

# The Information Retrieval Anthology

---

Martin Potthast<sup>1</sup>, Sebastian Günther<sup>2</sup>, Janek Bevendorff<sup>3</sup>, Jan Philipp Bittner<sup>2</sup>,  
Alexander Bondarenko<sup>2</sup>, Maik Fröbe<sup>2</sup>, Christian Kahmann<sup>1</sup>, Andreas Niekler<sup>1</sup>,  
Michael Völske<sup>3</sup>, Benno Stein<sup>3</sup>, Matthias Hagen<sup>2</sup>

<sup>1</sup>Leipzig University, <sup>2</sup>Martin-Luther-Universität Halle-Wittenberg, <sup>3</sup>Bauhaus-Universität Weimar

September 1, 2021

IR.webis.de

# The Information Retrieval Anthology

## Motivation

- ❑ General purpose academic search engines come with off-topic results
- ❑ No filtering: information overload
- ❑ Limited pagination past top-1000 results
- ❑ Dataset is a good starting point for development of further tools

*A collection and search tool to aid in IR-related research  
and to perform research on*

# The Information Retrieval Anthology

## Introduction

- ❑ The IR Anthology - available at [IR.webis.de](http://IR.webis.de)
- ❑ Web platform & search engine
- ❑ Based on the ACL Anthology
- ❑ Extensive modifications and additions
- ❑ Collections of IR-related data

# The Information Retrieval Anthology

## Browsing - Main Overview

IR Anthology

Full-text search...

q

The IR Anthology currently contains 53673 papers on the study of information retrieval.

### Conferences

Venue	2020+	2019 – 2010	2009 – 2000	1999 – 1990	1989 and older
AIRS	21	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04		
CHIIR/HCIR	21 20	19 18 17 16 13 12			
CIKM	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92	
CLEF	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00		
CORIA		19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04		
DESIRES	21	18			
ECIR	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97	
FIRE	21 20	19 18 17 16 15 14 13 12 11			
ICTIR	21 20	19 18 17 16 15 14 13 11	09 07		
IRFC		14 13 12 11 10			
ISMIR		19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00		
ICMR	20	19 18 17 16 15 14 13 12 11 10	08 07 06 05 04 03		
NTCIR	20	19 17 16 14 13 11 10	08 07 05 04 02 01	99	
RIAQ/OAIR		13 10	07 04 00	97 94 91	88 85
SIGIR	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79 78 73 71
SPIRE	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98	
TREC	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92	
WSDM	21 20	19 18 17 16 15 14 13 12 11 10	09 08		
WWW	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01		
Workshops	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 97 96	
Societies	ACM SIGIR   ACM SIGIR Tokyo   ARIA   ASIS&T   BCS IRSG   CLEF Initiative   GI FGIR   IRSI   ISMIR   MIR   PECAMP   ROMIP   WGI				
Schools	ASIRF   EARIA   ESSIR   RUSSIR				

### Journals

Venue	2020+	2019 – 2010	2009 – 2000	1999 – 1990	1989 and older
-------	-------	-------------	-------------	-------------	----------------

Table-based overview with venues, journals and grouping by year

# The Information Retrieval Anthology

## Browsing - Venues

IR Anthology



### Special Interest Group on Information Retrieval Conference

- 2020**
  - Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020 **383 papers**
- 2019**
  - Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019 **263 papers**
  - Proceedings of the Open-Source IR Replicability Challenge co-located with 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, OSIRRC@SIGIR 2019, Paris, France, July 25, 2019 **15 papers**
- 2018**
  - The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018 **252 papers**
- 2017**
  - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017 **256 papers**
- 2016**
  - Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016 **235 papers**
- 2015**
  - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015 **202 papers**
- 2014**
  - The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014 **230 papers**

Track listing of a single venue

# The Information Retrieval Anthology

## Browsing - Single Venue

IR Anthology

Full-text search...

Q

### Special Interest Group on Information Retrieval Conference (2020)

#### Contents

- Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020 **383 papers**

↑up

#### Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020

doi dblp	Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020
doi dblp	<b>The Next Generation of Neural Networks</b> Geoffrey E. Hinton
pdf doi dblp	<b>Proof by Experimentation? Towards Better IR Research</b> Norbert Fuhr
doi dblp	<b>Coopetition in IR Research</b> Ellen M. Voorhees

### Publications at one venue

# The Information Retrieval Anthology

## Browsing - Publication

IR Anthology

Full-text search...

