

Project DIS18

06 – Projects

Jüri Keller and Philipp Schaer

25-01-10 – Cologne, Germany

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



Technology
Arts Sciences
TH Köln



Provisional Schedule

	Datum	Notiz	
W1	18.10.24		
W2	25.10.24		Einführung zu Temporal IR
W3	01.11.24	Entfällt (Allerheiligen)	
W4	07.11.24		Queries and 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		Projektphase
W10	20.12.24		Projekte präsentieren
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		LongEval
W13	17.01.25		

LongEval CLEF 2024⁵ Lab

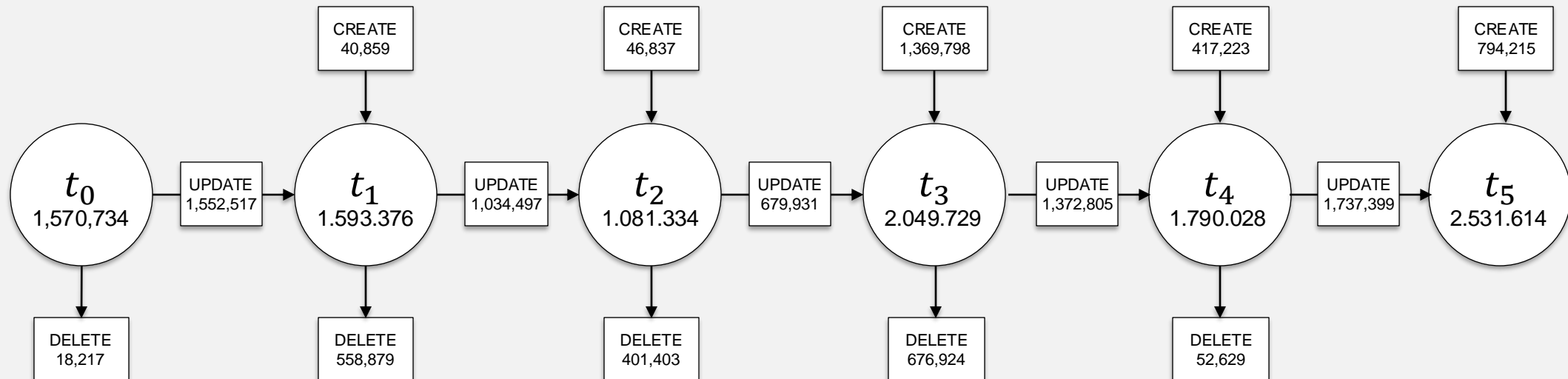


Longitudinal Evaluation of Model Performance

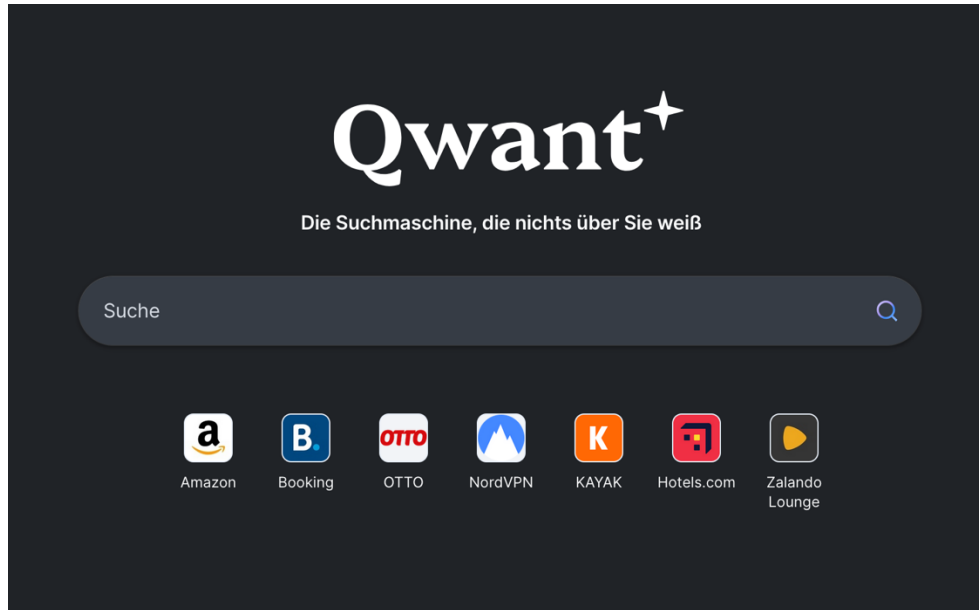
- Wie verhalten sich IR-Systeme wenn die Collection sich verändert?
- Wann müssen wir unser System updaten?

