

Project DIS18

01 – Introduction

Jüri Keller and Philipp Schaer

24-11-07 – Cologne, Germany

<https://ir.web.th-koeln.de>



Technology
Arts Sciences
TH Köln



Cologne Information Retrieval Research Group (CIR)



Prof. Dr. Philipp Schaer

- Information Retrieval, IR evaluation

Prof. Dr. Christin Kreutz (TH Mittelhessen)

- Scientometrics, metadata, recommender

Prof. Dr. Narjes Nikzad (GISMA University)

- Deep Learning, NLP, knowledge graphs

Dr. Timo Breuer

- Living Labs, reproducibility

Fabian Haak, M.Sc.

- Bias in query suggestions, language models

Björn Engelmann, M.Sc.

- Information extraction, table retrieval

Jüri Keller, M.Sc.

- Continuous and longitudinal IR evaluation

Andreas Kruff, M.Sc.

- Simulation of IR user interactions

Open Postions @ CIR

Hands-on experience in applied research

- Deepen skills you learned during study
- Programming (Python), data annotations, data cleaning, literature research, etc.
- Find spin-off topics for your thesis

Hard facts

- Flexible/student-friendly working hours
- 8-10 hours/week and €15 hourly
- 12 months contract, extension possible

Coffee and Kamelle

- Nice & friendly place to work + great team!
- apply@cir-group.org



Your Homework

TASK:

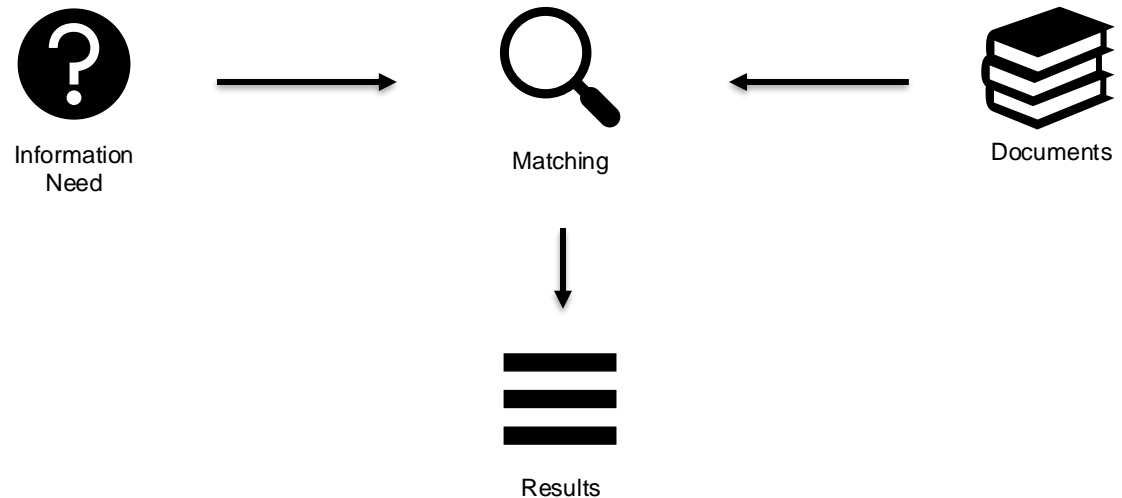
Bereiten Sie mindestens drei inhaltliche Fragen zum Text vor, die bei Ihnen nach der Lektüre offengeblieben sind.



