# Extending a Research-Paper Recommendation System with Scientometric Measures

Sophie Siebert, Siddharth Dinesh, and Stefan Feyer

## Motivation

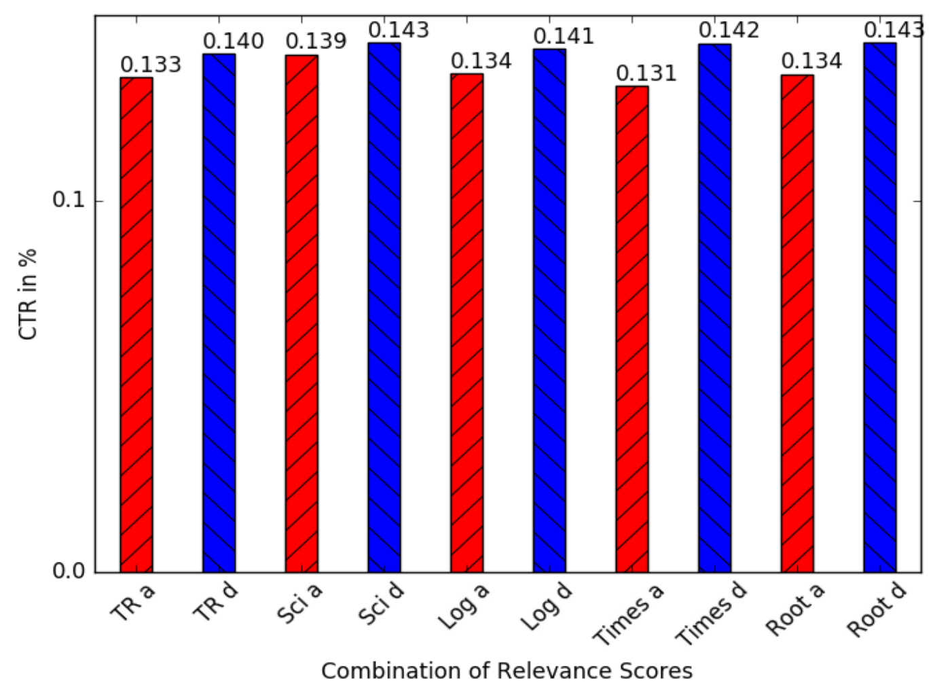
The number of academic publications doubles approximately every ten years. As a result it becomes more difficult for researchers to find relevant literature. To handle this information flood, recommender systems come into account. They identify the informational needs of researchers and recommend the best fitting literature. Since it is important to recommend papers which are relevant, it is necessary to improve recommender systems. In this paper we focus on the use of scientometrics to rank the recommendations. The assumption in this paper is that a paper with a good reputation is more worth reading, thus should be recommended. To measure the reputation we will use scientometrics. The scientific question is which scientometrics and their combination with the similarity ranking is most liked by the user and thus the best.

## System and Data

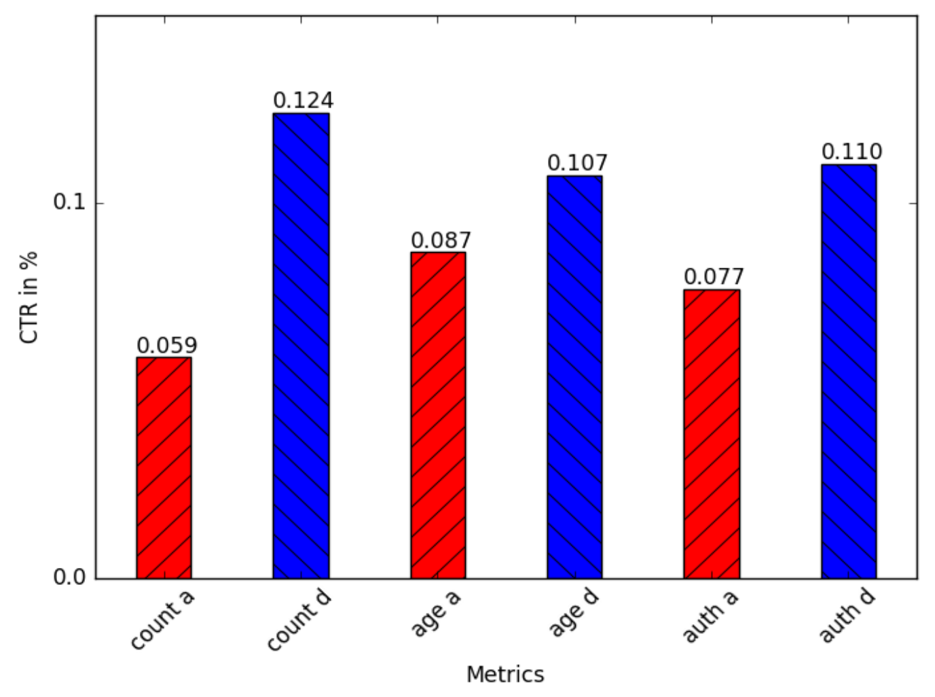
Mr. DLib (Machine Readable Digital Library, http://mr-dlib.org/) is a research-paper recommender system and for academic purposes only. It recommends papers similar to a given input paper. We cooperate with GESIS, that provides a framework for the user, as well as our data of 9.5 million documents. In total we analysed 38,740,893 recommendation with 53,441 clicks, which corresponds to a CTR of 0.138%.