Q

### Proof by Experimentation? Towards Better IR Research

Norbert Fuhr

**Anthology ID:** 2020.sigirconf\_conference-2020.2

**Volume:** [Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020](#)

**Year:** 2020

**Venue:** [sigirconf\\_conference](#)

**Publisher:** ACM

**Pages:** 2

**URL:** <https://doi.org/10.1145/3397271.3402426>

**DOI:** 10.1145/3397271.3402426

**DBLP:** [conf/sigir/Fuhr20](#)

**DBLP bib files:**

BibTeX

RIS

RDF N-Triples

RDF/XML

XML

Copy BibTeX to Clipboard

BibTeX

Google Scholar

Microsoft Academic

Semantic Scholar

Listing for one single publication

# The Information Retrieval Anthology

## Browsing - Author

IR Anthology

Full-text search...

Q

### Norbert Fuhr

2021

- doi dblp

**A Transparent Logical Framework for Aspect-Oriented Product Ranking Based on User Reviews**  
Firas Sabbah | Norbert Fuhr  
Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part I
- doi dblp

**Starting Conversations with Search Engines - Interfaces that Elicit Natural Language Queries**  
Andrea Papenmeier | Dagmar Kern | Daniel Hienert | Alfred Sliwa | Ahmet Aker | Norbert Fuhr  
CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021
- doi dblp

**Dataset of Natural Language Queries for E-Commerce**  
Andrea Papenmeier | Dagmar Kern | Daniel Hienert | Alfred Sliwa | Ahmet Aker | Norbert Fuhr  
CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021

2020

- doi dblp

**Proof by Experimentation? Towards Better IR Research**  
Norbert Fuhr

Search

Co-authors

Henrik Nottelmann9

Nicola Ferro8

Ian Soboroff6

Ulrich Pfeifer5

Thomas Rölleke4

show all...

Venues

SIGIR23

### Author profile



# The Information Retrieval Anthology

## Searching



### Entities in your query

Person

Norbert Fuhr

### Search results 1-10 for *fuhr evaluation*

Total results: 83 (retrieved in 1,775 ms)

#### Living Lab **Evaluation** for Life and Social Sciences Search Platforms - LiLAS

Schaer et al. · ECIR · 2021 ▾

**Fuhr** argues that **evaluation** initiatives should take a leading role in improving the current IR **evaluation** practice. We want to move beyond the traditional offline **evaluation** setup and bring together industry and practice **evaluation** techniques into the academic realm.

#### SIGIR 2013 workshop on modeling user behavior for information retrieval **evaluation**

Clarke et al. · SIGIR Forum · 2013 ▾

Tran and **Fuhr**, Markov Modeling for User Interaction in Retrieval. Working from within the framework of the Interactive Probability Ranking Principle, the authors derive Markov models from user interaction data gathered through logging and eye-tracking.

#### **Evaluation** Forum, CLEF 2000, Lisbon, Portugal, September 21-22, 2000,

CLEF · 2001 ▾

W. r. t. the schema above, for example the following queries can be issued: -Author sounds-like Norbert **Fuhr** asks for documents being authored by someone whose name sounds similar to Norbert **Fuhr**. -PubDate around-year 1999 asks for documents which have been published around year 1999.

## Searching the IR Anthology

# The Information Retrieval Anthology

## Author Disambiguation



fuhr evaluation author:sakai



### Entities in your query

Person	Person	Person
Norbert Fuhr	Tetsuya Sakai	Yoshifumi Sakai
	<u>author:sakai → author:"Tetsuya Sakai"</u>	author:sakai → author:"Yoshifumi Sakai"

### Search results 1-9 for *fuhr evaluation author:sakai*

Total results: 9 (retrieved in 1,663 ms)

#### SIGIR 2013 workshop on modeling user behavior for information retrieval **evaluation**

Clarke et al. · SIGIR Forum · 2013 ▾

Group members: Mikhail Ageev, Aleksandr Chuklin, Kathy Brennan, Gleb Gusev, Miguel Martinez, Iadh Ounis, Andrei Rikitienskii, Tetsuya **Sakai** (scribe), and Aleksei Tikhonov. • How do we validate **evaluation** measures?

#### Exploiting Stopping Time to Evaluate Accumulated Relevance

Ferrante and Ferro · ICTIR · 2020 ▾

Among others, **Fuhr** raised the issue of the scales adopted by **evaluation** measures and how they affect the validity of the computed statistics, like mean and variance. Ferrante et al.