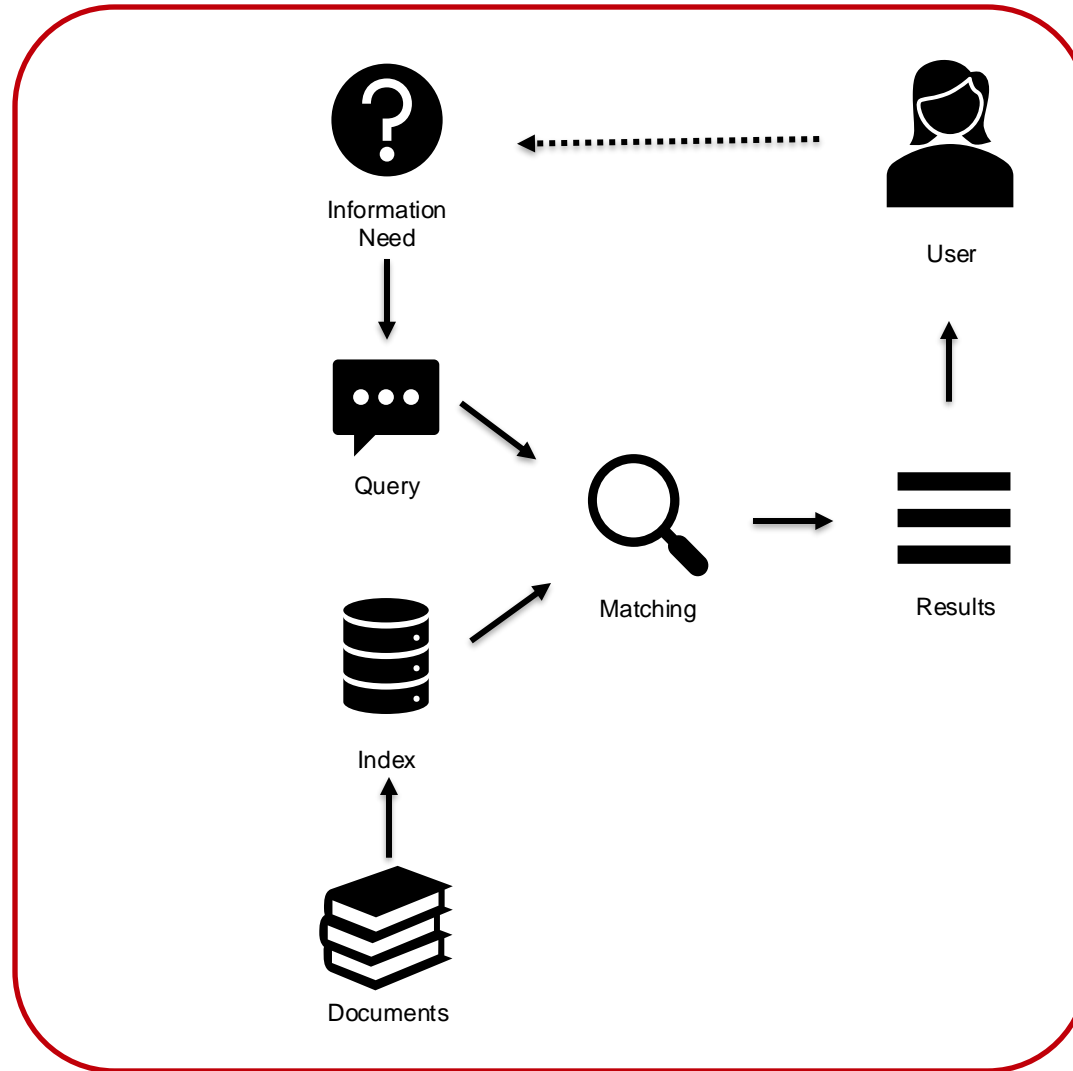
Information Retrieval Problem

“Information retrieval (IR) is **finding material** (usually documents) of an unstructured nature (usually