Projet IA & Cognition

RAPID

(Risk Analysis & Proactive Intelligence for Decision-Making)

Aziz Khalsi

Raef Khalifa

Dhafeur Ankoud

Fadi Kharroubi

Houssem Jabally

Mohamed Ben Hmida

Hadil Amamou

5 DS 3

2024/2025

Business: TenStep Tunisia

https://www.tenstep.tn/

16, Rue Ibn Zeidoun Ksar tej-Menzah V ,2080 Ariana, Tunisie, Tunis, Tunisia

Summary

Section 1. Project Context

1.Introduction

The *RAPID* project, initiated by the *Deep-Innovators* team, aims to revolutionize the field of risk management by introducing an AI-based solution that leverages historical data to enhance decision-making processes. The primary objective is to automate the generation of responses in risk management scenarios, thereby increasing both the speed and accuracy of analyses. The manual, error-prone processes currently employed in many organizations hinder effective decision-making and slow down the identification of risks, which this project seeks to resolve.

The team behind the project includes Khalsi Aziz, Raef Khalifa, Hadil Amamou, Dhafeur Ankoud, Houssem Jabally, Mohamed Ben Hmida, and Fadi Kharroubi, all contributing to developing a state-of-the-art AI solution for risk management.

2.Existing solution:

In the field of risk management, several software solutions are available on the market. However, none of these solutions leverage a content-based recommendation system like the one proposed in this project, which represents a true innovation. Below, we provide an overview of the main existing solutions.

Corporater: An enterprise risk management platform.

ServiceNow: A cloud solution for risk management.

RiskPro: Software designed for risk management in different sectors.

A1 Tracker: A tracking and risk management solution for businesses.

These solutions offer a variety of features for risk management, ranging from planning and monitoring to compliance and security. However, they do not rely on a personalized content-based recommendation system. In fact, most of these tools focus on data collection and analysis without utilizing knowledge graphs or advanced recommendation techniques.

3. Objectives:

The objective of this work is to use deep learning architectures linked to semantic technologies to respond to current problems in the field of risk management knowledge. This objective includes managing issues related to PRM terminologies in compliance with risk management standards, as well as formalizing the content of these practices, moving from a human-readable format to a computer-interpretable format.

Section 2. Probleme Statement

1. Problem Statement

Risk management is a critical component of modern business operations, but current methods often fail to meet the demands of today's fast-paced environments. Several key issues are associated with traditional risk management:

- **Inefficient Manual Methods**: The manual generation of responses to risks is slow and often inconsistent, leading to delays and inaccuracies in the decision-making process. Decision-makers must manually sift through historical data and other sources of information, which consumes valuable time and leaves room for errors.
- **Manual Overload and Lack of Anticipation**: The current approach overloads risk analysts with manual tasks, limiting their ability to anticipate risks and respond in a timely manner.
- **Underutilized Data**: One of the biggest challenges in traditional risk management is the failure to effectively use historical data. As a result, risk analysis is often based on incomplete information, and the context surrounding current risks is insufficiently addressed.

In summary, there is a pressing need for automation in risk management to reduce manual labor, improve the use of historical data, and speed up decision-making processes.

2. Challanges

To address this issue, several key challenges must be tackled:

1.Knowledge Extraction: How can we extract and structure relevant information from the book "Practice Standard for Project Risk Management 2017" to create a robust knowledge base.

2.Knowledge Graph Development: How can we design and implement an effective knowledge graph that captures the relationships between project management concepts, associated risks, and recommended best practices?

3.Recommendation System Design: How can we develop a recommendation system that uses the knowledge graph to generate personalized advice and relevant recommendations for users, based on their specific needs and the context of their projects?

4.Validation and Evaluation: How can we assess the accuracy and effectiveness of the recommendation system and ensure it provides useful and applicable recommendations to students and project managers?

By addressing these challenges, this project aims to provide an innovative and valuable tool for project and risk management, facilitating access to recommendations based on proven knowledge and recognized practices in the PMP field.

CRISP-DM Methodology

To develop the *RAPID* system, the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology will be used. This approach ensures that the AI objectives are fully aligned with business goals and that the models developed are robust and reliable:

1. **Understanding Business Objectives**: Ensuring that the AI objectives meet the company's needs and provide valuable outcomes.
2. **Effective Data Management**: Preparing historical risk data to ensure high-quality inputs for the modeling process.
3. **Flexible Modeling**: Experimenting with various models (NLP, RAG, ML) to find the most suitable ones for automating risk management processes.
4. **Continuous Evaluation**: Using iterative evaluation processes to test model performance and refine recommendations as needed.

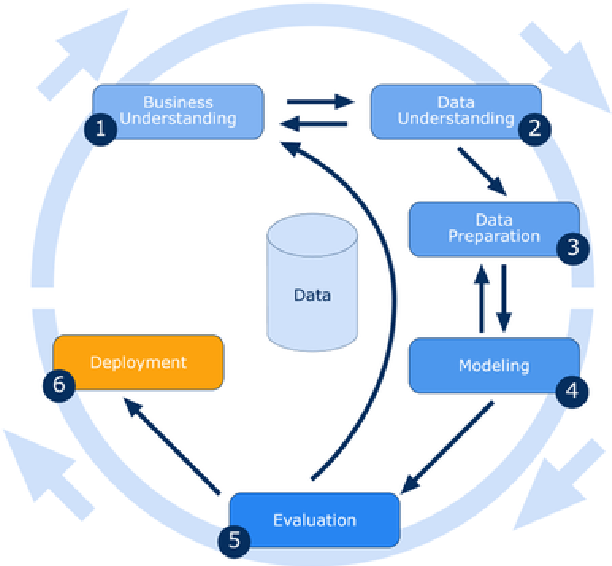


Figure 1 - CRISP-DM methodology

Business Objectives

The overarching business objectives of the *RAPID* project are aimed at enhancing the effectiveness and efficiency of risk management processes. These objectives focus on automating data extraction and response generation, maximizing the use of historical data, and improving the accuracy of responses:

- **Providing Common Vocabulary with a Complete Package Of Definitions :** This goal aims to provide a common vocabulary that includes a comprehensive set clear definitions. This approach ensures that all stakeholders, regardless of their field or experience, share a unified understanding of key terms.

- **Enhance Personalized Risk Management Recommendations :** The goal is to provide tailored recommendations for managing risks, ensuring that each user receives guidance specific to their unique risk profile, thereby improving decision-making and reducing uncertainty in risk scenarios.

- **Speed Up Response Generation for Risk Management :** The aim is to automate the process of generating risk management responses, reducing the time needed to analyze data and provide actionable insights, which enables faster decision-making in critical situations.

- **Improve Decision-Making Accuracy in Risk Management** : The objective is to enhance the accuracy of decisions by providing data-driven insights, ensuring that the recommendations are reliable and based on comprehensive knowledge extracted from the book, thus minimizing errors in risk assessment.

Data Science Objectives

In alignment with the business goals, several technical objectives have been outlined to ensure the success of the project. These include:

- **Develop a Vocabulary Extraction Model Using NLP for Consistent Terminology :** Extract definitions and key terms from the book to create a unified vocabulary, ensuring consistent understanding across all stakeholders.
    - **Data correspondence required for each business objective** Data on all relevant terms, definitions, and key concepts from the book is necessary to build a comprehensive vocabulary for risk management.
    - **Explanation of data sources and data exploitation** : The book is the sole source of data. NLP techniques will be used to extract definitions, terms, and context, ensuring that stakeholders share a unified understanding of key terminology.

- **Develop a Knowledge Graph-Integrated Content-Based Recommender System :** Extract key concepts and relationships from the book using NLP and create a knowledge graph to generate personalized risk management recommendations.
    - **Data correspondence required for each business objective** Concepts, relationships, and attributes from the book are needed to build a knowledge graph for personalized recommendations.
    - **Explanation of data sources and data exploitation** : The book's content is analyzed using NLP to extract concepts and their relationships. This information is organized in a knowledge graph that powers the recommender system, which aligns the recommendations with user preferences.

- **Implement Retrieval-Augmented Generation (RAG) for Automated Responses :** Use concepts and relationships from the book with RAG to automate risk response generation, providing timely, relevant answers for decision-making.

    - **Data correspondence required for each business objective**The extracted concepts and relationships from the book are required to support automated response generation.

    - **Explanation of data sources and data exploitation** : The knowledge graph created from the book serves as a data source for RAG. This model uses retrieved historical information and current risk scenarios to generate relevant responses efficiently.

- **Develop a Graph Neural Network (GNN) for Contextual Risk Analysis :** Utilize the knowledge graph with a GNN to derive contextual understanding of risks, improving the accuracy of decision-making in risk management.

    - **Data correspondence required for each business objective**Data representing the relationships between risk-related concepts from the book is essential to train the GNN.

