

Questions

- Can we use libraries that have already implemented concrete functions that are used in information retrieval models (e.g. scikit-learn TfidfVectorizer) or should we implement all of them ourselves?
 - Yes, we are allowed to use any available libraries
- What about PyTerrier (<https://github.com/terrier-org/pyterrier>)? This is an information retrieval framework and provides a lot of ready-to-use information retrieval models. Are we allowed to use such models or should we build them from scratch?
 - Yes, we can use PyTerrier
 - We should also look into Transformers
(<https://www.sbert.net/examples/applications/information-retrieval/README.html>)
- Do we need a frontend or cli? Would Jupyter notebook be enough as a kind of frontend to run retrieval models?
 - Nice to have, but not mandatory
- Do we need to write a project report? If yes, how many pages? What should be the content of the report, should we also explain the models or just evaluate our results?
 - He is gonna check with Pedro
- What is the overall goal of the project? Do we need to compare different models or approaches or just implement one?
 - Up to us
 - He said that it would be a good idea to compare classical approaches (e.g. BM25) and newer ones (from the last 1 or 2 years) (e.g. Neural Reranking)
 - “Old ones you read about in books vs. The new Kids on the block”
- What about task 2?
 - Up to us
 - We should start with Task 1 and see how far we get. Basically making sure Task 1 is fully functional before we consider doing Task 2
- Recommendations for task 1? Should we consider both text and formulas or is text on its own enough?
 - We should consider both to get better results

Next Meeting is set for the 24.11.2022 12:00 – 13:15! Please give me feedback whether you can attend or not so I can reschedule.

Until next meeting

- First draft of the system
- Complete Pipeline
- Numbers for one or two models
- Doesn't have to be perfect (90%), but complete draft
- Have numbers and technical problems ready for discussion
- If we have issues we should not wait to reach out