

# Towards Result Delta Prediction Based on Knowledge Deltas for Continuous IR Evaluation

QPP++ ECIR 2023

06 Apr. 2023

Gabriela Gonzalez-Saez

Petra Galuščáková

Lorraine Goeuriot

Philippe Mulhem

Alaa El-Ebshihy

Tobias Fink

Florina Piroi

David Iommi

# Agenda

- Introduction
- Knowledge & Result Delta
- Experiments & Results
- Summary

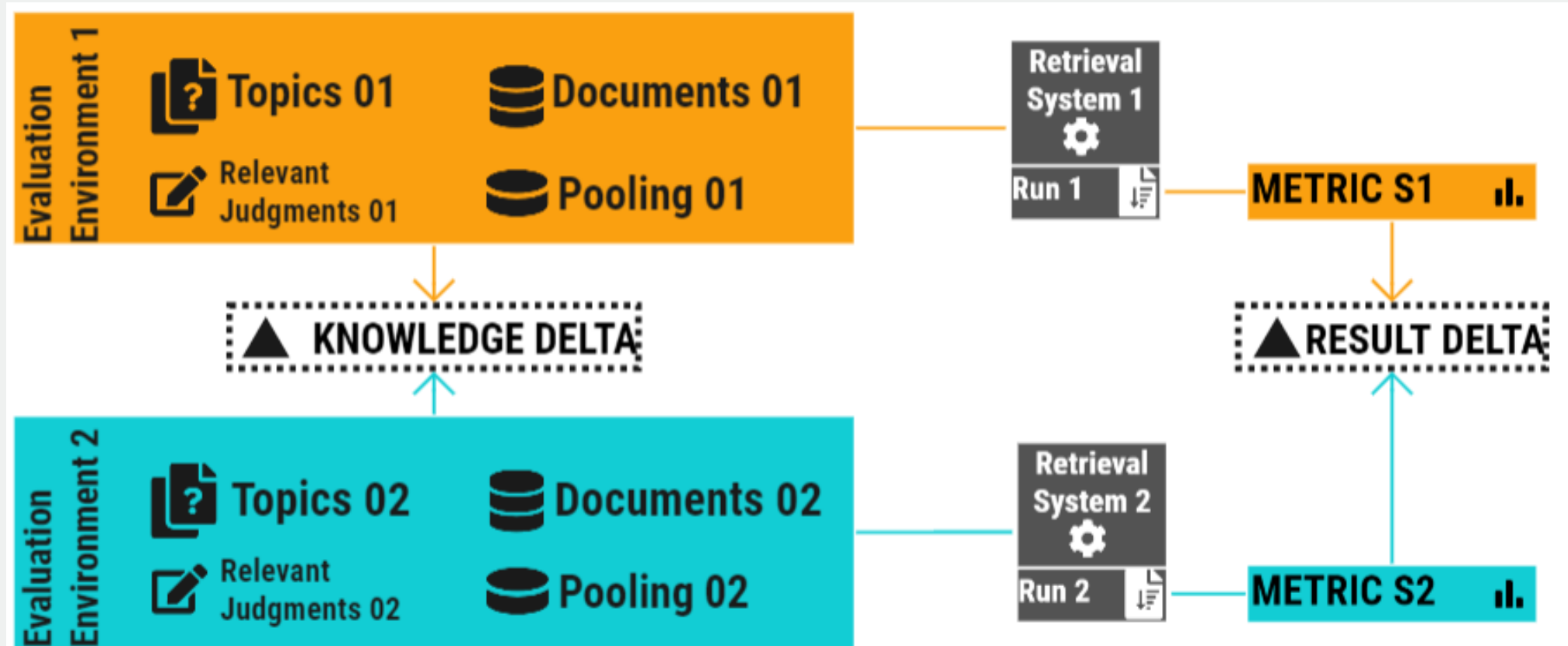
# Agenda

- **Introduction**
- Knowledge & Result Delta
- Experiments & Results
- Summary

# Introduction

## Motivation

The Evolution of the Evaluation Environments challenging in the continuous evaluation of IR Systems



