

Documentation:

After collecting the data, we start generating the summary. We have used LexRank summariser for the summarisation process. LexRank is an unsupervised text summarizing method that uses graph-based sentence centrality rating. The key premise is that sentences “recommend” additional sentences that are similar to them to the reader. As a result, if one statement sounds a lot like a lot of other sentences, it's probably a big deal. The significance of this sentence is enhanced by the significance of the sentences that precede it. As a result, in order to be highly ranked and included in a summary, a sentence must be similar to a large number of other sentences. There are mainly five steps used in LexRank:

1. Read the arguments and divide it into different sentences
2. Generate similarity matrix across sentences
3. Page ranking is done on this sentence matrix
4. Sorting the rank and selecting the top 2 sentences
5. Output the summarized text

LexRank employs a graph-based technique, which is useful for dealing with high-dimensional data. Because it is a combinatorial optimization problem, it becomes faster and less expensive than any other approach.

The ability to compare information from diverse data without any pre-processing is the main benefit of employing graphs, which makes these techniques robust to a wide range of issues. As a result, we've taken this strategy.

A similarity matrix is constructed across each sentence in each argument after the sentences in each argument have been divided. LexRank constructs a graph by creating a vertex for each sentence in the argument. The edges between these sentences are created using semantic similarity. LexRank uses cosine similarity of the TF-IDF vectors.

LexRank ranks each sentence in the argument using the PageRank algorithm. “PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of measuring its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references.” [1]

Each of the sentence in the argument is ranked, and then sorted in descending order. This page rank algorithm aids in the discovery of the most significant sentences. It assists in the recognition of the top two sentences in an argument, which then serve as the argument's summary.

As we are using pre-trained model LexRank, we have not used the training dataset.

Generating Output File:

We take the generated text and map it to the text id and write them in the desired format to **<dataset>_result.json** file.

Steps:

```
python lexrang_arg_gen.py -y <VALIDATION SET> -z <TEST SET>
```

Output generated by the classifier in for validation data: **val_result.json**

Output generated by the classifier in for test data: **test_result.json**

References:

[1] Wikipedia contributors. (2021, July 6). PageRank. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:53, July 10, 2021, from <https://en.wikipedia.org/w/index.php?title=PageRank&oldid=1032313046>