#### Evaluating Web Search with a Bejeweled Player Model

Zhang et al. · SIGIR · 2017 ▾

In Section 2, we introduce our proposed **evaluation** framework based on the BPM.

## Entity detection within queries

# The Information Retrieval Anthology

## Filtering



fuhr evaluation author:"Tetsuya Sakai"



### Entities in your query

Person

Norbert Fuhr

Person

Tetsuya Sakai

### Search results 1-9 for *fuhr evaluation author:"Tetsuya Sakai"*

Total results: 9 (retrieved in 1,665 ms)

#### How to Measure the Reproducibility of System-oriented IR

Breuer et al. · SIGIR · 2020 ▾

#### SIGIR 2013 workshop on modeling user behavior for information retrieval **evaluation**

Clarke et al. · SIGIR Forum · 2013 ▾

Group members: Mikhail Ageev, Aleksandr Chuklin, Kathy Brennan, Gleb Gusev, Miguel Martinez, Iadh Ounis, Andrei Rikitienskii, Tetsuya **Sakai** (scribe), and Aleksei Tikhonov. • How do we validate **evaluation** measures?

#### Exploiting Stopping Time to Evaluate Accumulated Relevance

Ferrante and Ferro · ICTIR · 2020 ▾

Among others, **Fuhr** raised the issue of the scales adopted by **evaluation** measures and how they affect the validity of the computed statistics, like mean and variance. Ferrante et al.

#### Evaluating Web Search with a Bejeweled Player Model

## Search operators

# The Information Retrieval Anthology

## Initial Work

- ❑ Create collection of IR-related publications
- ❑ Enable scholars to browse the collection
- ❑ Develop search and recommender tools
- ❑ Expand this concept to IR-related data
  - Summer schools
  - Research groups
  - Datasets

# The Information Retrieval Anthology

## Corpus Construction

- ❑ Main metadata source: DBLP XML dump
- ❑ Filter for relevant conferences/journals
- ❑ Full text sources:
  - Webis-CSP-15
  - Ingestion of local collections
  - Crawling of missing publications
- ❑ Aggregation into common `ir-anthology.bib`<sup>1</sup> (metadata only)
- ❑ Indexing into Elasticsearch for search

