

Week 5 Research Notebook

From the reading:

Language models are deeply influenced by external social and psychological factors but the words can still be evaluated empirically and a score assigned to them based on the certain types of attributes that can be associated with them. Such values tend to form patterns over a large corpus of texts and can be analysed and the text can be judged under much more formal parameters. This is definitely eye opening as in the previous reading it seemed that there are inherent biases in the languages but even such biases can be reduced and mathematically homogenized across large corpus.

From the research topic “Universities”:

There are multiple techniques to find out document similarity, and ranking of documents based on their similarity. The LSA approach, which is **Latent Semantic Analysis** seems most appropriate to be used in the evaluation of university brochure or sop documents and find out similarity between them to contrast it against the rankings that these universities hold in a worldwide ranking context.

QS world ranking of universities can be referred. And the ranking can be compared using correlation metrics such as **Spearman's rank correlation coefficient or Kendall's tau-b**.

"A Study on Document Similarity Measures" by S. Vijayarani and S. Umarani (2015): This paper provides an overview of different document similarity measures and compares their performance using a dataset of news articles.

"**Academic Ranking of World Universities**" by J. Liu, C. Cheng, and H. Huang (2005): This paper introduces the Academic Ranking of World Universities (ARWU) and discusses the methodology used to rank universities based on various indicators such as research output and academic reputation.

"Document Similarity Analysis for Identifying Plagiarism" by H. Taneja and N. Sharma (2016): This paper presents a method for identifying plagiarism in academic documents using document similarity analysis.

"**Ranking of World Universities: A Comparative Analysis of Three International Ranking Systems**" by A. Ale Ebrahim, N. Ahmed, and S. Taha (2013): This paper

compares the rankings of universities using three different international ranking systems and discusses the factors that influence university rankings.

"LSA-based Document Similarity Analysis for Online News Articles" by K. Ramanathan and M. Varadarajan (2018): This paper presents a method for analyzing the similarity of online news articles using Latent Semantic Analysis (LSA) and discusses the performance of different similarity measures.