

Semantic Search in Articles using NLP

Description

- Write a pipeline of searching some of words in English articles then extract hot keywords.
- You can split your data to testing, training, validation.
- Do pre-processing on the data just to be able to run it.
- Do optimization time for article to extract the relevant words.
- Feel free to use any technique of feature extraction.
- You can use Deep Learning Methods and show the diff and which one the best. (optional)

What to turn in:

1. Project documentation (details given below)
2. Report any resource you use
3. Your source code Jupyter Notebook with comments

Project Documentation:

Please ensure that the documentation is easy to read and follow. Python output presented as an unformatted screen dump is not acceptable. Do not include code in the documentation. Provide sufficient details for the reader to understand your process and reproduce your results or something similar. Vague statements like “I’ve used a Weka filter” without specifying which filter are not acceptable. It’s preferable to focus on technicalities rather than implementation details.

Document Structure:

1. Introduction: State the problem and your objectives
2. Data description: Describe the dataset (number of instances, features, division between training and testing, etc.)
3. Baseline experiments: State the goal, present experiments, and provide conclusions
4. Other experiments: For each, include goal, steps, results, and conclusion
5. Overall conclusion

Additional Requirements:

- List tools used (you may use any tool or programming language you’re comfortable with)
- List any external resources used
- Include captioned tables for clarity
- Compare your results to previously obtained benchmarks when possible

Reflection Questions:

After completing your experiments, answer the following: 1. What was the biggest challenge you faced during this project? 2. What do you think you have learned from the project?