

# A Comprehensive Dataset for Modern Learning to Rank Solutions (Abstract)

Domenico Dato<sup>1</sup>, Sean MacAvaney<sup>2</sup>, Franco Maria Nardini<sup>3</sup>, Raffaele Perego<sup>3</sup> and Nicola Tonellotto<sup>4</sup>

<sup>1</sup>Istella, Italy

<sup>2</sup>University of Glasgow, UK

<sup>3</sup>ISTI-CNR, Italy

<sup>4</sup>University of Pisa, Italy

## Abstract

In recent years, interest in neural Learning-to-Rank (LtR) approaches based on pre-trained language models has grown. These techniques have been demonstrated to be very effective at various ranking tasks, such as question answering and ad-hoc document ranking. The main reason for this success is the ability of deep neural networks to understand complex language patterns and learn to extract effective features from text. In the same time frame, feature-based LtR methods reached maturity, and research on this area focused primarily on specific aspects such as efficiency or diversification.

These two research areas progressed almost entirely disjointly and the effectiveness of neural LtR approaches compared to traditional feature-based LtR methods has not yet been well-established. A major reason that left the two areas well separated is the lack of publicly-available datasets enabling a direct comparison. LtR datasets providing query-document feature vectors do not contain the raw query and document text while the benchmarks often used for evaluating neural models, e.g., MS-MARCO, TREC Robust, etc., provide text but do not provide query-document feature vectors.

In this presentation, we introduce Istella22, a new dataset that enables such comparisons by providing both query/document text and strong query-document feature vectors used by an industrial search engine. The dataset, detailed in a resource paper that will be presented at ACM SIGIR 2022 [1], consists of a comprehensive corpus of 8.4M web documents, a collection of query-document pairs including 220 hand-crafted features, relevance judgments on a 5-graded scale, and a set of 2,198 textual queries used for testing purposes.

Istella22 enables a fair evaluation of traditional learning-to-rank and transfer ranking techniques on the same data. LtR models exploit the feature-based representations of training samples while pre-trained transformer-based neural rankers can be evaluated on the corresponding textual content of queries and documents. Through preliminary experiments on Istella22, we find that neural re-ranking approaches lag behind LtR models in terms of effectiveness. However, LtR models identify the scores from neural models as strong signals.

## References

- [1] D. Dato, S. MacAvaney, F. M. Nardini, R. Perego, N. Tonellotto, The Istella22 Dataset: Bridging Traditional and Neural Learning to Rank Evaluation, in: Proc. ACM SIGIR, 2022.


---

*IIR 2022: 12th Italian Information Retrieval Workshop, June 29 - June 30th, 2022, Milan, Italy*

✉ domenico@istella.ai (D. Dato); sean.macavaney@glasgow.ac.uk (S. MacAvaney); francomaria.nardini@isti.cnr.it (F. M. Nardini); raffaele.perego@isti.cnr.it (R. Perego); nicola.tonellotto@unipi.it (N. Tonellotto)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)