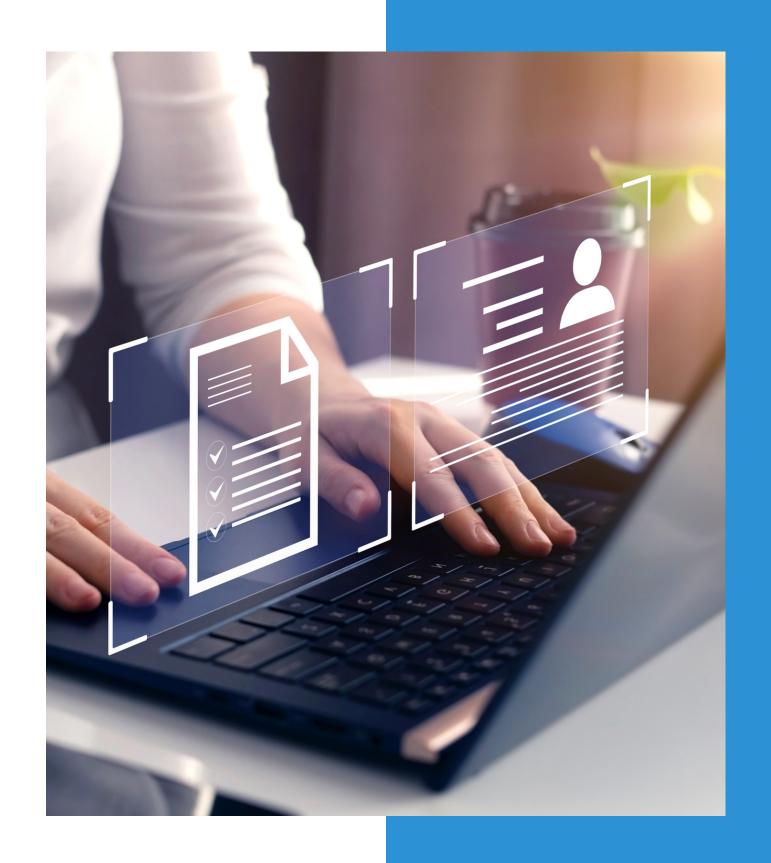
TUM WS 2023/24

APPROACHING INFORMATION SYSTEM CHALLENGES WITH NATURAL LANGUAGE PROCESSING

Topic B (rule based)

Extract rules for the data objects of the events from written regulatory process information

Leo Holzhauer | Dec 05 2023



PIPELINE

Suggestion

Splitting and Cleaning

Enhancement

Word **Embeddings**

Modelling

Evaluation

PIPELINE

Own adaptation

Preprocessing

Manual

Rephrasing of input to match process

Script

- Formatting
- Cleaning

Modelling

- Extracting process steps and variables
- Extracting constraints

Evaluation

- Calculating similarity scores
- Calculating final scores

COFFEE ROASTING

UNFCCC CDM

26 constraints

38* constraints

DATA

* w/o negations

AKTG

PATG

7 constraints

6 constraints

OPEN TASKS

Envisioned ideas to advance the project

Implement Extract matching rules process steps Generalize algorithm for **Extract variables** pattern matching Compose process steps, **Transform matches into** variables and matches into formalized structure constraints Generalize algorithm for **Define final score** similarity scores

OPEN QUESTIONS

Input appreciated

Rule based extraction of process steps?	1	
Relevance of negated constraints?	2	
Specifics of the calculation of similarity scores (threshold, "directly follows", structure)?	3	
Calculation of final score?	4	
Literature recommendations?	5	



Extract rules for the data objects of the events from written regulatory process information

Leo Holzhauer | Dec 05 2023