

# Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp

M. Neumann<sup>1</sup> C. Michels<sup>2</sup> P. Schaer<sup>1</sup> R. Schenkel<sup>2</sup>

<sup>1</sup>Department of Information Science  
TH Köln (University of Applied Sciences)

<sup>2</sup>Department of Computer Science  
University of Trier

Lernen. Wissen. Daten. Analysen. LWDA 2019

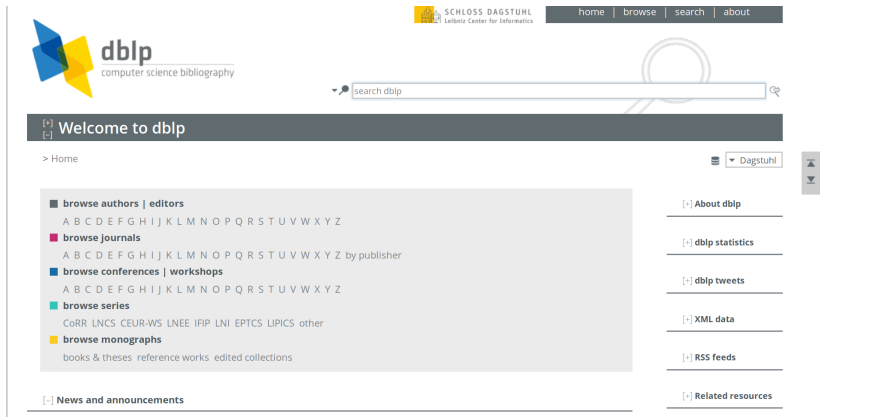
# Outline

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution

# Outline

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution

# The dblp Bibliography



The screenshot shows the dblp website homepage. At the top left is the dblp logo (a stylized 'd' made of blue and yellow triangles) and the text 'dblp computer science bibliography'. To the right is the 'SCHLOSS DAGSTUHL Leibniz Center for Informatics' logo and a navigation menu with 'home', 'browse', 'search', and 'about'. Below the logo is a search bar with the text 'search dblp'. A dark grey banner below the search bar contains the text 'Welcome to dblp'. Underneath, there is a '> Home' link and a 'Dagstuhl' dropdown menu. The main content area is a light grey box with several sections: 'browse authors | editors' with an alphabetical list 'A B C D E F G H I J K L M N O P Q R S T U V W X Y Z'; 'browse journals' with an alphabetical list 'A B C D E F G H I J K L M N O P Q R S T U V W X Y Z by publisher'; 'browse conferences | workshops' with an alphabetical list 'A B C D E F G H I J K L M N O P Q R S T U V W X Y Z'; 'browse series' with a list 'CoRR LNCS CEUR-WS LNEE IFIP LNI EPTCS LIPICS other'; and 'browse monographs' with a list 'books & theses reference works edited collections'. At the bottom of this box is a link for 'News and announcements'. On the right side, there is a vertical list of links: 'About dblp', 'dblp statistics', 'dblp tweets', 'XML data', 'RSS feeds', and 'Related resources'. A vertical navigation bar is visible on the far right.

<https://dblp.org/>

## "Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp."

Mandy Neumann et al. (2018)

> Home



Mandy Neumann, Christopher Michels, Philipp Schaer, Ralf Schenkel:





**Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp.** JCDL 2018: 45-48

# The dblp Bibliography

## "Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp."

Mandy Neumann et al. (2018)

> Home

-     Mandy Neumann, Christopher Michels, Philipp Schaer, Ralf Schenkel:  
**Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp.** JCDL 2018: 45-48

<https://doi.org/10.1145/3197026.3197069>

# The dblp Bibliography



Information retrieval

Search dblp

powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg

> Home

Dagstuhl

Venue search results

Likely matches

- International Semantic Web Conference (ISWC)
- Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)
- European Conference on Information Retrieval (ECIR)
- International Conference on Multimedia Retrieval (ICMR)  
also: ACM SIGMM International Workshop/Conference on Multimedia Information Retrieval (MIR)
- ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare (MIIRH @ MM)

show all 53 matches

Publication search results

found 10,883 matches

2019

- Razane Tajeddine:  
Private Information Retrieval from Coded Storage. Aalto University, Espoo, Finland 2019
- Kerstin Hartig:  
Entwicklung eines Information-Retrieval-Systems zur Unterstützung von Gefährdungs- und Risikoanalysen. Technical University of Berlin, Germany 2019
- David Maxwell:  
Modelling search and stopping in interactive information retrieval. University of Glasgow, UK 2019
- Omar Alonso:  
The Practice of Crowdsourcing. Synthesis Lectures on Information Concepts, Retrieval, and Services, Morgan & Claypool Publishers 2019
- Jiqun Liu, Chirag Shah:  
Interactive IR User Study Design, Evaluation, and Reporting. Synthesis Lectures on Information

