cornelissen_2019_a_computational_analysis_of_news_media_bias_a_south_african_case_study

Year

2019

Author(s)

Laurenz A. Cornelissen and Lucia I. Daly and Qhama Sinandile and Heinrich de Lange and Richard J. Barnett

Title

A Computational Analysis of News Media Bias: A South African Case Study

Venue

SAICSIT

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

Each topic was labelled using knowledge on the domain, these labels are to aid with interpretation, but they are always subjective and not wholly descriptive of the complexities of each topic
The topic labels are nevertheless intuitive to individuals who are aware of South African news and politics.

Topic
Mandela
Elections
Economics
ANC NEC
Higher Edu
Nkandla
Captured
ANC Leadership
Land
Parliament
Local Election
Cope
De Lille Scandal
Tshwane
EFF
VBS
ANC NW
Fransman
Unions
MOC
State Capture Inquiry
Zuma Gordhan
Public Services
EFF Parly
Life Esidimeni
USA

Motivation

these labels are to aid with interpretation

Topic modeling

LDA

Topic modeling parameters

Nr of topics: 38

Nr. of topics

38

Label

Manually assigned single or multi word labels

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: News

Dataset: News

Problem statement

News media in South Africa is assumed to be unbiased and objective in their reporting of the news.

Indeed, editors are required to uphold an objective and balanced view with no favour to external political or corporate interests. This assumption of objectivity is tested on a large scale by computationally analysing 30 000 articles published by five media houses: News24, SABC, EWN, ENCA, and IOL. Using topic modelling, 38 topics are extracted from the corpus, and sentiment is computed for each topic. The study highlights various cases of both over and under-reporting by media houses on particular topics. We also identify various tonality biases by media houses.

Corpus

Origin: News24, SABC, EWN, ENCA, and IOL

Nr. of documents: 30 843

Details:

Document

News article content

Pre-processing

- Removal of non-ASCII characters
- contractions are converted (ex. isn't into is not)
- replacing URLs
- extra white spaces after removing large HTML sections were also removed.
- All apostrophes following nouns, such as "Africa's" were removed
- A threshold was set at 710 character length and 112-word count, any sentence above both these thresholds was removed from the corpus.
- punctuation marks and stop-words were removed
- Words with less than 3 characters were also removed
- Tokenisation
- tokens were lemmatised and only nouns were retained

```
@inproceedings{cornelissen_2019_a_computational_analysis_of_news_media_bias_a_so
uth_african_case_study,
    author = {Laurenz A. Cornelissen and Lucia I. Daly and Qhama Sinandile and
Heinrich de Lange and Richard J. Barnett},
    booktitle = {Proceedings of the South African Institute of Computer
```

```
Scientists and Information Technologists 2019},
    date-added = {2023-04-13 22:32:47 +0200},
    date-modified = {2023-04-13 22:32:47 +0200},
    doi = {10.1145/3351108.3351134},
    month = {sep},
    publisher = {{ACM}},
    title = {A Computational Analysis of News Media Bias},
    url = {https://doi.org/10.1145%2F3351108.3351134},
    year = 2019}
```

#Thesis/Papers/FS