Newspaper Scraping

Frequency Analysis: 10 articles each from TOI and The Wire

Q1) What did the articles cover?

In the Times of India articles, the words' protest', 'Assam, 'caa', 'vote', 'bjp', and 'congress' occur the most number of times, which tells us that the major point of coverage in these articles was the Assam elections and the possible implementation of the Citizenship Amendment Act in the state. Since the word protest is one of the most frequent, it probably means that the people in the state are not very happy with the Act and are protesting against it. The words' dissent', 'student', 'Guwahati are also reasonably frequent, which again points to the fact that these articles revolved around possible rallies and protests against the CAA. Hence we can probably make out that the ten selected articles are merely reports about the events happening around the country regarding the CAA, especially in Assam.

On running the frequency analysis on ten random articles of The Wire with the keyword "CAA", the most occurring words come out to be 'India, 'bjp', 'democracy', 'Bengal, 'Muslim, and 'modi'. Over here, the coverage seems to revolve around the implementation of the CAA and what it could mean for the democracy. Since the articles also include 'Bengal and 'Muslim, they probably talk about the protests in Bengal and the impact that the Act will have on the Muslims living in the country. The words' economy' and 'constitution' also make an appearance. This could mean that the ten articles selected were generally analytical in nature, maybe discussing the possible pros and cons of the implementation of the Citizenship Amendment Act.

Q2) Who was featured in these articles?
As already mentioned previously, the words' bjp' and 'congress' are the most occurring. This means that these two political parties are the prominent points of discussion in the articles. This also makes sense as they are mainly about the Assam elections, so it is natural to feature the two major parties contesting them. The articles may also feature various politicians because 'rahul', 'gandhi', 'gadkari', 'banerjee' also occur fairly regularly. The words' Muslim and 'shaheen bag' are also frequent, and thus we can make out that they could be talking about the protests and Shaheen Bagh and also about Muslims.

The Wire articles also mainly feature the political party 'bjp', and also the government of India. Since the words 'Uttar Pradesh', 'Bengal are very frequent which means that the coverage includes the various states of India, and how the CAA may have been received by the general population in those states. The words' social and 'communities' are frequent, which may suggest that articles talk about the general public and the feeling among the masses regarding the CAA.

Q3) Who was mainly quoted in these articles?
The frequency analysis does not provide us with enough information to be able to answer this question. And this is where the limitations of using computational tools to analyse social sciences come in. Unless we read all the 20 articles, we will never be able to the actual content they contain or all the

people that are being talked about. Also, merely looking at the frequency analysis and trying to make conclusions is not the best way of analysing the articles because these frequencies are devoid of any context. Either we will need more advanced computational tools, or we read all the articles to answer the questions with more conviction and accuracy.

By: Hitesh Goel
2020115003
CHD

(the frequency analysis dcuments and the corresponding graphs are contained in the respective folders of the zip file).