Refine list




refine by author

Philipp Mayr (85)  
Norbert Fuhr (64)  
C. J. van Rijsbergen (64)  
Nicholas J. Belkin (64)  
W. Bruce Croft (62)  
Jian-Yun Nie (62)  
Fabio Crestani (60)  
Massimo Melucci (59)  
Gareth J. F. Jones (58)  
ChengKiang Zhai (57)  
15,718 more options

refine by venue

SIGIR (522)  
CoRR (368)  
Multimedia Information Retrieval (341)  
Inf. Process. Manage. (299)  
Information Storage and Retrieval (264)  
-----

# The dblp Bibliography

[+] An Nguyen 0001   

> Home > Persons

 by year  Dagstuhl

## [-] Person information

- *affiliation*: University of Sydney, Australia












## [-] Other persons with the same name

- An Nguyen
- An Nguyen 0002 — University of Texas at Arlington, Heracleia Human Centered Computing Laboratory
- An Nguyen 0003 — University of Southampton, UK
- An Nguyen 0004 — Kim Tu Dien Multilingual Data Center, Ho Chi Minh, Vietnam
- An Nguyen 0005 — University of Queensland, Brisbane, School of English, Media Studies and Art History

## [+] Other persons with a similar name

## [-] 2010 - today

### 2017

- [2]     Peter Eades, Seok-Hee Hong, An Nguyen, Karsten Klein:  
**Shape-Based Quality Metrics for Large Graph Visualization.** J. Graph Algorithms Appl. 21(1): 29-53 (2017)
- [1]     Andrew Kennedy, Karsten Klein, An Nguyen, Florence Ying Wang:  
**The Graph Landscape: using visual analytics for graph set analysis.** J. Visualization 20(3): 417-432 (2017)
- [c3]     An Nguyen, Seok-Hee Hong:  
**k-core based multi-level graph visualization for scale-free networks.** PacificVis 2017: 21-25

## [-] Refine list

showing all 5 records

### refine by search term

### refine by type

- Journal Articles (only)
  - Conference and Workshop Papers (only)
- select all | deselect all

### refine by coauthor

Karsten Klein 0001 (4)  
Seok-Hee Hong (3)





# The dblp Bibliography




## Journal of Librarianship and Information Science

> Home > Journals





JOLIS @ SAGE Publications

- Volume 51: 2019
- Volume 50: 2018
- Volume 49: 2017
- Volume 48: 2016
- Volume 47: 2015
- Volume 46: 2014
- Volume 45: 2013
- Volume 44: 2012
- Volume 43: 2011
- Volume 42: 2010
- Volume 41: 2009
- Volume 40: 2008
- Volume 39: 2007
- Volume 38: 2006
- Volume 37: 2005
- Volume 36: 2004
- Volume 35: 2003





### LWDA 2019: Berlin, Germany

-     Robert Jäschke, Matthias Weidlich:  
**Proceedings of the Conference "Lernen, Wissen, Daten, Analysen", Berlin, Germany, September 30 - October 2, 2019.** CEUR Workshop Proceedings 2454, CEUR-WS.org 2019 [contents]

### LWDA 2018: Mannheim, Germany




-     Rainer Gemulla, Simone Paolo Ponzetto, Christian Bizer, Margret Keuper, Heiner Stuckenschmidt:  
**Proceedings of the Conference "Lernen, Wissen, Daten, Analysen", LWDA 2018, Mannheim, Germany, August 22-24, 2018.** CEUR Workshop Proceedings 2191, CEUR-WS.org 2018 [contents]

### LWDA 2017: Rostock, Germany

-     Michael Leyer:  
**Lernen, Wissen, Daten, Analysen (LWDA) Conference Proceedings, Rostock, Germany, September 11-13, 2017.** CEUR Workshop Proceedings 1917, CEUR-WS.org 2017 [contents]



### LWDA 2016: Potsdam, Germany

# The dblp Bibliography



LWDA 2019: Berlin, Germany   

> Home > Conferences and Workshops > LWDA

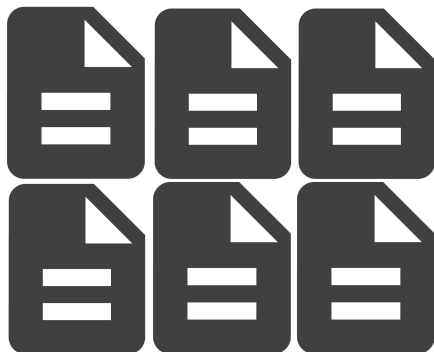
 Dagstuhl

-     Robert Jäschke, Matthias Weidlich:  
**Proceedings of the Conference on "Lernen, Wissen, Daten, Analysen", Berlin, Germany, September 30 - October 2, 2019.** CEUR Workshop Proceedings 2454, CEUR-WS.org 2019