LongEval

- Dokumente und Nutzende verändern sich
- Retrieval-Systeme sind dieser Veränderung ausgesetzt
- LongEval spiegelt diese Veränderungen wider



Zwei Tasks



Task 1: LongEval-WebRetrieval

- Web Search Retrieval
- Websites

The world's largest collection of open access research papers

Search 316M papers from around the world

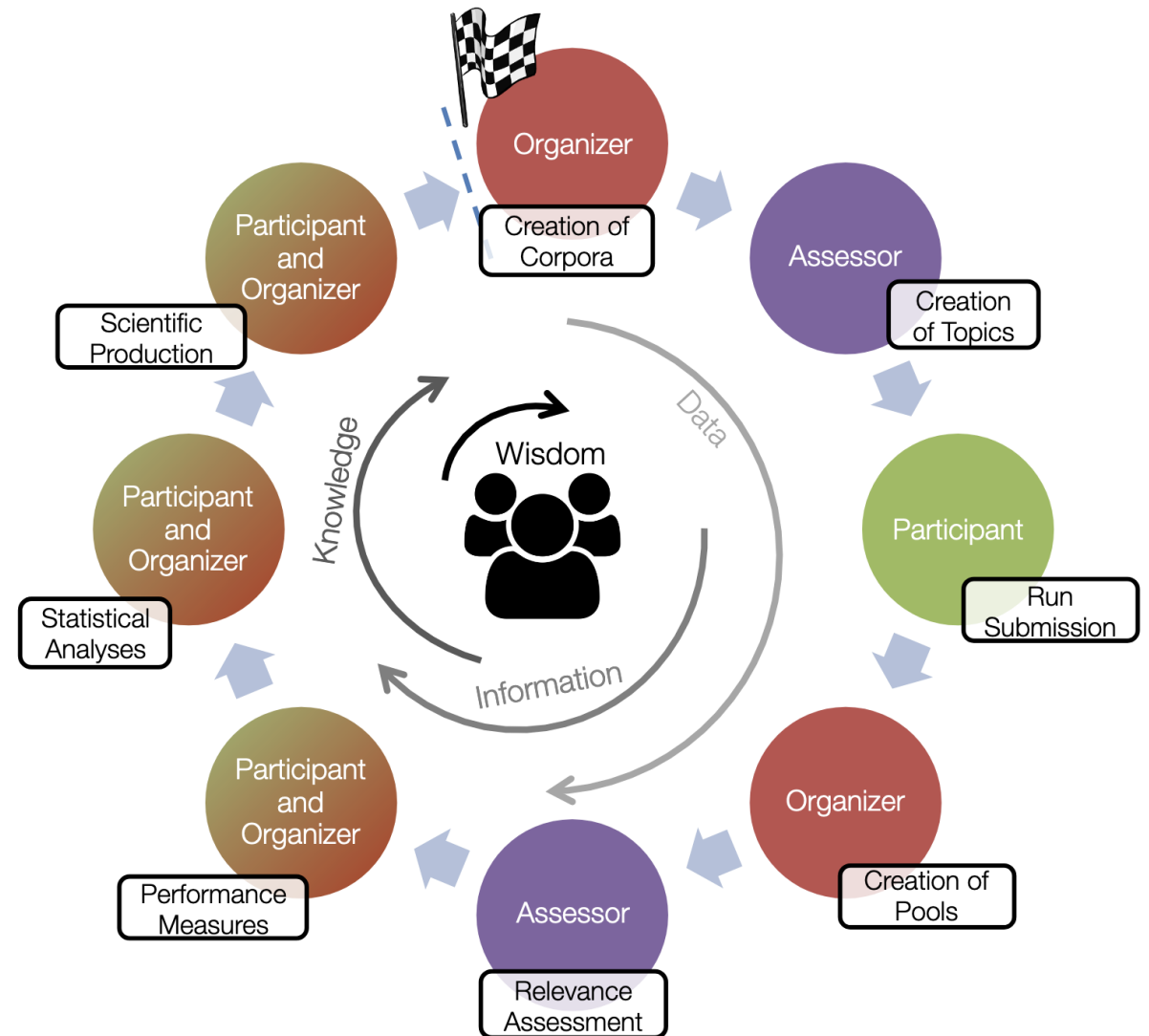
SEARCH

Task 2: LongEval-SciRetrieval

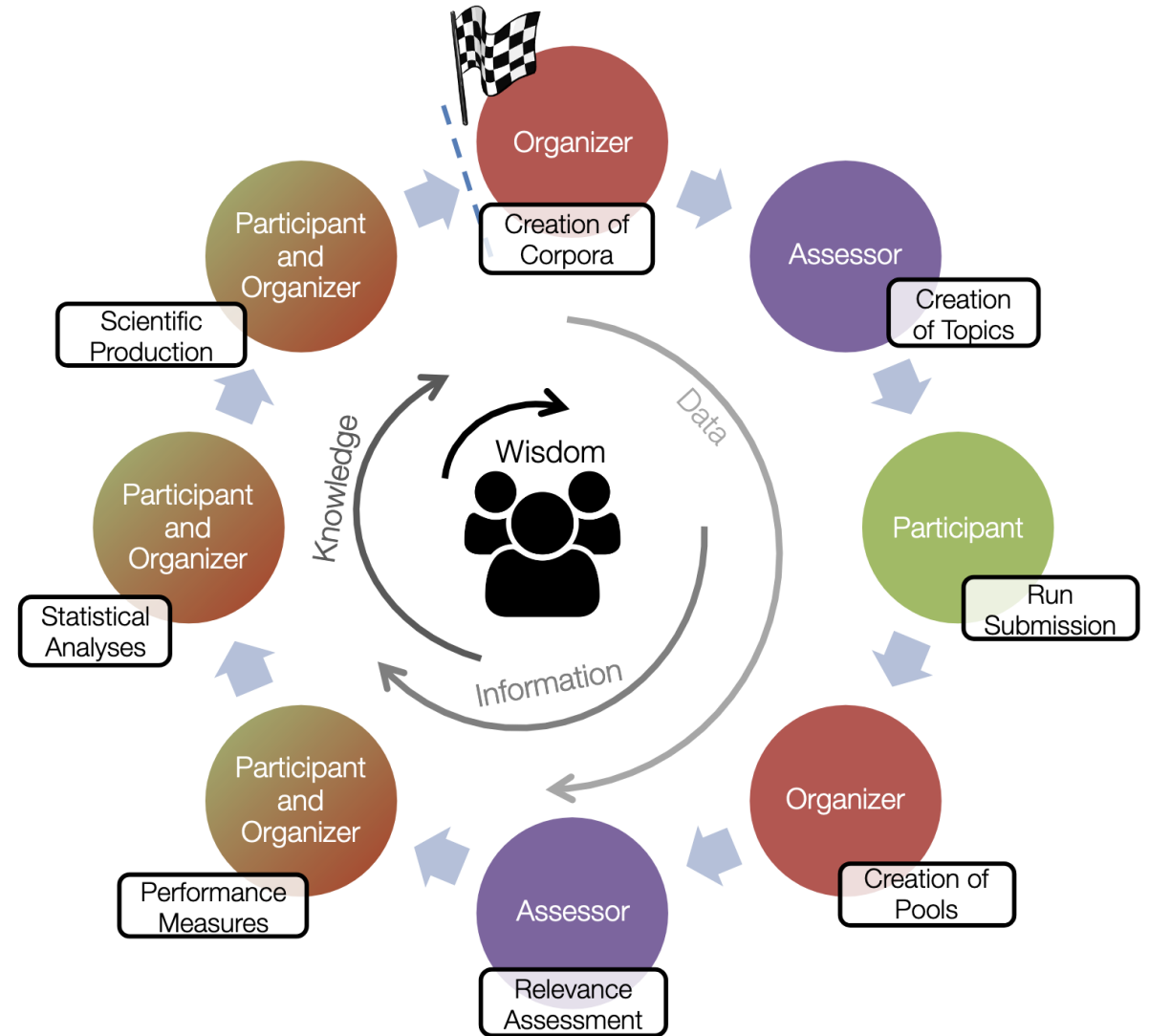
