

6th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech) 2025

collocated with the 48th International ACM SIGIR Conference on Research and Development
in Information Retrieval

Ralf Krestel¹, Hidir Aras², Linda Andersson³, Florina Piroi⁴, Allan Hanbury⁴ and
Dean Alderucci⁵

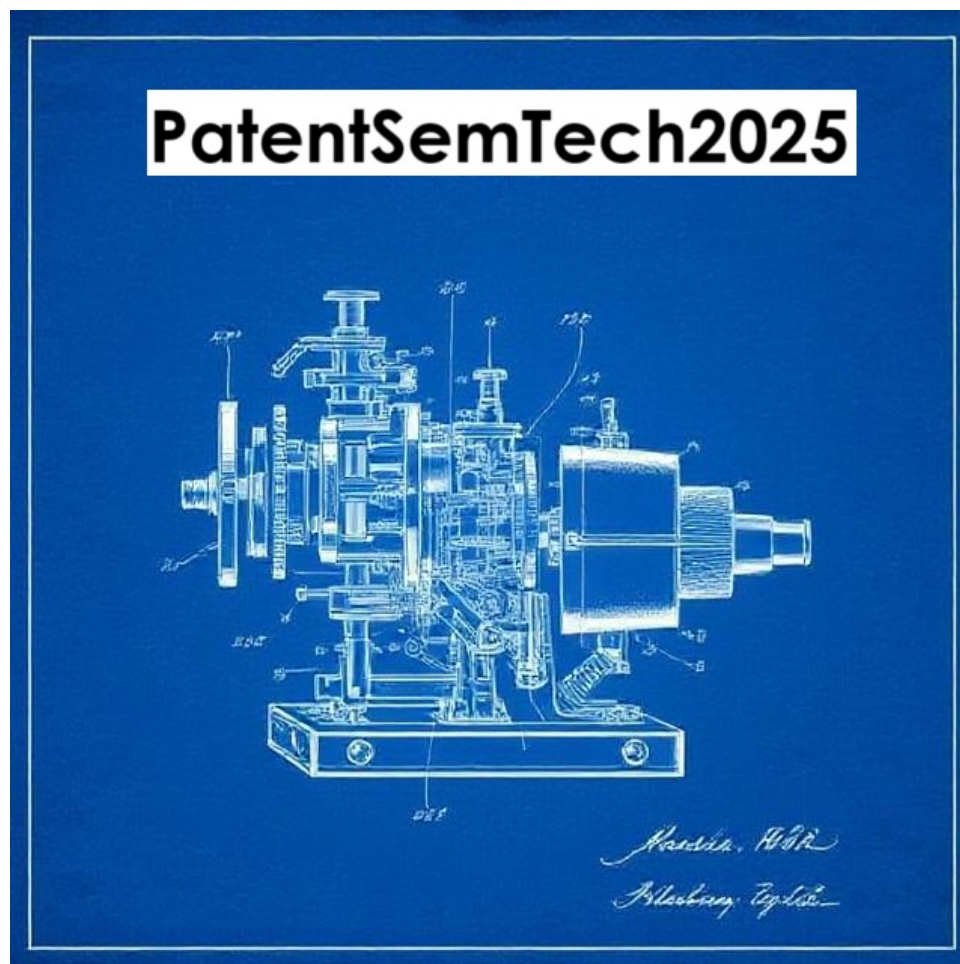
¹ZBW – Leibniz Information Centre for Economics & Kiel University, Düsternbrooker Weg 120, 24105, Kiel, Germany

²FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344
Eggenstein-Leopoldshafen, Germany

³Artificial Researcher IT GmbH, Taubstummengasse 11 (i2c), 1040, Wien Austria

⁴Institute of Information Systems Engineering, TU Wien, Favoritenstr. 9-11/194-04, Vienna, Austria

⁵Center for AI and Patent Analysis, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA



6th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech) 2025

✉ rkr@informatik.uni-kiel.de (R. Krestel)

🆔 0000-0002-5036-8589 (R. Krestel)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Preface

The sixth edition (PatentSemTech2025) of the workshop series *Patent Text Mining and Semantic Technologies* was held as a full-day event in conjunction with the SIGIR 2025 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 analyze compared to corpora comprising other text genres. Working with patent data, in addition to 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 that encloses solutions for real-world applications. (4) They are complementary knowledge to scientific literature, mostly not published elsewhere.

With the PatentSemTech2025 workshop, we continue our series of workshops launched in 2019, aiming to establish long-term collaboration and a two-way communication channel between the IP industry and academia from relevant fields. Therefore, the 6th PatentSemTech workshop was organized as a full-day event with 9 presentations of research papers that were accepted after peer review out of 19 submissions. We had 4 long paper presentations and 5 short paper presentations; the latter were also presented in the joint workshops' poster presentation. In addition, Iryna Gurevych (TU Darmstadt, Germany) gave a keynote speech on "How to InterText? Elevating NLP to the cross-document level". In an expert panel moderated by Alexander Klenner (European Patent Office, The Netherlands) six experts discussed how generative AI is changing the patent domain and experiences were exchanged with the workshop participants on how to apply AI for patent-related tasks.

Germany, Austria, USA, July 2025

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

- Linda Andersson (Artificial Researcher IT GmbH, Vienna, Austria)
- Hidir Aras (FIZ Karlsruhe, Germany)
- Sebastian Björkqvist (IP Rally, Helsinki, Finland)
- Rene Hackl-Sommer (DeepL SE, Germany)
- Allan Hanbury (TU Wien, Austria)
- Christoph Hewel (Paustian & Partner, Germany)
- Alexander Klenner-Bajaja (European Patent Office, Netherlands)
- Ralf Krestel (ZBW Kiel, Germany)
- Florian Matthes (TU Munich, Germany)
- Florina Piroi (TU Wien, Austria)
- Michail Salampasis (IHU, Thessaloniki, Greece)
- Anthony Trippe (Patinformatics, Ireland)
- Karin Verspoor (RMIT University, Melbourne, Australia)
- Lei Zhang (FIZ Karlsruhe, Germany)

Website

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