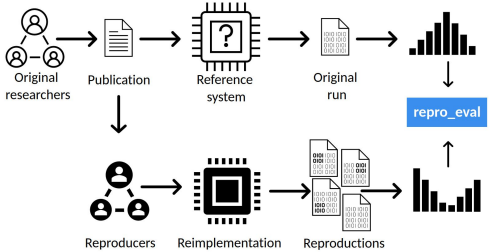
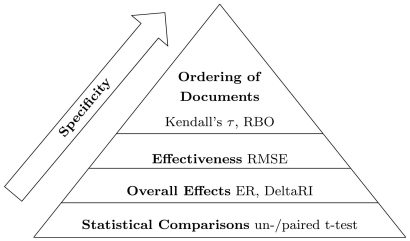
Reproducibility Issues



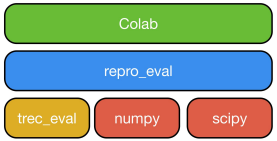
Given the original **publication** and the corresponding **run** file only, how do we know our **reimplementation** delivers the correct **reproductions**?

Reproducibility Measures

In previous studies, we introduced a set of reproducibility measures with different **levels of specificity** [SIGIR20].



repro_eval compiles these measures and is provided as a **Python package**. It can be run from the command line or interactively with **Google Colab**.



repro_eval

A Python Interface to Reproducibility Measures of System-oriented IR Experiments

Timo Breuer, Nicola Ferro, Maria Maistro, Philipp Schaer

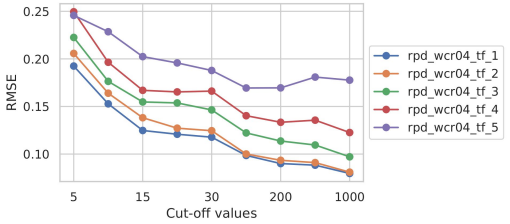
ECIR '21, Monday March 29, 13:45–16:45 UTC+1 (Session 3C)

GitHub Repository

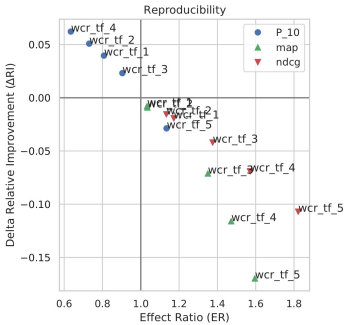
Google Colab Demo

University of Copenhagen
Università degli Studi di Padova
Technology Arts Sciences TH Köln

Reproducibility Analysis



The example plots help us to get a better understanding of our reproductions. The **Root-Mean-Square-Error** (RMSE) between topic scores illustrates the **reproduction quality** across the cut-off ranks.



At the level of overall effects, the **Effect Ratio & Delta Relative Improvement** are a valuable tool helping to explore the **space of reproduction**.

References

[SIGIR20] How to Measure the Reproducibility of System-oriented IR Experiments; Timo Breuer, Nicola Ferro, Norbert Fuhr, Maria Maistro, Tetsuya Sakai, Philipp Schaer, Ian Soboroff; SIGIR 2020