    - **Explanation of data sources and data exploitation** : The book provides all relevant concepts and relationships, which are transformed into a knowledge graph. This graph is then used as input for the GNN, enabling the model to derive context-aware insights to improve decision accuracy.

Section 6. Metrics

Metrics

To measure the effectiveness of the AI-driven risk management system, several KPIs have been established:

- **Accuracy of Data Extraction**: This KPI measures the accuracy of the system's ability to extract relevant data. The target is to achieve an accuracy rate of 90% in data extraction processes, which will directly impact the quality of the risk analysis.
- **Success Rate of Retrievals**: A cosine similarity score between newly identified risks and risks retrieved from historical data will be measured. The goal is to achieve a similarity score of 0.85 or higher, ensuring that the system retrieves highly relevant data.
- **Response Processing Time**: The time required to generate a response after receiving a request will be minimized. The objective is to reduce the processing time to less than 30 seconds per response, significantly improving the speed of risk management.

- **Generated Response Score**: Expert evaluations of the relevance and usefulness of AI-generated responses will be used to score the system's performance. The goal is to achieve a minimum rating of 4.5 out of 5 in these evaluations, ensuring high-quality output.
- **Interpretation of data validation :** To build a knowledge graph that visualizes the relationships between concepts, A knowledge graph characterized by key metrics such as:
  - Total Nodes
  - Total Relationships
  - Average Degree
  - Number of Relationship Types
  - Graph Density

Section 7. Content-Based Recommender System

1.Introduction:

A content-based recommender system suggests items like book by analyzing their attributes and matching them to a user's past preferences. Unlike collaborative systems, which use data from similar users, content-based systems focus on the item's content to deliver personalized recommendations.

2.How content-based recommender systems are built:

Here's a step-by-step approach to building a Content-Based Recommender System:

1. Data Collection :

- User Data: Collect user interaction history (articles read, videos watched), explicit preferences (ratings, favorites), or implicit signals (clicks, time spent).
- Item Data: Gather features of the items to be recommended (title, description, keywords, categories).

2. Feature Extraction :

- Use techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or embeddings (e.g., Word2Vec) to transform the text of items (title, description) into numerical vectors.
- For other types of content (images, videos), extract relevant features such as dominant colors, objects, or metadata.

3. User Profile Creation :

- Build a user profile by aggregating the features of items the user has interacted with or liked. This profile can be represented as the average or weighted sum of the item feature vectors.

4. Similarity Measurement :

- Use metrics like cosine similarity or Euclidean distance to calculate the similarity between the user profile and items the user hasn't interacted with. Higher similarity indicates a higher likelihood of recommendation.

5. Recommendation Generation :

- Rank the items based on their similarity to the user profile. The items with the highest similarity scores are recommended to the user.

6. Refinement and Feedback :

- Implement feedback loops to adjust recommendations based on new user interactions or explicit feedback to refine the system over time.

3.Knowledge graph:

A Knowledge Graph is a structured representation of information that connects entities (such as people, places, objects, or concepts) through semantic relationships. Each entity is represented as a node in the graph, and the edges (or links) between the nodes represent relationships between these entities. This approach organizes complex information in a way that highlights how different entities are interrelated, facilitating data discovery, search, and navigation.

Key components of a Knowledge Graph:

1.Entities: The objects or concepts being represented (e.g., books, movies, people).

2.Relationships: The semantic connections between entities (e.g., an author writes a book, an actor stars in a movie).

3.Attributes: Additional information about each entity (e.g., the title of a book, the release year of a movie).

4.Integration of the Knowledge Graph into a Recommendation System:

A Knowledge Graph enriches a content-based recommender system by providing a deeper and more structured view of the relationships between items and users. Here is how it can be integrated:

1. Enriching item characteristics :

The Knowledge Graph can provide additional relationships and contexts around items. For example, to recommend movies, the graph can connect movies by common actors, similar genres, or even by more abstract concepts like cultural influence.

If a user likes a specific movie, the Knowledge Graph can suggest other movies that share actors, directors, or themes.

2. Discovering implicit relationships :

Thanks to the semantic relationships in the Knowledge Graph, indirect connections between items can be found. For example, a user who likes a book can be recommended a movie based on the same book, or another work by the same author, even if these connections are not obvious at first glance in a classical model.

3. Creating richer user profiles :

The user profile can be enriched by the entities and relationships of the Knowledge Graph. If a user shows interest in certain authors or actors, the system can infer additional preferences based on the relationships in the graph.

For example, if a user likes several science fiction movies, the Knowledge Graph can identify subgenres or themes specific to this type of content to refine recommendations.

4. Semantic-based recommendations :

Unlike a simple comparison of textual features, a Knowledge Graph allows reasoning on the relationships between items. This allows recommending items by taking into account semantics, thus offering more relevant and sophisticated suggestions.

For example, a recommended book may not have similar terms to those already read by the user, but may have relevant semantic relationships, such as a theme shared with previous books via the

Knowledge Graph.

5. Explainability of recommendations :

One of the advantages of using a Knowledge Graph is that it makes recommendations more explainable. For example, a system might explain a recommendation by saying, "This book is recommended to you because it is by the same author as the one you recently read," or "This actor has appeared in several movies that you like."

Analytical Approach

The *RAPID* system will use several cutting-edge techniques to achieve its objectives:

- **Data Extraction**:

  NLP techniques such as Spacy will be used to extract relevant risk management data from text. This includes cleaning and splitting text for easier processing.

- **Construction of the Conceptual Graph:**

  Once the knowledge has been extracted, it will be organized in a knowledge graph, using deep learning (DL) architectures. This graph will represent risk management concepts, their relationships, properties of objects, as well as as the associated rules and axioms.

  a. **Tools and techniques used:**

     RDF (Resource Description Framework) and OWL (Web Ontology Language): To structure data in the form of a graph.

- **Rule Generation and Recommendation System :**

  The recommendation algorithm will explore the knowledge graph to infer suggestions adapted to user requests. From the graph repository and Generated rule bases, recommendations will be formulated in real time.

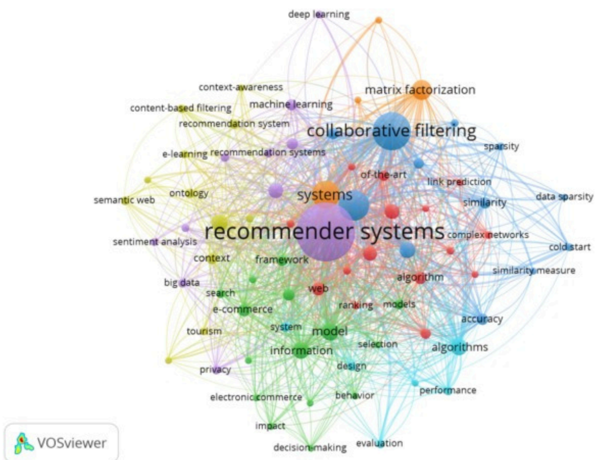


Figure 2 - the knowledge graph

- **Vector Creation**:

  Embedding techniques such as SentenceTransformers will be used to convert text into vectors, making it easier to compare and analyze.

- **RAG System**:

  The RAG system will use ChromaDB, a vector database, to retrieve relevant information. This data will then be passed through a generative model like Llama 3.1:8b to create detailed responses.

