Praktikum Wirtschaftsinformatik
10. December 2025
Prof. Dr. Mirjam Minor
M.Sc. Tolga Tel

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

In this exercise, we will further evaluate the descriptive texts using different metrics. We will also try to remove the mistakes that were found during evaluation.2222
You can find all the documentation on https://haystack.deepset.ai/

Deadline: 12.01.2026, 10am

Exercise 1 (Automatic Evaluator) — 6 Points

a. **Write** Python Code to automatically extract all element names from a BPMN file. **(1 Point)**
   - **Input: .bpmn file**
   - **Output: Names of Tasks(, Events and Gateways)**
b. **Choose** one of the methods from last exercise sheet (Structured Outputs, Text Annotation or generating an annotated text) to get a list of elements from a BPMN description. **(1 Point)**
   - **Input: Descriptive Text of a BPMN-model (and the corresponding .bpmn)**
   - **Output: Names of Tasks(, Events and Gateways)**
c. **Write** a component, that matches the elements extracted from the BPMN and the text and automatically computes the precision and recall values. **(2 Points)**
   - **Input: Two lists with element names (from the text and the model)**
   - **Output: Precision and recall values for these lists**
d. **Run** a pipeline containing the elements from a)-c) for 3 different **descriptive texts** of BPMN models. Present the scores in your presentation. **(2 Points)**

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Exercise 2 (Adaptation) **9 Points**

a. **Create** a pipeline for a RAG system, that receives **a mistake of a given text** and tries to remove the mistakes. It should contain following components: **(6 Points)**
   - A **Document Store** containing rules to update a text. For example:
     i. Task x is missing in the description of the process model. Please add the task at the correct position.
     ii. Task y is hallucinated, please remove it from the description of the process model.
   - A **custom retriever**, that receives the **error in a descriptive text as its input** and retrieves the correct feedback from the document store using rules. (if element_type = task and status = missing then…)
   - A (Chat)**PromptBuilder** that receives a query containing a BPMN model, its textual description, and the instruction from the document store to apply feedback. It instructs the LLM to change the textual description according to the feedback given.
   - A **generator** that works with the prompt builder.

b. **Run** this pipeline for 3 different BPMN-text pairs where the text contains errors like missing tasks. Verify the results by checking if the error was fixed correctly, and if new errors were added to the texts. Present the results in your presentation. **(3 Points**)
   - **Input: A descriptive text with the corresponding .bpmn file and a feedback list.**
   - **Output: A descriptive text where the mistake has hopefully been corrected.**

Exercise 3 (LLM as a Judge) **5+4* Points**

a. **Create** a Pipeline that takes **a descriptive text of a BPMN model** as its input and evaluates its readability by instructing an LLM to give the text a score. **(2 Points)**
   - **Input: Descriptive Text of a BPMN-model**
   - **Output: A score (maybe 0-5, or 0-10, be creative)**

b. **Run** the pipeline using 3 different BPMN models from past exercises. For each, evaluate a generated descriptive text and a descriptive text written by yourself (ground truth) and compare the scoring results. Present the scores in your presentation. **(3 Points)**

c. **Find** another useful metric to evaluate the descriptive texts (apart from precision/recall and the metric used in exercise 3a.) and **use** an LLM to receive scoring results for this metric for the BPMN models used in exercise 3b. Present the scores in your presentation. **(4* Points)**