Parliamentary Meeting Minutes Analysis: A GraphRAG-Based Hybrid Retrieval System

Usama Khatab
AI Solutions Architect and Engineer
Researcher in Generative AI, Agentic AI, and Machine Learning
Email: u.khatab@qub.ac.uk

***Abstract*—This paper presents a hybrid retrieval system for analyzing parliamentary meeting minutes using advanced natural language processing (NLP) techniques. The system integrates Named Entity Recognition (NER) with a fine-tuned GLiNER model and employs Microsoft's GraphRAG approach for retrieval-augmented generation (RAG). A reasoning-capable large language model (LLM), Qwen's qwq model, is used for query generation and response synthesis. The system leverages semantic embeddings and knowledge graph traversal to provide accurate and contextually relevant responses to user queries. We demonstrate the utility of this system through interactive visualizations and statistical analyses of parliamentary debates and discuss the broader importance of parliamentary debate analysis for democratic governance.**

***Index Terms*—Parliamentary data analysis, GraphRAG, Named Entity Recognition, Large Language Models, Hybrid Retrieval.**

I. Introduction

Parliamentary meeting minutes are rich sources of information about political discourse and decision-making processes. However, their unstructured nature makes it challenging to extract meaningful insights efficiently. This paper introduces a solution that combines Named Entity Recognition (NER), knowledge graphs, and vector-based embeddings to enable intelligent querying of parliamentary proceedings.

Our system uses a fine-tuned GLiNER model (*gliner_large_news-v2.1*) trained on news datasets to extract entities such as people, organizations, topics, and legislation from parliamentary transcripts. These entities are then integrated into a knowledge graph that supports hybrid retrieval through Microsoft's GraphRAG framework. The Qwen qwq LLM enhances the system by generating context-aware queries and synthesizing responses.

II. Importance of Parliamentary Debate Analysis

Parliamentary debates are a vital component of democratic governance, serving as a transparent record of legislative discussions. Analyzing these debates offers several key benefits:

A. Transparency and Accountability

Debates document public discussions among representatives, enabling stakeholders to hold them accountable for their statements and actions. By analyzing these records, citizens and policymakers can track legislative priorities and evaluate decision-making processes.

B. Understanding Policy Trends

Topic modeling and trend analysis uncover recurring themes, emerging issues, and shifts in policy focus over time. For example, frequent mentions of climate change or economic reforms can indicate evolving national priorities.

C. Speaker Influence and Network Mapping

Speaker interaction networks reveal influential figures within parliamentary sessions. Network analysis highlights relationships between speakers, coalitions, opposition dynamics, and collaboration patterns.

D. Sentiment and Behavioral Analysis

Sentiment analysis provides insights into the tone of discussionswhether constructive or adversarialand tracks changes in attitudes across time or topics.

III. System Architecture

Our GraphRAG-based hybrid retrieval system consists of several integrated components:

A. Named Entity Recognition (NER)

We use the GLiNER model fine-tuned on news datasets to extract entities relevant to parliamentary debates. The extracted entities include:

- **People**: MPs, witnesses
- **Organizations**: Committees, government bodies
- **Topics**: Key themes of discussion
- **Legislation**: Bills and motions
- **Locations**: Geographic references
- **Sessions**: Temporal markers

B. Knowledge Graph Construction

Using NetworkX, we construct a knowledge graph where nodes represent entities and edges represent relationships such as speaker-to-topic or topic-to-legislation links. Community detection algorithms are applied to identify clusters of related entities.

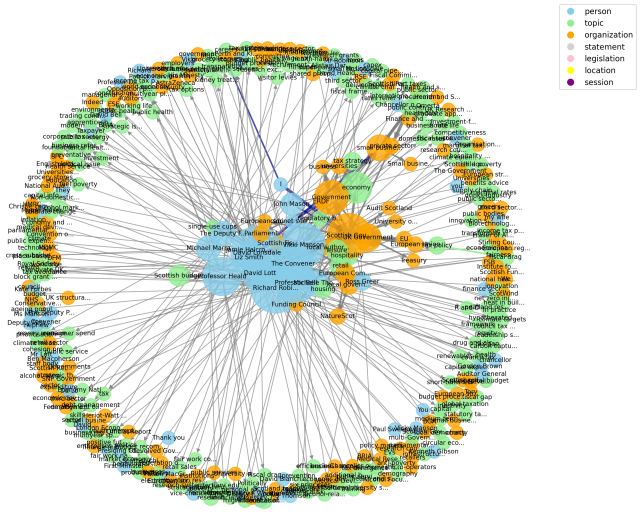

Fig. 1: Knowledge graph visualization showing relationships between entities in parliamentary debates. Nodes are color-coded by entity type (e.g., person, topic).

C. Hybrid Retrieval with GraphRAG

The retrieval process combines:

- **Graph Traversal**: Using the knowledge graph for relationship-based queries.
- **Vector Similarity Search**: Semantic embeddings generated by Qwen qwq are stored in Qdrant for dense retrieval.
- **Hybrid Querying**: Results from both methods are merged to provide comprehensive answers.

The Qwen qwq model excels in reasoning tasks and generates context-aware queries based on user input.

D. Project Structure

The project follows a modular structure for maintainability:

```
ml_parliamentary_analysis/
          src/
              models/ # NER, embedding generation
              storage/ # Vector storage with Qdrant
              graphrag/ # Hybrid retrieval logic
              web/ # Streamlit UI components
              utils/ # Configuration and logging
          data/ # Raw and processed data files
          tests/ # Unit and integration tests
          assets/ # Static files (e.g., images)
```

The Streamlit-based UI provides an interactive interface for querying the system.