# Agenda

- **Introduction**
- **Knowledge & Result Delta**
- Experiments & Results
- Summary

# Knowledge Delta and Result Delta

- **Knowledge Delta ( $K\Delta$ )**: differences between text representation
  - $Kd\Delta$ : between documents representations
  - $Kq\Delta$ : between queries representations

# Knowledge Delta and Result Delta

- **Knowledge Delta ( $K\Delta$ ):** differences between text representation
  - $Kd\Delta$ : between documents representations
  - $Kq\Delta$ : between queries representations
- **Results Delta ( $R\Delta$ )<sup>1</sup>:** differences between IR systems performance
  - $Rs\Delta$ : different IR systems in the same EE
  - $Re\Delta$ : same IR systems in different EEs
  - $Rse\Delta$ : different IR systems in different Ees

<sup>1</sup> G. N. González-Sáez, P. Mulhem, L. Goeuriot, Towards the evaluation of information retrieval systems on evolving datasets with pivot systems, in: K. S. Candan, B. Ionescu, L. Goeuriot, B. Larsen, H. Müller, A. Joly, M. Maistro, F. Piroi, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer International Publishing, Cham, 2021, pp. 91–102.

# Goal

Understand the impact of  $K\Delta$  on  $R\Delta$

- Correlation
- Prediction model
- Start with  $Kd\Delta$  and  $Re\Delta$



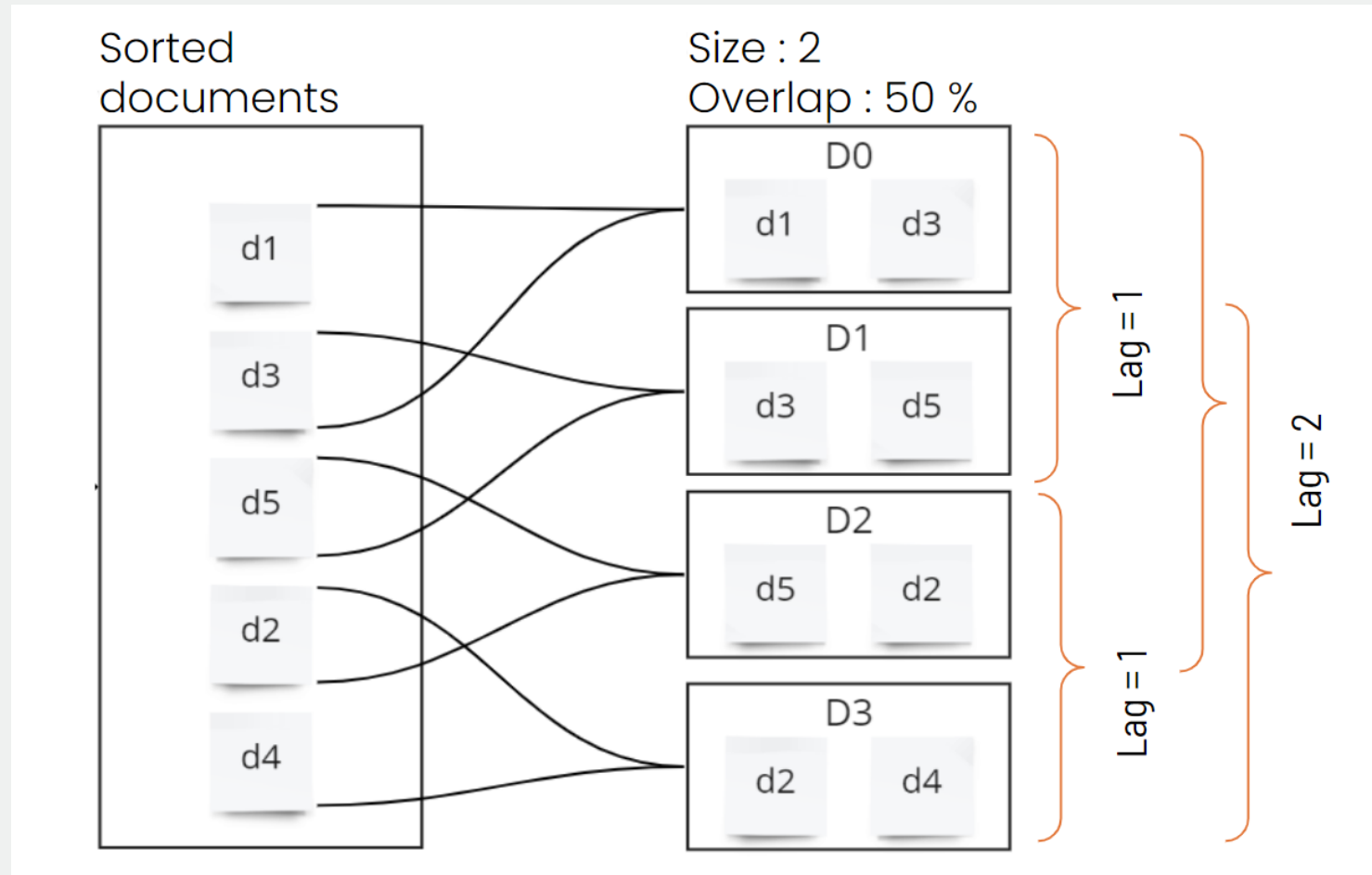
# Agenda

- Introduction
- Knowledge & Result Delta
- Experiments & Results
- Summary

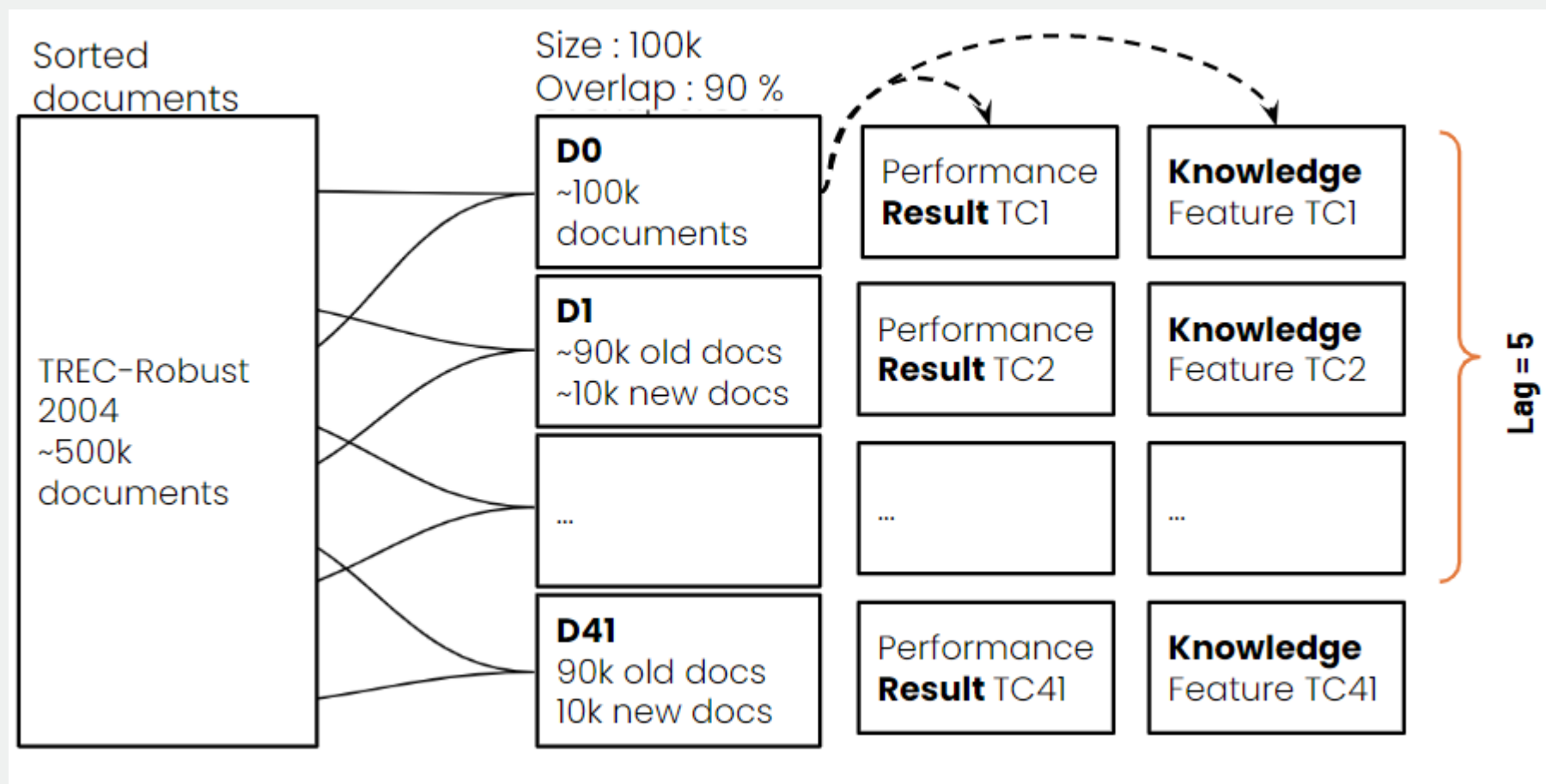
# Experimental Setup

- **Robust '04 test collection**
  - 41 sub-collections
  - 100k documents
  - 90% of overlap
  - 249 topics
- **K $\Delta$ : QPP Features**
  - Averaged Term Weight Variability (**avVAR**)
  - Averaged Collection Query Similarity (**avSCQ**)
- **R $\Delta$ : Performance metric BM25 system**
  - Mean Average Precision (**MAP**) per each sub-collection

