

# Project DIS18

## 06 – Projects

Jüri Keller and Philipp Schaer

24-12-20 – 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
<b>W10</b>	<b>20.12.24</b>		<b>Projekte präsentieren</b>
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		LongEval
W13	17.01.25		

# Result Assessment Tool (RAT)



1. Studie erstellen
2. Ergebnisse sammeln
3. Bewerten & Analysieren
4. Ergebnisse exportieren

The screenshot shows the RAT web interface. At the top, there's a navigation bar with 'Dashboard', 'Studies', and 'New Study' links, and a user profile 'a@bc.de'. Below this, there are three main sections: 'RESULTS COLLECTION' (Status: Study running. 1960 / 1960) with 'Start' and 'Pre-Test' buttons; 'SEARCH TASKS & QUESTIONS' (No questions yet. Create one!) with an 'Add Question' button; and 'PARTICIPANTS' (No participants yet. Invite one!) with an 'Invite' button. The main content area is titled 'STUDY SUMMARY' and includes a 'Settings' button. The study title is 'Überschneidung und SEO-Merkmale bei pseudo-medizinischen Suchanfragen'. The description states: 'Die Suchbegriffe für diese Studie wurden mit dem Keyword Planner aus dem initialen Begriff Krebsdiät generiert. Die Studie soll folgende Fragen beantworten: (a) Wie unterscheiden sich die Ergebnisse bei ähnlichen bzw. verwandten Suchbegriffen? (b) Wie hoch ist die Überschneidung von Suchergebnissen in unterschiedlichen Suchmaschinen? (c) Wie hoch ist der Anteil von Webseiten, die SEO-Merkmale aufweisen?'. Below this, there are three columns: 'STUDY TYPE' (Relevance Assessment), 'RESULT TYPE' (Organic Results), and 'RESULT COUNT' (10). Under 'SEARCH ENGINES', there are buttons for 'Google' and 'Bing'. At the bottom, 'SEARCH QUERIES (SHOWING 5 OUT OF 98. SEE ALL)' are listed: 'eiweiß öl diät krebs', 'ernährungsplan bei chemo', 'rezepte zur unterstützung einer ketogenen ernährung für krebspatienten', 'ernährungsplan krebspatienten', and 'breuss diät krebs'.

# Result Assessment t0

- Guideline: Zusätzliches Feld für Websites die nicht erreichbar waren
- ...

# GUIDELINE

**Highly relevant (2):** Die Website entspricht voll und ganz dem Informationsbedürfnis, welches durch das Topic ausgedrückt wird, d. h. sie beantwortet die Frage im Topic. Die Website muss nicht alle Informationen zum Topic enthalten, aber sie muss für sich genommen eine Antwort auf die Frage geben.

**Relevant (1):** Die Website beantwortet einen Teil der Frage, müsste aber mit anderen Informationen kombiniert werden, um eine vollständige Antwort zu erhalten.

**Not relevant (0):** alles andere

# TASK:

1. Annotieren Sie bis zum **20.12.24** die Dokumente für Ihre Topics in RAT nach der Guideline.
2. Bereiten sie die Zwischenpräsentation Wintersemester 24 vor:
  - 5 min pro Gruppe (4-5 Slides)
  - Was habt ihr gemacht: **Methodik**
  - Wie gut hat es funktioniert: **Ergebnisse**
  - Wie könnte man das in einem Retrieval System nutzen: **Future Work**

# 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	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
<b>W10</b>	<b>20.12.24</b>		<b>Projekte präsentieren</b>
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		Start 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	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
<b>W10</b>	<b>20.12.24</b>		<b>Projekte präsentieren</b>
		Weihnachten	
W11	03.01.25		Home Office
W12	10.01.25		Start LongEval
W13	17.01.25		
	24.01.25	Prüfungen	
	31.01.25	Prüfungen	
	07.02.25	Prüfungen	

**10.02. oder 07.02?**  
**&**  
**1 Termin im März**  
**24.03?**

# Provisional Schedule

	Datum	Notiz	
W1	04.04		
W2	11.04	<i>ECIR</i>	
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>	