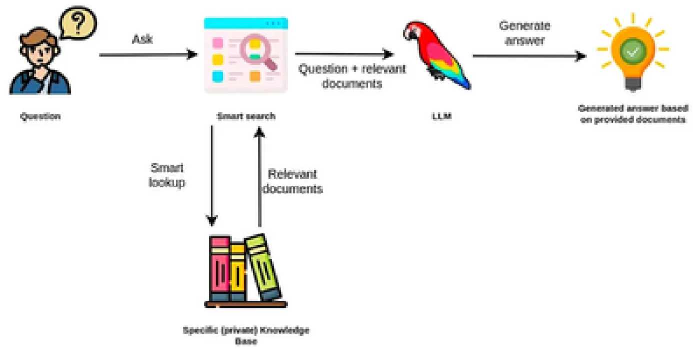


Figure 3 - RAG description

Section 9. Corpus Preprocessing Steps with Clear, Defined Outputs

Corpus Preprocessing Steps with Clear, Defined Outputs

1. Text Extraction :

- **Definition:** To extract raw text from PDF documents.
- **Method:** Using PDFreader, we extracted both the text and figures (such as images and tables) from the documents.
- **Output:** Unprocessed text data split into chapters and figure placeholders to indicate where images and tables occur in the document.

```
Saving chapters and printing snippets...
--- Chapitre 1 Introduction Snippet ---
INTRODUCTION
P roject Management Institute (PMI) practice standards are guides to the use of a tool, technique, or process
identifi ed in A Guide to the Project Management Body of Knowledge ( PMBOK ® G uide – Fourth Edition) or
other PMI standards.
-------------------------

--- Chapitre 2 Principles And Concepts Snippet ---
PRINCIPLES AND CONCEPTS
2.
-------------------------

--- Chapitre 3 Introduction To Project Risk Management Processes Snippet ---
INTRODUCTION TO PROJECT RISK MANAGEMENT PROCESSES
3.
-------------------------

--- Chapitre 4 Plan Risk Management Snippet ---
PLAN RISK MANAGEMENT
4.
-------------------------

--- Chapitre 5 Identify Risks Snippet ---
IDENTIFY RISKS
5.
-------------------------

--- Chapitre 6 Perform Qualitative Risk Analysis Snippet ---
PERFORM QUALITATIVE RISK ANALYSIS
6.
-------------------------

--- Chapitre 7 Perform Quantitative Risk Analysis Snippet ---
PERFORM QUANTITATIVE RISK ANALYSIS
7.
-------------------------

--- Chapitre 8 Plan Risk Responses Snippet ---
PLAN RISK RESPONSES
T he Plan Risk Responses process determines effective response actions that are appropriate to the priority
of the individual risks and to the overall project risk.
-------------------------

--- Chapitre 9 Monitor And Control Risks Snippet ---
MONITOR AND CONTROL RISKS
The effectiveness of Project Risk Management depends upon the way the approved plans are carried out.
-------------------------
```

Figure 4- Output of unprocessed text data split into chapters

2. Chapter Identification and Figure Detection :

- **Definition:** To segment text into chapters and identify figures within each chapter.
- **Method:** A summary-based approach was used to extract chapters and tag figure pages (such as tables and images) for further processing.
- **Output:** Text divided into chapters with markers indicating where figures appear.

```
Extracting figure pages from pages 1 to 12...
Figure pages found: [2, 6, 17, 23, 27, 29, 32, 33, 38, 41, 44, 49, 53]
Figure pages saved to 'output/figure_pages.txt'.

Extracting full text from pages 13 to 124...
Full text saved to 'output/extracted_text.txt'.

Dividing text into 9 chapters...
 - Chapitre 1 Introduction: 2430 words
 - Chapitre 2 Principles And Concepts: 1797 words
 - Chapitre 3 Introduction To Project Risk Management Processes: 1729 words
 - Chapitre 4 Plan Risk Management: 2196 words
 - Chapitre 5 Identify Risks: 1526 words
 - Chapitre 6 Perform Qualitative Risk Analysis: 1673 words
 - Chapitre 7 Perform Quantitative Risk Analysis: 1939 words
 - Chapitre 8 Plan Risk Responses: 2582 words
 - Chapitre 9 Monitor And Control Risks: 15397 words
```

Figure 5- Text divided into chapters with markers

3. Advanced OCR and Figure Descriptions :

- **Definition:** To create descriptive text for figures using OCR and LLAMA3.2.
- **Method:** We applied advanced OCR techniques (PyTesseract) in combination with large language models (LLAMA 3.2) to generate descriptions for tables, images, and other figures within the documents.
- **Output:** Descriptive text associated with each figure, to be inserted back into the text where figures are referenced.

Preference Factors	
1	Equally Preferred
2	Mildly Preferred
3	Moderately Preferred
4	Greatly Preferred
5	Always Preferred

Input Matrix (Preference Factors)				
	Cost	Time	Scope	Quality
Cost	1.00	0.25	0.33	0.20
Time	4.00	1.00	1.00	0.25
Scope	3.00	1.00	1.00	0.25
Quality	5.00	4.00	4.00	1.00

Note: Preference Factors input into the Dark Gray Area. Principal Diagonal is 1.0 by definition. Other cells calculated as 1 / preference factor for same objectives.

Calculated Factors (Preference Factor / Column Total)					Weighting Factors
	Cost	Time	Scope	Quality	Average of Row
Cost	0.08	0.04	0.05	0.12	0.1
Time	0.31	0.16	0.16	0.15	0.2
Scope	0.23	0.16	0.16	0.15	0.2
Quality	0.38	0.64	0.63	0.59	0.6
Sum	13.00	6.25	6.33	1.70	1.0

Based on the provided text, I'll break down the Figure: Preference Factors and Input Matrix. **Preference Factors** The table lists four preference factors: 1. **Equally Preferred**: Indicates that an objective is equally important to all others. 2. **Mildly Preferred**: Suggests that one objective is slightly more important than others. 3. **Moderately Preferred**: Implies that one objective is more significant than others, but not as crucial as those considered "Greatly Preferred". 4. **Always Preferred**: Indicates an objective is of utmost importance. **Input Matrix (Preference Factors)** The Input Matrix appears to be a table used for risk management or decision-making purposes. It's likely related to prioritizing objectives based on their relative importance (preference factors). Here's how it works: * The first row and column are labeled "Time", which might represent the timeframe within which decisions must be made. * The principal diagonal (from top-left to bottom-right) is filled with 1.0s, indicating that each objective has a preference factor of 1.0 for itself (i.e., it's always preferred when considering its own importance). * Other cells in the table contain values calculated as 1 / preference factor for the same objectives. This suggests that the preference factors are normalized to create a relative scale. * The numbers in the remaining cells represent the relative weight of each objective compared to others. **Calculated Factors (Preference Factor / Column Total)** The bottom row, "Calculated Factors", provides a summary of the preference factors for each column (objective). Here's what we can infer: 1. **Cost**: Has a calculated factor of 0.04, indicating that this objective is relatively less important compared to others. 2. **Time**: Has a calculated factor of 0.05, suggesting it's slightly more important than Cost. 3. **Scope**: The calculated factor is 0.16 for both the Time and Scope columns, implying these objectives are similarly important. In summary, this figure provides a matrix for evaluating and prioritizing multiple objectives (Cost, Time, Scope) based on their relative importance or preference factors. The calculated factors help to normalize and compare the weights of each objective, facilitating informed decision-making in a risk management context.

Figure 6- Descriptive text associated with each figure

4. Figure Description Integration

- **Definition:** To integrate figure descriptions into the document.
- **Method:** The extracted text was processed to insert the descriptions generated by the OCR and language models at the appropriate figure markers.
- **Output:** Complete text with figures described and integrated into their respective positions.

```
Mapping figure descriptions to chapters...
Added page_101_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_101_img_2_description.txt to chapitre_9_monitor_and_control_risks
Added page_102_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_106_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_107_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_107_img_2_description.txt to chapitre_9_monitor_and_control_risks
Added page_108_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_112_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_113_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_116_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_14_img_1_description.txt to chapitre_1_introduction
Added page_18_img_1_description.txt to chapitre_1_introduction
Added page_29_img_1_description.txt to chapitre_3_introduction_to_project_risk_management_processes
Added page_35_img_1_description.txt to chapitre_4_plan_risk_management
Added page_39_img_1_description.txt to chapitre_5_identify_risks
Added page_41_img_1_description.txt to chapitre_5_identify_risks
Added page_44_img_1_description.txt to chapitre_6_perform_qualitative_risk_analysis
Added page_45_img_1_description.txt to chapitre_6_perform_qualitative_risk_analysis
Added page_50_img_1_description.txt to chapitre_7_perform_quantitative_risk_analysis
Added page_53_img_1_description.txt to chapitre_7_perform_quantitative_risk_analysis
Added page_56_img_1_description.txt to chapitre_8_plan_risk_responses
Added page_61_img_1_description.txt to chapitre_8_plan_risk_responses
Added page_65_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_89_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_89_img_2_description.txt to chapitre_9_monitor_and_control_risks
Added page_90_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_91_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_92_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_93_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_95_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_97_img_1_description.txt to chapitre_9_monitor_and_control_risks
Added page_97_img_2_description.txt to chapitre_9_monitor_and_control_risks
```

5. Text Manipulation :

- **Normalization:** The extracted text was normalized by converting it into lowercase, removing punctuation, and standardizing text formats (e.g., dates, numbers).

```
Summary of stop words removed for each chapter:
 - Chapitre 1 Introduction: are, to, the, of, a
 - Chapitre 2 Principles And Concepts: and, t, his, the, to
 - Chapitre 3 Introduction To Project Risk Management Processes: to, and, all, are, is
 - Chapitre 4 Plan Risk Management: and, of, the, t, he
 - Chapitre 5 Identify Risks: and, of, the, a, be
 - Chapitre 6 Perform Qualitative Risk Analysis: and, of, the, t, he
 - Chapitre 7 Perform Quantitative Risk Analysis: and, of, the, the, a
 - Chapitre 8 Plan Risk Responses: t, he, that, are, to
 - Chapitre 9 Monitor And Control Risks: and, the, of, the, the
```

