Metadata Annotations of Experimental Data with the `ir_metadata` Schema

Timo Breuer, Jüri Keller, Philipp Schaer

https://www.ir-metadata.org/

Dagstuhl Seminar 23031, 16th January 2023

Technology Arts Sciences TH Köln

IR GROUP

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Motivation & Contribution

Research goal?

Platform?

Implementation?

Data?

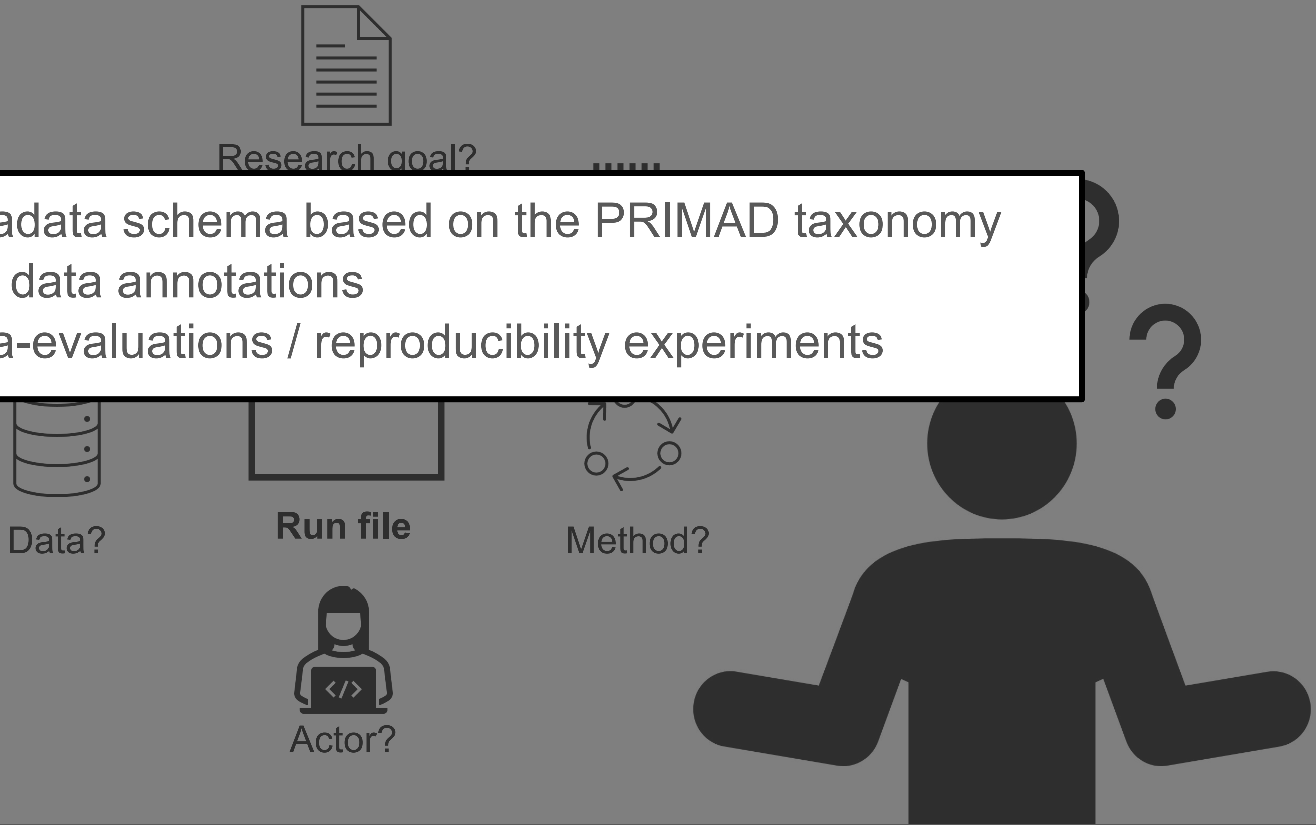
Run file

Method?

Actor?

