

Proceedings of the 1st International Workshop on Semantic Technologies for Scientific, Technical and Legal Data co-located with Extended Semantic Web Conference 2023

Rima Dessi¹, Mehwish Alam², Francesco Osborne³ and Hidir Aras¹

¹FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Germany

²Télécom Paris, Institut Polytechnique de Paris, France

³The Open University, Milton Keynes, United Kingdom

1. Organizing Committee

- Rima Türker Dessi, FIZ Karlsruhe, Germany.
- Mehwish Alam, Télécom Paris, Institut Polytechnique de Paris, France.
- Francesco Osborne, The Open University, Milton Keynes, United Kingdom.
- Hidir Aras, FIZ Karlsruhe, Germany.

2. Program Committee

- Mehwish Alam, Télécom Paris, Institut Polytechnique de Paris, France.
- Ahmad Alrifai, FIZ Karlsruhe, Germany.
- Hidir Aras, FIZ Karlsruhe, Germany.
- Davide Buscaldi, Sorbonne Paris North University, France.
- Pablo Calleja, Polytechnic University of Madrid, Spain.
- Danilo Dessi, GESIS, Germany.
- Paul Groth, the Netherlands.
- Rene Hackl-Sommer, DeepL, Germany.
- Nils Holzenberger, Télécom Paris, Institut Polytechnique de Paris, France.
- Francesco Osborne, The Open University, Milton Keynes, United Kingdom.
- Tassilo Pellegrini, University of Applied Sciences St. Pölten, Austria.
- Angelo Salatino, KMi, Open University, UK.
- Martin Theobald, University of Luxembourg, Luxembourg.
- Rima Türker Dessi, FIZ Karlsruhe, Germany.
- Ian Wetherbee, Google, United States.
- Lei Zhang, FIZ Karlsruhe, Germany.

SemTech4STLD'23: Semantic Technologies for Scientific, Technical and Legal Data, May 28, 2023, Hersonissos, Greece

✉ rima.dessi@fiz-karlsruhe.de (R. Dessi); mehwish.alam@telecom-paris.fr (M. Alam);

francesco.osborne@open.ac.uk (F. Osborne)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

3. Preface

The rapid growth of online available scientific, technical, and legal data such as patents, reports, articles, etc. has made the large-scale analysis and processing of such documents a crucial task. Today, scientists, patent experts, inventors, and other information professionals (e.g., information scientists, lawyers, etc.) contribute to this data every day by publishing articles, writing technical reports, or patent applications. It is a challenging task to process, analyze, and explore these documents due to their length, the use of domain-specific vocabulary, and the complexity introduced by targeting various scientific fields and domains. These semi-structured types of documents cover unstructured textual parts and structured parts such as tables, mathematical formulas, diagrams, and domain-specific information such as chemical names, bio-sequences, etc. Such kind of information brings complexity in processing such documents.

In order to benefit from the scientific-technical knowledge present in such documents, e.g., for decision-making or for professional search and analytics, there is an urgent need for analyzing, enriching, and linking such 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 in order 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 organized workshop¹ was to provide a meeting forum for people from academia as well as industry to come together and discuss topics such as the application of Semantic Web Technologies 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 "Making Legal Knowledge Accessible to Machines: Challenges and Opportunities" by Sabrina Kirrane, Assistant Professor at the University of Vienna, Austria. An invited talk was also given on "A Coordinated Ecosystem of Systems and Models for Maintenance and Publication of Data, Metadata, and Legal Documents in the European Commission" by Armando Stellato. These talks led to very useful discussions within the community.

Overall, the workshop's success can be demonstrated by the high number of participants and the number of submissions. 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 helped participants build a new network and encourage future projects related to the mentioned topics. We definitely plan to organize the 2nd edition of this workshop.

June 2023

Rima Dessi, Mehwish Alam, Francesco Osborne, Hidir Aras

¹<https://rima-turker.github.io/SemTech4STLD/>