# Session 3.3

##### Session 3.3 (SEMANTiCS)

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

#### Chair: TBA

## **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. Hence, we propose an approach based on a majority vote over multiple sources to increase the reliability of the query result. For this work, we focus on TPF interfaces, since they are the least expressive interfaces and hence our approach can easily be adopted for more expressive interfaces, e.g. SPARQL endpoints. The presented methods to capture the overlap between sources of a federation have shown to generate useful overlap profiles with a maximum deviation of less than five percent. Even if the identification of redundant data is NP-hard we presented an approximation with a significant reduction in requested endpoints. Further, we have shown that ORAQL is granularly tunable towards reliability and can beat a state-of-the-art baseline system in terms of coverage and reliability.

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

### 

### Ontotext [SP]

| TBAOntotext |
| --- |

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

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

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

Purpose:

Sparql is a highly expressive query language for knowledge graphs; yet, formulating precise Sparql queries can be challenging for users non-expert users. A potential solution is translating natural questions into Sparql queries, known as Sparql generation. This paper addresses the challenges of translating natural language questions into Sparql queries for different knowledge graphs.

Methodology:

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.

Findings:

We conducted several experiments on benchmark datasets and showed that our approach outperforms the state-of-the-art methods by a large margin. Our approach achieves a significant improvement in F1 score of 4.4% and 3.0% for the QALD-10 and QALD-9 datasets, respectively.

Value:

Our CoT-Sparql approach contributes to the semantic web community by simplifying access to knowledge graphs for non-expert users. In particular, CoT-Sparql enable non-expert end-users to query knowledge graphs in natural languages, where CoT-Sparql converts user natural languages queries into Sparql queries, which can be executed via the knowledge graph’s Sparql endpoint.

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