Motivation & Contribution

- Metadata schema based on the PRIMAD taxonomy
- Run data annotations
- Meta-evaluations / reproducibility experiments



PRIMAD - the logical plan of the metadata schema

Report from Dagstuhl Seminar 16041

Reproducibility of Data-Oriented Experiments in e-Science

Edited by
Juliana Freire[1], Norbert Fuhr[2], and Andreas Rauber[3]

1 New York University, US, juliana.freire@nyu.edu
2 Universität Duisburg-Essen, DE, norbert.fuhr@uni-due.de
3 TU Wien, AT, rauber@ifs.tuwien.ac.at

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16041
cibility of Data-Oriented Experiments in e-Science". In many subfields of compu
experiments play an important role. Besides theoretic properties of algorithms or m
effectiveness and performance often can only be validated via experimentation. In
cases, the experimental results depend on the input data, settings for input p
potentially on characteristics of the computational environment where the exp
signed and run. Unfortunately, most computational experiments are specified
papers, where experimental results are briefly described in figure captions; the
the results is seldom available.

has serious implications. Scientific discoveries do not happen i
the result of sequences of smaller, less significant s
ented, reproducible, and generalizable, i
dering the ability of others
absence of repr

WORKSHOP REPORT

Increasing Reproducibility in IR: Findings from the Dagstuhl Seminar on "Reproducibility of Data-Oriented Experiments in e-Science"

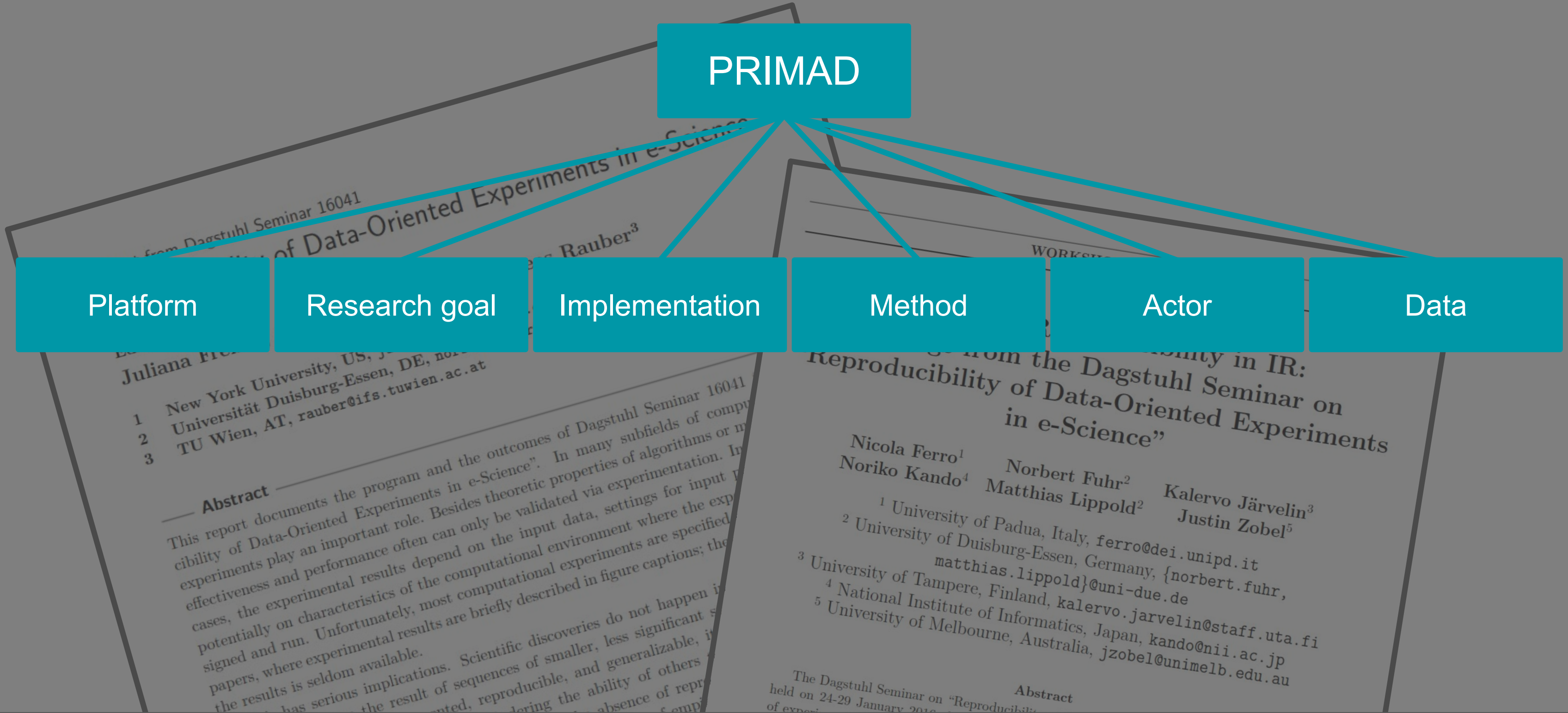
Nicola Ferro[1] Norbert Fuhr[2] Kalervo Järvelin[3]
Noriko Kando[4] Matthias Lippold[2] Justin Zobel[5]

