

# Modelling and Explaining IR System Performance Towards Predictive Evaluation

Guglielmo Faggioli

University of Padua

Italy

`guglielmo.faggioli@unipd.it`

## Abstract

Information Retrieval (IR) systems play a fundamental role in many modern commodities, including search engines (SEs), digital libraries, recommender systems, and social networks.

The IR task is particularly challenging because of the volatility of IR systems performance: users' information needs change daily, and so do the documents to be retrieved and the concept of what is relevant to a given information need. Therefore, the empirical offline evaluation of an IR system is a costly and slow post-hoc procedure, that happens after the system deployment. Given the challenges linked to empirical IR evaluation, predicting a system's performance before its deployment would add significant value to the development of an IR system.

In this manuscript, we place the cornerstone for the prediction of IR performance, by considering two closely related areas: the modeling of IR systems performance and the Query Performance Prediction (QPP). The former area allows us to identify those features that impact the most on the performance and that can be used as predictors, while the latter provides us with a starting point to instantiate the predictive task in IR.

Concerning the modeling of IR performance, we first investigate one of the most popular statistical tools, ANOVA. In particular, we compare traditional ANOVA with a recent approach, bootstrap ANOVA, and observe the different conclusions that can be achieved using these two different statistical tools [Faggioli and Ferro, 2021]. Secondly, using ANOVA, we study the concept of topic difficulty and observe that the topic difficulty is not an intrinsic property of the information need but stems from the formulation used to represent the topic [Culpepper et al., 2022]. Finally, we show how to use Generalized Linear Models (GLMs) as an alternative to the traditional linear modeling of IR performance [Faggioli et al., 2022]. We show how GLMs provide more powerful inference with comparable stability.

Our analyses on the QPP domain start with developing a predictor used to select among a set of reformulations for the same information need, the best-performing one for the systematic review task [Di Nunzio and Faggioli, 2021]. Secondly, we investigate how to classify queries as either semantic or lexical to predict whether neural models will perform better than lexical ones. Finally, given the challenges shown in the evaluation of the previous approaches, we devise a new evaluation procedure, dubbed sMARE [Faggioli et al., 2021]. sMARE allows moving from a single point estimation of the performance to a distributional one, allowing us to achieve improved comparisons between QPP models and more precise analyses.

---

**Awarded by:** University of Padova, Padova, Italy **on** 20 March 2023.

**Supervised by:** Nicola Ferro.

**Available at:** <https://www.research.unipd.it/handle/11577/3472979>.

## Selected Publications

J. Shane Culpepper, Guglielmo Faggioli, Nicola Ferro, and Oren Kurland. Topic Difficulty: Collection and Query Formulation Effects. *ACM Trans. Inf. Syst.*, 40(1):19:1–19:36, 2022. doi: 10.1145/3470563. URL <https://doi.org/10.1145/3470563>.

Giorgio Maria Di Nunzio and Guglielmo Faggioli. A Study of a Gain Based Approach for Query Aspects in Recall Oriented Tasks. *Applied Sciences*, 11(19), 2021. ISSN 2076-3417. doi: 10.3390/app11199075. URL <https://www.mdpi.com/2076-3417/11/19/9075>.

Guglielmo Faggioli and Nicola Ferro. System Effect Estimation by Sharding: A Comparison Between ANOVA Approaches to Detect Significant Differences. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part II*, volume 12657 of *Lecture Notes in Computer Science*, pages 33–46. Springer, 2021. doi: 10.1007/978-3-030-72240-1\\_3. URL [https://doi.org/10.1007/978-3-030-72240-1\\_3](https://doi.org/10.1007/978-3-030-72240-1_3).

Guglielmo Faggioli, Oleg Zendel, J. Shane Culpepper, Nicola Ferro, and Falk Scholer. An Enhanced Evaluation Framework for Query Performance Prediction. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part I*, volume 12656 of *Lecture Notes in Computer Science*, pages 115–129. Springer, 2021. doi: 10.1007/978-3-030-72113-8\\_8. URL [https://doi.org/10.1007/978-3-030-72113-8\\_8](https://doi.org/10.1007/978-3-030-72113-8_8).

Guglielmo Faggioli, Nicola Ferro, and Norbert Fuhr. Detecting Significant Differences Between Information Retrieval Systems via Generalized Linear Models. In Mohammad Al Hasan and Li Xiong, editors, *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, pages 446–456. ACM, 2022. doi: 10.1145/3511808.3557286. URL <https://doi.org/10.1145/3511808.3557286>.