## Results and Conclusion

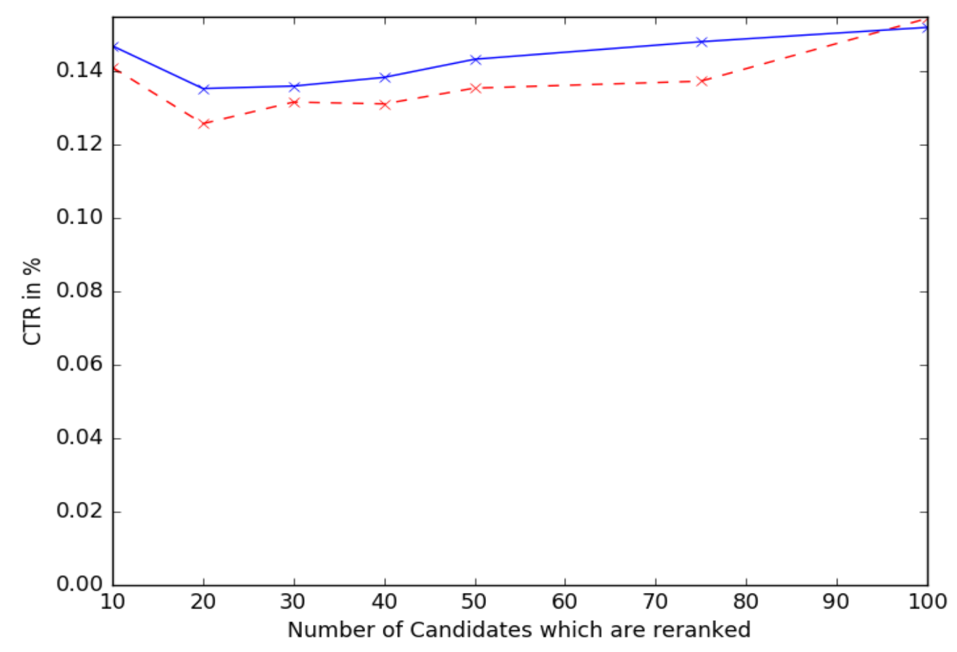
Scientometrics do improve the ranking of documents in a recommendation system compared to a text relevance only approach. However, this improvement is rather small. We sorted the rankings both ascending (a) and descending (d). The surprising findings were the good scoring of the 'scientometric only' with ascending ordering and that the normalization of the metrics worsened the CTR. The former might origin from an insufficient coverage.



Combination of scores
- TR: text relevance only
- Sci: scientometrics only
- Log: TR · log(Sci)
- Times: TR · Sci
- Root: TR · √Sci



Metrics
- count: absolute readership count
- age: count normalized by age of paper in years
- auth: count normalized by number of authors