[1] University of Padua, Italy, ferro@dei.unipd.it
[2] University of Duisburg-Essen, Germany, {norbert.fuhr, matthias.lippold}@uni-due.de
[3] University of Tampere, Finland, kalervo.jarvelin@staff.uta.fi
[4] National Institute of Informatics, Japan, kando@nii.ac.jp
[5] University of Melbourne, Australia, jzobel@unimelb.edu.au

Abstract

The Dagstuhl Seminar on "Reproducibili
held on 24-29 January 2016
of experi

PRIMAD - the logical plan of the metadata schema



...from Dagstuhl Seminar 16041

...of Data-Oriented Experiments in e-Science

...Rauber[3]

Juliana Fre...

1 New York University, US, j...
2 Universität Duisburg-Essen, DE, nor...
3 TU Wien, AT, rauber@ifs.tuwien.ac.at

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16041 ... cibility of Data-Oriented Experiments in e-Science". In many subfields of comp... experiments play an important role. Besides theoretic properties of algorithms or m... effectiveness and performance often can only be validated via experimentation. In... cases, the experimental results depend on the input data, settings for input p... potentially on characteristics of the computational environment where the exp... signed and run. Unfortunately, most computational experiments are specified... papers, where experimental results are briefly described in figure captions; the... the results is seldom available.

... Scientific discoveries do not happen i... ...has serious implications. ...the result of sequences of smaller, less significant s... ...nted, reproducible, and generalizable, i... ...ering the ability of others... ...absence of repr... ...of emp...

WORKSH...

...bility in IR:
...from the Dagstuhl Seminar on
Reproducibility of Data-Oriented Experiments
in e-Science”