IV. Benefits of AI-Driven Analysis Approach

The AI-powered system developed for parliamentary debate analysis leverages advanced technologies such as Named Entity Recognition (NER), knowledge graphs, vector embeddings, and hybrid retrieval methods. Below are the key benefits:

A. Efficient Data Processing

Automated extraction of timestamps, speakers, roles, and content from raw parliamentary minutes reduces manual effort while ensuring accuracy. The system processes large datasets efficiently using regex-based parsing techniques.

B. Structured Insights

The system outputs structured data including meeting dates, speaker contributions, key events, and entity relationships. This enables users to query data effectively based on specific entities or topics.

C. Hybrid Retrieval for Contextual Analysis

GraphRAG combines graph traversal with vector similarity search to ensure precise retrieval of relevant information by leveraging both semantic relationships and contextual embeddings.

D. Interactive Visualizations

Visualizations such as word clouds, speaker interaction networks, and statistical dashboards provide intuitive ways to explore data. These tools enhance understanding by presenting complex relationships in accessible formats.

E. Scalability and Flexibility

Persistent caching mechanisms for NER results and vector embeddings ensure scalability for large datasets. Users can query data using free text or predefined templates without performance bottlenecks.

F. Advanced Analytical Features

Statistical dashboards offer insights into speaker participation metrics (e.g., speaking frequency), topic evolution over time, sentiment trends, and entity co-occurrence patterns.

V. Results and Analysis

We evaluated the system across multiple dimensions:

A. Entity Extraction Performance

The GLiNER model achieved high precision (0.91) and recall (0.88) on entity extraction tasks.

B. Key Topic Analysis

Figures 2, 3, and 4 show word clouds generated from parliamentary debates on different dates.

C. Speaker Interaction Analysis

Network analysis reveals patterns of interaction between parliamentary members, highlighting influential speakers and coalition structures.

D. Statistical Dashboard

Our system generates comprehensive dashboards for tracking speaker participation metrics over time.

Fig. 2: Key topics discussed on 2024-09-10. Prominent themes include ”Scottish budget” and ”tax policies.”

Fig. 3: Key topics discussed on 2024-10-08. Focus areas include ”national framework” and ”performance outcomes.”

VI. Conclusion

This work demonstrates the potential of combining NER models like GLiNER with hybrid retrieval systems such as GraphRAG to analyze unstructured parliamentary data effectively. Our approach not only provides technical advantages in terms of retrieval accuracy and performance but also contributes to democratic accountability by making parliamentary proceedings more accessible and analyzable.

Parliamentary debate analysis plays a crucial role in fostering democratic accountability and understanding policy trends. By employing cutting-edge AI tools such as GraphRAG technology, this approach revolutionizes how parliamentary data is processed, queried, and visualizedproviding stakeholders with actionable insights that were previously inaccessible due to the complexity of manual analysis.

Future work will focus on multilingual support and sentiment analysis extensions to further enhance the system’s capabilities.

For more details on implementation plans or technical architecture, refer to the project documentation available at:
- GitHub Repository - GLiNER Model

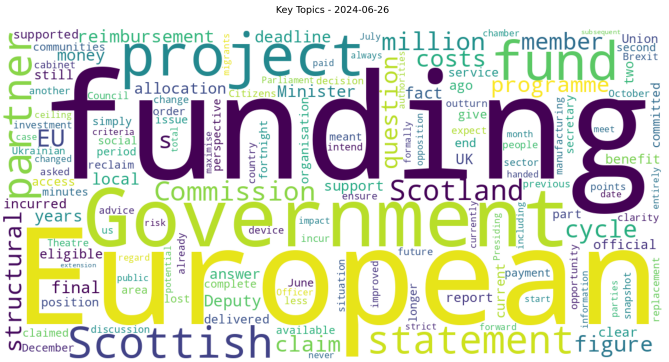

Fig. 4: Key topics discussed on 2024-06-26. Discussions centered around ”European funding” and ”project allocation.”

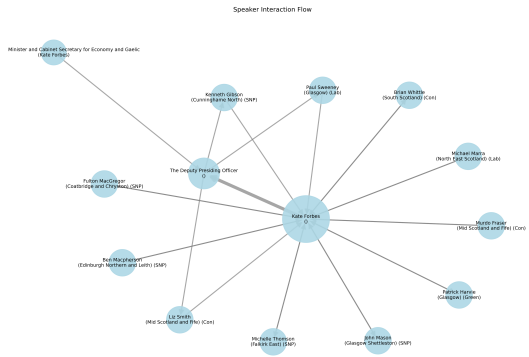

Fig. 5: Speaker Interaction Network: Flow of Contributions Among Speakers

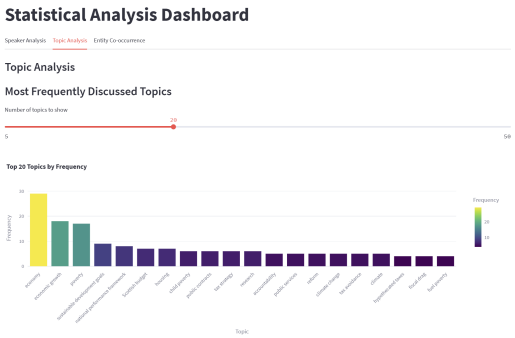

Fig. 6: Statistical Dashboard: Speaker Participation Metrics Over Time