

# A Tool for Enhancing the Understanding of Information Retrieval System Components for Educational Purposes

*Thomas Wilhelm-Stein, Maximilian Eibl*

Technische Universität Chemnitz

09107 Chemnitz, Germany

{thomas.wilhelm-stein, maximilian.eibl}@informatik.tu-chemnitz.de

## **Abstract**

We present a web-based tool to support the understanding of the information retrieval process first hand and to gain practical knowledge for students. Our focus is the arrangement and configuration of IR components like stemmers, indexing algorithms or feedback cycles and their evaluation.

**Keywords:** Information retrieval, Teaching, Web application, Components

## **1 General idea**

An Information Retrieval System (IRS) can be understood as a process chain incorporating different components for different tasks like stemmers and stop-words removal for pre-processing, indexing and searching algorithms for processing, and blind relevance feedback for post-processing topics (Kürsten 2012: 323). Experienced researchers have a more or less deep understanding, which components will work well together and which will not. A novice researcher needs to develop this understanding, which is a cumbersome process of trial and error.

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14<sup>th</sup> International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19<sup>th</sup>–21<sup>st</sup> May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 590–592.

The here proposed system eases this learning task by providing a web-based graphical user interface, which enables learners to freely select, configure, and arrange the IRS-components. Employing evaluation corpora like from CLEF learners can immediately see the impact of their configuration on the retrieval quality. Thus, learners easily gain deep insight in the dependencies of retrieval components without the need to cope with complex programming and laborious evaluations.

## 2 System overview

The system is build as a single page web application using a RESTful API. Specifically we use Jersey (jersey.java.net) as backend and AngularJS (angularjs.org) as frontend. Students should be able to use it without installing additional software except for a modern browser. The system consists of three main sections: managing and inspection of collections, configuration of experiments and exploration of results.

Existing research focused on single components, like VIRlab or IR-Components. VIRlab (Fang 2014), a web-based platform, addresses retrieval models by providing easy access to term and document statistics. IR-Components (López-García 2011) is a development framework and desktop application, which aims to support an easier setup and new components on a more coarse level.

There is a working prototype at <https://tu-chemnitz.de/cs/mi/xtrieval>. The system is used for education in lectures at the Technische Universität Chemnitz. We are currently evaluating the system within a university course on information retrieval and are planning to provide access for other universities.

## References

- López-García, R., & Ccheda, F. (2011). A Technical Approach to Information Retrieval Pedagogy. In: *Teaching and Learning in Information Retrieval* (pp. 89–105). Berlin, Heidelberg: Springer.

- Fang, H., Wu, H., Yang, P., & Zhai, C. (2014). Virlab: A web-based virtual lab for learning and studying information retrieval models. In: *Proceedings of the 37<sup>th</sup> international ACM SIGIR conference on Research & development in information retrieval* (pp. 1249 f.). ACM.
- Kürsten, J. (2012). *A Generic Approach to Component-Level Evaluation in Information Retrieval*. Dissertation. Universitätsverlag Chemnitz.