

PROTOTYPING OPEN WEB SEARCH APPLICATIONS WITH TIRA: A CASE STUDY IN RESEARCH-ORIENTED TEACHING

Maik Fröbe* Theresa Elstner† Harrison Scells† Benno Stein‡ Martin Potthast†,§

*Friedrich-Schiller-Universität Jena †Leipzig University ‡Bauhaus-Universität Weimar §ScaDS.AI

Introduction. The Open Search Foundation (OSF) aims to promote competition in the search market by enabling a wide range of independent search applications to strengthen Europe’s digital sovereignty. We present a new teaching concept for information retrieval (IR) courses at universities that mutually benefits students’ practical education and the rapid prototyping of open (web) search engines [3]. Students can gain practical experience by developing and evaluating all components of a search application. We use TIRA [1] to ensure that the artifacts created throughout the course, i.e., domain-specific test collections and retrieval approaches, are reusable and reproducible so that the community might pick them up and successful prototypes might eventually count towards the search applications maintained by the OSF.

Teaching Open Search Prototyping. Building search engines for new domains requires new test collections to guide the development. As part of teaching, we must consider that students can only invest a certain amount of time into this task. Suitable test collections comprise three components: (1) a collection of documents to be searched, (2) a set of topics defining user information needs, and (3) corresponding relevance assessments for documents in the collection. The document collection can be sampled from existing corpora, especially the envisioned Open Web Index built by the OpenWebSearch.EU project. This leaves the creation of topics and relevance assessments as remaining tasks. Both are crucial to reliable search engine evaluation and the information retrieval curriculum, and students are highly receptive to learning about it [4]. Student-built test collections offer a valuable learning opportunity to them and a valuable research opportunity to “crowdsource” independent search engine evaluations. A particularly salient opportunity is the ongoing evaluation using existing test collections, which may need more freshness or breadth (number of topics) and depth (number of relevance assessments per topic).

Project-Based Test Collection Development. Figure 1 presents our methodology for teaching student projects to create a test collection while catering to their learning needs [3].

We designed a milestone-based system where students create a test collection over one university teaching semester. Our methodology aims to develop a useable test collection from a pre-existing document collection (e.g., a selected subset of the Open Web Index with no topics or relevance assessments). Students work in groups to create different parts of the test collection, working toward three milestones. In Milestone 1, students analyze the document collection, and each student constructs one information need. In Milestone 2,

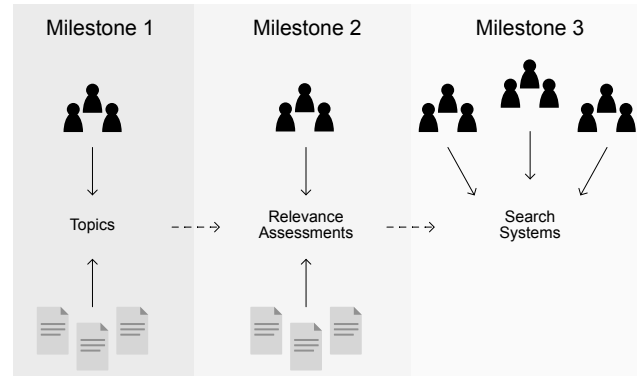


Figure 1: Our methodology for developing test collections during collaborative group student projects yielding topics, relevance assessments, and strong baseline search systems.

students create relevance assessments for their information needs by pooling and judging the documents retrieved by 10 diverse retrieval systems from TIREx [2]. Milestone 3 combines all groups’ topics and relevance assessments into a hidden final test collection. Students then develop prototypes using their topics, submitting them to a shared leaderboard in a competition to develop the most effective system on the hidden joint test set. This way, students learn the fundamentals of information retrieval evaluation and use this knowledge to prototype effective search engines. The joint test collection and reasonably strong baselines can henceforth serve for further research or as a starting point for shared tasks. Moreover, students are invited to open source their search engines under the open web search licensing scheme.

Future Work. Going forward, joint test collections built by students will have to be evaluated, too. We envision an augmentation and comparison to expert-supplied topics and relevance judgments in a quantitative and qualitative topic (difficulty) assessment. A comprehensive teaching concept documentation will also be developed.

REFERENCES

- [1] M. Fröbe et al. Continuous Integration for Reproducible Shared Tasks with TIRA.io. In *Proc. of ECIR 2023*, 2023.
- [2] M. Fröbe et al. The Information Retrieval Experiment Platform. In *Proc. of SIGIR 2023*, 2023.
- [3] T. Elstner et al. Shared tasks as tutorials: A methodical approach. In *Proc. of AAAI 2023*, 2023.
- [4] M. Sanderson and A. Warner. *Training Students to Evaluate Search Engines*. 2011.