## DB Research Papers

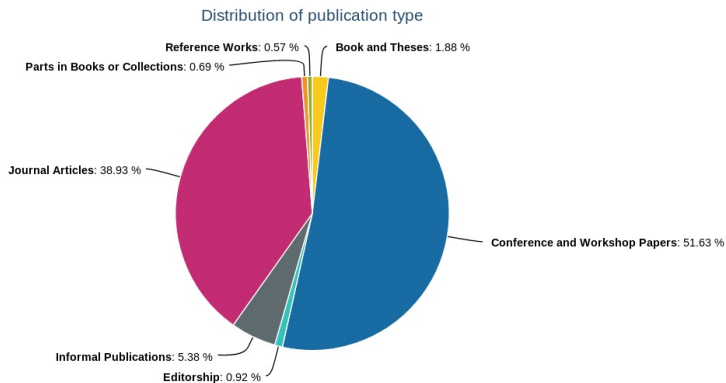
-     Triet Doan, Lena Wiese, Sven Bingert, Ramin Yahyapour:  
**A Graph Database for Persistent Identifiers.** 1-5
-     Lan Jiang, Gerardo Vitagliano, Felix Naumann:  
**A Scoring-based Approach for Data Preparator Suggestion.** 6-9
-     Mohammad Mahdavi, Felix Neutatz, Larysa Visengeriyeva, Ziawasch Abedjan:  
**Towards Automated Data Cleaning Workflows.** 10-19
-     Dennis Marten, Holger Meyer, Andreas Heuer:  
**Database Support for Automotive Analysis.** 20-24
-     Mark Lukas Möller, Melke Klettke, Uta Störi:  
**Keeping NoSQL Databases Up to Date - Semantics of Evolution Operations and their Impact on Data Quality.** 25-37
-     Stefanie Scherzinger:  
**Have your Students build their own Mini Hive in just Eight Weeks.** 38-41
-     Deter K. Schwab, Maximilian S. Langohr, Jonas Röhrig, Damian F. Vöhringer, Andreas M. Wahl, Klara Meyer-Wegener