text) that **satisfies an information need** from within large collections (usually stored on computers).”

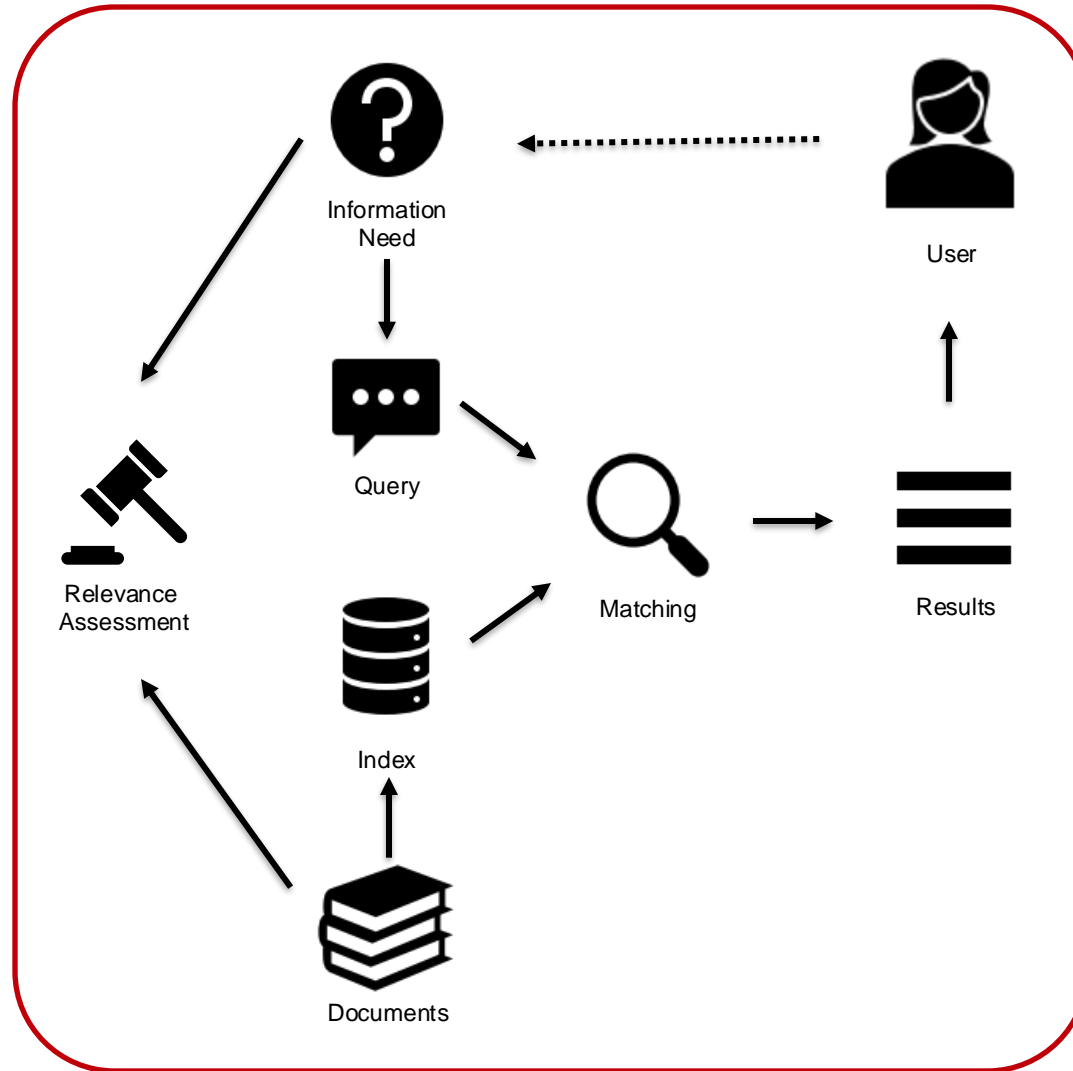
– Manning et al. (2008)