Figure 7- Extracted Text

- **Tokenization:** Text was split into individual tokens (words or phrases) to allow for further linguistic analysis.
- **Lemmatization:** Each word was reduced to its base or dictionary form (lemma) to avoid duplication of similar words with different inflections.

```
All chapters have been lemmatized and saved.

Summary of lemmatization applied to each chapter:
 - Chapitre 1 Introduction: standards -> standard, guides -> guide, targeted -> target, audiences -> audience, projects -> project
 - Chapitre 2 Principles And Concepts: introduces -> introduce, ideas -> idea, required -> require, projects -> project, following -> follow
 - Chapitre 3 Introduction To Project Risk Management Processes: PROCESSES -> process, projects -> project, undertakings -> undertaking, based -> base, a
ssumptions -> assumption
 - Chapitre 4 Plan Risk Management: Objectives -> objective, processes -> process, executed -> execute, activities -> activity, requires -> require
 - Chapitre 5 Identify Risks: Objectives -> objective, Risks -> risk, managed -> manage, completed -> complete, aims -> aim
 - Chapitre 6 Perform Qualitative Risk Analysis: Objectives -> objective, assesses -> assess, evaluates -> evaluate, characteristics -> characteristic, r
isks -> risk
 - Chapitre 7 Perform Quantitative Risk Analysis: Objectives -> objective, provides -> provide, based -> base, plans -> plan, considering -> consider
 - Chapitre 8 Plan Risk Responses: RESPONSES -> response, determines -> determine, actions -> action, risks -> risk, takes -> take
 - Chapitre 9 Monitor And Control Risks: depends -> depend, approved -> approve, plans -> plan, carried -> carry, executed -> execute

Lemmatized chapters saved to 'output/lemmatized_chapters' directory.
```

Figure 8- lemmatized chapters

- **Part-of-Speech Tagging (POS):** Grammatical tags were assigned to each word (e.g., noun, verb, adjective) to help understand its role in the sentence.

--- Chapitre 1 Introduction ---

	Word	POS Tag	Frequency
0	project	NOUN	102
1	management	PROPN	86
2	project	PROPN	80
3	risk	NOUN	71
4	management	NOUN	67

--- Chapitre 2 Principles And Concepts ---

	Word	POS Tag	Frequency
0	risk	NOUN	71
1	project	NOUN	65
2	project	PROPN	32
3	management	PROPN	30
4	risk	PROPN	27

--- Chapitre 3 Introduction To Project Risk Management Processes ---

	Word	POS Tag	Frequency
0	risk	NOUN	114
1	project	NOUN	60
2	management	PROPN	49
3	risk	PROPN	43
4	project	PROPN	42

Figure 9- POS Tagging

- **Output:** A fully processed and cleaned corpus, where each token is normalized, lemmatized, and tagged for part-of-speech.

6. Word Frequency Analysis and Semantic Filtering :

- **Definition:** To identify important and noisy words in each chapter.
- **Steps:**
    - Frequent Words: Words occurring more than 10 times in a chapter were identified as frequent words and considered domain-specific.
    - Infrequent Words: Words occurring 10 times or less were evaluated based on their semantic similarity to frequent words.
    - Semantic Similarity: Using models like sentence-transformers, we computed similarity scores. Infrequent words with a similarity score above 0.4 were kept as important words, while others were marked as noisy.
- **Output:** A filtered list of important words (both frequent and semantically related infrequent words) and noisy words for each chapter.

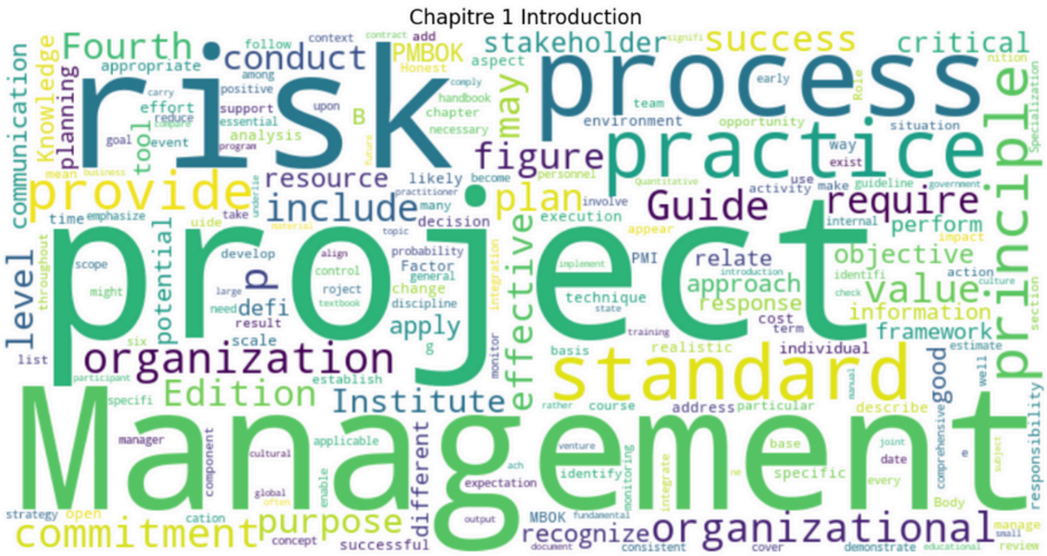


Figure 10 - List of important words

Chapitre_1_Introduction

Noisy Words (Sample)		Important Words (Sample)	
Word	**Similarity Score**	**Word**	**Similarity Score**
roject	0.338719	introduction	0.489965
identifi	0.351460	guide	0.519123
uide	0.322987	use	0.524136
target	0.389802	technique	0.575393
audience	0.374284	manager	0.801691

Figure 11 - list of noisy/important words

7. Concept Extraction and Relationship Identification :

- **Definition:** To identify and extract concepts and their relationships from the text.
- **Method:** Various grammar rules were applied to detect concepts and relationships. For instance:
  - Attribute Extraction: Concepts associated with adjectives or noun compounds (e.g., “risk management” or “high-level strategy”).
  - Relationship Extraction: Using subject-verb-object constructions (e.g., "risk affects outcomes") and possessive relationships (e.g., "management's role in risk").
- **Output:** A structured set of single-word and multi-word concepts, with their relationships prioritized to avoid overlaps and ensure context is preserved.

```
--- Chapitre_1_introduction ---
Most Pertinent Concepts:
               Word  Frequency
         management         24
            process         15
    risk management         11
           practice         11
              guide          7
          principle          7
         commitment          6
       organization          6
 project management          6
             figure          5
           resource          5
           approach          5
              level          5
              value          5
            project          4
          framework          4
                way          4
            edition          3
          knowledge          3
            section          3
Concept frequencies saved to output/concepts\Concepts_Chapitre_1_introduction.csv
Looking for lemmatized file: output/lemmatized_chapters\Chapitre_2_principles_and_concepts.txt

--- Chapitre_2_principles_and_concepts ---
Most Pertinent Concepts:
                    Word  Frequency
                 project          7
              management          6
                    risk          6
                     cid          6
             stakeholder          5
                 process          5
                  effect          4
                 manager          4
                attitude          4
                   level          3
                    plan          3
                  action          3
            project risk          3
 project risk management          2
                practice          2
                     use          2
               condition          2
```

Figure 12 - list of most pertinent concepts

```
Sample Summary for Chapitre_7_Perform_Quantitative_Risk_Analysis:
```

	Concept	Frequency	Attributes	Relationship	Related Concept
0	risk analysis	14	analysis, effective, model, new, objective, ov...	require project_example	project example
1	risk analysis	14	analysis, effective, model, new, objective, ov...	provide risk_response	risk response
2	impact	8	overall, potential, probability, risk	perform analysis	analysis
3	impact	8	overall, potential, probability, risk	perform analysis	analysis
4	datum	7	accurate, analysis, expert, present, quality, ...	occur reason	reason

```
Saved summary for Chapitre_8_Plan_Risk_Responses to output/summary\Summary_Chapitre_8_Plan_Risk_Responses.csv

Sample Summary for Chapitre_8_Plan_Risk_Responses:
```

	Concept	Frequency	Attributes	Relationship	Related Concept
0	plan	18	accordance, additional, analysis, analyze, app...	develop address	address
1	response	10	account, ach, additional, approach, arrive, bu...	implement risk	risk
2	response	10	account, ach, additional, approach, arrive, bu...	identify resource	resource
3	response	10	account, ach, additional, approach, arrive, bu...	include action	action
4	response	10	account, ach, additional, approach, arrive, bu...	develop address_risk_relate	address risk relate

Figure 13 - sample summary

Section 10. Interpretation of data validation

Interpretation of data validation :