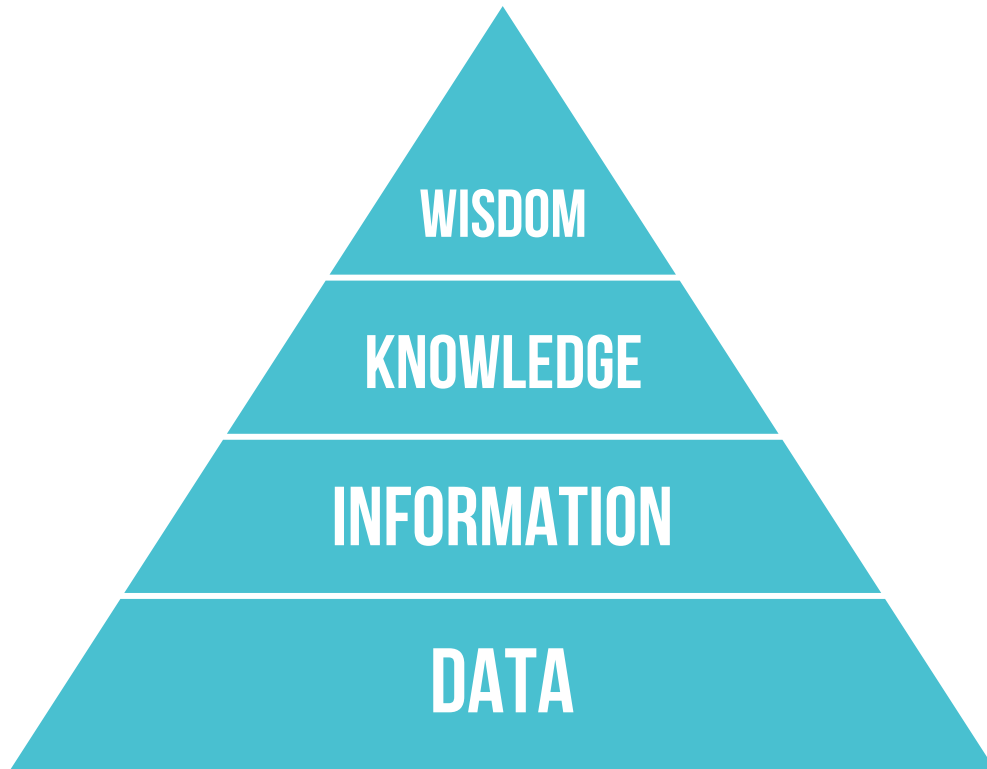
- Scientific Retrieval
- Wissenschaftliche Literatur

LongEval Lab Shared Task

- Ein Problem viele Lösungen
- Lab Organisatoren stellen einen Datensatz zur Verfügung
- Teilnehmende reichen Lösungen ein