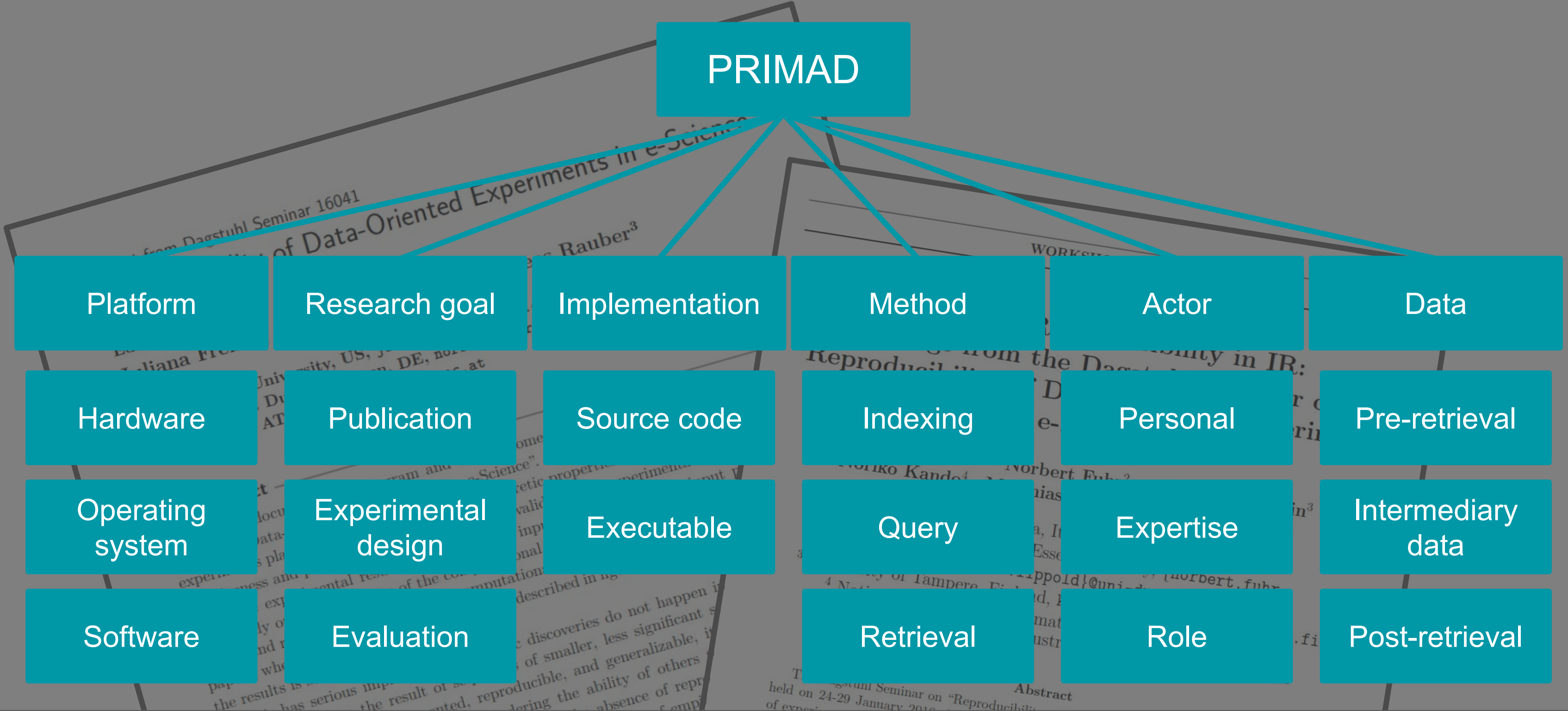
Nicola Ferro[1] Norbert Fuhr[2] Kalervo Järvelin[3]
Noriko Kando[4] Matthias Lippold[2] Justin Zobel[5]

[1] University of Padua, Italy, ferro@dei.unipd.it
[2] University of Duisburg-Essen, Germany, {norbert.fuhr, matthias.lippold}@uni-due.de
[3] University of Tampere, Finland, kalervo.jarvelin@staff.uta.fi
[4] National Institute of Informatics, Japan, kando@nii.ac.jp
[5] University of Melbourne, Australia, jzobel@unimelb.edu.au

Abstract

The Dagstuhl Seminar on “Reproducibili...
held on 24-29 January 201...
of exp...

PRIMAD - the logical plan of the metadata schema



- PRIMAD
  - Platform
    - Hardware
    - Operating system
    - Software
  - Research goal
    - Publication
    - Experimental design
    - Evaluation
  - Implementation
    - Source code
    - Executable
  - Method
    - Indexing
    - Query
    - Retrieval
  - Actor
    - Personal
    - Expertise
    - Role
  - Data
    - Pre-retrieval
    - Intermediary data
    - Post-retrieval

Metadata annotations of run files

```
307       Q0        497476        1         0.9931        bm25
307       Q0        469928        2         0.9674        bm25
307       Q0        125806        3         0.9623        bm25
307       Q0        504815        4         0.9453        bm25
307       Q0        392547        5         0.9223        bm25
...
```