# Maintaining the dblp Bibliography



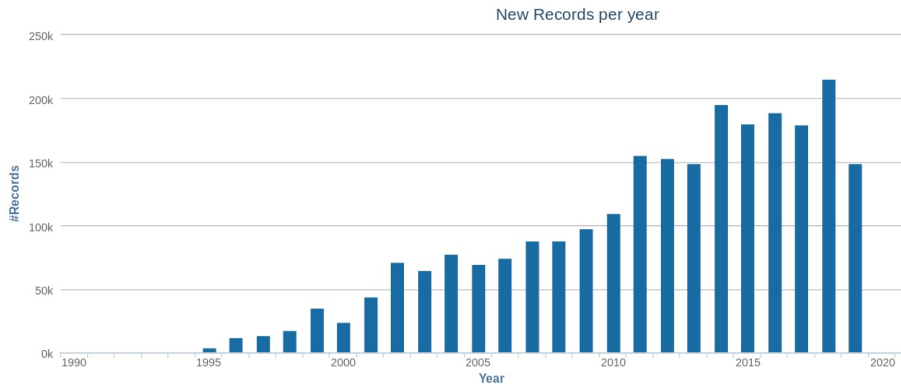
4.7m

# Maintaining the dblp Bibliography

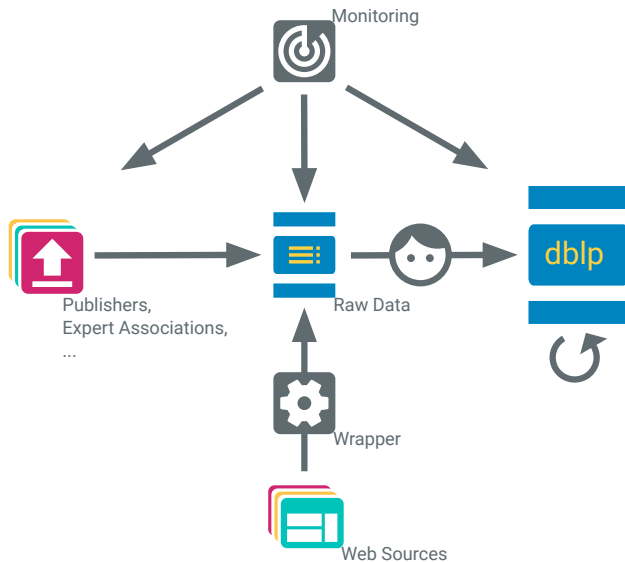


# Maintaining the dblp Bibliography

New entries to the database per year: conference and workshop papers



# Maintaining the dblp Bibliography



# Keeping the Archive up to Date – Obstacles

- limited resources
- conference proceedings
- implicit relevance decisions





# Keeping the Archive up to Date – Obstacles

- limited resources
- conference proceedings
- implicit relevance decisions



# Keeping the Archive up to Date – Obstacles

- limited resources
- conference proceedings
- implicit relevance decisions



# Keeping the Archive up to Date – Obstacles

- limited resources
- conference proceedings
- implicit relevance decisions



# Keeping the Archive up to Date – Obstacles

- limited resources
- conference proceedings
- implicit relevance decisions



# Outline

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution

# Research Question

How can we find a prioritization mechanism for conference series with regard to their expected urgency for the data acquisition process at a given point in time?

→ Ranking problem: have the conferences for which an update is expected next ranked highest

# Outline

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution

# Method

In fact, we are trying to *model* the relevance decisions of our archive curators.





# Method

Basis: temporal patterns

- base calculations on patterns from historical data
- use relation between past event dates and dates of entry to dblp



2014

January	February	March	April
1 2 3 4 5	1 2	1 2	1 2 3 4 5 6
6 7 8 9 10 11 12	3 4 5 6 7 8 9	3 4 5 6 7 8 9	7 8 9 10 11 12 13
13 14 15 16 17 18 19	10 11 12 13 14 15 16	10 11 12 13 14 15 16	14 15 16 17 18 19 20
20 21 22 23 24 25 26	17 18 19 20 21 22 23	17 18 19 20 21 22 23	21 22 23 24 25 26 27
27 28 29 30 31	24 25 26 27 28	24 25 26 27 28 29 30	28 29 30
		31	

May	June	July	August
1 2 3 4	1	1 2 3 4 5 6	1 2 3
5 6 7 8 9 10 11	2 3 4 5 6 7 8	7 8 9 10 11 12 13	4 5 6 7 8 9 10
12 13 14 15 16 17 18	9 10 11 12 13 14 15	14 15 16 17 18 19 20	11 12 13 14 15 16 17
19 20 21 22 23 24 25	16 17 18 19 20 21 22	21 22 23 24 25 26 27	18 19 20 21 22 23 24
26 27 28 29 30 31	23 24 25 26 27 28 29	28 29 30 31	25 26 27 28 29 30 31
	30		

September	October	November	December
1 2 3 4 5 6 7	1 2 3 4 5	1 2	1 2 3 4 5 6 7
8 9 10 11 12 13 14	6 7 8 9 10 11 12	3 4 5 6 7 8 9	8 9 10 11 12 13 14
15 16 17 18 19 20 21	13 14 15 16 17 18 19	10 11 12 13 14 15 16	15 16 17 18 19 20 21
22 23 24 25 26 27 28	20 21 22 23 24 25 26	17 18 19 20 21 22 23	22 23 24 25 26 27 28
29 30	27 28 29 30 31	24 25 26 27 28 29 30	29 30 31



# 2015

January	February	March	April
1 2 3 4			1 2 3 4 5
5 6 7 8 9 10 11	2 3 4 5 6 7 8	2 3 4 5 6 7 8	6 7 8 9 10 11 12
12 13 14 15 16 17 18	9 10 11 12 13 14 15	9 10 11 12 13 14 15	13 14 15 16 17 18 19
19 20 21 22 23 24 25	16 17 18 19 20 21 22	16 17 18 19 20 21 22	20 21 22 23 24 25 26
26 27 28 29 30 31	23 24 25 26 27 28	23 24 25 26 27 28 29 30 31	27 28 29 30
May	June	July	August
	1 2 3 4 5 6 7	1 2 3 4 5	1 2
4 5 6 7 8 9 10	8 9 10 11 12 13 14	6 7 8 9 10 11 12	3 4 5 6 7 8 9
11 12 13 14 15 16 17	15 16 17 18 19 20 21	13 14 15 16 17 18 19	10 11 12 13 14 15 16
18 19 20 21 22 23 24	22 23 24 25 26 27 28	20 21 22 23 24 25 26	17 18 19 20 21 22 23
25 26 27 28 29 30 31	29 30	27 28 29 30 31	24 25 26 27 28 29 30 31
September	October	November	December
1 2 3 4 5 6	1 2 3 4		1 2 3 4 5 6
7 8 9 10 11 12 13	5 6 7 8 9 10 11	2 3 4 5 6 7 8	7 8 9 10 11 12 13
14 15 16 17 18 19 20	12 13 14 15 16 17 18	9 10 11 12 13 14 15	14 15 16 17 18 19 20
21 22 23 24 25 26 27	19 20 21 22 23 24 25	16 17 18 19 20 21 22	21 22 23 24 25 26 27
28 29 30	26 27 28 29 30 31	23 24 25 26 27 28 29 30	28 29 30 31

