corpus of 250 universities with their mission statements and other metadata like if the university is public/private, belongs to a specific region like EU, North America, South America, Asia, Australia etc and whether it is an old or a new university. This corpus is provided here <https://drive.google.com/file/d/1KGexDVG-NSGHs0ehG0zrH7yIYGxIISWf/view>.

It is important to group the universities together based on the Age band because when we're vectorising and finding similarity between two mission statements, then we could have two mission statements matching for two universities where one could be merely 50 years old (say) having rank 500, whereas the other could be 200 years old having rank 90. Mission statements being equal, the correlation between ranking would show no correlation in this case, but this would be incorrect as a university that is 200 years old shouldn't be grouped with a university that is 50 years old.

Grouping based on Age band, private/public ie type of university and region should be meaningful as in the above research it was found that universities from European region had a crisp and brief Mission statement that other non-english speaking countries.

Other research papers also highlight this:

Bart, C., & Baetz, M. (1998). The relationship between mission statements and firm performance: an exploratory study. *Journal of Management Studies*, 35(6), 0022-2380. <https://doi.org/10.1111/1467-6486.00121>

Our current approach is to search for 15 universities individually , meaningfully group them, and implement a technique each. Then integrate the technique and compare and contrast.

LDA and LSA are both topic modelling techniques. One drawback of topic modelling is bag of word assumption issue. Hierarchical clustering would be a different approach. We will also do sentiment analysis to discuss and review university ranking based on sentiments.