3rd International Workshop on Natural Language Processing for Knowledge Graph Creation (NLP4KGC 2024)

Edlira Vakaj[1], Rizou Stamatia[2], Nandana Mihindukulasooriya[3], Sanju Tiwari[4], Fernando Ortiz-Rodríguez[4] and Ryan Mcgranaghan[5]

[1]*Birmingham City University, Birmingham, UK*

[2]*Singular Logic, Athens, Greece*

[3]*IBM Research, New York, USA*

[4]*Universidad Autónoma de Tamaulipas, Tamaulipas, Mexico*

[5]*NASA Jet Propulsion Laboratory, California, United States*

Abstract

Knowledge Graphs (KG) are emerging and becoming increasingly popular. Building a domain knowledge graph from a large amount of text is a challenging task which requires a tremendous amount of work, including entity recognition, entity disambiguation, and relationship extraction. In this context, the 3rd NLP4KGC workshop brings together academics, industry experts, and enthusiasts from the natural language processing (NLP) and knowledge graph (KG) creation fields in a common forum. The objective is to foster collaboration and knowledge sharing among researchers, developers, and practitioners working at the cutting edge of NLP and KG creation. After two successful previous editions, the 3rd NLP4KGc is co-located with the 20th International Conference on Semantic Systems (SEMANTiCS 2024). After a rigorous review process, 10 papers were accepted for oral presentations at the event.

Workshop Website: https://sites.google.com/view/3rdnlp4kgc/

Previous Proceedings: https://ceur-ws.org/Vol-3510/, https://doi.org/10.1145/3543873.3589746

Keywords

natural language processing, knowledge graphs, semantics, large language models, graph neural networks, responsible AI

1. Introduction

Drawing from the success of the 2ndNLP4KGc workshop at the SEMANTICs 2023 conference we propose the continuation in the 3rdNLP4KGc workshop. This year the main focus of the proposal aims to explore the intersection of Knowledge Graphs (KGs), Natural Language Processing (NLP), and Responsible AI, emphasizing how KGs enhance the trustworthiness of large language models. Building a domain knowledge graph from a large amount of text requires a tremendous amount of work, including entity recognition, entity disambiguation and relationship extraction. Because of this, more work has been done on automated ways to generate knowledge graphs from text. Recent efforts in NLP (Natural Language Processing) development have shown that semantic deep neural networks can learn the complex syntactic and semantics of the natural language and thus, give more potential for automation even in the most complex domains i.e., legal documents. New approaches to KG development use a combination of extraction methods and state-of-the-art NLP techniques. We welcome applications that analyze and synthesize large volumes of scientific information (especially scientific literature, notes documents, and dataset metadata) and can use that information to create ontologies and semantic representations that better organize them. We also welcome contributions that address automation components of maintenance of ontologies and knowledge graphs.

This workshop invites contributions on methods and approaches of knowledge and data extraction from text, as well as theoretical and practical aspects of using semantic deep NLP for KG creation and the use of such KG for Graph Neural Network (GNN) tasks.

2. Keynote

Speaker: Professor Amit Sheth, University of South Carolina, USA
Title: "The EMPWR Platform: Data and Knowledge-Driven Processes for the Knowledge Graph Lifecycle"
Abstract: The unparalleled volume of generated data has heightened the need for approaches to manage and translate them into decision-making and actionable insights. While contemporary data-driven and GenAI systems are popular for handling large volumes of diverse data, they are stymied due to the lack of grounding or the use of a world model reflecting domain knowledge and facts. In this keynote, using the example of our EMPWR platform for comprehensive KG development and lifecycle support, I will describe a hybrid approach involving symbolic and GenAI systems to create, combine, align, and maintain large KGs. I will conclude by reviewing the current state of AI/NLP and laying out how KG plays a central role in supporting AGI: Alignment, Grounding, and Instructability for a more robust AI.

3. Accepted Papers

- **On the pertinence of LLMs for ontology learning**
- **Accessing the Capabilities of KGs and LLMs in Mapping Indicators within Sustainability Reporting Standards**
- **Towards the automation of Knowledge Graph construction using Large Language Models**
- **Assessing SPARQL capabilities of Large Language Models**
- **Converting Fire Safety Regulations to SHACL Shapes Using Natural Language Processing**
- **Ontology Learning from Text: an Analysis on LLM Performance**
- **Ontology-based Dataset Discovery in the BUILDSPACE Data Management Platform**
- **Pruning Cycles in UMLS Metathesaurus: A Neuro Symbolic AI Approach**
- **Breaking Down Financial News Impact: A Novel AI Approach with Geometric Hypergraphs**
- **Automated Generation of Competency Questions Using Large Language Models and Knowledge Graphs**

4. Organizers

Dr Edlira Vakaj   is leading the Natural Language Processing AI Lab and is an Associate Professor of Neuro-Symbolic AI at Birmingham City University, UK.

Rizou Stamartina   works as a researcher/project manager at the RD Department of Singular Logic. She has been project coordinator, technical coordinator and/or exploitation manager of several EU projects (BUILDSPACE, SRI-ENACT, 5G-MEDIA, Water4Cities, Broker@Cloud, I-DONT-FALL).

Nandana Mihindukulasooriya   Dr Nandana Mihindukulasooriya is a Senior Research Scientist at IBM Research, New York, USA. Prior to that he was a postdoctoral fellow at MIT-IBM Watson AI Lab.

Sanju Tiwari   Dr. Sanju Tiwari (CEO and Founder of ShodhGuru Research Labs, India) and a Senior Researcher at TIB Hannover Germany. She is former recipient of DAAD Post-Doc-Net AI Fellow of Germany for 2021 and visited different German Universities.

Fernando Ortiz-Rodríguez   Fernando Ortiz-Rodriguez is a Full Professor, the Manufacturing Research and Education Center Director, and the Head of the Artificial Intelligence and Innovation Lab at Tamaulipas Autonomous University.

Ryan Mcgranaghan   Ryan McGranaghan is a senior Data Scientist at the NASA Jet Propulsion Laboratory, where he works with the Machine Learning and Instrument Autonomy (MLIA) group to apply data science techniques robustly and responsibly to the Earth and Space Sciences.

5. Scientific Advisor

Professor Mohamed Gaber, Birmingham City University, United Kingdom.

6. Programme Committee Members

- **Sima Iranmanesh, Birmingham City University, United Kingdom - Web and Communication**
- Amna Dridi, Birmingham City University, United Kingdom
- Hansi Hettiarachchi, Lancaster University, United Kingdom
- Thomas Beach, Cardiff University
- Hadeel Sadanay, Birmingham City University
- Vladimir Alexiev, Ontotext, Bulgaria
- Jagdev Bhogal, Birmingham City University, United Kingdom
- Ellie Young, Common Action, USA
- Hossein Ghomeshi, Senior ML Researcher at Shipamax, United Kingdom
- Ryan Mcgranaghan, NASA Jet Propulsion Laboratory, USA
- Katarina Karii, Lead Ontologist Inter IKEA Systems B.V., Finland
- He Tan, Jönköping University, Sweden
- Diellza Elshani, University of Stuttgart, Germany
- Ronakkumar Panchal, Gujarat Technology University, India

7. Acknowledgement