# 2016

January	February	March	April
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

May	June	July	August
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

September	October	November	December
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

2017

January	February	March	April
1	1 2 3 4 5	1 2 3 4 5	1 2
2 3 4 5 6 7 8	6 7 8 9 10 11 12	6 7 8 9 10 11 12	3 4 5 6 7 8 9
9 10 11 12 13 14 15	13 14 15 16 17 18 19	13 14 15 16 17 18 19	10 11 12 13 14 15 16
16 17 18 19 20 21 22	20 21 22 23 24 25 26	20 21 22 23 24 25 26	17 18 19 20 21 22 23
23 24 25 26 27 28 29	27 28	27 28 29 30 31	24 25 26 27 28 29 30
30 31			

May	June	July	August
1 2 3 4 5 6 7	1 2 3 4	1 2	1 2 3 4 5 6
8 9 10 11 12 13 14	5 6 7 8 9 10 11	3 4 5 6 7 8 9	7 8 9 10 11 12 13
15 16 17 18 19 20 21	12 13 14 15 16 17 18	10 11 12 13 14 15 16	14 15 16 17 18 19 20
22 23 24 25 26 27 28	19 20 21 22 23 24 25	17 18 19 20 21 22 23	21 22 23 24 25 26 27
29 30 31	26 27 28 29 30	24 25 26 27 28 29 30	28 29 30 31

September	October	November	December
1 2 3	1	1 2 3 4 5	1 2 3
4 5 6 7 8 9 10	2 3 4 5 6 7 8	6 7 8 9 10 11 12	4 5 6 7 8 9 10
11 12 13 14 15 16 17	9 10 11 12 13 14 15	13 14 15 16 17 18 19	11 12 13 14 15 16 17
18 19 20 21 22 23 24	16 17 18 19 20 21 22	20 21 22 23 24 25 26	18 19 20 21 22 23 24
25 26 27 28 29 30	23 24 25 26 27 28 29	27 28 29 30	25 26 27 28 29 30 31
	30 31		



2018

January  
1 2 3 4 5 6 7  
8 9 10 11 12 13 14  
15 16 17 18 19 20 21  
22 23 24 25 26 27 28  
29 30 31

February  
1 2 3 4  
5 6 7 8 9 10 11  
12 13 14 15 16 17 18  
19 20 21 22 23 24 25  
26 27 28

March  
1 2 3 4  
5 6 7 8 9 10 11  
12 13 14 15 16 17 18  
19 20 21 22 23 24 25  
26 27 28 29 30 31

April  
1  
2 3 4 5 6 7 8  
9 10 11 12 13 14 15  
16 17 18 19 20 21 22  
23 24 25 26 27 28 29  
30

May  
1 2 3 4 5 6  
7 8 9 10 11 12 13  
14 15 16 17 18 19 20  
21 22 23 24 25 26 27  
28 29 30 31

June  
1 2 3  
4 5 6 7 8 9 10  
11 12 13 14 15 16 17  
18 19 20 21 22 23 24  
25 26 27 28 29 30

July  
1  
2 3 4 5 6 7 8  
9 10 11 12 13 14 15  
16 17 18 19 20 21 22  
23 24 25 26 27 28 29  
30 31

August  
1 2 3 4 5  
6 7 8 9 10 11 12  
13 14 15 16 17 18 19  
20 21 22 23 24 25 26  
27 28 29 30 31

September  
1 2  
3 4 5 6 7 8 9  
10 11 12 13 14 15 16  
17 18 19 20 21 22 23  
24 25 26 27 28 29 30

October  
1 2 3 4 5 6 7  
8 9 10 11 12 13 14  
15 16 17 18 19 20 21  
22 23 24 25 26 27 28  
29 30 31

November  
1 2 3 4  
5 6 7 8 9 10 11  
12 13 14 15 16 17 18  
19 20 21 22 23 24 25  
26 27 28 29 30

December  
1 2  
3 4 5 6 7 8 9  
10 11 12 13 14 15 16  
17 18 19 20 21 22 23  
24 25 26 27 28 29 30  
31



# Method

- Example conference:
  - interval: 1 year
  - usual month: June
  - usual delay: 3 months
- expected: September 2019
- 180 other conferences also due in September
- base scoring: delay between expected and current date

# Method

## Ranking factors and data sets:

- conference rating
- citation counts
- activity indicator
- internationality
- author prominence
- conference size
- log



# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts
- activity indicator
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts
- activity indicator
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp



# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence: dblp data
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence: dblp data
- conference size
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

## Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence: dblp data
- conference size: dblp data
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence: dblp data
- conference size: dblp data
- log