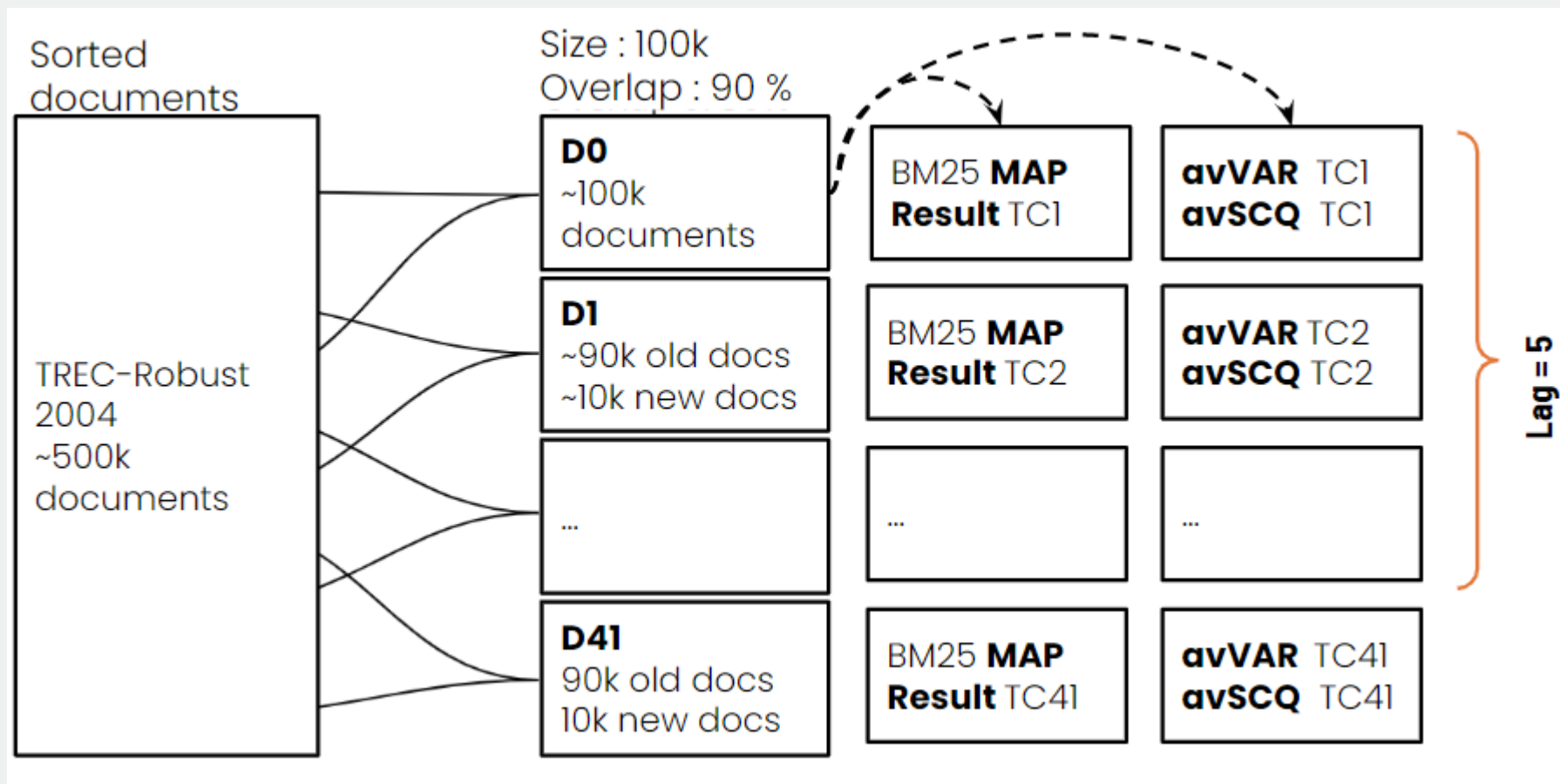
# Controlled sub-collections



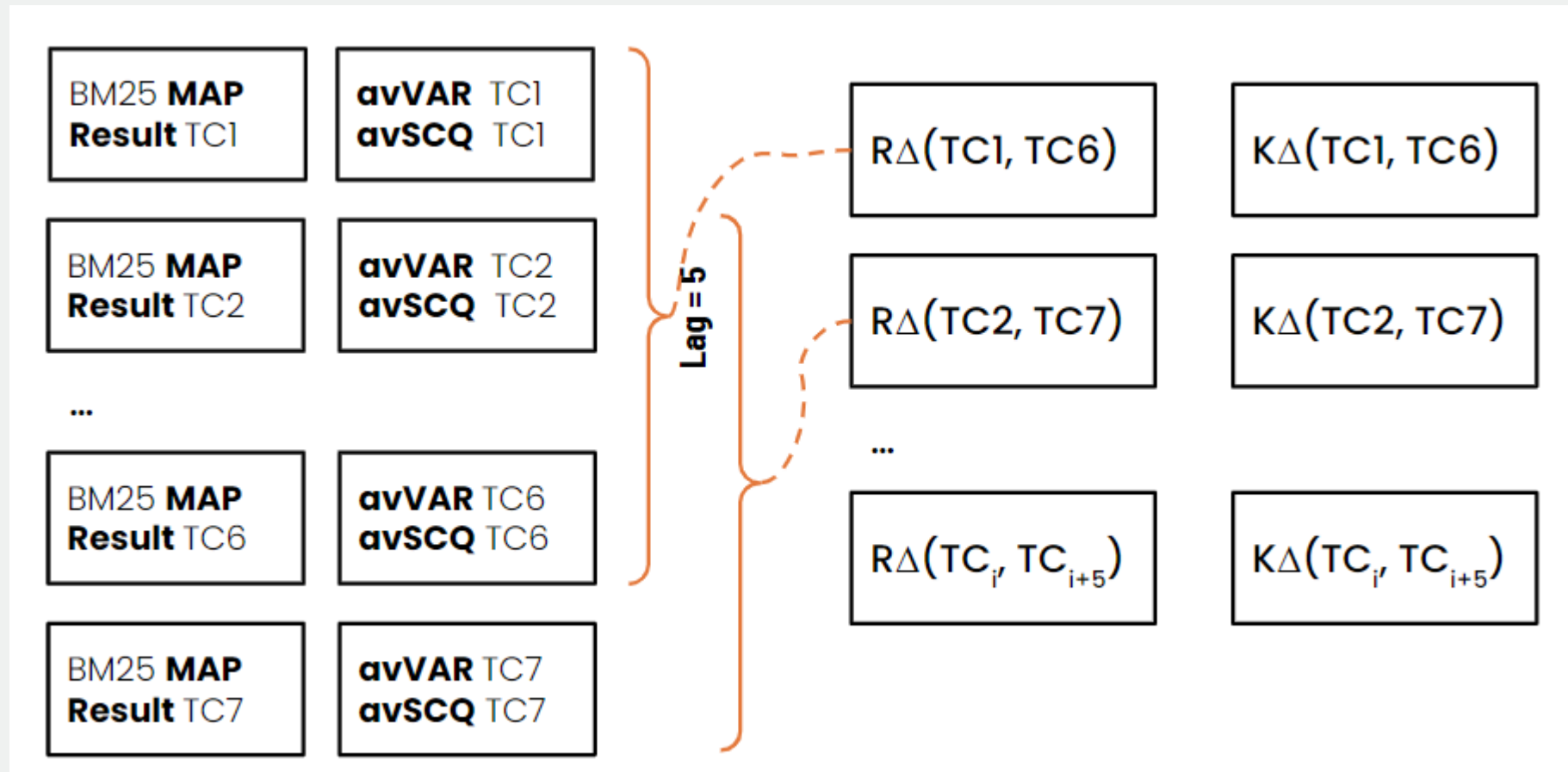
# Controlled sub-collections



# Controlled sub-collections

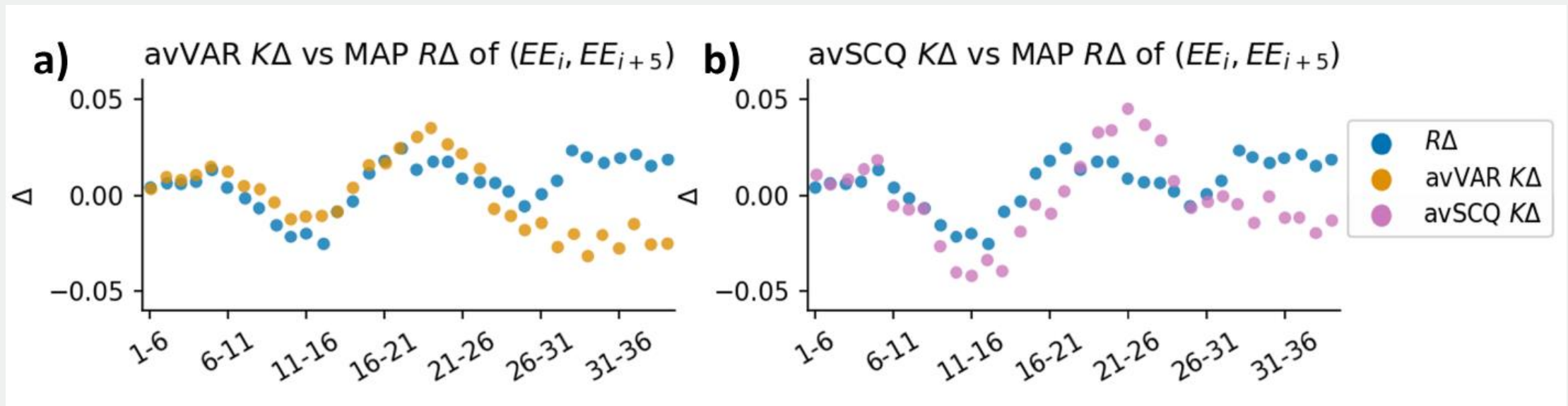


# R $\Delta$ & K $\Delta$ Correlation



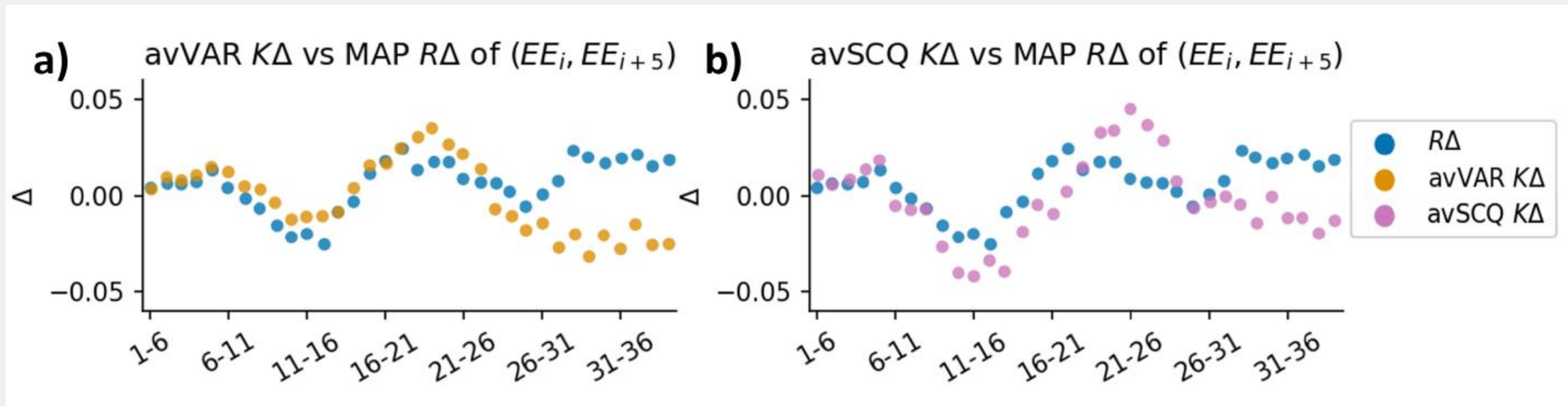
# Results

Pearson Correlation ( $r$ )  $R\Delta$  &  $K\Delta$



# Results

Pearson Correlation ( $r$ )  $R\Delta$  &  $K\Delta$



$r = 0.5$

$r = 0.12$



# Agenda

- Introduction
- Knowledge & Result Delta
- Experiments & Results
- Summary

# Summary and Future work

- We propose the definition of Knowledge Delta ( $K\Delta$ ) and Results Delta ( $R\Delta$ ) for IR systems
- We present a first attempt to quantify the impact of  $K\Delta$  on  $R\Delta$ 
  - Pearson Correlation between avVAR and avSCQ representations and  $\Delta$ MAP
- We plan to build a prediction model ( $K\Delta \sim R\Delta$ )
  - Machine learning
  - Time series analysis
- We plan to apply the model on the LongEval<sup>1</sup> collection.

<sup>1</sup> P. Galuščáková, R. Deveaud, G. Gonzalez-Saez, P. Mulhem, L. Goeuriot, F. Piroi, M. Popel, Longeval-retrieval: French-english dynamic test collection for continuous web search evaluation, arXiv preprint arXiv:2303.03229 (2023).

# Thanks

Work supported by

**French Agence Nationale de la Recherche (ANR)**

**Austrian Science Fund (FWF)**



**FWF**

Der Wissenschaftsfonds.