

Project Report D1.2-D1.3

Arsenina Anna, Pinigin Arseny

D1.2

Dataset collection

- Dataset that we wanted to use is not available anymore
- We found another dataset, but we have to modify original task a little bit
- <https://zenodo.org/record/3250095>

Read papers and distribute tasks

- Read <https://www.aclweb.org/anthology/I17-1095.pdf>
- Read <http://ceur-ws.org/Vol-1391/144-CR.pdf>
- Anna will perform preprocessing of the documents
- Arseny will implement Word Vector-based Dynamic Time Warping based on preprocessed documents

D1.3

Anna:

- Preprocess documents
- Transform them into vectors

Arseny:

- Implement DTW method for paragraphs
- Implement DTW method for documents

Plans for D1.4

- Add efficient embeddings from gensim
- Test implemented methods on the collected dataset with DTW
- Think about improvements of the model
- Evaluate the final result