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Method

Ranking factors and data sets:

- conference rating: CORE; Martins et al.<sup>1</sup>
- citation counts: Microsoft Academic Graph (MAG)
- activity indicator: dblp data
- internationality: dblp data
- author prominence: dblp data
- conference size: dblp data
- log: dblp data

---

<sup>1</sup>W. S. Martins et al. "Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science"  dblp

# Evaluation

Gold standard: pseudo-relevance:

- distance in months between current month and month of ingestion into dblp
- inverted to give higher values to more recent entries

# Outline

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution**

# Main Results

Overview on ndcg-100 values for each evaluated month with the average.

system	may	jun	jul	aug	sep	oct	nov	avg
base	0.306	0.504	0.583	0.451	0.463	0.457	0.559	0.475
active	0.304	0.486	<b>0.655</b>	<b>0.564</b>	<b>0.582</b>	<b>0.556</b>	<b>0.601</b>	<b>0.536</b>
rate	<b>0.354</b>	<b>0.529</b>	<b>0.615</b>	<b>0.466</b>	<b>0.466</b>	<b>0.518</b>	<b>0.628</b>	<b>0.511</b>
size	<b>0.309</b>	0.493	0.575	0.444	0.463	<b>0.461</b>	<b>0.566</b>	0.473
intl	0.177	0.352	0.394	0.363	0.387	0.319	0.538	0.361
affil	0.289	0.489	0.577	0.435	0.461	<b>0.484</b>	<b>0.563</b>	0.471
cite	0.305	0.495	0.574	0.443	0.460	0.456	0.552	0.469
prom	<b>0.312</b>	0.452	0.571	<b>0.481</b>	<b>0.534</b>	<b>0.519</b>	<b>0.591</b>	<b>0.494</b>
log	<b>0.311</b>	0.494	0.583	<b>0.476</b>	<b>0.466</b>	0.456	<b>0.577</b>	<b>0.480</b>



# Main Results

Overview on ndcg-100 values for each evaluated month with the average.

system	may	jun	jul	aug	sep	oct	nov	avg
base	0.306	0.504	0.583	0.451	0.463	0.457	0.559	0.475
active	0.304	0.486	<b>0.655</b>	<b>0.564</b>	<b>0.582</b>	<b>0.556</b>	<b>0.601</b>	<b>0.536</b>
rate	<b>0.354</b>	<b>0.529</b>	<b>0.615</b>	<b>0.466</b>	<b>0.466</b>	<b>0.518</b>	<b>0.628</b>	<b>0.511</b>
size	<b>0.309</b>	0.493	0.575	0.444	0.463	<b>0.461</b>	<b>0.566</b>	0.473
intl	0.177	0.352	0.394	0.363	0.387	0.319	0.538	0.361
affil	0.289	0.489	0.577	0.435	0.461	<b>0.484</b>	<b>0.563</b>	0.471
cite	0.305	0.495	0.574	0.443	0.460	0.456	0.552	0.469
prom	<b>0.312</b>	0.452	0.571	<b>0.481</b>	<b>0.534</b>	<b>0.519</b>	<b>0.591</b>	<b>0.494</b>
log	<b>0.311</b>	0.494	0.583	<b>0.476</b>	<b>0.466</b>	0.456	<b>0.577</b>	<b>0.480</b>

# Main Results

Overview on ndcg-100 values for each evaluated month with the average.

system	may	jun	jul	aug	sep	oct	nov	avg
base	0.306	0.504	0.583	0.451	0.463	0.457	0.559	0.475
active	0.304	0.486	<b>0.655</b>	<b>0.564</b>	<b>0.582</b>	<b>0.556</b>	<b>0.601</b>	<b>0.536</b>
rate	<b>0.354</b>	<b>0.529</b>	<b>0.615</b>	<b>0.466</b>	<b>0.466</b>	<b>0.518</b>	<b>0.628</b>	<b>0.511</b>
size	<b>0.309</b>	0.493	0.575	0.444	0.463	<b>0.461</b>	<b>0.566</b>	0.473
intl	0.177	0.352	0.394	0.363	0.387	0.319	0.538	0.361
affil	0.289	0.489	0.577	0.435	0.461	<b>0.484</b>	<b>0.563</b>	0.471
cite	0.305	0.495	0.574	0.443	0.460	0.456	0.552	0.469
prom	<b>0.312</b>	0.452	0.571	<b>0.481</b>	<b>0.534</b>	<b>0.519</b>	<b>0.591</b>	<b>0.494</b>
log	<b>0.311</b>	0.494	0.583	<b>0.476</b>	<b>0.466</b>	0.456	<b>0.577</b>	<b>0.480</b>

# Main Results

Overview on ndcg-100 values for each evaluated month with the average.

