



# PatentSemTech 2024

5th Workshop on Patent Text Mining and Semantic Technologies

**@SIGIR'24**

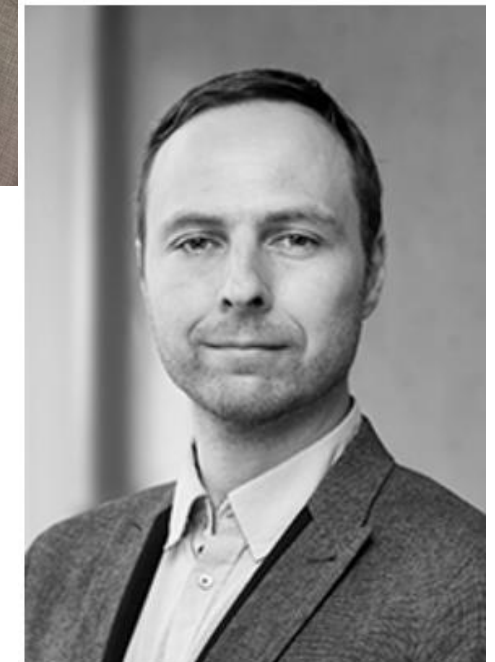
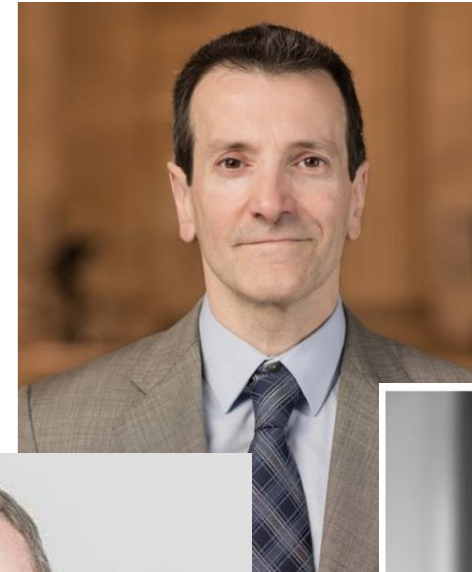
The 47<sup>th</sup> International ACM SIGIR Conference on Research and Development in Information Retrieval

July 14-18, 2024 (Washington D.C., USA)

Hidir Aras, FIZ Karlsruhe, July 18 2024, Washington D.C.

# ORGANIZERS

- Linda Andersson (Artificial Researcher, Austria)
- Florina Piroi, Allan Hanbury (TU Vienna, Austria)
- Hidir Aras (FIZ Karlsruhe, Germany)
- Ralf Krestel (ZBW Kiel, Germany)
- Dean Alderucci (CMU, USA)





# AIMS AND GOALS

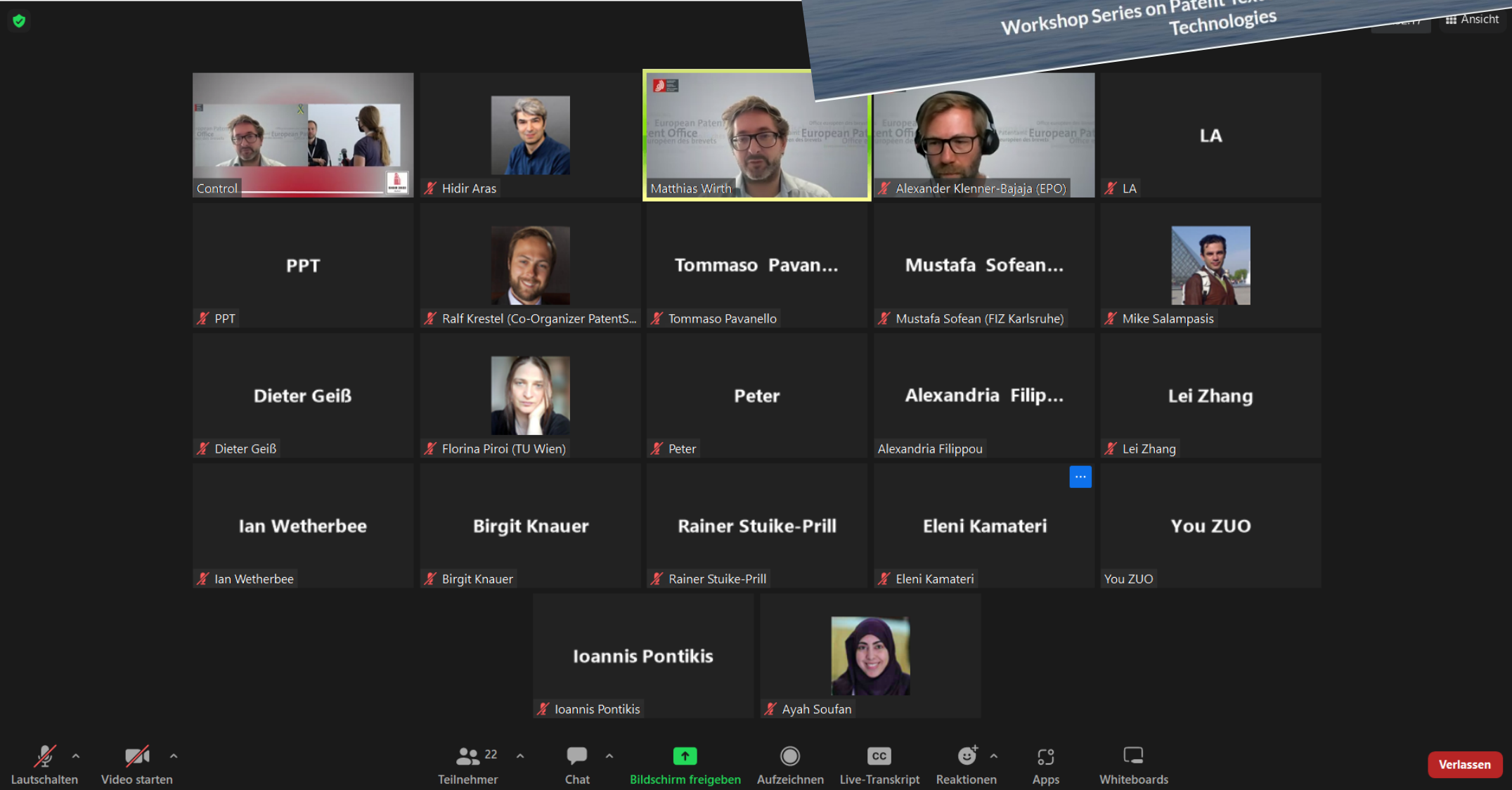
- Invite IP industry and academia to discuss problems / challenges related to the patent domain
- Show scientists that patent data provides also rich opportunities for research
- Explore new methods and technologies and apply them to relevant problems in the patent domain such as:
  - Text mining, NLP
  - Language models, Knowledge graphs
  - ML/DL and LLMs
  - etc.
- Help to enhance applications for patent retrieval and analytics
- Create research resources such as datasets, benchmarks, etc.