LongEval Lab Shared Task



cir_run_5:
BM25 + time boost

Dokumente nach alter boosten

- Neue Dokumente sind relevant
- Alte Dokumente sind relevant

cir_run_5:
BM25 + time boost

Dokumente nach alter boosten

- Neue Dokumente sind relevant
- Alte Dokumente sind relevant

cir_run_3:
BM25 + qrel boost

Relevante Dokumente sind weiter relevant

- Bekannte Query – Dokumenten Paare
- Alter der Dokumente

cir_run_5:
BM25 + time boost

Dokumente nach alter boosten

- Neue Dokumente sind relevant
- Alte Dokumente sind relevant

cir_run_3:
BM25 + qrel boost

Relevante Dokumente sind weiter relevant

- Bekannte Query – Dokumenten Paare
- Alter der Dokumente

cir_run_4:
BM25 + RF

Relevanzfeedback

- Terme von alten relevanten Dokumenten
- Querys erweitern
- Alter der Dokumente

System	t	P@10	bpref	nDCG	MRR
BM25	t_3	0.1624	0.4373	0.3638	0.3660
BM25	t_4	0.1370	0.3572	0.2817	0.3162
BM25	t_5	0.1076	0.2791	0.2106	0.3116
+monoT5	t_3	0.1776*	0.4571*	0.3839*	0.3929
+monoT5	t_4	0.1591*	0.3719	0.3081*	0.3847*
+monoT5	t_5	0.1246*	0.2862*	0.2303*	0.3551*
<i>cir_run_5:</i> BM25 + time boost	t_3	0.1007*	0.4178*	0.2758*	0.2774*
	t_4	0.1380	0.3583	0.2850	0.3308
	t_5	0.1065	0.2798	0.2116	0.3210*
<i>cir_run_3:</i> BM25 + qrel boost	t_3	0.1870*	0.4515*	0.3910*	0.4394*
	t_4	0.1975*	0.3815*	0.3536*	0.4922*
	t_5	0.1349*	0.2881*	0.2419*	0.4063*
<i>cir_run_4:</i> BM25 + RF	t_3	0.1758*	0.4819*	0.3955*	0.3997*
	t_4	0.1608*	0.3958*	0.3197*	0.3821*
	t_5	0.1230*	0.2990*	0.2293*	0.3373*

Run

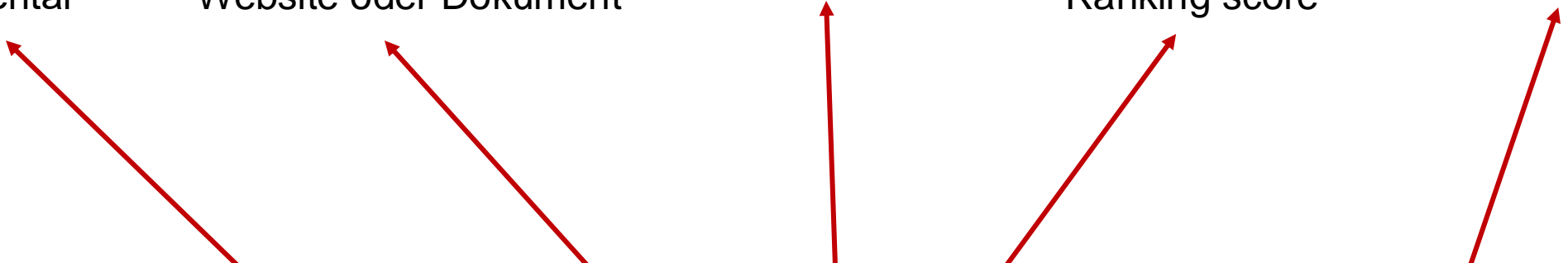
Query ID
"Car Rental"

Doc ID
Website oder Dokument

Rang

Score
Ranking score

Run Name



q062210081	Q0	doc062200206592	0	23.77288566678163	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200109914	1	22.661697446651136	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200101758	2	22.469785136101127	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200101893	3	22.347536313638575	IRC_BM25+Bo1.WT
q062210081	Q0	doc062210912628	4	22.23395874634665	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200602177	5	21.832272792365636	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200308829	6	20.805521032004638	IRC_BM25+Bo1.WT
q062210081	Q0	doc062200503293	7	20.714209794229724	IRC_BM25+Bo1.WT

