

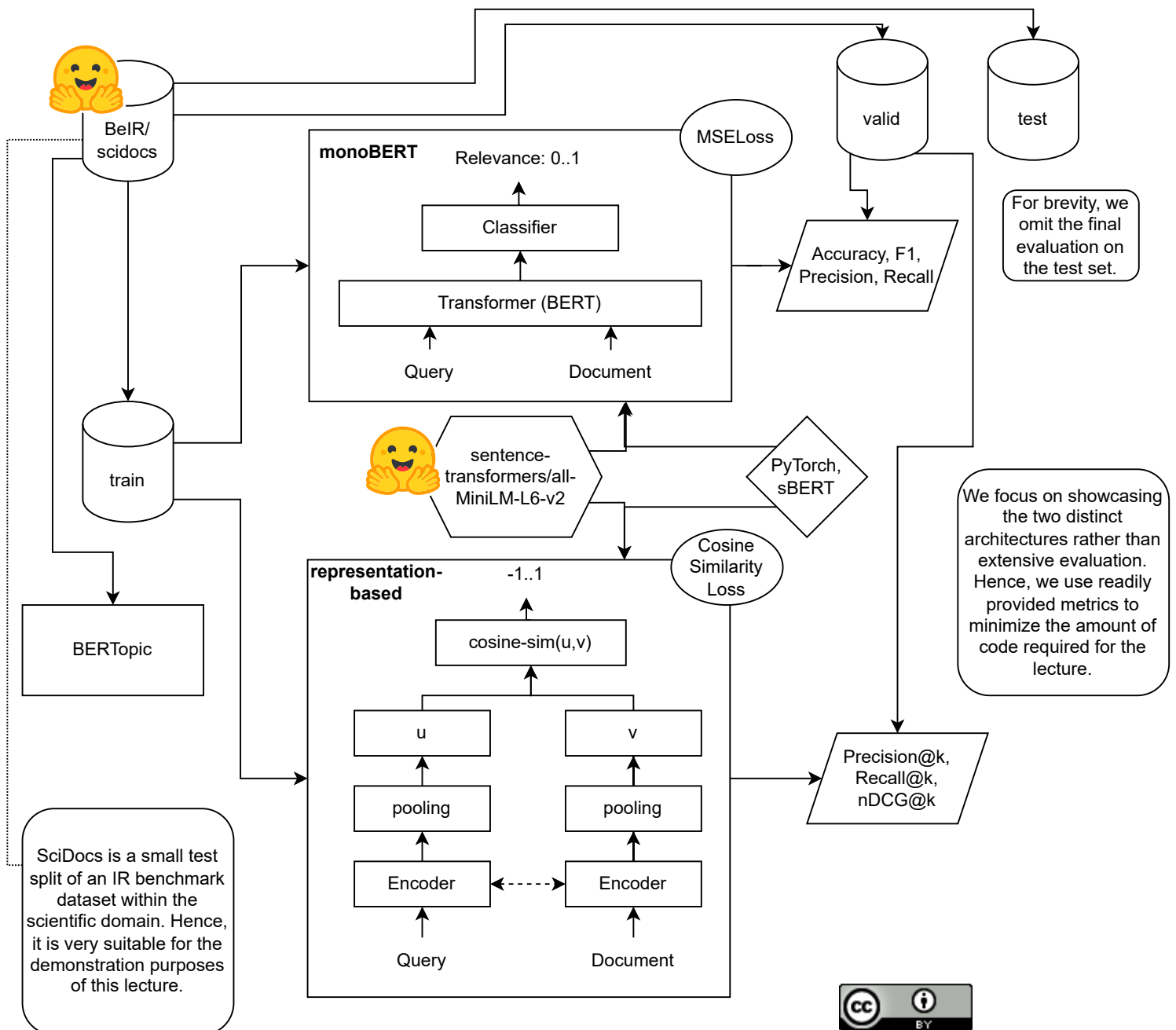
Advanced IR - Transformers4IR Concept

By Markus Reiter-Haas (Role: Advanced IR Lecturer)

This document illustrates a potential design description for the advanced IR project based on last year's Hands-on Lecture: <https://www.kaggle.com/code/markusreiterhaas/advanced-information-retrieval-7-transformers4ir>

sentence-transformers/all-MiniLM-L6-v2 The aim of the lecture is to teach how the Transformer architecture can be used in information retrieval (IR). It interweaves theoretical grounding through written (Markdown) text with a practical example in code. The code is structured as a (small) IR project, which fine-tunes pretrained Transformers (i.e., base model: sentence-transformers/all-MiniLM-L6-v2) to learn a relevance function.

For the experiment, we use an openly available dataset (BeIR/scidocs from HuggingFace). First, we briefly analyze the dataset with BERTopic. Then, we implement both a Cross-encoder and Bi-encoder using PyTorch and sBERT libraries. After training on the GPU, we analyze their prediction/ranking performance (e.g., F1, nDCG@k) over several epochs. Finally, we discuss how the two architecture could be combined and extended for future work.



Advanced IR - Transformers4IR
Concept by [Markus Reiter-Haas](#) is licensed under a [Creative Commons Attribution 4.0 International License](#).
The Creative Commons licence does not apply to the HuggingFace logo.