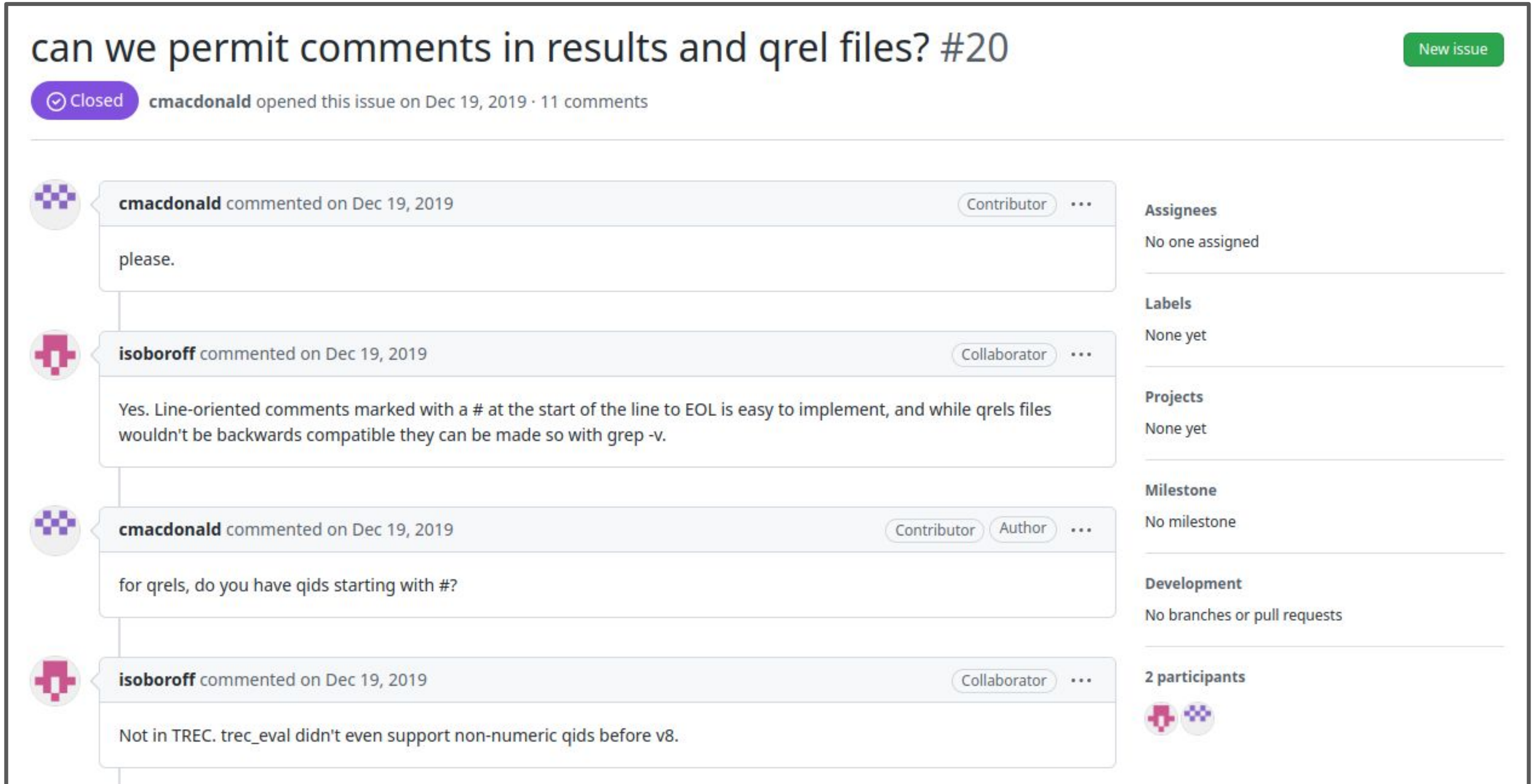
Metadata annotations of run files

```
# ir_metadata.start
# platform:
#    ...
# research goal:
#    ...
# implementation:
#    ...
# method:
#    ...
# actor:
#    ...
# data:
#    ...
# ir_metadata.end
307     Q0      497476      1       0.9931      bm25
307     Q0      469928      2       0.9674      bm25
307     Q0      125806      3       0.9623      bm25
307     Q0      504815      4       0.9453      bm25
307     Q0      392547      5       0.9223      bm25
...
```

Metadata annotations of run files

```
platform:
     hardware:
          cpu:
               model: 'Intel Xeon Gold 6144 CPU @ 3.50GHz'
               architecture: 'x86_64'
               operation mode: '64-bit'
               number of cores: 16
          ram: '64 GB'
     operating system:
          kernel: '5.4.0-90-generic'
          distribution: 'Ubuntu 20.04.3 LTS'
     software:
          libraries:
               python:
                    - 'scikit-learn==0.20.1'
                    - 'numpy==1.15.4'
               java:
                    - 'lucene==7.6'
          retrieval toolkit:
               - 'anserini==0.3.0'
```

Software support - trec_eval



can we permit comments in results and qrel files? #20

cmacdonald opened this issue on Dec 19, 2019 · 11 comments

cmacdonald commented on Dec 19, 2019

please.

isoboroff commented on Dec 19, 2019

Yes. Line-oriented comments marked with a # at the start of the line to EOL is easy to implement, and while qrels files wouldn't be backwards compatible they can be made so with grep -v.

cmacdonald commented on Dec 19, 2019

for qrels, do you have qids starting with #?

isoboroff commented on Dec 19, 2019

Not in TREC. trec_eval didn't even support non-numeric qids before v8.

