Proceedings of the 3rd International Workshop on Semantic Technologies and Deep Learning Models for Scientific, Technical and Legal Data co-located with the Extended Semantic Web Conference 2025

Rima Dessi[1], Joy Jeenu[2], Danilo Dessi[3], Francesco Osborne[4], and Hidir Aras[2]

[1] Higher Colleges of Technology, Sharjah, UAE
[2] FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Germany
[3] College of Computing and Informatics, University of Sharjah, Sharjah, UAE
[4] Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom
rdessi@hct.ac.ae, joy.jeenu@fiz-karlsruhe.de, ddessii@sharjah.ac.ae,
francesco.osborne@open.ac.uk, hidir.aras@fiz-karlsruhe.de

1 Organizing Committee

- Rima Dessi, Higher Colleges of Technology, Sharjah, UAE.
- Joy Jeenu, FIZ Karlsruhe, Germany.
- Danilo Dessi, College of Computing and Informatics, University of Sharjah, Sharjah, UAE.
- Francesco Osborne, The Open University, Milton Keynes, United Kingdom.
- Hidir Aras, FIZ Karlsruhe, Germany.

2 Program Committee

- Rubén Alonso, R2M Solution Srl, Italy.
- Ahmad Alrifai, FIZ Karlsruhe, KIT-AIFB, Germany.
- Pablo Calleja, UPM Ontology Engineering Group, Spain.
- Mathieu D'Aquin, LORIA, University of Lorraine, France.
- Susmita Gangopadhyay, GESIS Leibniz Institute for the Social Sciences, Germany.
- Rene Hackl-Sommer, DeepL SE, Germany.
- Inma Hernandez, University of Seville, Spain.
- Fabian Hoppe, Vrije Universiteit Amsterdam, The Netherlands.
- M. Taimoor Khan, GESIS Leibniz Institute for the Social Sciences, Germany.
- Mirko Marras, University of Cagliari, Italy.
- Giacomo Medda, University of Cagliari, Italy.
- Angelo Antonio Salatino, The Open University, United Kingdom.
- Lise Stork, Vrije Universiteit Amsterdam, The Netherlands.
- Sabine Wehnert, Otto-von-Guericke-Universität Magdeburg, Germany.
- Lei Zhang, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany.

3 Preface

The rapid expansion of publicly available scientific, technical, and legal documents, such as patents, reports, and research articles, has made large-scale processing and analysis increasingly vital. Every day, researchers, patent analysts, inventors, and other professionals (e.g., information scientists, legal experts) contribute to this growing body of data through publications, technical reports, and patent filings. However, analyzing and navigating these documents presents significant challenges

due to their considerable length, the use of specialized vocabulary, and their coverage of diverse scientific domains. These documents are often semi-structured, combining unstructured text with structured elements such as tables, mathematical expressions, diagrams, and domain-specific content like chemical compounds or biological sequences. This combination of formats and domain complexity significantly increases the difficulty of processing them effectively.

To benefit from the scientific and technical knowledge present in such documents, such as for decision-making or professional search and analytics, there is an urgent need to analyze, enrich, and link this data by employing state-of-the-art Semantic Web technologies and AI methods. However, as they are heterogeneous and are written using domain-specific terminology, applying the existing semantic technologies is not straightforward. To address the challenges mentioned above, Semantic Web Technologies, Natural Language Processing (NLP) techniques, and Deep Neural Networks (DNN) must be leveraged to provide efficient and effective solutions for creating easily accessible and machine-understandable knowledge of science and industry.

To this end, the goal of the SemTech4STLD workshop[5] was to provide a meeting forum for academics and industry professionals to come together and discuss topics such as the application of Semantic Web Technologies and Deep Learning Models to scientific, technical, and legal data. Further, the primary objective of the workshop was to promote collaboration among the participants and exchange ideas. The workshop started with a keynote entitled "Evaluation Challenges in Using Generative AI for Science and Technical Contents" by Prof. Dr. Paul Groth.

Overall, the workshop's success can be demonstrated by the high number of participants. Further, during the workshop, many participants joined the discussions, asked questions, and exchanged ideas about the application of Semantic Web Technologies and Machine Learning models on Scientific, Technical, and Legal Data. We believe this workshop has helped participants build a new network and has encouraged future projects. We definitely plan to organize the 4th edition of this workshop.

June 2025   Rima Dessì, Joy Jeenu, Danilo Dessì, Francesco Osborne, and Hidir Aras

Contents

Keynote Talk *Prof. Dr. Paul Groth*

Keynote on **Evaluation Challenges in Using Generative AI for Science & Technical Content** .

Keynote Abstract:
Foundation Models show impressive results in a wide range of tasks on scientific and legal content from information extraction to question answering and even literature synthesis. However, standard evaluation approaches (e.g., comparing to ground truth) often do not seem to work. Qualitatively, the results look great, but quantitative scores do not align with these observations. In this talk, I discuss the challenges we have faced in our lab in evaluation. I then outline potential routes forward.

Papers presented at SemTech4STLD

Evaluating LLMs for Named Entity Recognition in Scientific Domain with Fine-Tuning and Few-Shot Learning. *Davide Buscaldi, Danilo Dessì, Francesco Osborne, Davide Piras and Diego Reforgiato Recupero*

---

[5] https://semtech4stld.github.io/

Taming Hallucinations: A Semantic Matching Evaluation Framework for LLM-Generated Ontologies. *Nadeen Fathallah, Steffen Staab and Alsayed Algergawy*

Benchmarking Large Language Models for Sustainable Development Goals Classification: Evaluating In-Context Learning and Fine-Tuning Strategies. *Andrea Cadeddu, Alessandro Chessa, Vincenzo De Leo, Gianni Fenu, Enrico Motta, Francesco Osborne, Diego Reforgiato Recupero, Angelo Salatino and Luca Secchi*

Biomedical Entity Linking with Triple-aware Pre-Training. *Xi Yan, Cedric Möller and Ricardo Usbeck*

Enabling Natural Language Access to BIM Models with AI and Knowledge Graphs. *Andrea Ibba, Rubén Alonso and Diego Reforgiato Recupero*

Context-Aware Explanations: Leveraging Knowledge Graphs for Adaptive Explainability in Dynamic Environments. *Erick Mendez Guzman, Rima Dessí, Alya Alshaami, Amna Alowais, Hamda Alhammadi, Nada Alzarooni, Weam N.A Jarbou and Zarak Khan*

Leveraging Knowledge Graphs and Generative AI for Augmented Research Paper Retrieval. *Rima Dessí, Erick Mendez Guzman, Alya Alshaami, Amna Alowais, Hamda Alhammadi, Nada Alzarooni, Weam N.A Jarbou and Zarak Khan*