Run

Doc ID

Website oder Dokument

Query ID

“Car Rental”

q062210081	Q0	doc062200206592	0	23
q062210081	Q0	doc062200109914	1	22
q062210081	Q0	doc062200101758	2	22
q062210081	Q0	doc062200101893	3	22
q062210081	Q0	doc062210912628	4	22
q062210081	Q0	doc062200602177	5	21
q062210081	Q0	doc062200308829	6	20
q062210081	Q0	doc062200503293	7	20

Google Car rental

Alle Bilder Orte Websites zu Orten Videos News Web Mehr Suchfilter

Near me Jetzt geöffnet Top-Bewertung Avis Cologne Airport Budget Enterprise Germany

Gesponsert

Rentalcars.com
http://www.rentalcars.com

RentalCars.com: Car Rentals
Best Prices Guaranteed — Compare & Save on your Car Rental now. World's Biggest Online Car Rental Service

Budget Car Rental
No Credit Card Fees Free Booking Amendments

Hertz Car Rental
No Credit Card Fees Best Prices Guaranteed

Thrifty Car Rental
Best Prices Guaranteed 24 Hour Call Centre

Enterprise Car Rental
Best Prices Guaranteed Book on Safe & Secure App

Gesponsert

Europcar
https://www.europcar.de

Europcar Auto Mieten | Mietwagenstationen weltweit
Erkunden Sie die Europcar Flotte und sichern Sie sich tolle Angebote. Buchen und losfahren

Gesponsert

CHECK24
https://mietwagen.check24.de › car_rental › preisvergleich

CHECK24: Car Rental Vergleich
Größter Mietwagenvergleich — Car Rental beim Testsieger CHECK24 - Mietwagen ab 5€ - "sehr gut" beim TÜV Service Test!

Rentalcars.com
https://www.rentalcars.com › Mietwagen › Deutschland

Billiger Mietwagen Köln
Vergleichen Sie Mietwagen in Köln und finden Sie die günstigsten Angebote von allen großen Autovermietungen. Buchen Sie Ihren Mietwagen heute auf dem ...

KAYAK
https://www.kayak.de › Günstige-Mietwagen-in-Koeln...

Mietwagen Köln ab 17 €/Tag - Autovermietung auf ...
Du suchst günstige Mietwagen in Köln? Finde Preise von CARO , Europcar, FLIZZR, National, Thrifty und keddy by Europcar. Aktuelle Preise: Economy 23 €/Tag.

TASKs:

Bis 16.01.25:

1. Bildet Teams von 2 bis 4 Personen
2. Erarbeitet eine konkrete Idee für ein Retrieval-System und skizziert sie auf ca. einer halben Seite und sendet diese per E-Mail an uns

Bis 17.02.25:

- Implementiert einen „langen dünnen Man“

Alle zwei Wochen:

- Annotiert eure Querys in RAT

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	Allerheiligen	
W4	07.11.24		Queries and 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		Projektphase
W10	20.12.24		Projekte präsentieren
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		LongEval
W13	17.01.25		
	24.01.25	Prüfungen	
	31.01.25	Prüfungen	
	07.02.25	Prüfungen	

Provisional Schedule

	Datum	Notiz	
W1	18.10.24		
W2	25.10.24		Einführung zu Temporal IR
W3	01.11.24	Allerheiligen	
W4	07.11.24		Queries and 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		Projektphase
W10	20.12.24		Projekte präsentieren
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		LongEval
W13	17.01.25		
	24.01.25	Prüfungen	
	31.01.25	Prüfungen	
	07.02.25	Prüfungen	

HACKATHON:
17.02 9:30 – 17:30
&
25.03 9:30 – 17:30

Provisional Schedule

	Datum	Notiz	
W1	04.04		
W2	11.04		
W3	18.04	Karfreitag	
W4	25.04		
W5	02.05	<i>Submission Runs</i>	LongEval Runs einreichen
W6	09.05		
W7	16.05	Projektwoche	
W8	23.05	<i>Submission Notebooks</i>	LongEval Bericht einreichen
W9	30.05		Abschlusspräsentation
W10	06.06	<i>entfällt</i>	
W11	13.06	<i>entfällt</i>	
W12	20.06	<i>entfällt</i>	
W13	27.06	<i>entfällt</i>	