---

<sup>1</sup><https://github.com/ir-anthology/ir-anthology-data>

# The Information Retrieval Anthology

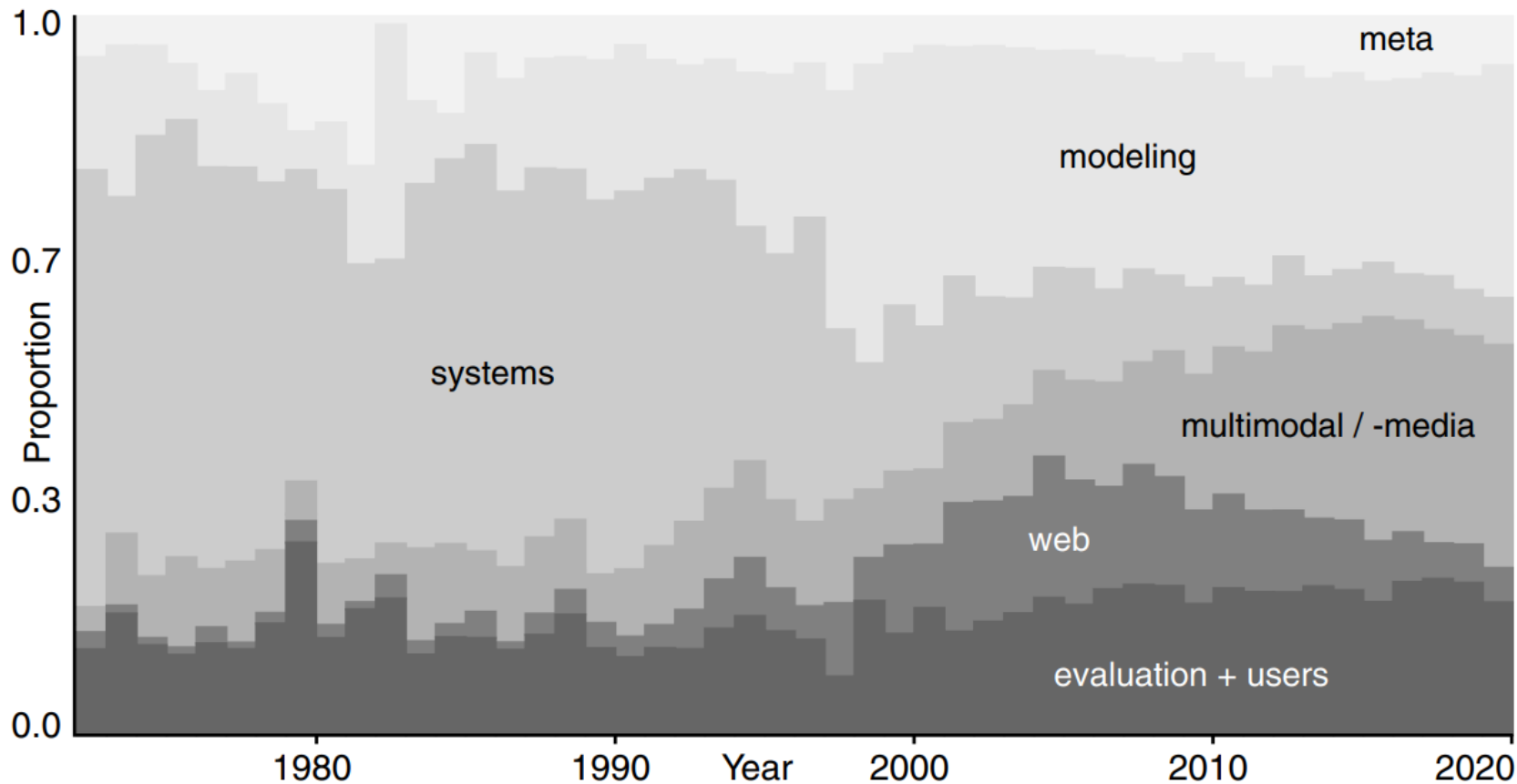
## Research

Topic Keywords (top 5)	Topic Category	Trend
metric, function, score, rank, ranking	evaluation+users	evergreen
student, health, group, study, course	evaluation+users	evergreen
similarity, algorithm, cluster, distance, space	modeling	evergreen
search, user, session, task, behavior	evaluation+users	neogreen
question, answer, tag, expert, qa	evaluation+users	neogreen
image, video, music, multimedia, photo	multimodal/media	neogreen
feature, classify, learning, training, label	modeling	rising
recommend, user, item, preference, filtering	modeling	rising
review, product, opinion, sentiment, trust	multimodal/media	rising
privacy, security, email, app, attack	multimodal/media	rising
ad, advertising, game, worker, market	multimodal/media	rising
location, city, sensor, mobility, region	multimodal/media	rising
event, news, tweet, medium, blog	multimodal/media	rising
network, graph, node, community, edge	multimodal/media	rising
evaluation, collection, test, retrieval, judgment	evaluation+users	peaked
document, topic, sentence, text, summarization	modeling	peaked
word, language, translation, text, speech	modeling	peaked
term, query, document, retrieval, model	modeling	peaked
page, web, search, engine, query	web	peaked
service, web, ontology, application, integration	web	peaked
index, xml, query, tree, database	systems	peaked
peer, server, load, latency, memory	systems	settled
program, system, software, programming, design	systems	settled
entity, copy, knowledge, fee, permission	meta	
research, workshop, tutorial, researcher, book	meta	

