# LLMs and NLP

##### Session 1.3 (SEMANTiCS)

#### Time: Wednesday, September 18, 2024 - 10:40 to 12:00

#### Chair: Angelo Salatino

## **Talks**

### Zero-shot Topic Classification of Column Headers: Leveraging LLMs for Metadata Enrichment

This work proposes a novel approach that leverages LLMs for topic classification of column headers using a controlled vocabulary, presenting a practical application of LLMs and RAG systems within the Semantic Web domain. This approach has the potential to facilitate automated metadata enrichment, thereby enhancing dataset retrieval and the Findability, Accessibility, Interoperability, and Reusability (FAIR) of research data on the Web.

|  |  |
| --- | --- |
| Margherita Martorana | Tobias Kuhn |
| Lise Stork | Jacco van Ossenbruggen |

### Multilingual linguistic word sense disambiguation for semantic annotations

Knowledge models, constructed from vocabularies and ontologies, establish a formal basis to enable semantic annotations, which can then support use cases like semantic search and recommendations. In such a scenario, we face the challenges of word sense disambiguation (WSD), multiword expressions (MWE), and multilinguality (of models and content). We will present the research project carried out jointly by Semantic Web Company and Lexicala by K Dictionaries, including our approach and methodologies along with the preliminary results of our experiments on converging language resources, knowledge graphs, and large language models.

|  |  |  |
| --- | --- | --- |
| Robert David | Ilan Kernerman Lexicala | Assaf Siani |

### Improving term networks through the detection of semantic perspectives

We present an industry use case in which we discuss the practical use of a formal list of definitions. The list of definitions will be used as a basis for a knowledge graph that will serve knowledge panels when interacting with information in the company and that will help improve question answering using large language models. We evaluate whether the list is complete, whether the terms are relevant and whether there might be ambiguity in these terms. For this purpose we have analyzed a large body of company documents that represent how employees use these terms in practice, as well as a body of formal documents that represent how these terms are used in the industry in general. This has lead to the finding that some of the terms may not be ambiguous, but can be interpreted differently. There is a constant balance in a term list between completeness, level of abstraction and relevance. Determining which terms may lead to confusion because of multiple interpretations is a relevant step forward in creating usable knowledge graphs.

|  |  |
| --- | --- |
| Maya Sappelli | Hans Fugers |
| Bart Kleijngeld | Marijn Siebel |

### Enhancing Answers Verbalization using Large Language Models

This study investigates the verbalization of answers generated by knowledge graph question answering (KGQA) systems using large language models. In user-centric applications, such as dialogue systems and voice assistants, answer verbalization is an essential step to enhance the quality of interactions. We experimented with multiple large language models to verbalize answers from knowledge-based question-answering systems. In particular, we fine-tuned the LLM models based on different inputs, including SPARQL queries and triples to determine which model performs best for answer verbalization. We found that fine-tuning language models and introducing additional knowledge such as SPARQL queries, achieve state-of-the-art results in verbalizing answers from KGQA systems.

|  |  |
| --- | --- |
| Daniel Vollmers | Parth Sharma |
| Hamada Zahera | Axel-Cyrille Ngonga Ngomo |