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone
No milestone

Development
No branches or pull requests

2 participants

Software support - repro_eval

- repro_eval==0.4.0 https://github.com/irgroup/repro_eval
- Metadata handling and (semi-)automatic annotations
- Meta-analysis based on metadata fields

Software support - repro_eval

- repro_eval==0.4.0 https://github.com/irgroup/repro_eval
- **Metadata handling and (semi-)automatic annotations**
- Meta-analysis based on metadata fields

```
from repro_eval.metadata import MetadataHandler

run_path='./run.txt',
metadata_path='./metadata.yaml'
metadata_handler = MetadataHandler(run_path, metadata_path)
metadata_handler.write_metadata(complete_metadata=True)
```

Software support - repro_eval

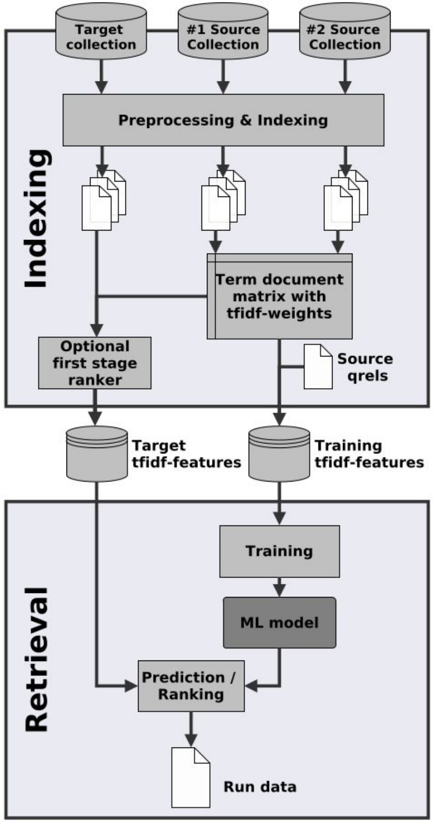
- repro_eval==0.4.0 https://github.com/irgroup/repro_eval
- Metadata handling and (semi-)automatic annotations
- **Meta-analysis based on metadata fields**

```
from repro_eval.metadata import MetadataAnalyzer, PrimadExperiment

run_path ='./run.txt'
dir_path ='./runs/'

metadata_analyzer = MetadataAnalyzer(run_path)
experiments = metadata_analyzer.analyze_directory(dir_path)

primad_type = 'priMad'
run_candidates = experiments.get(primad_type)

primad_experiment = PrimadExperiment(primad=primad_type,rep_base=run_candidates,...)
primad_experiment.evaluate()
```

Meta-evaluations / reproducibility experiments

Cross-collection relevance feedback
by Grossman and Cormack:

1. Derive tfidf training samples from source collection(s)
2. Train topic-based relevance classifier
3. Rank target collection



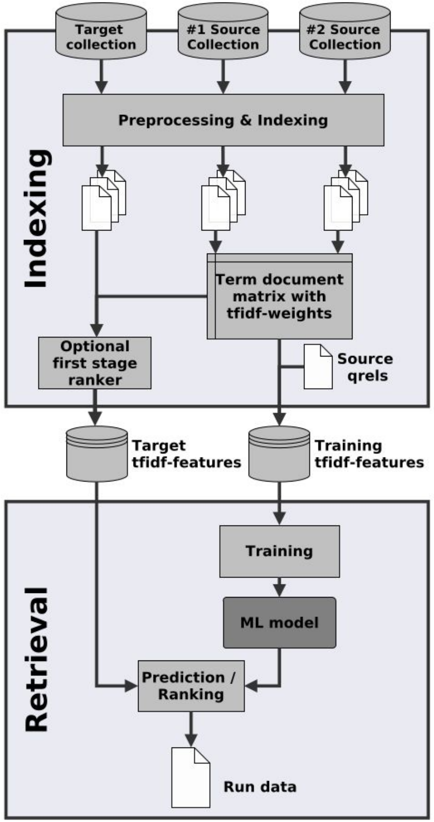
***MRG_UWaterloo and WaterlooCormack Participation in the TREC 2017 Common Core Track**;*
Grossman and Cormack; TREC Common Core 2017
***Simple techniques for cross-collection relevance feedback**;*
Yu, Xie, and Lin; ECIR 2019
***How to measure the reproducibility of system-oriented IR experiments**;*
Breuer, Ferro, Fuhr, Maistro, Sakai, Schaer, Soboroff; SIGIR 2020