1. **Definition:** To build a knowledge graph that visualizes the relationships between concepts.
2. **Output:** A knowledge graph characterized by key metrics such as:
   - Total Nodes: 573
   - Total Relationships: 747
   - Average Degree: 2.51
   - Number of Relationship Types: 206
   - Graph Density: 0.0044
   - Sample nodes include concepts like "project budget" and "management context," while relationships include types like "approval obtain" and "risk management."

```
INFO: Total Nodes Extracted: 573
INFO: Total Relationships Extracted: 747

===== Graph Metrics =====
Total Nodes: 573
Total Relationships: 718
Average Degree: 2.51
Number of Relationship Types: 206
Number of Connected Components: 45
Size of Largest Connected Component: 476
Graph Density: 0.0044
=========================


Sample Nodes:
   ConceptID                   Concept  Frequency       Attributes
0          1       part project budget          0    part, project
1          2      specify level impact          0            level
2          3     approval obtain order          0  approval, obtain
3          4  change management context         0       management
4          5        scope project risk          0          project

Sample Relationships:
   StartConceptID  EndConceptID RelationshipType
0             556           524         handbook
2             128           286         organize
3             128           372          include
4             128           144            write
5             406           377         underlie
```

Figure 14 - interpretation of data validation

1. Validation :

1. Concept and Relationship Extraction Validation

- **Nodes (573 Total)**: We successfully extracted 573 unique concepts, which are represented as nodes in the knowledge graph. These concepts include key domain-specific terms such as "project budget" and "management context."
  - **Validation**: A thorough review of the node list was performed to ensure that the identified concepts are relevant, distinct, and correctly reflect the underlying text. This step involved cross-referencing the extracted concepts with the source material to confirm that important terms were captured without redundancy or omission.
- **Relationships (747 Total)**: A total of 747 relationships were extracted, linking the identified concepts. These relationships represent interactions such as "approval obtain" and "risk management."
  - **Validation**: Each relationship was verified by checking the context from which it was extracted to ensure logical connections between the concepts. This involved sampling relationships and ensuring that they accurately represent meaningful links between the associated nodes.

2. Structural Validation

- **Average Degree (2.51)**: Each node has, on average, 2.51 connections, suggesting that the graph is moderately connected.
  - **Validation**: The degree distribution was examined to confirm that central concepts, such as "risk" and "management," are appropriately connected, while less central concepts maintain fewer relationships. This distribution is in line with domain expectations, where certain key concepts act as hubs.
- **Relationship Types (206)**: There are 206 distinct relationship types, which reflect the diversity of interactions between concepts in the text.
  - **Validation**: We reviewed the list of relationship types to ensure that the variations are meaningful and align with the domain-specific language used in the source material. Similar relationships were checked for consistency in labeling, preventing unnecessary fragmentation into multiple types.

3. Graph Density Validation

- **Graph Density (0.0044)**: The graph is relatively sparse, which is expected for a domain-specific knowledge graph. A low density indicates that while the graph captures relevant relationships, it avoids over-linking and excessive noise.
  - **Validation**: The sparsity was examined to ensure that it is appropriate for the domain. We validated that key concepts are connected without over-saturating the graph with weak or irrelevant links, thus maintaining clarity and interpretability.

1. Conclusion :

The validation process confirms that the constructed knowledge graph accurately reflects the extracted concepts and their relationships. Key concepts are well-represented, and the structural properties of the graph are aligned with domain expectations. This validation ensures that the graph is ready for further use in applications such as Graph Neural Networks (GNNs) or visualization tools for deeper analysis.

Section 11. A prototype of the graph obtained

Knowledge Graph Construction – Prototype :

1. **Definition :** The extracted concepts and their relationships were converted into a graph, where nodes represent concepts and edges represent relationships.

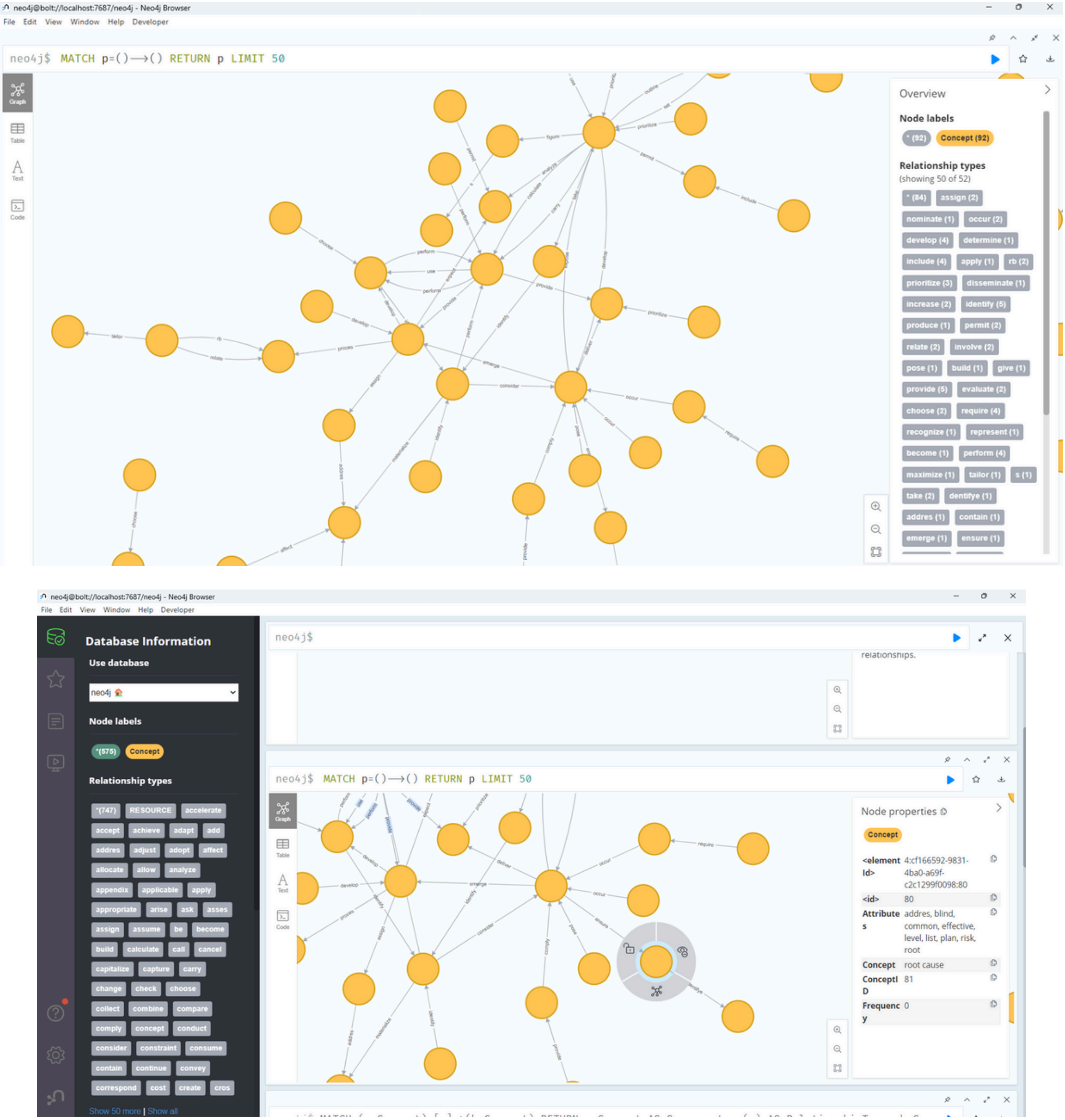


Figure 15 - graph representation

1. Embedding Generation with Node2Vec :

Node embeddings generated by Node2Vec are assigned as the x attribute in the PyTorch Geometric Data object. This attribute (data.x) serves as the initial node feature matrix, representing each node's position in the embedding space.

```
Computing transition probabilities: 100% 573/573 [00:00<00:00, 4227.21it/s]
```

2. Preprocess Data for PyTorch Geometric :

The NetworkX graph is converted into a PyTorch Geometric Data object using from_networkx(), enabling integration with PyTorch's deep learning tools for graph processing.

