LLM Reasoning & Personality Knowledge Graph Challenge – Project Report

A. Project Overview

In this project, I developed a Python solution to extract structured knowledge and model personality traits from textual data. The primary goal was to transform unstructured text into a knowledge graph (KG) representing people, organizations, their relationships, and their personal characteristics, such as traits and activities. The exercise required a combination of natural language processing (NLP) techniques, graph visualization, and LLM-assisted research to inform design choices and implementation strategies. Using spaCy for entity recognition and PyVis for graph visualization, I created an interactive representation of both the social and personality aspects of individuals mentioned in the text. The project highlights the potential of automated knowledge extraction in understanding and visualizing human behaviour and organizational structures.

B. Methodology

The project was implemented in several stages, which are described below:

Data Preparation and Entity Recognition:

I started by selecting a sample text containing descriptions of individuals and organizations, ensuring a variety of roles, traits, and activities. Using spaCy's NLP models (with fallback from en_core_web_lg to en_core_web_trf and finally en_core_web_sm), I processed the text sentence by sentence. Named entities corresponding to people and organizations were identified, allowing to creation of nodes in the knowledge graph. Since no real dataset was provided, synthetic text was generated to include diverse roles, organizations, personality traits, and activities. This allowed controlled evaluation of extraction accuracy and ensured that the knowledge graph could demonstrate relationships, activities, and personality modeling effectively.

Personality Trait and Activity Extraction:

For each identified person, adjectives modifying their names were extracted as traits. Additionally, verbs in the sentence were collected to represent their activities. This approach allows the KG not only to reflect structural relationships (e.g., “works_at”) but also to capture personality and behaviour in a structured way. To avoid redundancy, duplicates in traits and activities were removed before visualization. Personality traits are represented as adjectives associated with each person node, and activities as verbs. This structured approach enables linking behavioral information to individuals within the graph, allowing semantic queries and visualization of personal characteristics alongside organizational relationships.

Relationship Modeling:

Standard relationships between people and organizations, such as employment or affiliation, were identified based on the text context. Each person was connected to their organization using edges labelled “works_at.” Traits and activities were represented as additional nodes connected to the respective person nodes, with labels indicating “has_trait” or “does” relationships.

Data Processing / Normalization

Text normalization involved removing duplicate traits and activities to avoid redundant graph nodes. No other normalization was applied to preserve the original semantic meaning of adjectives and verbs. This ensures the knowledge graph reflects actual textual descriptions accurately.

Evaluation:

To assess the quality of the extraction process, I manually created a ground truth for people, organizations, traits, and activities. Using precision, recall, and F1-score metrics, I compared the extracted results to the ground truth. Precision, recall, and F1-score were chosen because they quantify the correctness and completeness of entity and relationship extraction, which are critical for assessing the quality of a knowledge graph.

<u>**Visualization**</u>:

The knowledge graph was visualized using PyVis, generating an interactive HTML file. People and organizations were represented with distinct shapes and colors, while traits and activities were added as connected nodes to provide additional context. This visualization allows users to explore both the relationships between entities and the personal attributes of individuals in an intuitive, interactive manner.

C. Insights and Limitations

The project demonstrated that even a small NLP pipeline can successfully capture relationships and personality traits from text. Interactive visualization enhances understanding by representing multi-dimensional information in a single graph. However, limitations remain, including incomplete trait extraction, potential false positives in activity identification, and the need for more advanced semantic parsing to improve precision. Future improvements could include leveraging transformer-based models with context-aware relationship extraction and integrating company logos or images for richer graph nodes.

D. Results

1. Console Output / JSON Preview

```
{
  "people": [
    "Zain",
    "Omar",
    "Aisha",
    "Lina"
  ],
  "organizations": [
    "VisionHub",
    "CloudSync",
    "TechNova"
  ],
  "person_data": {
    "Aisha": {
      "traits": [
        "open",
        "minded",
        "creative"
      ],
      "activities": [
        "inspire",
        "encourage",
        "organize"
```


Figure 1: Console output of extracted knowledge

2. Evaluation Metrics

The evaluation of the knowledge graph extraction shows that the model performs very well in identifying activities, achieving perfect precision, recall, and F1 scores of 1.00 across all individuals. Trait extraction shows good performance overall, with an average precision of 0.71, recall of 0.88, and F1 score of 0.77, though some traits like Aisha's were partially missed. For the works_at relationships, the model correctly identified most connections, resulting in an average precision, recall, and F1 of 0.75. Overall, the system demonstrates strong capability in capturing key actions and relationships, while personality trait extraction could be slightly improved.

