# Implementation of an innovative Conceptual Graph-Based Recommendation System


Supervised by Mrs WIEM ZAOUGA

Elaborated by Mzoughi Mohamed ali,Braiki Farah, Allouche Mohamed Aziz, namani zitouni Nada,Guetat Louay,Jalleli Oussama

Project AI & Cognition


## Introduction

Our project aims to address the complex challenge of developing an effective recommender system based on graph embeddings. In the era of information overload, users often struggle to find relevant content tailored to their preferences. The goal of our research is to leverage advanced graph neural network techniques, natural language processing (NLP), and machine learning to improve recommendation accuracy and provide users with more personalized and context-aware suggestions. This involves navigating the challenges of sparse and dynamic graph structures, as well as incorporating additional node attributes for a richer understanding of the relationships within the graph. In parallel, proactive risk management aligned with the principles of PMBOK 6th edition will be integrated throughout the project lifecycle, ensuring a comprehensive and anticipatory approach.

## Data Comprehension

Data comprehension is a critical aspect of our project, involving a multifaceted approach to extract, process, and understand information from the PMBOK 6th edition document. Our initial step includes the conversion of the PDF document into text and image data. The text data is extracted from each page, capturing the rich content of the PMBOK guide. Simultaneously, images are saved to preserve visual information such as diagrams, graphs, and figures embedded in the document.

### Data Preprocessing

**Data cleaning** **Data transformation**

### Text data processing:

- Text cleaning
- Stop words
- Text sectioning using regex expressions: Regular expressions are employed to address any inconsistencies, such as instances of repeated characters or four-digit numbers that may not contribute to the meaningful information.

PyMuPDF **Transformers** Σ

### Image processing:

- Images are processed to identify and extract Regions of Interest (ROIs) containing relevant visual content. We employ edge detection techniques using the Canny algorithm, followed by contour analysis to identify and filter out areas of interest. These ROIs are then saved as separate images, each associated with a unique UF'

PDF → OpenCV

## Feature extraction :

**1.Synonyms and Definitions:**

1. Leveraging **spaCy's en_core_web_md** model for precise NLP.
2. Identifying synonyms and definitions to capture nuanced meanings.

**2.Concepts, Objects, and Predicates:**

1. Custom parsing algorithms for entity extraction.
2. Fuzzy string matching using **fuzzywuzzy** for robust relationship recognition.

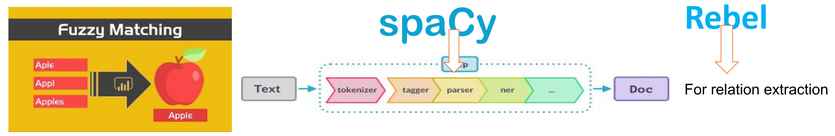
**3.References:**

1. Integration of associated URLs for visual context.
2. Enriching the dataset with both textual and visual references.

**Transformer-based Models:**

1. Utilizing transformers library for advanced pattern recognition.
2. Enhancing system capabilities for nuanced content analysis.

Our feature extraction process combines these elements, providing a comprehensive understanding of the PMBOK guide's content and laying the foundation for a powerful and context-aware recommendation system.




| | Index | Concepts | Text | Refrences | def | Keywords | Synonyms | Detected_Auxiliary | Type_relation | Concept_of_type_relation |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 11.1 | plan risk management | plan risk management is the process of defining... | captures/page_0_image_1.png | The purpose of this risk management process is... | ['project risk management process', 'plan comp... | ['The PLAN RISK MANAGEMENT is a critical compo... | ['enterprise environmental factors', 'risk man... | studies | risks |
| 2 | 11.1.1.1 | project charter | described in section 4.1.3.1. the project char... | sections/4.1.3.1_page117.png | This section describes the project charter. It... | ['project charter documents highlevel'] | ['The PROJECT CHARTER is the most recent proje... | [] | has part | project team |
| 3 | 11.1.1.2 | project management plan | described in section 4.2.3.1. in planning proj... | sections/4.2.3.1_page122.png | In planning project risk management, the risk ... | ['planning project risk management', 'approved... | ['The PROJECT MANAGEMENT PLAN is designed to m... | ['plan risk management', 'risk management plan'] | facet of | project risk |
| 4 | 11.1.1.3 | project documents | project documents that can be considered as in... | sections/13.1.3.1_page650.png | This chapter describes the stakeholder registr... | ['projects stakeholders provides overview', 'd... | ['The project documents are part of the Projec... | [] | has part | project team |

## Conclusion and perspective

In conclusion, our project successfully tackled the challenge of developing a robust recommender system, utilizing advanced NLP techniques and GNN models. The feature extraction process, covering synonyms, definitions, and essential entities, lays a strong foundation for nuanced content understanding within the PMBOK guide.

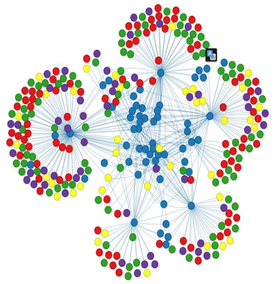
Looking ahead, our perspectives involve the integration of transformer-based models for refined content analysis. We aim to explore dynamic graph structures and evolve the GNN architecture to adapt to different PMBOK editions, ensuring continuous optimization. Envisioning a chatbot interface aligned with PMBOK 6th edition principles adds a real-time, interactive dimension to our system, transforming it into an integral part of project management workflows. These perspectives reflect our commitment to advancing the system in line with evolving project management practices.

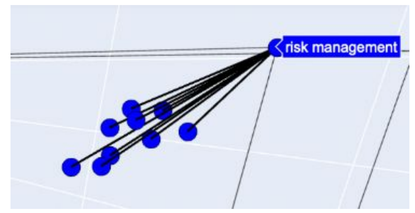
## Detailed Aproach

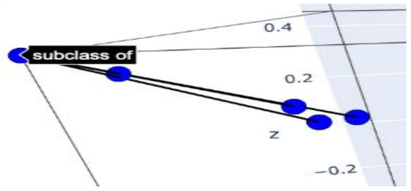
### Node and edge embeddings:

- **Node2Vec**
- **Bert embeddings (bert uncased)**

### Graph construction:

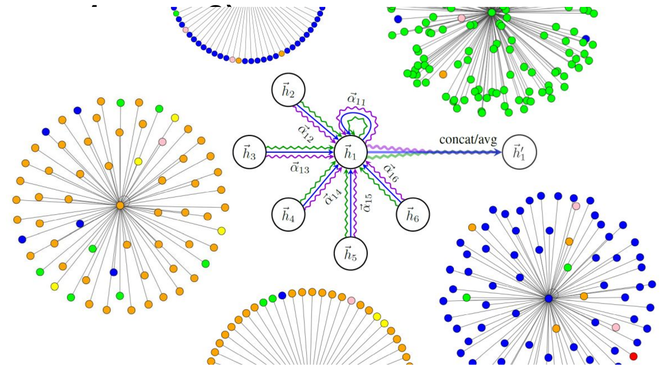





### GNN architecture :

**2 Graph Attention Network (GAT) Layers**




## Evaluation

**Used RMSE criteria for 20% of data test**

```
Mean Squared Error on Test Set: 2.2422869205474854
```