3. Define and Train the GNN Model :

We used a Graph Neural Network (GNN) to efficiently learn from and model our knowledge graph, which consists of nodes representing concepts (e.g., project management terms) and edges representing relationships between these concepts. GNNs are designed to handle such graph-structured data, enabling the model to learn contextualized representations of concepts by aggregating information from their connected neighbors (i.e., related concepts). This enriches the node representations, allowing for better query answering by more accurately capturing how concepts are interrelated in the knowledge graph.

```
Epoch 179, Loss: 0.7183
Epoch 180, Loss: 0.7060
Epoch 181, Loss: 0.6990
Epoch 182, Loss: 0.6963
Epoch 183, Loss: 0.6952
Epoch 184, Loss: 0.6949
Epoch 185, Loss: 0.6950
Epoch 186, Loss: 0.6956
Epoch 187, Loss: 0.6970
Epoch 188, Loss: 0.7006
Epoch 189, Loss: 0.7076
Epoch 190, Loss: 0.7229
Epoch 191, Loss: 0.7382
Epoch 192, Loss: 0.7557
Epoch 193, Loss: 0.7401
Epoch 194, Loss: 0.7259
Epoch 195, Loss: 0.7097
Epoch 196, Loss: 0.7025
Epoch 197, Loss: 0.6988
Epoch 198, Loss: 0.6978
Epoch 199, Loss: 0.6979
Model saved to /content/gnn_model.pth
```

Explanation of GNN Layers and Their Benefits :

1. GCNConv Layer (Graph Convolutional Layer) :

- **The Graph Convolutional Network (GCN)** layer is the primary building block of the GNN.
- **How it Works**: Each node updates its representation by aggregating (or "mixing") the features of its neighboring nodes. This means that, in each layer, a node not only retains its own information but also incorporates information from its connected nodes.
- **Benefits:**
  - Helps capture local interactions (relationships) between concepts.
  - Allows nodes to learn richer feature representations based not only on themselves but also on their neighbors. This is especially useful in query answering, where the meaning of a concept often depends on its context and connections.

2. ReLU Activation Function :

- **ReLU** (Rectified Linear Unit) is applied after each convolution step.
- **How it Works:** ReLU introduces non-linearity to the model by passing all positive values as they are and turning negative values to zero.
- **Benefits:**
  - Adds non-linearity, allowing the model to learn more complex patterns in the data.
  - Helps the model avoid the "vanishing gradient" problem during training, which ensures faster and more stable training.

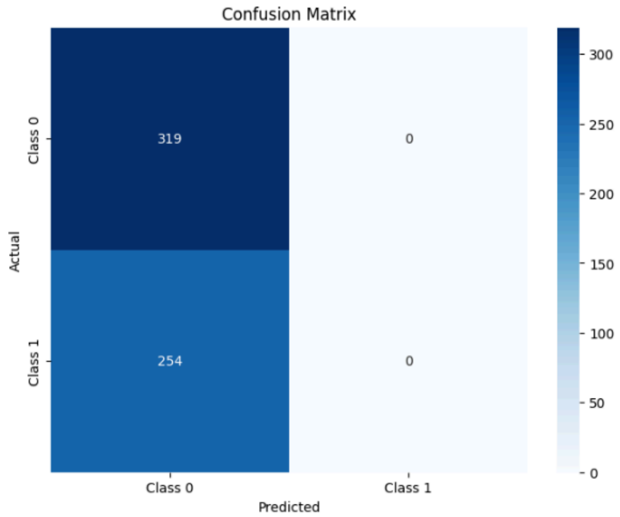
3. Multiple GCN Layers :

- **The network is composed of multiple GCN** layers stacked together.
- **How it Works:** Stacking layers allows each node to progressively gather information from further (multi-hop) neighbors. With two GCN layers:
  - **1st layer:** Allows nodes to aggregate information from their immediate neighbors (directly connected nodes).
  - **2nd layer:** Allows nodes to gather information from two-hop neighbors (i.e., connections of the neighbors).
- **Benefits:**
  - By stacking layers, the model captures relationships and dependencies that extend beyond just direct neighbors, learning richer patterns across the entire graph.

4. Evaluation Metrics :

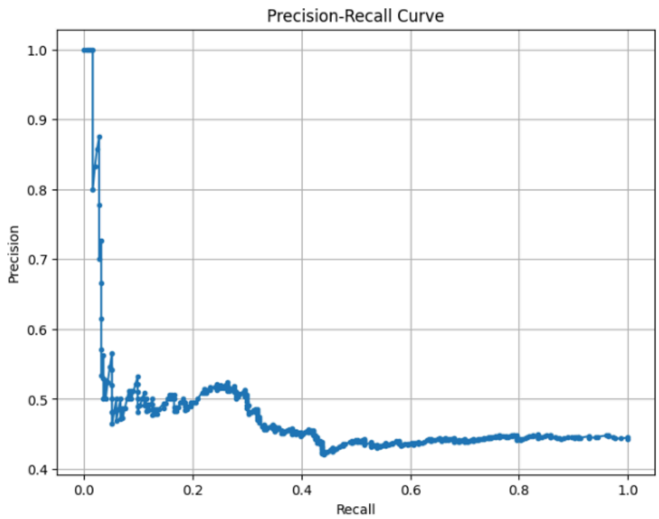
1. Confusion Matrix

- **The Confusion Matrix** shows that all predictions are for Class 0 (319 correct predictions for Class 0, 254 correct predictions for Class 1 classified as Class 0), while no predictions are made for Class 1.
- **Observation:** The model is equally towards predicting Class 0 and predicted Class 1 , indicating a significant class balance in the model's learning.



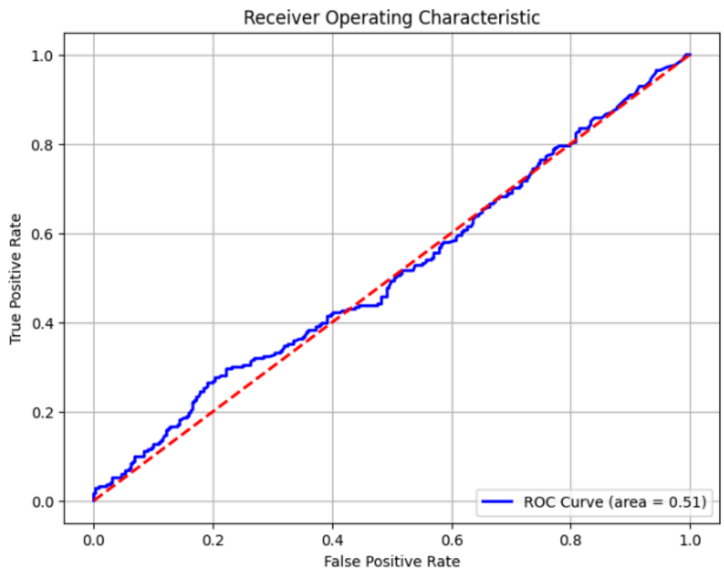
2. Precision-Recall Curve

- **The Precision-Recall Curve** shows fluctuating precision values, with overall low precision as recall improves.
- **Observation:** The curve suggests that the model has a balanced precision and recall, especially at higher recall values. This indicates high performance in managing positive class predictions, which matches the observation in the confusion matrix where no Class 1 predictions are made.



3. ROC Curve

- **The ROC Curve** is close to the diagonal line (AUC = 0.51), meaning the model's performance is far from random guessing.
- **Observation:** With an AUC of 0.51, the model shows discriminative ability between the positive (Class 1) and negative (Class 0) classes. This further reinforces that the model has high predictive power, especially for Class 0.



5. Query Processing and Response Generation :

In this section, we explain the process used to handle user queries and generate responses using the knowledge graph and a combination of Graph Neural Networks (GNN) and Text Embedding techniques. The goal is to serve relevant responses from the knowledge graph based on the user's input query.

1. Embedding Nodes and Queries

- **Node Features:** Each node in the knowledge graph is represented by text attributes (e.g., descriptions). We used TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to create embeddings for these node attributes, capturing the importance of words within the node attribute descriptions.

Example:

- For the node "part project budget", the node's attribute features would be a vectorized representation of a description like:
  > "Budget Allocation Details on how the budget components are allocated..."
- **Query Embedding:** User queries are also embedded using the same TF-IDF vectorizer, which converts the textual query into a vector in the same feature space as the node attributes.

2. Finding the Most Relevant Node

- **GNN Inference:** After embedding the user's query, it is fed into a GNN model that has been trained on the knowledge graph. The model, leveraging its structure and node relationships, captures both local and global information in the graph. This allows it to dynamically understand not just individual node embeddings but also the contextual relevance between nodes.
- **Why a GNN Model?** Unlike standalone similarity measures, a GNN model enables more nuanced reasoning by integrating structural insights and relationships into the embeddings. This approach goes beyond direct cosine similarity, making it possible to identify the most relevant node by considering both its content (attributes) and its connections with surrounding nodes.
- The node that the GNN model predicts as the most relevant is retrieved based on its proximity to the user query in the learned representation space. The attributes of this node are then used to form a well-contextualized response.