Meta-evaluations / reproducibility experiments

Researchers	Type	Venue
GC	Original experiment	TREC 2017
YXL	Reproduction	ECIR 2019
BFFMSSS		SIGIR 2020

Run dataset: https://zenodo.org/record/5997491

MRG_UWaterloo and WaterlooCormack Participation in the TREC 2017 Common Core Track*;*
Grossman and Cormack; TREC Common Core 2017
Simple techniques for cross-collection relevance feedback*;*
Yu, Xie, and Lin; ECIR 2019
How to measure the reproducibility of system-oriented IR experiments*;*
Breuer, Ferro, Fuhr, Maistro, Sakai, Schaer, Soboroff; SIGIR 2020

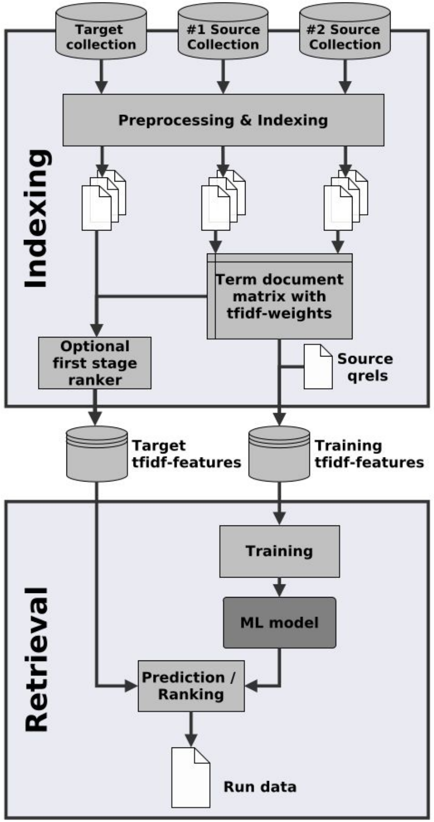


Meta-evaluations / reproducibility experiments

P’R’I’M’A’D

Measure	GC	YXL	BFFMSSS
AP	0.3711	0.4018	0.3612
KTU	1.0000	0.0086	0.0051
RBO	1.0000	0.1630	0.5747
RMSE	0.0000	0.1911	0.1071
p-value	1.0000	0.1009	0.7885

MRG_UWaterloo and WaterlooCormack Participation in the TREC 2017 Common Core Track*;*
Grossman and Cormack; TREC Common Core 2017
Simple techniques for cross-collection relevance feedback*;*
Yu, Xie, and Lin; ECIR 2019
How to measure the reproducibility of system-oriented IR experiments*;*
Breuer, Ferro, Fuhr, Maistro, Sakai, Schaer, Soboroff; SIGIR 2020



Future work

- Reduce labeling effort
  - Automatic run annotations
  - Integration into existing retrieval toolkits
- Sanity checks; prioritizing important fields
  - Verification of checksums, completeness, …
  - Assign requirement levels according to RFC2119
- Public database
  - Find baselines, conduct meta-evaluations, …
- How can we make it a community standard?
  - Collaborations with shared task organizers?
  - How can we motivate IR practitioners to annotate their experimental data? Reward?
  - TREC Deep Learning track as pioneering example

Thank you!

Website:	https://www.ir-metadata.org/
arXiv:	https://arxiv.org/abs/2207.08922
ACM DL:	https://dl.acm.org/doi/10.1145/3477495.3531738