D1

Anna Batra, Sam Briggs, Junyin Chen, Hilly Steinmetz
Department of Linguistics, University of Washington
{batraa, briggs3, junyinc, hsteinm}@uw.edu

Abstract

This deliverable contains a skeleton of the paper and a list of team members. This is an example citation (Jurafsky and Martin, 2022). According to Jurafsky and Martin (2022), this is another example citation.

1 Introduction

2 Engines

3 System Overview

4 Approach

5 Results

6 Discussion

7 Conclusion

8 Appendix A: Workload distribution

- Anna Batra set up the Github repository, turned in D1
- Junyin Chen got the team together and set up a communication channel
- Sam Briggs set up the Overleaf file and sent out a when-to-meet to schedule weekly meetings
- Hilly Steinmetz edited the Overleaf file to prepare it for D1.

9 Appendix B: Code repository and additional software and data used in your system

The repository for our project can be found on Github at github.com/LING-575-Summarization/Summarization.

We plan to use the following software for the project:

- Python
- Anaconda (for virtual environment)
- NLTK (for parsing, NER, and more)
- spaCy (for parsing, NER, and more)
- Scikit-learn (for machine learning models)
- Pytorch (for large LMs)
- Hugging Face's Transformer library (for auto-tokenization and find-tune trained model)

References

Dan Jurafsky and James H. Martin. 2022. *Speech and Language Processing*, 3rd edition (draft) edition. Online.

Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *NAACL-ANLP 2000 Workshop: Automatic Summarization*.