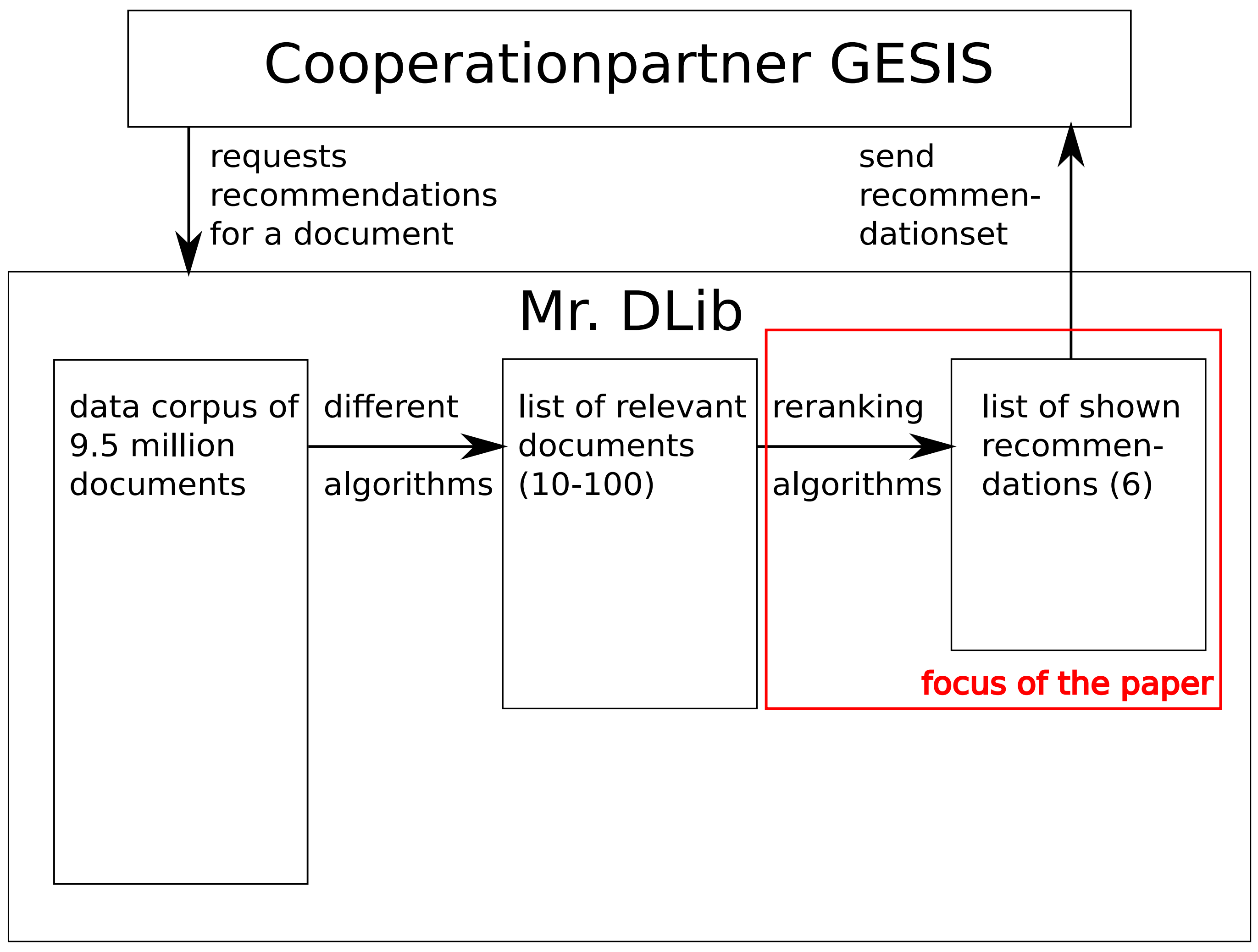


Re-ranked candidates are the first x documents from the relevance algorithms to rerank.

## Algorithm and ranking approaches

As shown is the figure we have a list of at most 100 documents, with attached relevance scores. Our ranking algorithm is randomly generated from the following variables:

1) How much influence has the relevant score
2) Which scientometric scheme is applied
3) How many documents are considered for the re-ranking


Cooperationpartner GESIS
requests recommendations for a document
send recommen-dationset
Mr. DLib
data corpus of 9.5 million documents
different algorithms
list of relevant documents (10-100)
reranking algorithms
list of shown recommen-dations (6)
focus of the paper

In this paper we use readership data from Mendeley to rank the documents. We got readership data for 1,694,373 documents, which is a coverage of 17.82%.

## Future Work

In the future we will integrate JabRef into Mr. Dlib and thus collect more data. Furthermore the combinations of rankings and relevance algorithms has to be investigated to find out if the rankings are stable and which combinations work best.

To improve the ranking the following steps can be done:
- Improve the coverage of the metrics by calculating author metrics and apply them back to the papers with sum or average
- Improve the coverage with a fallback mechanic, which will choose a higher ranking with higher coverage
- Improve the evaluation by logging clicks for exports
- collect and calculate metrics based on citation data

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin
DAAD Deutscher Akademischer Austausch Dienst
German Academic Exchange Service
gesis Leibniz-Institut für Sozialwissenschaften
Mr. DLib Machine-readable Digital Library
NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics