

Preface

The fourth edition of the workshop series *Patent Text Mining and Semantic Technologies* (PatentSemTech'23) was held as a full-day event in conjunction with the SIGIR 2023 conference. The workshop focused on research and new developments from relevant fields such as Natural Language Processing, Text and Data Mining and Semantic Technologies applied to Patent Retrieval and Patent Analytics. One important focus of the workshop was to address the adaptation of existing NLP, MP/DL tools for search and analytics due to the complexity of patent documents being a lengthy, heterogeneous type of scientific text covering diverse scientific subject areas, such as chemistry, pharmacology, etc. Thus, patent data is more difficult to analyse compared to corpora comprising general language texts. Working with patent data, besides its challenging aspects, does bring a richness of facets to be exploited with text-mining and semantic analysis methods as well: (1) It constitutes a huge corpus of scientific-technical documents for a variety of technological domains. (2) They are rich in available meta-data such as spatial data, bibliographic data, classifications, temporal data, etc. (3) Patents describe essential scientific-technical knowledge enclosing solutions for real-world applications. (4) They are complementary knowledge to scientific literature, e.g. chemical and physical properties, bio-science knowledge for drug-target-interaction, which appears first in patents, mostly not published elsewhere.

With the PatentSemTech2023 workshop we continued our series of workshops launched in 2019, aiming to establish a long-term collaboration and a two-way communication channel between the IP industry and academia from relevant fields. Therefore, the 4th PatentSemTech workshop was organized as a full-day event with research paper presentations (2 long and 5 short) that were accepted after peer-reviewing, a hands-on summarization session and an open panel discussion around the topics “LLMs and Patent data” as well as “Knowledge Graphs for Patent Data”.

Germany, Austria, USA, July 2023

Ralf Krestel,
Hidir Aras,
Linda Andersson,
Florina Piroi,
Allan Hanbury,
Dean Alderucci

Copyright © 2023 for this text by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Organizers

- Ralf Krestel (Hasso Plattner Institute, University of Potsdam, German)
- Hidir Aras (FIZ Karlsruhe, Germany)
- Linda Andersson (Artificial Researcher IT GmbH, Vienna, Austria)
- Florina Piroi (Data Science Studio, Vienna, Austria)
- Allan Hanbury (TU Wien, Austria)
- Dean Alderucci (Carnegie Mellon University, Pittsburgh, USA)

Program Committee

- Christoph Hewel (Paustian & Partner, Germany)
- Rene Hackl-Sommer (DeepL SE, Germany)
- Anthony Trippe (Patinformatics, Ireland)
- Paul Groth (University of Amsterdam, Netherlands)
- Hans-Peter Zorn (inovex GmbH, Karlsruhe, Germany)
- Michael Natterer (Dennemeyer Octimine GmbH, Germany)
- Karin Verspoor (RMIT University, Melbourne, Australia)
- Ian Wetherbee (Google Inc., Sunnyvale, United States)
- Michail Salampasis (International Hellenic University, Thessaloniki, Greece)
- Simone Ponzetto (University of Mannheim, Germany)
- Dieter Franz Kogler (University College Dublin, Ireland)
- Alexander Klenner-Bajaja (European Patent Office, Netherlands)
- Tobias Fink (TU Wien, Austria)
- Ron Daniel (AI4Science.com, United States)

Website

Further information on the topics, schedule, and further developments of the PatentSemTech workshop can be found at the website:

<http://ifs.tuwien.ac.at/patentsemtech/>