3. Response Generation:

- Based on the most relevant node, its attributes (descriptions from the knowledge graph) are used to directly answer the user's query. **For instance:**
  - If the query is about budgeting processes, the node most closely associated with this concept (e.g., "part project budget") is selected along with its detailed attributes (e.g., how budget allocations are made and monitored).
- For a query related to risks, a node discussing risk analysis or risk mitigation would be selected.

```
Question: what is quantitative risk
Relevant Concept: risk analysi attention
Response: comprehensive risk analysis advanced quantitative assessment techniques indepth risk evaluation methods
```

```
Question: definition of risk mangement
Relevant Concept: correspond
Response: Defense strategies proactive and reactive approaches for handling security threats Implementation methodologies agile and waterfall methodolog
ies for effective execution of projects Occurrence tracking realtime monitoring and analysis of incidents for swift response Strategic planning longterm
vision development and goalsetting Process optimization continuous improvement of workflows and operations for efficiency Project management frameworks
Scrum Kanban and PMBOK for structured project execution Relevance metrics key performance indicators and metrics for measuring impact and effectiveness
Task prioritization prioritization techniques such as MoSCoW Eisenhower Matrix and ABC analysis for efficient task management
```

Section 13. Retrieval-Augmented Generation (RAG) for Query Answering

1. Chunking and Splitting Text

- **Reason for Chunking:** When working with large sections of text like chapters or documents, it's inefficient to process the entire text at once due to model token limits and potential loss of relevance within smaller sections. To handle this, the text is split into manageable chunks that are easier for the model to process.
- **Token Limits & Overlap:** Each chunk is structured to not exceed a predefined token limit (e.g., 150 tokens) to ensure compatibility with language models like GPT. Additionally, an overlap-based strategy is employed to include overlapping parts of text between chunks (e.g., 50 tokens overlap). This helps retain context across adjacent chunks, preventing loss of meaning.

```
                Chapter                                           Content  \
0  Chapitre_1_introduction  introduction p roject Management Institute PMI...
1  Chapitre_1_introduction  introduction p roject Management Institute PMI...
2  Chapitre_1_introduction  introduction p roject Management Institute PMI...
3  Chapitre_1_introduction  introduction p roject Management Institute PMI...
4  Chapitre_1_introduction  introduction p roject Management Institute PMI...

                                              Chunks
0
1  introduction p roject Management Institute PMI...
2   \n\n here 's a breakdown of the component men...
3   * * handbook * * : these might be detailed gu...
4   this recognition be crucial as it set the sta...
```

2. Embedding Text Chunks

-

- **Embedding Technique:** Each chunk of text is transformed into a dense numeric vector through OpenAI's **text-embedding-ada-002 model**. This model captures the semantics of the text, allowing it to be represented in a high-dimensional vector space, ideal for comparing how similar two pieces of text are.
- **Why Embeddings?:** Text embeddings encode the meanings of words and sentences into fixed-length numerical representations. These can be directly compared using similarity metrics (such as **cosine similarity**). This is essential for finding relevant answers during query processing because the embeddings efficiently determine how closely a chunk of text matches the user's query.
- **Choice of Model: The ada-002 embedding model** is used because of its balance of computational efficiency and state-of-the-art performance in producing semantic embeddings. It's optimized to create

compact vector representations of text for efficient similarity comparisons.

```
Text chunk: holder become clear change although Project Risk Management process form integral part overall project management plan budget term resource
cost time specifi risk management activity establish order well track control necessary defend correspond expenditure throughout project cost treating r
isk include appropriately project budget risk management plan describe part project budget evaluate allocate manage risk management plan defi monitoring
method ensure corresponding expenditure track appropriately well condition approve budget risk management modifi ed way project management process progr
essive elaboration risk management activity need repeat throughout project risk management plan defi normal frequency repeat process well specifi except
ional condition corresponding action initiate correspond risk management activity integrate project management plan two category success criterion risk
management success project general success Project Risk Management Project Management Institute Practice Standard Project Risk Management Project Relate
d Criteria assess success Project Risk Management stakeholder must agree acceptable level result
Metadata: {'chapter': 'Chapitre_4_plan_risk_management', 'index': 53}
----------------------------------------
Text chunk: MONITOR CONTROL RISKS effectiveness Project Risk Management depend upon way approve plan carry plan execute correctly review update regularl
y carry correctly invest effort reward future project benefi project experience Purpose objective Monitor Control Risks Process primary objective risk m
onitor control track identifi ed risk monitor residual risk identify new risk ensure risk response plan execute appropriate time evaluate effectiveness
throughout project life cycle addition track manage risk response action effectiveness Project Risk Management process review provide improvement manage
ment current project well future one F risk set risk contingent response defi ned corresponding set trigger condition specifi ed responsibility action o
wner ensure condition effectively monitor corresponding action carry defi ne timely manner Plan risk Responses process complete approve unconditional re
sponse action include defi ned current project management plan fi rst action risk monitoring control check whether case take appropriate action necessar
y
Metadata: {'chapter': 'Chapitre_9_monitor_and_control_risks', 'index': 148}
```

3. Query-Based Retrieval

- **Storing Embeddings Using ChromaDB:** Once embedded, the text chunks, alongside their metadata (e.g., chapter names and indices), are stored in **ChromaDB.** This is a high-performance, **persistent embedding database** designed for fast and scalable query-time retrieval. It allows us to retrieve the most similar text chunks in response to user queries by efficiently ranking them based on similarity.
- **Query Processing:** When a user submits a query, it's embedded using the **same embedding model** and vector representation as the text chunks. The embedding for the query is then used to retrieve the most **relevant text chunks** from ChromaDB based on cosine similarity, ensuring accuracy and relevance in the document retrieval process.

4. Generating Contextual Answers

- **Combining Retrieved Information:** The top retrieved chunks from ChromaDB are combined into a cohesive context that covers multiple aspects related to the query. This combined context is crucial, as the model needs more surrounding information to answer complex questions effectively.
- **Using the Llama-3 Model for Answers:** The retrieved context is passed to **Ollama's Llama-3 model**. This large language model, which has access to billions of parameters, is trained to understand and synthesize the retrieved context to generate precise and relevant responses. **The Llama-3** model is particularly suited for **multi-hop reasoning** and answering questions based on more advanced and detailed domain knowledge, such as project risk management.

5. Interactive Query Session

- **Why Interactivity Matters:** The interactive query session allows users to directly engage with the knowledge base in a dynamic manner. Instead of predefined outputs, users can ask queries and receive **real-time AI-generated responses** based on what was stored in the knowledge graph-like **ChromaDB.**
- **Dynamic Retrieval and AI Integration:** Each query triggers a new search within ChromaDB to retrieve the most relevant chunks. The process mirrors a **closed-loop information retrieval system** where both the retrieval and response generation components are constantly linked for better query answering.
- The interactive session runs until the user manually exits, making it a continuous tool for information exploration and question answering.

```
Enter your query:  How does risk management integrate into the overall project scope?

Llama Model's Answer:
According to the context, risk management integrates into the overall project scope by:

1. Being an integral part of the overall project management plan.
2. Being integrated with project management to ensure that potential risks are identified and addressed throughout the project lifecycle.
3. Involving ongoing monitoring and assessment of risk as well as continuous improvement of process and procedure.
4. Being included in the project budget and risk management plan, which describes part of the project budget and evaluates, allocates, and manages risk.
5. Having a normal frequency of repeat throughout the project, with exceptions being made for exceptional conditions that require corresponding actions
to initiate.

In essence, risk management is an integral component of the overall project scope, involving ongoing monitoring, assessment, and improvement to ensure t
he achievement of success criteria and stakeholder agreement on acceptable levels of result.

Context used for the answer:
 likely to take an active role in the risk management process which can lead to a more effective and efficient approach integrate with Project Managemen
t risk management should be integrate with project management to ensure that potential risk be identify and address throughout the project lifecycle thi
s integration enable organization to proactively manage risk reduce the likelihood of surprise or unforeseen consequence by integrate risk management in
to the project planning phase organization can develop mitigation strategy and contingency plan early in the process risk management success the goal of
risk management be to achieve success which mean minimize potential loss and maximize opportunity a successful risk management strategy should involve o
ngoing monitoring and assessment of risk as well as continuous improvement of process and procedure Scale risk Effort to Project the risk effort require
should be scale to the project s size complexity and scope this ensure

holder become clear change although Project Risk Management process form integral part overall project management plan budget term resource cost time sp
ecifi risk management activity establish order well track control necessary defend correspond expenditure throughout project cost treating risk include
appropriately project budget risk management plan describe part project budget evaluate allocate manage risk management plan defi monitoring method ensu
re corresponding expenditure track appropriately well condition approve budget risk management modifi ed way project management process progressive elab
oration risk management activity need repeat throughout project risk management plan defi normal frequency repeat process well specifi exceptional condi
tion corresponding action initiate correspond risk management activity integrate project management plan two category success criterion risk management
success project general success Project Risk Management Project Management Institute Practice Standard Project Risk Management Project Related Criteria
assess success Project Risk Management stakeholder must agree acceptable level result

 emerge or modification to exist risk mitigation plan management Plan this plan outline the overall approach to manage risk within the project it set th
e stage for the risk management process and define the scope role and responsibility of team member risk response Monitoring this section involve monito
r the effectiveness of implement risk response plan and track the status of risk mitigation status review regular review be conduct to assess the curren
t state of risk and ensure that plan response be on track these review help identify area where adjustment may be necessary audit periodic audits be per
form to verify that risk management process be be follow correctly and to identify any gap or opportunity for improvement additional Major Trigger this
section highlight the importance of have trigger in place to alert stakeholder when a significant event occur these trigger can prompt additional risk i
dentification analysis and response planning

Retrieved From (metadata):
{'chapter': 'Chapitre_1_introduction', 'index': 20}
{'chapter': 'Chapitre_4_plan_risk_management', 'index': 53}
{'chapter': 'Chapitre_9_monitor_and_control_risks', 'index': 307}
```

Key Benefits of This RAG Approach:

- **Modular and Efficient:** By splitting and embedding chunks, we leverage token efficiency and ensure that each user query pulls only the most relevant parts of the text.
- **Contextually Rich Responses:** The process of combining relevant chunks as context ensures that the responses generated by **Llama-3 are** grounded in real, retrieved text, making the answers more

trustworthy.

