

Preface

The second edition of the workshop series *Patent Text Mining and Semantic Technologies* (PatentSemTech'21) was held as a full-day online event in conjunction with the SIGIR 2021 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 PatentSemTech2021 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 2nd PatentSemTech workshop was organized as a full-day event with research paper presentations (3 long and 4 short) that were accepted after peer-reviewing, 2 keynote speakers (Osmat Jefferson, lens.org, Australia; Noriko Kando, National Institute of Informatics, Japan), expert short talks and a panel discussion around the topic “Artificial Intelligence and Patent Analysis: Friends or Foes?” with 5 invited speakers from patent institutes, universities and industry. In a demo session, academic, start-up and open-source IP text mining tools were presented in 3 separate demos.

Germany, Austria, USA, July 2021

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

Copyright © 2021 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 (Betten & Resch, Germany)
- Julian Risch (deepset GmbH, Germany)
- Florian Matthes (TU Munich, Germany)
- Rene Hackl-Sommer (FIZ Karlsruhe, Germany)
- Anthony Trippe (Patinformatics, Ireland)
- Sam Arts (KU Leuven)
- Paul Groth (University of Amsterdam, Netherlands)
- Hans-Peter Zorn (inovex GmbH, Karlsruhe, Germany)
- Michael Natterer (Dennemeyer Octimine GmbH, Germany)

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/>