

Cheap IR Evaluation: Fewer Topics, No Relevance Judgements, and Crowdsourced Assessments

Kevin Roitero
University of Udine, Udine, Italy
roitero.kevin@spes.uniud.it

Abstract

To evaluate Information Retrieval (IR) effectiveness, a possible approach is to use test collections, which are composed of a collection of documents, a set of description of information needs (called topics), and a set of relevant documents to each topic. Test collections are modelled in a competition scenario: for example, in the well known TREC initiative, participants run their own retrieval systems over a set of topics and they provide a ranked list of retrieved documents; some of the retrieved documents (usually the first ranked) constitute the so called pool, and their relevance is evaluated by human assessors; the document list is then used to compute effectiveness metrics and rank the participant systems. Private Web Search companies also run their in-house evaluation exercises; although the details are mostly unknown, and the aims are somehow different, the overall approach shares several issues with the test collection approach.

The aim of this work is to: (i) develop and improve some state-of-the-art work on the evaluation of IR effectiveness while saving resources, and (ii) propose a novel, more principled and engineered, overall approach to test collection based effectiveness evaluation.

In this thesis we focus on three main directions: the first part details the usage of few topics (i.e., information needs) in retrieval evaluation and shows an extensive study detailing the effect of using fewer topics for retrieval evaluation in terms of number of topics, topics subsets, and statistical power. The second part of this thesis discusses the evaluation without relevance judgements, reproducing, extending, and generalizing state-of-the-art methods and investigating their combinations by means of data fusion techniques and machine learning. Finally, the third part uses crowdsourcing to gather relevance labels, and in particular shows the effect of using fine grained judgement scales; furthermore, explores methods to transform judgements between different relevance scales.

Awarded by: University of Udine, Udine, Italy **on** 19 March 2020.

Supervised by: Professor Stefano Mizzaro.

Available at: <https://kevinroitero.com/resources/kr-phd-thesis.pdf>.

Selected Publications

- Lei Han, Kevin Roitero, Ujwal Gadiraju, Cristina Sarasua, Alessandro Checco, Eddy Maddalena, and Gianluca Demartini. The Impact of Task Abandonment in Crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, pages 1–1, 2019a. doi: 10.1109/TKDE.2019.2948168. URL <https://ieeexplore.ieee.org/abstract/document/8873609>.
- Lei Han, Kevin Roitero, Eddy Maddalena, Stefano Mizzaro, and Gianluca Demartini. On Transforming Relevance Scales. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, pages 39–48. ACM, 2019b. doi: 10.1145/3357384.3357988. URL <https://doi.org/10.1145/3357384.3357988>.
- Kevin Roitero, Eddy Maddalena, Gianluca Demartini, and Stefano Mizzaro. On Fine-Grained Relevance Scales. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 675–684. ACM, 2018. doi: 10.1145/3209978.3210052. URL <https://doi.org/10.1145/3209978.3210052>.
- Kevin Roitero, Andrea Brunello, Giuseppe Serra, and Stefano Mizzaro. Effectiveness Evaluation without Human Relevance Judgments: a Systematic Analysis of Existing Methods and of their Combinations. *Information Processing & Management*, 57(2):102149, 2020a. doi: 10.1016/j.ipm.2019.102149. URL <https://doi.org/10.1016/j.ipm.2019.102149>.
- Kevin Roitero, J. Shane Culpepper, Mark Sanderson, Falk Scholer, and Stefano Mizzaro. Fewer topics? A million topics? Both?! On topics subsets in test collections. *Information Retrieval Journal*, 23(1):49–85, 2020b. doi: 10.1007/s10791-019-09357-w. URL <https://doi.org/10.1007/s10791-019-09357-w>.