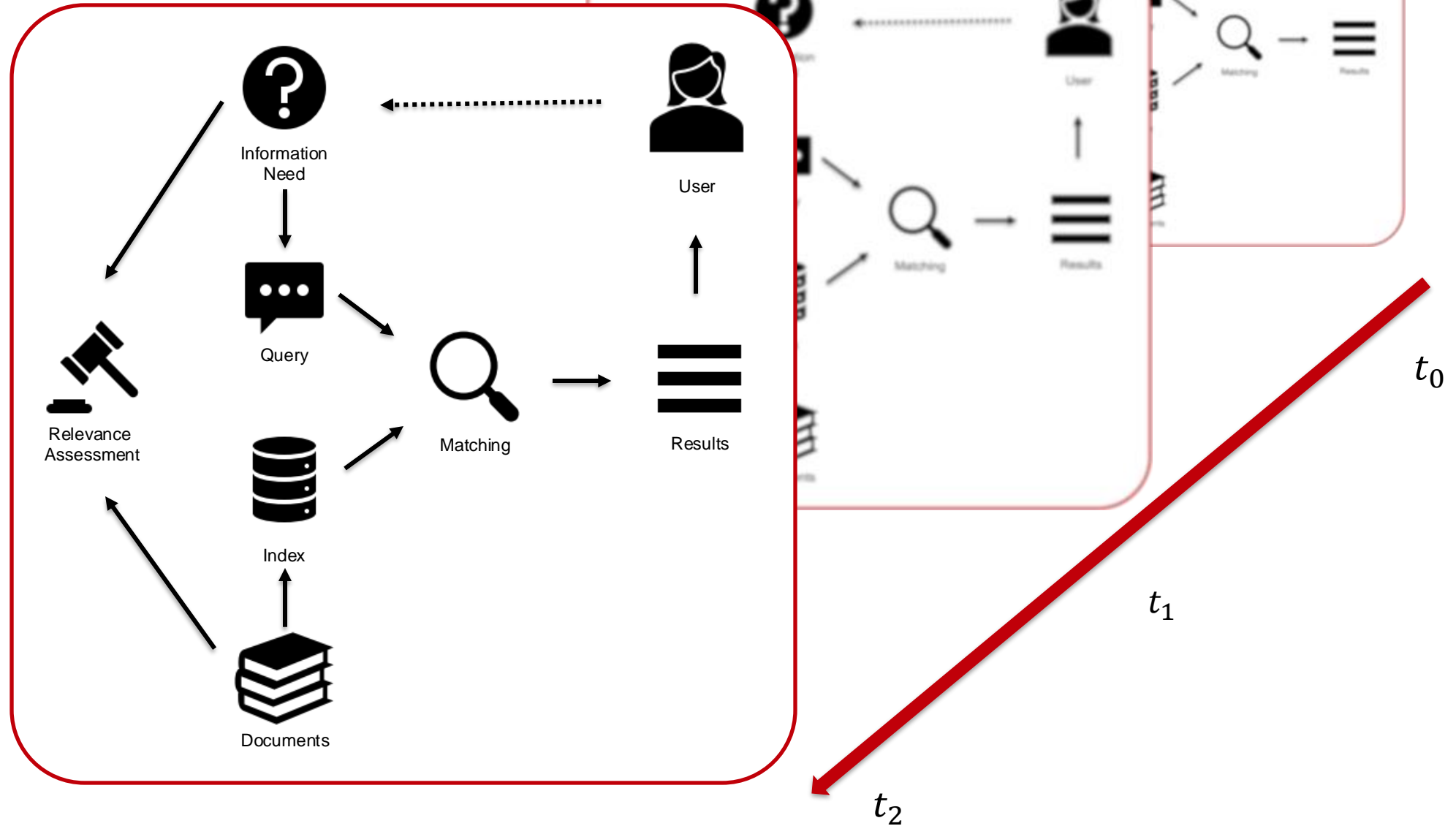
Problem



Problem



Problem



LongEval Shared Task

- *Longitudinal Evaluation of Model Performance*
- Two tasks: retrieval and classification
- Classic web search
- Two languages: French and English
- Over time!

