Assignment 3: Knowledge Graph Population: Development and Evaluation of Entity Extractors

Kalemkeridis Evangelos Rafail

1 Task 1: Entity Extractor Development

Two systems were developed: a rule-based extractor and an LLM-based extractor.

1.1 Rule-Based Extractor

The rule-based extractor uses a static, manually curated dictionary.

- **Strength:** Perfect recall inside its domain.
- **Limitation:** Cannot generalize beyond the dictionary.

1.2 LLM-Based Extractor (Gemini API)

This extractor uses Gemini with a structured JSON schema, ensuring parseable and reliable output.

2 Task 2: Extractor Evaluation

2.1 Evaluation Dataset

A dataset of 12 texts was annotated, containing 29 gold entities.

2.2 Performance Metrics

Table 1: Performance Comparison of Entity Extractors

Extractor	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F_1 Score**
Rule-Based	29	3	0	0.906	1.000	0.951
LLM-Based (Gemini)	22	3	7	0.880	0.759	0.815

3 Task 3: LLM Evaluator (Judge) Development

3.1 LLM-as-a-Judge System

The judge validates extracted terms and outputs structured JSON verdicts.

Table 2: Agreement Metrics for LLM-as-a-Judge

Metric	Value	n
Accuracy	0.923	26
Cohen's Kappa	0.000	26

3.2 Agreement Metrics

4 Conclusion

The project implemented a functioning pipeline for Knowledge Graph population. The rule-based extractor excelled in-domain, while the LLM-based extractor demonstrated strong generalisation. The LLM Judge provided transparency and reviewability.