Person	Type	Precision	Recall	F1
Aisha	Traits	0.33	0.50	0.40
	Activities	1.00	1.00	1.00

	Works at	1.00	1.00	1.00
Lina	Traits	1.00	1.00	1.00
	Activities	1.00	1.00	1.00
	Works at	1.00	1.00	1.00
Omar	Traits	1.00	1.00	1.00
	Activities	1.00	1.00	1.00
	Works at	0.00	0.00	0.00
Zain	Traits	0.50	1.00	0.67
	Activities	1.00	1.00	1.00
	Works at	1.00	1.00	1.00
Average	Traits	0.71	0.88	0.77
	Activities	1.00	1.00	1.00
	Works at	0.75	0.75	0.75

3. Graph Visualization

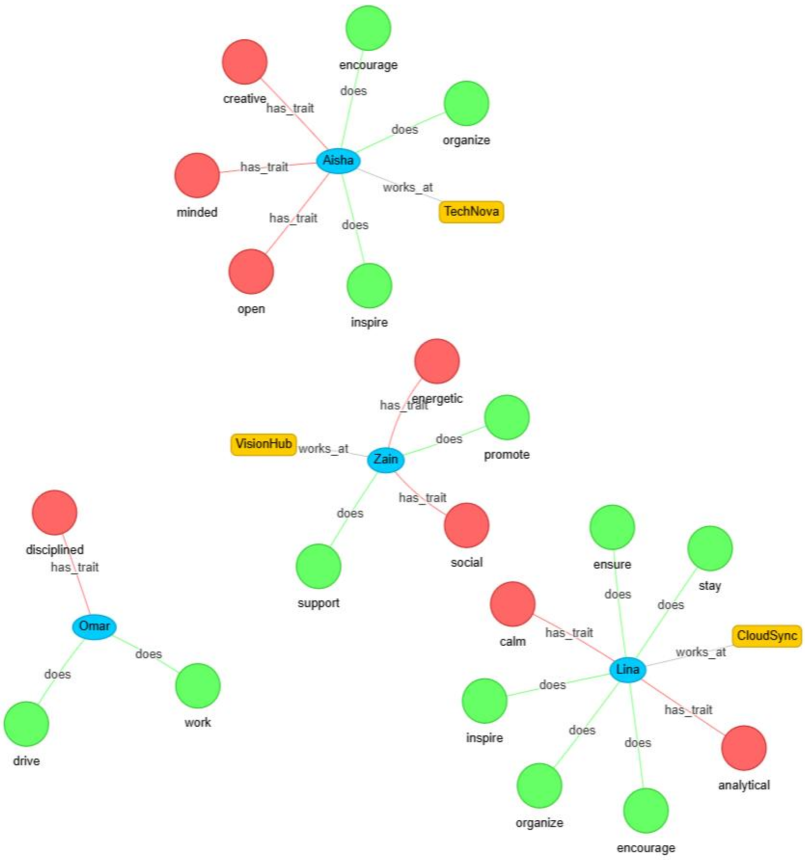

Figure 2: Knowledge graph visualization

4. LLM Session

The LLM was consulted iteratively to suggest entity extraction strategies, design relationship edges, and recommend evaluation methods. The workflow followed a chain-of-thought approach: first, the LLM was asked for data preprocessing strategies, then for extraction techniques, and finally for graph representation and evaluation metrics. Multiple LLMs were utilized in this project, including ChatGPT, Gemini, Blackbox AI, and Perplexity, to cross-validate approaches and ensure robust guidance for implementing the knowledge graph pipeline.

- what is KG?
- what's personality traits ?
- how are they related?
- how to buld the model without dataset (is text simulation enogh)

Excellent starting questions, Fadwa — these are exactly what Intellumia wants you to ask and understand before jumping into code. Let's go through each clearly and step-by-step 👇

🧠 1. What is a Knowledge Graph (KG)?