Topics by keyword identified within the corpus

# The Information Retrieval Anthology

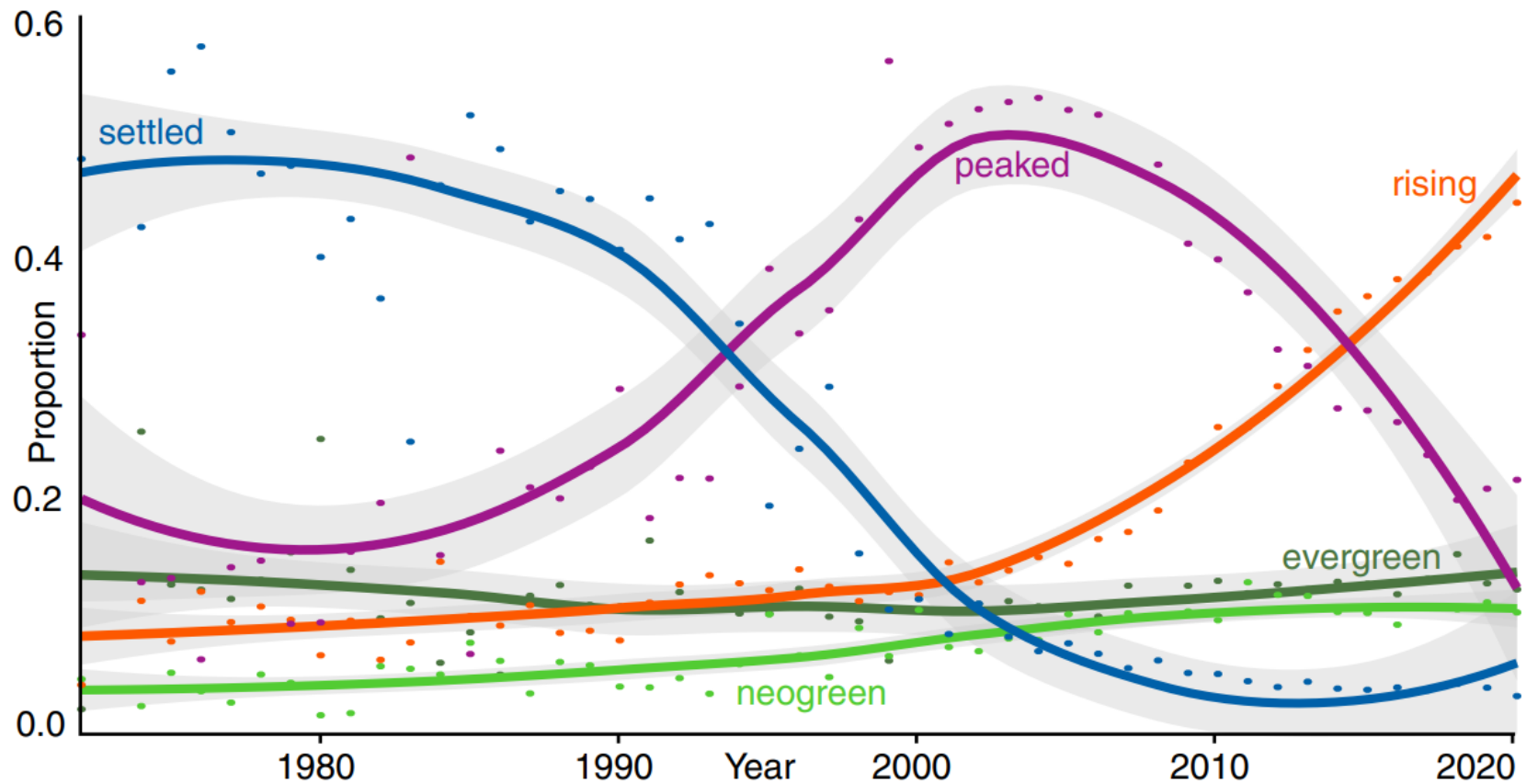
## Research



Topic groups aggregated over time

# The Information Retrieval Anthology

## Research



Topics aggregated by trend over time



# The Information Retrieval Anthology

## Ongoing Work

- ❑ Integrating new data sources & types
- ❑ Improving usability (navigation, UI)
- ❑ Detection and rectification of data issues
- ❑ Researching into search & recommendation using keyqueries

# The Information Retrieval Anthology

## Future Work

- ❑ Improve coverage
- ❑ Addition of new metadata sources
- ❑ Collection of full texts
- ❑ Extension to other data types (code, datasets, slides, shared tasks, etc.)
- ❑ Improvements to the ChatNoir search
- ❑ Shared tasks around the corpus using TIRA
- ❑ Long-term sustainability through community contributions

# The Information Retrieval Anthology

## Summary

- ❑ A platform to browse and search IR-related publications
- ❑ Centered around a metadata and full text collection
- ❑ Goal: easier access to relevant literature
- ❑ Basis for development of future tools and research

# The Information Retrieval Anthology

## Conclusion

- ❑ IR Anthology: [IR.webis.de](http://IR.webis.de)
- ❑ ChatNoir: [IR.chatnoir.eu](http://IR.chatnoir.eu)
- ❑ Feedback: <https://twitter.com/IRanthology>
- ❑ Bugs & suggestions: <https://github.com/ir-anthology/ir-anthology/issues>

Thank you!