

# Information Retrieval: Extending the Theory of Accessibility Measures

Scientific Report

## Purpose of the STSM

In Information Retrieval (IR), accessibility analysis relies on high computational cost experiments. In a recent paper, we have initiated the exploration of a theoretical approach to Retrievability that, in case of boolean IR models, predicts analytically the retrievability of a document, therefore avoiding the need to compute it experimentally. However, this approach assumes that all the queries made using all the possible combination of terms in the collection of documents are likely to be submitted. To better approximate this assumption, the visit to Dr. Evangelous Kanoulas and his colleagues, active researchers on the test collection based evaluation topic at the University of Amsterdam, aimed to extend the retrievability theory on keywords based domains (e.g. legal and medical) including keyword query distributions, in order to better approximate the users' behavior when using a search engines in specific domains.

## Description of the Work Carried Out during the STSM

The STSM focused on Keyword Search and User Interaction. Concretely, we explored two keyword-related domains, Legal and e-Health, and measured the retrievability of a document by simulating a user's keywords-based search. The work of the STSM has been carried out with a series of discussions with Dr. Evangelous Kanoulas and his colleagues where, through the analysis of the newly generated data, we have better understood the limitations of the current theory of Accessibility Measures and initiated the exploration of some new ideas. One of them was, for example, how to use retrievability information on search engines that use relevance feedback in recall oriented domains. These search engines usually make use of this feedback when constructing a query model to extend the given topic in order to retrieve additional relevant documents at successive iterations. It has been observed that the number of relevant documents retrieved on the first interactions increases until reaching a plateau that lasts till a topic shift happens in the query model due the massive negative feedback, thereby allowing the system to retrieve some additional relevant documents. Here, we aimed to

understand, in case these documents were poorly retrievable, for which queries they would have got retrieved.

### **Description of the New Main Results Obtained**

The STSM established interaction paths between the information retrieval group at the Vienna University of Technology and the Information and Language Processing Systems (ILPS) group at the University of Amsterdam. Moreover, we extended the theory of Accessibility Measures.

### **Future Collaboration with the Host Institution**

The STSM has established a collaboration between the two institutions, and, without a doubt, it will continue.

### **Foreseen publications/articles resulting or to result from the STSM (if applicable)**

A paper was written and sent to the 2<sup>nd</sup> ACM International Conference on the Theory of Information Retrieval and a second one will be submitted to an up and coming conference.