5th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech)

collocated with the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval

Ralf Krestel[1], Hidir Aras[2], Linda Andersson[3], Florina Piroi[4], Allan Hanbury[5] and Dean Alderucci[6]

[1]*ZBW – Leibniz Information Centre for Economics & Kiel University, Düsternbrooker Weg 120, 24105, Kiel, Germany*
[2]*FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany*
[3]*Artificial Researcher IT GmbH, Taubstummengasse 11 (i2c), 1040, Wien Austria*
[4]*RSA FG Studio Data Science, Thurngasse 8/16, 1090, Vienna Austria*
[5]*Institute of Information Systems Engineering, TU Wien, Favoritenstr. 9-11/194-04, Vienna, Austria*
[6]*Center for AI and Patent Analysis, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
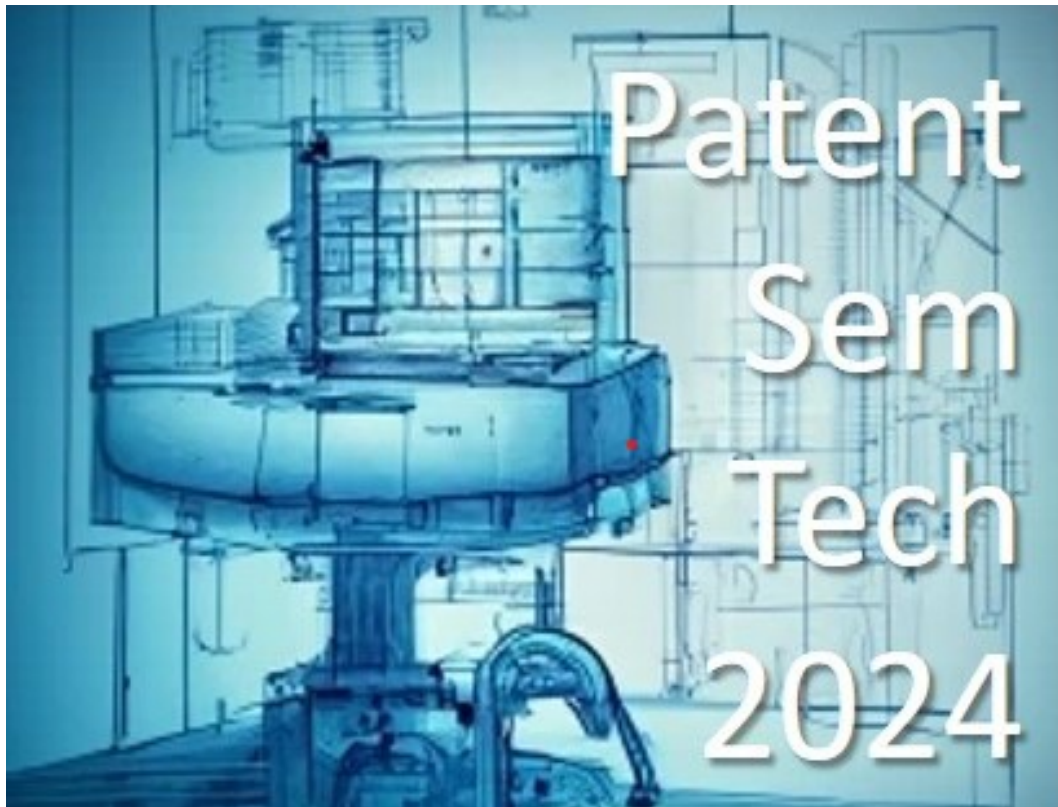
5th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech) 2024

✉ rkr@informatik.uni-kiel.de (R. Krestel)
0000-0002-5036-8589 (R. Krestel)

Preface

The fifth edition (PatentSemTech2024) of the workshop series *Patent Text Mining and Semantic Technologies* was held as a full-day event in conjunction with the SIGIR 2024 conference. As in the previous editions, the workshop focused on new developments and research in patent retrieval and patent analytics. An important focus of the workshop was to address the adaptation of existing deep learning models, e.g. large language models, for the patent domain, covering diverse scientific subject areas, such as chemistry, pharmacology, etc. In general, patent data is more difficult to analyse compared to corpora comprising other text genres. 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 PatentSemTech2024 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 5th PatentSemTech workshop was organized as a full-day event with 10 research paper presentations that were accepted after peer-review out of 17 submissions. 6 long papers were presented as oral presentations while 4 short papers were presented as posters. In addition, Matthew Wahlrab, CEO of RapidAlpha, gave a keynote speech on "Unlocking Strategic Growth: The Role of AI Technology in Intellectual Property". In an open discussion on "How to transform research insights into products?", the workshop participants exchanged ideas and reported their experience with applying AI in the patent domain. The workshop closed with Linda Andersson looking back at 5 successful PatentSemTech workshops and how the field has developed over these years.

Germany, Austria, USA, July 2024

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

Organizers

- Ralf Krestel (ZBW - Leibniz Information Centre for Economics & Kiel University, Germany)
- Hidir Aras (FIZ Karlsruhe, Germany)
- Linda Andersson (Artificial Researcher IT GmbH, Vienna, Austria)
- Florina Piroi (TU Wien & Data Science Studio, Vienna, Austria)
- Allan Hanbury (TU Wien, Austria)
- Dean Alderucci (Carnegie Mellon University, Pittsburgh, USA)

Program Committee

- Alexander Klenner-Bajaja (European Patent Office, Netherlands)
- Anthony Trippe (Patinformatics, Ireland)
- Christoph Hewel (Paustian & Partner, Germany)
- Eric Müller-Budack (TIB Hannover, Germany)
- Florina Piroi (TU Wien, Austria)
- Hans-Peter Zorn (inovex Gmbh, Karlsruhe, Germany)
- Hidir Aras (FIZ Karlsruhe, Germany)
- Karin Verspoor (RMIT University, Melbourne, Australia)
- Lei Zhang (FIZ Karlsruhe, Germany)
- Linda Andersson (Artificial Researcher IT GmbH, Vienna, Austria)
- Michail Salampasis (International Hellenic University, Thessaloniki, Greece)
- Ralf Krestel (ZBW Kiel, Germany)
- Rene Hackl-Sommer (DeepL SE, Germany)
- Simone Ponzetto (University of Mannheim, Germany)
- Tobias Fink (TU Wien, Austria)

Website

Further information on the topics, schedule, and further developments of the PatentSemTech workshop can be found on the website: http://ifs.tuwien.ac.at/patentsemtech/