system	may	jun	jul	aug	sep	oct	nov	avg
base	0.306	0.504	0.583	0.451	0.463	0.457	0.559	0.475
active	0.304	0.486	<b>0.655</b>	<b>0.564</b>	<b>0.582</b>	<b>0.556</b>	<b>0.601</b>	<b>0.536</b>
rate	<b>0.354</b>	<b>0.529</b>	<b>0.615</b>	<b>0.466</b>	<b>0.466</b>	<b>0.518</b>	<b>0.628</b>	<b>0.511</b>
size	<b>0.309</b>	0.493	0.575	0.444	0.463	<b>0.461</b>	<b>0.566</b>	0.473
intl	0.177	0.352	0.394	0.363	0.387	0.319	0.538	0.361
affil	0.289	0.489	0.577	0.435	0.461	<b>0.484</b>	<b>0.563</b>	0.471
cite	0.305	0.495	0.574	0.443	0.460	0.456	0.552	0.469
prom	<b>0.312</b>	0.452	0.571	<b>0.481</b>	<b>0.534</b>	<b>0.519</b>	<b>0.591</b>	<b>0.494</b>
log	<b>0.311</b>	0.494	0.583	<b>0.476</b>	<b>0.466</b>	0.456	<b>0.577</b>	<b>0.480</b>

# Main Results

Overview on ndcg-100 values for each evaluated month with the average.

system	may	jun	jul	aug	sep	oct	nov	avg
base	0.306	0.504	0.583	0.451	0.463	0.457	0.559	0.475
active	0.304	0.486	<b>0.655</b>	<b>0.564</b>	<b>0.582</b>	<b>0.556</b>	<b>0.601</b>	<b>0.536</b>
rate	<b>0.354</b>	<b>0.529</b>	<b>0.615</b>	<b>0.466</b>	<b>0.466</b>	<b>0.518</b>	<b>0.628</b>	<b>0.511</b>
size	<b>0.309</b>	0.493	0.575	0.444	0.463	<b>0.461</b>	<b>0.566</b>	0.473
intl	0.177	0.352	0.394	0.363	0.387	0.319	0.538	0.361
affil	0.289	0.489	0.577	0.435	0.461	<b>0.484</b>	<b>0.563</b>	0.471
cite	0.305	0.495	0.574	0.443	0.460	0.456	0.552	0.469
prom	<b>0.312</b>	0.452	0.571	<b>0.481</b>	<b>0.534</b>	<b>0.519</b>	<b>0.591</b>	<b>0.494</b>
log	<b>0.311</b>	0.494	0.583	<b>0.476</b>	<b>0.466</b>	0.456	<b>0.577</b>	<b>0.480</b>

# Main Results

Comparison of ndcg values on different cut-offs.

system	ndcg10	ndcg20	ndcg100	ndcg200
base	0.566	0.533	0.475	0.396
active (A)	<b>0.615</b>	<b>0.588</b>	<b>0.536*</b>	<b>0.517**</b>
rate (R)	<b>0.567</b>	<b>0.576</b>	<b>0.511**</b>	<b>0.478**</b>
size	<b>0.577</b>	<b>0.540</b>	0.473	<b>0.398</b>
intl	0.304	0.330	0.361	0.281
affil	0.543	0.523	0.471	<b>0.400</b>
cite	0.566	0.533	0.469	0.395
prom (P)	0.535	<b>0.553</b>	<b>0.494</b>	<b>0.485***</b>
log	0.560	0.528	<b>0.480</b>	<b>0.407*</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs.

system	ndcg10	ndcg20	ndcg100	ndcg200
base	0.566	0.533	0.475	0.396
active (A)	<b>0.615</b>	<b>0.588</b>	<b>0.536*</b>	<b>0.517**</b>
rate (R)	<b>0.567</b>	<b>0.576</b>	<b>0.511**</b>	<b>0.478**</b>
size	<b>0.577</b>	<b>0.540</b>	0.473	<b>0.398</b>
intl	0.304	0.330	0.361	0.281
affil	0.543	0.523	0.471	<b>0.400</b>
cite	0.566	0.533	0.469	0.395
prom (P)	0.535	<b>0.553</b>	<b>0.494</b>	<b>0.485***</b>
log	0.560	0.528	<b>0.480</b>	<b>0.407*</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs.

system	ndcg10	ndcg20	ndcg100	ndcg200
base	0.566	0.533	0.475	0.396
active (A)	<b>0.615</b>	<b>0.588</b>	<b>0.536*</b>	<b>0.517**</b>
rate (R)	<b>0.567</b>	<b>0.576</b>	<b>0.511**</b>	<b>0.478**</b>
size	<b>0.577</b>	<b>0.540</b>	0.473	<b>0.398</b>
intl	0.304	0.330	0.361	0.281
affil	0.543	0.523	0.471	<b>0.400</b>
cite	0.566	0.533	0.469	0.395
prom (P)	0.535	<b>0.553</b>	<b>0.494</b>	<b>0.485***</b>
log	0.560	0.528	<b>0.480</b>	<b>0.407*</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs.