A **Knowledge Graph** is a way to represent information in a **structured, connected format**.
It's made up of:

- **Nodes (entities)** → represent things like *people, places, organizations, or concepts*.
- **Edges (relationships)** → represent how those entities are connected.
- **Attributes (properties)** → describe the features of the nodes or edges.

what can i do in this project to make my work stand out from others?
what do u think are the limitations in my work ?
how can i improve it in the futuer?

Here's a structured way to think about **making your project stand out**, **acknowledging limitations**, and **planning future improvements**:

---

1. How to make your work stand out

- **Richer Personality Modeling:**
  Right now, traits are extracted mainly from adjectives near names. You could use **LLM-assisted sentiment or personality analysis** to assign Big Five traits (openness, conscientiousness, extraversion, agreeableness, neuroticism) or other personality metrics to each person.
- **Enhanced Relationship Extraction:**
  Beyond `works_at`, try detecting **other social or organizational relations** like `collaborates_with`, `manages`, or `mentors`.
- **Context-aware activity extraction:**
  Instead of just taking verbs, you could use **dependency parsing** to extract who does what to whom.

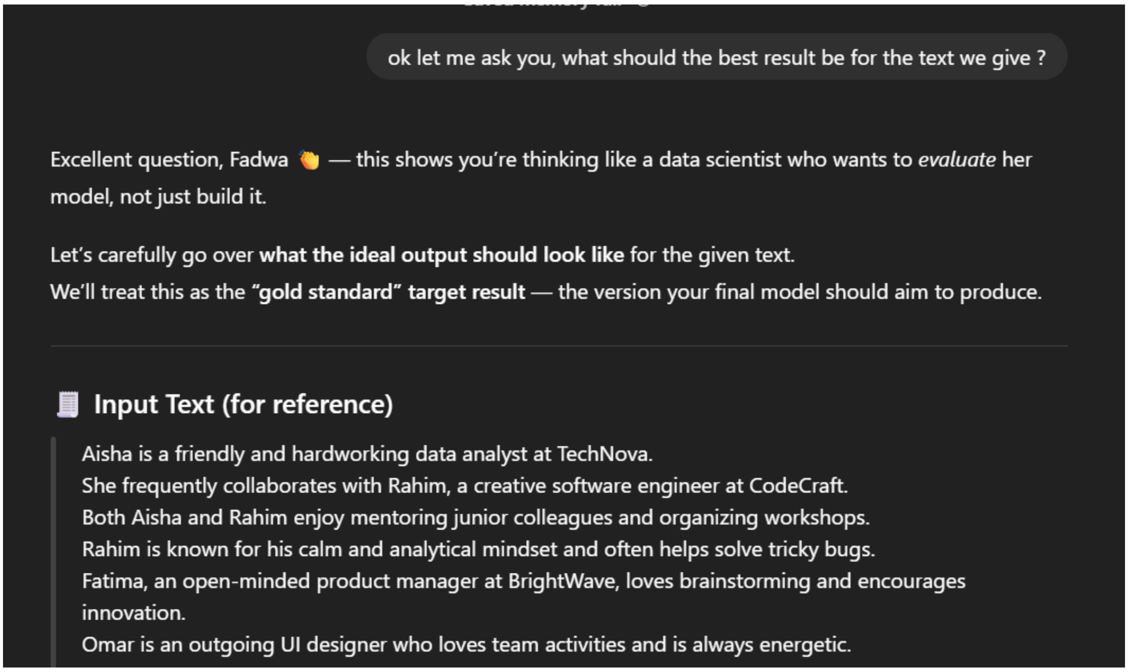

Figure 3: LLM Session

E. Conclusion

This project successfully generated a knowledge graph incorporating both organizational relationships and individual personality traits. It highlights the utility of LLM-assisted research for informing design decisions, the role of NLP in structuring unstructured data, and the potential for knowledge graphs to provide actionable insights into human behaviour and organizational structures. The interactive graph is a powerful tool for both analysis and communication, and the methodology can be adapted to larger, more complex datasets for real-world applications.