LongEval CLEF 2024 Lab

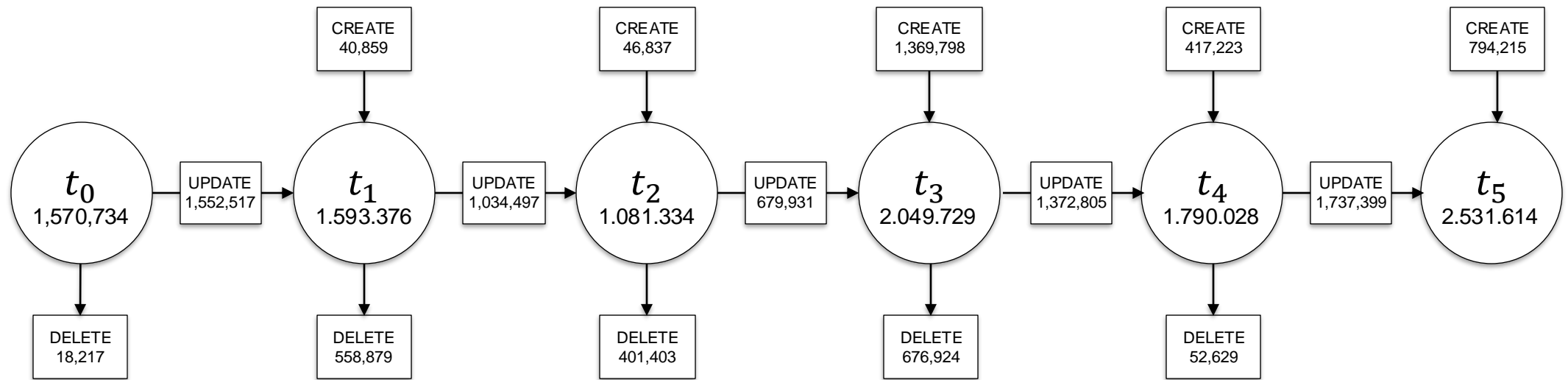


LongEval CLEF 2023 Lab

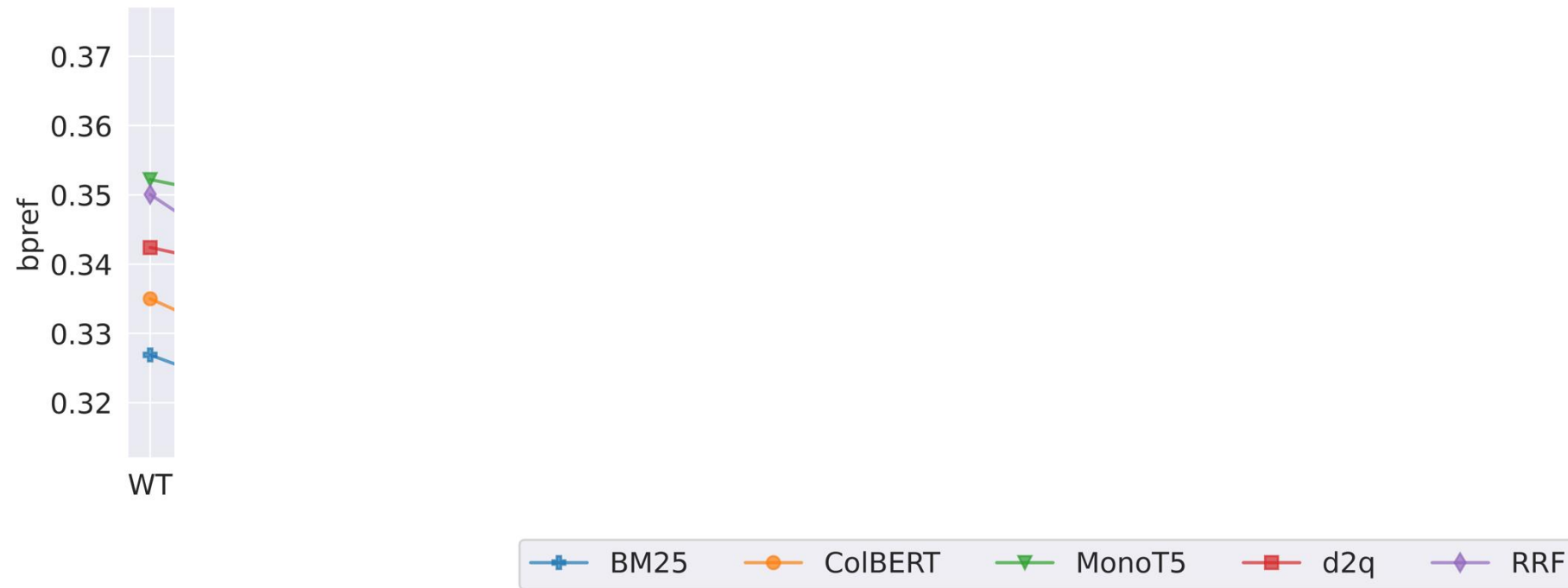


[Description](#) [Dates](#) [Organizers](#) [Tasks](#) [Data](#) [Submissions](#)