system	ndcg10	ndcg20	ndcg100	ndcg200
base	0.566	0.533	0.475	0.396
active (A)	<b>0.615</b>	<b>0.588</b>	<b>0.536*</b>	<b>0.517**</b>
rate (R)	<b>0.567</b>	<b>0.576</b>	<b>0.511**</b>	<b>0.478**</b>
size	<b>0.577</b>	<b>0.540</b>	0.473	<b>0.398</b>
intl	0.304	0.330	0.361	0.281
affil	0.543	0.523	0.471	<b>0.400</b>
cite	0.566	0.533	0.469	0.395
prom (P)	0.535	<b>0.553</b>	<b>0.494</b>	<b>0.485***</b>
log	0.560	0.528	<b>0.480</b>	<b>0.407*</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$



# Main Results

Comparison of ndcg values on different cut-offs for combined factors.

system	ndcg20	ndcg30	ndcg100	ndcg200
base	0.533	0.511	0.475	0.396
R × A	<b>0.628*</b>	<b>0.629***</b>	<b>0.561***</b>	<b>0.547***</b>
R × P	<b>0.582</b>	<b>0.603**</b>	<b>0.550**</b>	<b>0.498***</b>
R × A × P	<b>0.587</b>	<b>0.606**</b>	<b>0.574***</b>	<b>0.542***</b>
A × P	<b>0.547</b>	<b>0.532</b>	<b>0.529**</b>	<b>0.509***</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs for combined factors.

system	ndcg20	ndcg30	ndcg100	ndcg200
base	0.533	0.511	0.475	0.396
<b>R × A</b>	<b>0.628*</b>	<b>0.629***</b>	<b>0.561***</b>	<b>0.547***</b>
R × P	<b>0.582</b>	<b>0.603**</b>	<b>0.550**</b>	<b>0.498***</b>
R × A × P	<b>0.587</b>	<b>0.606**</b>	<b>0.574***</b>	<b>0.542***</b>
A × P	<b>0.547</b>	<b>0.532</b>	<b>0.529**</b>	<b>0.509***</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs for combined factors.

system	ndcg20	ndcg30	ndcg100	ndcg200
base	0.533	0.511	0.475	0.396
R × A	<b>0.628*</b>	<b>0.629***</b>	<b>0.561***</b>	<b>0.547***</b>
R × P	<b>0.582</b>	<b>0.603**</b>	<b>0.550**</b>	<b>0.498***</b>
<b>R × A × P</b>	<b>0.587</b>	<b>0.606**</b>	<b>0.574***</b>	<b>0.542***</b>
A × P	<b>0.547</b>	<b>0.532</b>	<b>0.529**</b>	<b>0.509***</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Main Results

Comparison of ndcg values on different cut-offs for combined factors.

system	ndcg20	ndcg30	ndcg100	ndcg200
base	0.533	0.511	0.475	0.396
R × A	<b>0.628*</b>	<b>0.629***</b>	<b>0.561***</b>	<b>0.547***</b>
R × P	<b>0.582</b>	<b>0.603**</b>	<b>0.550**</b>	<b>0.498***</b>
R × A × P	<b>0.587</b>	<b>0.606**</b>	<b>0.574***</b>	<b>0.542***</b>
A × P	<b>0.547</b>	<b>0.532</b>	<b>0.529**</b>	<b>0.509***</b>

\*\*\* =  $p < 0.001$ , \*\* =  $p < 0.01$ , \* =  $p < 0.05$

# Interpretation

Best performing factors in terms of information quality:

- **credibility:**
  - expressed through ratings
- **currency:**
  - expressed through activity
- **popularity:**
  - expressed through prominence and logs

# Limitations

- ratings available for only about 20% of the data
- log data hard to interpret

# Summary

- We can use information quality-related features to rank conferences for data ingestion routines.
- All proposed features outperform the baseline derived from ingestion delays.
- Outlook
  - separate workshops
  - extend approach to journals etc.
  - Learning to Rank

# Discussion

Thank you for your attention!  
Feel free to ask any questions now!

## Contact us:

mandy.neumann@th-koeln.de

michelsc@uni-trier.de

philipp.schaer@th-koeln.de

schenkel@uni-trier.de

Visit <https://dblp.org>

Read the paper: <https://doi.org/10.1145/3197026.3197069>



# Table of contents

- 1 Motivation
- 2 Research Question
- 3 Method
- 4 Our Results/Contribution