# HOW IT STARTED ... 2019 IN KARLSRUHE / SEMANTICS 2019



# HOW IT CONTINUED ... 2021 ONLINE VIA ZOOM ... DURING COVID PANDEMIC ☹️

PatentSemTech 2021  
Workshop Series on Patent Text Mining and Semantic Technologies





# PATENTSEMTECH @SIGIR'22 / HYBRID EVENT IN MADRID





# PATENTSEMTECH @SIGIR'23 IN TAIPEI / TAIWAN





# WHO CAME ...?



**Mr. Henry (Jamie) Holcombe**  
United States Patent and Trademark  
Office (USPTO)

## Invited Keynote Speakers



**Noriko Kando Ph.D (Library and  
Information Science)**  
National Institute of Informatics (NII)



**Prof. Osmat Jefferson PhD MInt Law**  
Queensland University of Technology (QUT)  
Director of Product Development, Cambia

## Panel Discussion

*"Artificial Intelligence and Patent Analysis: Friends or Foes?"*



**Dolores Modic**  
IPLodB project



**Christoph Hewel**  
BETTEN & RESCH



**Alexander Klenner-  
Bajaja**  
EPO



**Irene Kitsara**  
WIPO



**Tanja Sovic**  
TU Wien



# WORKSHOP SCHEDULE – PART 1



09:00 - 09:10

Welcome address

09:10 - 10:30

**Multi-Task Classification Model for Multilingual Patents**

Yongxin Peng, Xinyu Tong, Yueyun Zhang and Yonghe Lu (Sun Yat-sen University, China)

**Improving Patent Classification using AI-Generated Summaries**

Naoya Yoshikawa and Ralf Krestel (ZBW, Germany)

**Large Language Model Informed Patent Image Retrieval**

Hao-Cheng Lo, Jung-Mei Chu, Jieh Hsiang and Chun-Chieh Cho (National Taiwan University, JCIPRNET)

10:30 - 11:00

**Coffee Break**

11:00 - 11:30

**Semi-Supervised Learning Methods for Patent Classification Using Search-Optimized Graph-Based Representations**

Ekaterina Kotliarova and Sebastian Björkqvist (IPRally Technologies Oy, Finland)

11:30 - 12:30

**Keynote: Matthew Wahlrab (CEO RapidAlpha)**

**Unlocking Strategic Growth: The Role of AI Technology in Intellectual Property**

12:30 - 13:30

**Lunch Break**

# WORKSHOP SCHEDULE – PART 2



13:30 - 14:30

## Logic Mill - A Knowledge Navigation System

Sebastian Erhardt, Mainak Ghosh, Erik Buunk, Michael E. Rose and Dietmar Harhoff (Max Planck Institute for Innovation and Competition, Germany)

## Automated Patent Landscaping using Deep Learning

Sebastian Erhardt (Max Planck Institute for Innovation and Competition, Germany)

14:30 - 15:00

## Poster Session (4x5min pitch + 40min running into the coffee break)

### ClaimCompare: A Data Pipeline for Evaluation of Novelty Destroying Patent Pairs

Arav Parikh and Shiri Dori-Hacohen (University of Connecticut, U.S.)

### Automatic Identification of Patent Claim Types: Enhancing Efficiency in Patent Analysis

Rima Dessi, Hidir Aras and Mark Prince (FIZ Karlsruhe, Germany; CAS - Chemical Abstracts Service U.S.)

### Comparing Complex Concepts with Transformers: Matching Patent Claims Against Natural Language Text

Matthias Blume and Christoph Hewel (IP Aptly, Inc., U.S./Germany)

### Employing Active Learning for Training a DL-Model for Citation Identification in Patent Text

Farag Saad, Hidir Aras and Mark Prince (FIZ Karlsruhe, Germany; CAS - Chemical Abstracts Service U.S.)

15:00 - 15:30

## Coffee Break (during poster session)



# WORKSHOP SCHEDULE – PART 3



---

15:30 - 16:30	Open discussion: How to transform research insights into products? (60min) / XAI (Explainable AI) for Patent Analytics?
16:30 - 16:50	Looking back at 5 PatentSemTech Workshops (20min)
16:50 - 17:00	Workshop closing (10min)

---