

# Internship Casper De Smet: first task

## Task description

Casper will implement several text embedding baselines to label scientific articles with fields of interest.

The fields of interest are listed in **fields\_of\_interest\_v1.py**

The articles are included in **article\_export.json**

The baselines should take as input a single article, and should output a ranking of the fields of interest for that article. The algorithm should create embedding representations for both the articles and the fields of interest, and perform the ranking using an unsupervised distance metric such as cosine similarity.

These baselines should include both relevant word embedding-based methods such as fastText (<https://fasttext.cc/>) as well as state-of-the-art embeddings such as (<https://www.sbert.net/>).

The code will be hosted in a private GitHub repository which is shared with <https://github.com/PieterFivez>