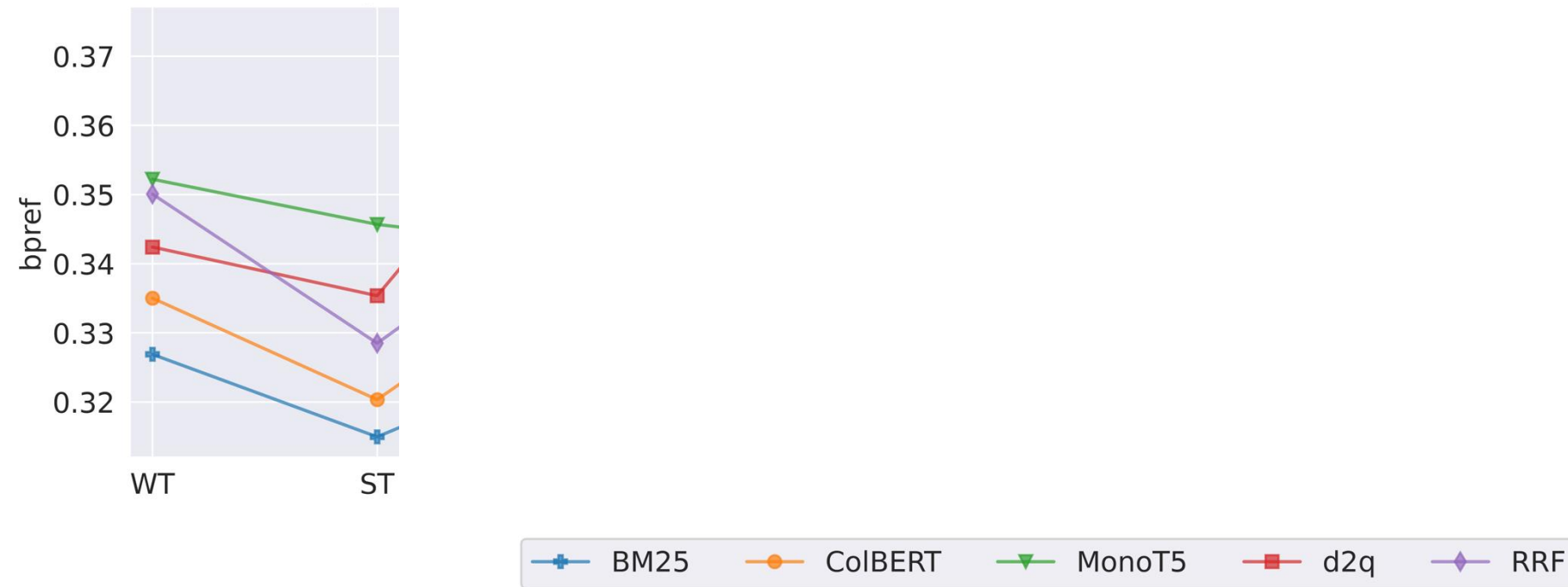
LongEval Shared Task



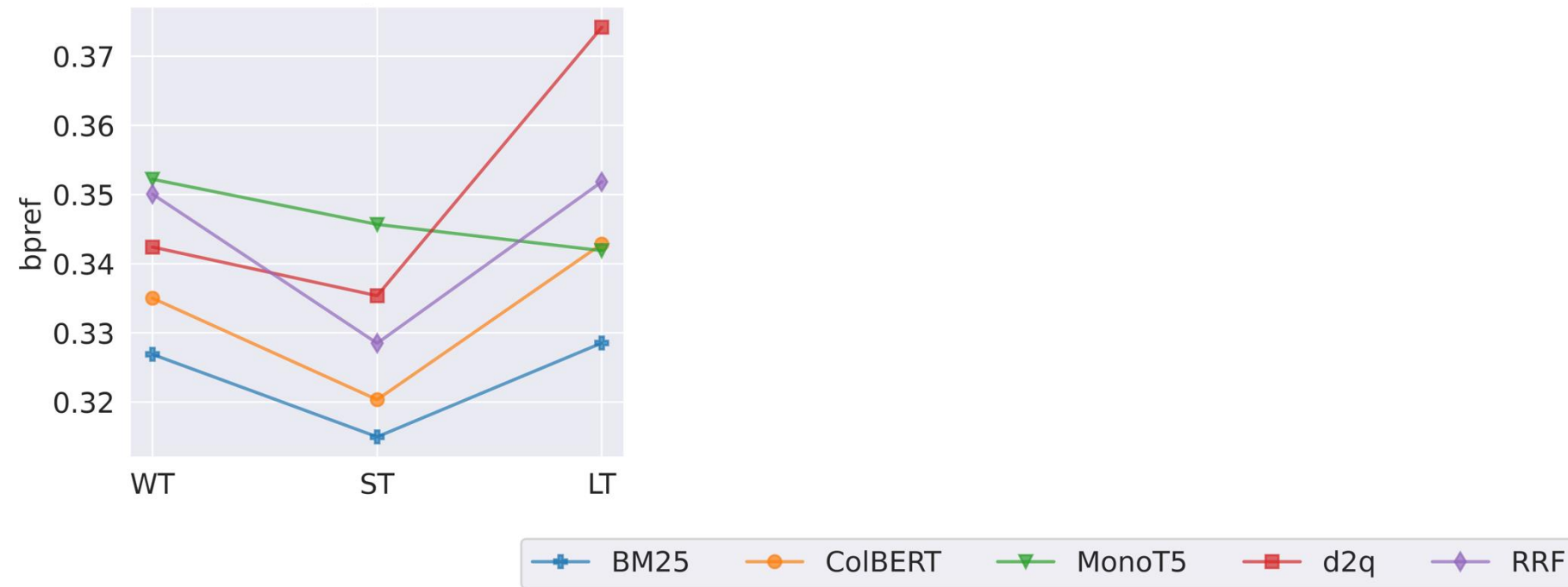
Longitudinal Evaluation



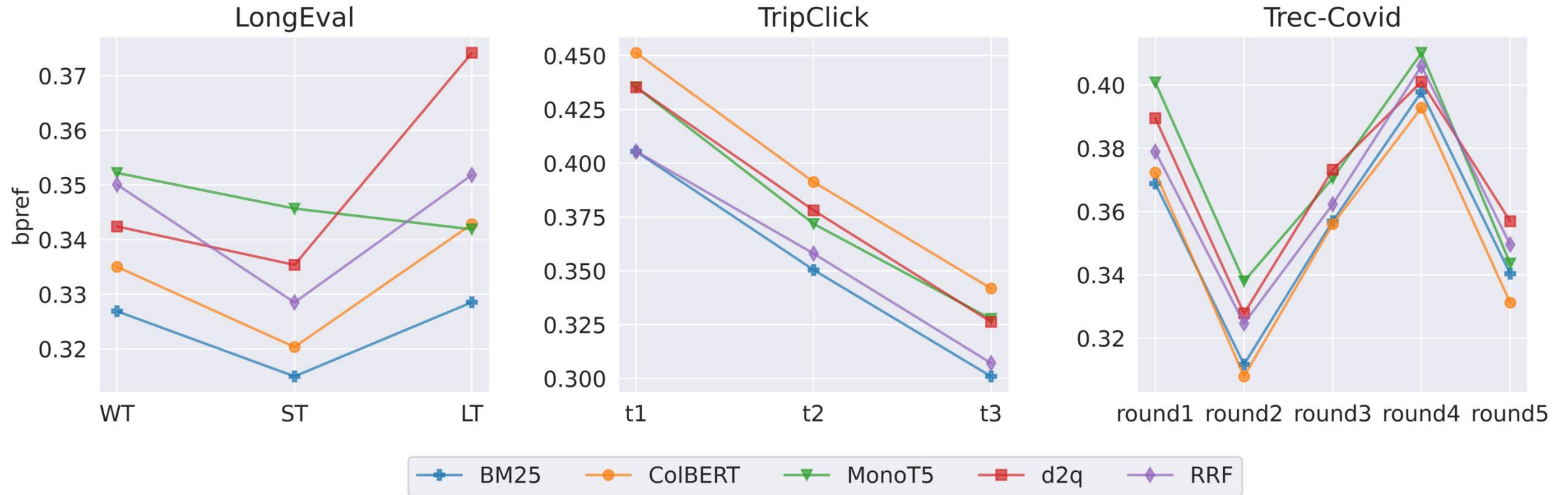
Longitudinal Evaluation



Longitudinal Evaluation



Longitudinal Evaluation



- Effektivität verändert sich über die Zeit.
- Das Ranking der Systeme verändert sich über die Zeit.

Evaluationen sind nicht zwangsläufig zeitlich verlässlich

Temporal Queries

Definition
Word

Fax Number

New Year

Presidential
Election

EM 2024

Tsunami

Informatik

Oldest
person alive

- “13.8% of queries contain explicit time”
- “17.1% of queries have temporal intent implicitly provided”

trick or treat

Tom Hanks
movie

Köln weather

Secure
Password

Halloween

Turkey

EM
Germany

Christmas

Olympics

Recipes

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitungstellung
2024 Oktober

Benzinpreis

Tom Hanks
Film

Wiesen

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags

Olympia

**Vulkanausbruch
Island**

Fax Nummer

Aura

 Zeitemstellung
2024 Oktober

Benzinpreis

 Tom Hanks
Film

Wiesen

Definition IR

 EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
		Vulkanausbruch Island		

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitung
2024 Oktober

Benzinpreis

Tom Hanks
Film

Wiesen

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
		Vulkanausbruch Island		

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitung
2024 Oktober

Benzinpreis

Tom Hanks
Film

Wiesen

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
	Zeitungstellung 2024 Oktober	Vulkanausbruch Island		

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitungstellung
2024 Oktober

Benzinpreis

Tom Hanks
Film

Wiesen

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
	Zeitungstellung 2024 Oktober	Vulkanausbruch Island		

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitungstellung
2024 Oktober

Benzinpreis

Tom Hanks
Film**Wiesen**

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
	Zeitungstellung 2024 Oktober	Vulkanausbruch Island	Wiesen	

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitungstellung
2024 Oktober

Benzinpreis

Tom Hanks
Film

Wiesen

Definition IR

EM
Deutschland

Temporality

1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität
Informatik	EM 2024	EM in Deutschland	Turkey	Köln Wetter
Secure Password	Letzte US Wahl	Tsunami	Star Wars	Top Hashtags
Definition IR	Zeitumstellung 2024 Oktober	Vulkanausbruch Island	Wiesen	Benzinpreis

