Project Deliverable 1.5

Team: Arseny Pinigin, Anna Arsenina

Link to the colab: https://colab.research.google.com/drive/1uMwCbWJ_3ZEi14jBGQ8qt9Bb8ytfUp6f

Deliverable results:

- New part of the model – Doc2Vec – is implemented and added to the system

- Crucial problem with dataset is found, now we are thinking about the solution

- But still model can be evaluated on a smaller dataset, tests on 10 documents showed good performance

We will try to find a dataset for evaluation, the problem is that most of them are closed by some privacy policies, but anyway the semantic text alignment model is implemented.