- **Use of Advanced Embedding Techniques:** The embeddings allow not just keyword matching, but **semantic search,** meaning that similar ideas are retrieved even if exact word matches aren't used.

Section 14. Enhancing RAG with Knowledge Graphs for Improved LLM Responses

1. Querying the Knowledge Graph

- **Query Matching**: When a user submits a query, the graph is queried to find nodes that match the keywords in the query. Nodes are ranked by their frequency to prioritize the most relevant ones.
- **Fallback Mechanism**: If no directly relevant nodes are found, the system falls back to the most frequent or central nodes in the knowledge graph, ensuring that information from the graph is always utilized.

2. Combining ChromaDB and Knowledge Graph Context

- The most relevant chunks retrieved from ChromaDB are combined with the contextual knowledge from the graph:
  - **ChromaDB** provides unstructured document chunks, while the knowledge graph offers structured relationships between key concepts.
- This combined context helps the language **model (Llama-3)** form better answers by considering both document-based and graph-based information.

3. Generating Better Responses

- **The Llama-3 model** generates answers based on this hybrid context, which improves the relevance and accuracy of responses. Relations between concepts in the graph help guide the model, resulting in more precise, **knowledge-driven answers**.

```
Enter your query:  How does risk management integrate into the overall project scope?

Llama Model's Answer:
Based on the provided context, risk management integrates into the overall project scope in the following way:

1. During the planning phase, risk management is integrated with project management to identify potential risks early on. This enables organizations to
develop mitigation strategies and contingency plans, thereby reducing the likelihood of surprises or unforeseen consequences.
2. The risk management plan describes part of the project budget, which outlines the overall approach to managing risk within the project. This sets the
stage for the risk management process and defines the scope, role, and responsibility of team members.
3. Risk response monitoring involves tracking the effectiveness of implemented risk responses and the status of risk mitigation plans. Regular reviews a
re conducted to assess the current state of risk and ensure that the plan remains on track.
4. The project scope includes risk management contingencies and potential impact assessments, which ensures that risks are considered when defining the
project deliverables, boundaries, exclusions, and success criteria.

In summary, risk management is an integral part of the overall project scope, as it helps organizations proactively manage risks, develop mitigation str
ategies, and track the effectiveness of implemented responses.

========== How the Knowledge Graph Helped ==========
The knowledge graph provided additional context by identifying 3 relevant nodes.
Here are the nodes and their attributes used to help form the answer:
--- Knowledge Graph Nodes and Relations Included ---
Node: project, Attributes: Strategic Project Goals Clearly articulated objectives that support the overall vision and mission of the public organization
aligning with key performance indicators and longterm sustainability targets. Detailed Project Scope An exhaustive outline specifying all project delive
rables boundaries exclusions and success criteria incorporating risk management contingencies and potential impact assessments. Resource Allocation Stra
tegy A comprehensive plan for the optimal distribution of financial human and technological resources ensuring alignment with organizational priorities
and capacity constraints.

Relates to context refl via "carry"

Relates to response via "The process by which a project initiates and influences a response often in the context of organizational management and operat
ional effectiveness."
```

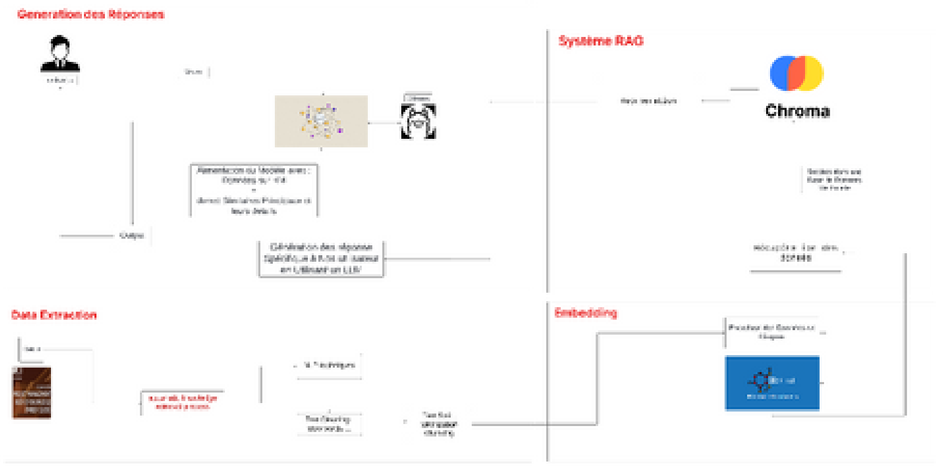
4. Interactive Query System

- Users interact with the system by submitting queries. Each query retrieves relevant chunks from ChromaDB and nodes from the **Knowledge Graph**, giving the model a holistic understanding of both documents and structured knowledge to generate responses.

```
--- Knowledge Graph Context ---

Node: project, Attributes: Strategic Project Goals Clearly articulated objectives that support the overall vision and mission of the public organization
aligning with key performance indicators and longterm sustainability targets. Detailed Project Scope An exhaustive outline specifying all project delive
rables boundaries exclusions and success criteria incorporating risk management contingencies and potential impact assessments. Resource Allocation Stra
tegy A comprehensive plan for the optimal distribution of financial human and technological resources ensuring alignment with organizational priorities
and capacity constraints.

Relates to context refl via "carry"

Relates to response via "The process by which a project initiates and influences a response often in the context of organizational management and operat
ional effectiveness."

========== Retrieved From (ChromaDB metadata): ==========
{'chapter': 'Chapitre_1_introduction', 'index': 20}
{'chapter': 'Chapitre_4_plan_risk_management', 'index': 53}
{'chapter': 'Chapitre_9_monitor_and_control_risks', 'index': 307}
```


Conclusion

The *RAPID* project presents a novel AI-based solution to automate risk management, leveraging PMBOOK and advanced language models. By addressing the limitations of manual risk analysis, the system enhances decision-making processes, providing fast, accurate, and contextually relevant responses. With clear business and data science objectives, a robust methodology, and a focus on performance metrics, *RAPID* is poised to revolutionize how organizations manage risk in complex environments.



workflow

References

Content-Based Recommender Systems

- Lops, P., de Gemmis, M., & Semeraro, G. (2011). "Content-based recommender systems: State of the art and trends." *Recommender Systems Handbook*, 73-105.
- Ricci, F., Rokach, L., & Shapira, B. (2011). "Introduction to recommender systems handbook." *Recommender Systems Handbook*, 1-35.

Knowledge Graphs

- Ehrlinger, L., & Wöß, W. (2016). "Towards a definition of knowledge graphs." *SEMANTiCS (Posters, Demos, SuCCESS)*, 50(1050), 1-4.
- Hogan, A., Blomqvist, E., Cochez, M., et al. (2021). "Knowledge graphs." *ACM Computing Surveys (CSUR)*, 54(4), 1-37.

Natural Language Processing (NLP)

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). "Introduction to Information Retrieval." *Cambridge University Press*.
- Jurafsky, D., & Martin, J. H. (2021). "Speech and Language Processing." *Pearson Education*.

Retrieval-Augmented Generation (RAG)

- Lewis, P., Oguz, B., Rinott, R., et al. (2020). "Retrieval-augmented generation for knowledge-intensive NLP tasks." *Advances in Neural Information Processing Systems*, 33, 9459-9474.

Embedding Techniques

- Reimers, N., & Gurevych, I. (2019). "Sentence-BERT: Sentence embeddings using Siamese BERT-networks." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.

Liste des acronymes

CBRS Content Based Recommender System

LLM Large Language Model

NER Named Entity Recognition

NLP Natural Language Processing

PMP Project Management Professional

PRM Project Risk Management

T&T Tools and Techniques

Les figures