# Querying and QA

##### Session 3.3 (SEMANTiCS)

#### Time: Thursday, September 19, 2024 - 10:30 to 12:00

#### Chair: Kossi Amouzouvi

## **Talks**

### Overlap and Quality Aware Query Processor for Federations of Triple Fragment Interfaces

The increasing numbers of available data sources have led to increased data redundancy and hence novel challenges for federations. Typically, federation engines query all endpoints that provide relevant data for a given query. However, considering the overlap, a subset of these sources might already be sufficient to obtain a complete answer. Further, we deliberately might not wish to include all sources in the evaluation and make a decision based the reliability of a source. We therefore present ORAQL (an Overlap and Reliability Aware Query Processing Layer), an approach that exploits statistics capturing the overlap between sources to choose a subset of the available sources in the federation to compute a complete answer while minimizing redundant answers. Moreover, a user-provided reliability goal is taken into account.

| Tobias Zeimetz | Katja Hose | Ralf Schenkel |
| --- | --- | --- |

### 

### Get LLMs in production with Graph RAG and Entity Linking [SP]

I will start of presenting three varieties of Graph RAG: what they are good for and what types of knowledge graphs they require. Next, I will elaborate the need of complementary AI models, particularly such for entity linking (EL) - the task to associate named entity references in text with concrete identifiers. While comprehensive EL is needed to improve the performance of NLQ, RAG, and information extraction, LLMs are not a good fit for it - they are slow, inaccurate and expensive. Many implement EL with vector databases, but accuracy is far behind purpose made AI models. I will wrap up with an overview of our inventory of EL models that offer state-of-the-art accuracy across different domains.

| Atanas Kiryakov Ontotext |
| --- |

### Cortex: An Experimentation for a e-Health Data Hub

The Cortex project initiated by the Service de Santé des Armées (French Health Services for military forces) aims to bring together all the data available within the different services and spread across many information systems. The solution proposed was the construction of a knowledge graph based on the ontological modeling of the medical domain to create a shared vocabulary across the different services with integration of international standard for the medical terminology. The use of a semantic interface between the different data systems and the user querying the data provides a unique end-point for all information system in a transparent manner. The system also integrates an access management system which differentiate the kind of information accessible depending on the profile of the user. In future works, the project will extend to integrate data from the hospital information system to offer more complete epidemiological analysis.

| Pauline Armary | Brice Sommacal |
| --- | --- |

### Generating SPARQL from Natural Language Using Chain-of-Thoughts Prompting

We propose CoT-Sparql, our approach to generate Sparql queries from input questions. Our approach employs Chain-of-thoughts prompting that guides large language models through intermediate reasoning steps and facilitates generating precise Sparql queries. Furthermore, our approach incorporates entities and relations from the input question, and one-shot example in the prompt to provide additional context during the query generation process.

| Hamada Zahera | Manzoor Ali | Mohamed Ahmed Sherif |
| --- | --- | --- |
| Diego Moussallem | Axel-Cyrille Ngonga Ngomo |  |