Olympia

Vulkanausbruch
Island

Fax Nummer

Aura

Zeitumstellung
2024 Oktober

Benzinpreis

Tom Hanks
Film

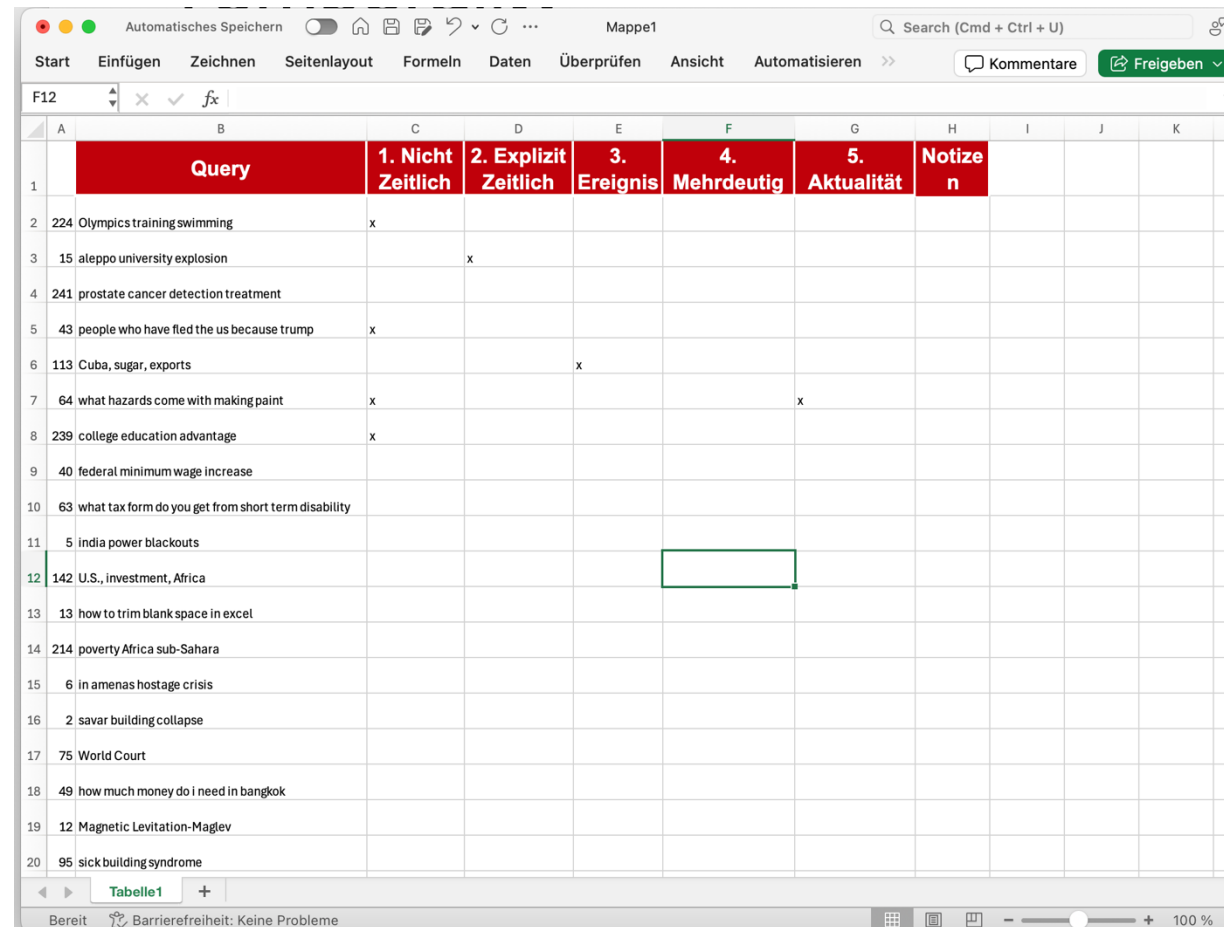
Wiesen

Definition IR

EM
Deutschland

TASK: Topics untersuchen

Klassifizieren Sie die Topics in die verschiedenen Kategorien. Lassen sich alle Topics klassifizieren? Fehlen vielleicht Kategorien? Notieren Sie sich Ihre Beobachtungen.



	Query	1. Nicht Zeitlich	2. Explizit Zeitlich	3. Ereignis	4. Mehrdeutig	5. Aktualität	Notizen
224	Olympics training swimming	x					
15	aleppo university explosion		x				
241	prostate cancer detection treatment						
43	people who have fled the us because trump	x					
113	Cuba, sugar, exports			x			
64	what hazards come with making paint	x				x	
239	college education advantage	x					
40	federal minimum wage increase						
63	what tax form do you get from short term disability						
5	india power blackouts						
142	U.S., investment, Africa						
13	how to trim blank space in excel						
214	poverty Africa sub-Sahara						
6	in amenas hostage crisis						
2	savar building collapse						
75	World Court						
49	how much money do i need in bangkok						
12	Magnetic Levitation-Maglev						
95	sick building syndrome						

How to pass this project

Hard requirements

- Be. Active. And. Participate. In. The. Project.
- Active means to code, label, document, plan, organize, ...
- Milestone presentations after 1st semester
- Do final presentation at the end of the course (2nd semester)
- Write a final term paper / project documentation (and maybe submit it to CLEF).

Soft requirements

- Be. Active. And. Have. Fun. In. The. Project.
- Learn a lot new and interesting stuff about Python, Data Labeling, project management, test collection design, search engines, and other DIS-related topics.

Provisional Schedule

	Datum	Notiz	
W1	18.10.24		
W2	25.10.24		Einführung zu Temporal IR
W3	01.11.24	Entfällt (Allerheiligen)	
W4	08.11.24		Gute und schlechte Topics
W5	15.11.24		Topics selbst erstellen
W6	22.11.24	Projektwoche	
W7	29.11.24		Relevanzbewertungen und Projekte
W8	06.12.24		Projektphase
W9	13.12.24		...
W10	20.12.24		...
		Weihnachten	
W11	03.01.25		Start LongEval
W12	10.01.25		
W13	17.01.25		

Provisional Schedule

	Datum	Notiz	
W1	24.03		
W2	07.04		
W3	14.04		
W4	21.04		
W5	28.04		
W6	05.05	<i>Submission Runs</i>	LongEval Runs einreichen
W7	12.05		
W8	19.05		
W9	26.05	<i>Submission Notebooks</i>	LongEval Bericht einreichen
W10	02.06		Abschlusspräsentation
W11	09.06		entfällt
W12	16.06		entfällt
W13	23.06		entfällt