The Impact of Query Variability and Relevance Measurement Scales on Information Retrieval Evaluation

Falk Scholer
RMIT University

ABSTRACT

Information retrieval makes extensive use of test collections for the measurement of search system effectiveness. Broadly speaking, this evaluation framework includes three components: search queries; a collection of documents to search over; and relevance judgements. In this talk, we'll consider two aspects of this process: queries, and relevance scales. Test collections typically use a single query to represent a more complex search topic or information need. However, different people may generate a wide range of query variants when instantiating information needs. We'll consider the implications of this for the evaluation of search systems, and the potential benefits and costs of incorporating variant queries into a test collection framework. Relevance judgements are used to indicate whether the documents returned by a retrieval system are appropriate responses for the query. They can be made using a variety of different scales, including ordinal (binary or graded) and techniques such as magnitude estimation. We'll examine a number of different approaches, and explore their benefits and drawbacks for judging relevance for retrieval evaluation.

BIOGRAPHY

Falk Scholer is a Professor in the Data Science discipline of the School of Computing Technologies at RMIT University in Melbourne, Australia. His research is in the area of information access and retrieval, focusing on understanding how systems such as search engines can assist users to resolve their information needs, and how their effectiveness can be measured. He also works on issues of fairness, accountability, transparency and ethics of systems and algorithms as part of the ARC Centre of Excellence in Automated Decision Making and Society, and on misinformation, fake news and fact-checking with the RMIT FactLab research hub. Falk is the Deputy Director of the RMIT Centre for Information Discovery and Data Analytics (CIDDA), which brings together experts across different academic disciplines, schools and colleges including computing technologies, science, maths and statistics, engineering, and business. He also teaches a range of courses, including on web development and programming, data science, HCI, and databases, and is the Program Manager for the postgraduate Master of Data Science. Falk also has a keen interest in research ethics and integrity, and chairs the STEM College Human Ethics Advisory Network (CHEAN).