

## Invited Talks

### **Panel: Artificial Intelligence and Patent Analysis: Friends or Foes?**

#### **Challenges for Patent Practitioners to Apply AI in their Workflows**

**Christoph Hewel**

Patent Lawyer  
BETTEN & RESCH  
C.Hewel@bettenpat.com

#### **Abstract**

Patent practice has a long history. As a consequence, the internal structure of patent law firms and their external interaction with clients, patent offices and courts is well established. Furthermore, also the workflows in patent prosecution are precisely defined. Such workflows in particular concern drafting and filing a patent application, prosecuting the application in the examination proceedings at a patent office until patent grant, and sometimes post grant proceedings (like revocation and litigation). It comes thus with no surprise that applying disruptive technologies like AI implies a huge hurdle for the patent industry. In the panel discussion I will present my view as a patent attorney of the concrete obstacles and some ideas of how they might be overcome. Such obstacles can especially be found in the internal structure of law firms and their business model. In particular, due to a time-based revenue model and the rather conservative nature of patent practitioners, there is high reluctance to invest (and at least in short-term loose) time trying new and potentially poorly conceived technologies. It therefore appears advisable to attempt adapting AI-based software solutions to the patent practitioner's needs and nature, in order to increase the level of confidence: Solutions which are custom-tailored to the patent-prosecution workflows and which imply a proven effect of gaining time for the patent practitioner. This does not only require advances in AI technology but also a profound understanding of the patent-prosecution workflows.

---

Copyright © 2021 for these texts by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## **IPLodB: Using Linked Open Data in the Innovation Field: Opportunities Unveiled and Problems Encountered**

**Dolores Modic**

IPLodB Project, Nord University  
dolores.modic@nord.no

### **Abstract**

The short talk addresses the linked open data (LOD) approach for enabling access to (linked) patent information. We also touch upon the IPLodB project, which takes advantage of two datasets that follow the LOD principles and are published by two reputable organizations, the European Patent Office and the Springer Nature. These two datasets represent the core on which we started building a new patent-centric LOD sub-cloud. Hence, we will look at AI and patent analysis from a linked open data perspective and try to discuss its technological impact for future developments.

## **Artificial Intelligence Opportunities in the Patent Grant Process: An IP Office Perspective**

**Alexander Klenner-Bajaja**

Head of Data Science  
European Patent Office  
aklenner@epo.org

### **Abstract**

Patents have much to offer in terms of artificial intelligence challenges. The filing of a patent application sounds like a enumeration of machine learning tasks: An application needs to be routed to the correct team (classification), it needs to be translated (neural machine translation) and last but not least it needs to be precisely classified within the CPC (Classification again). What happens next is a search for prior art: An information retrieval task that also benefits already today from machine learning. The information in patents is stored in figures (computer vision) and unstructured text (natural language processing), which makes it even more interesting to apply latest deep learning breakthroughs to solve challenges around patents. The citation graph of all prior art is waiting to be explored by graph neural networks. However patents are also different: they are written in a legal and technical language that uses different syntax and different terminology compared to the internet in general (i.e. usual off the shelf trained models). The drawings are not those of cats and dogs, but of technical nature, in black and white. In this talk some of the challenges will be highlighted and we show how they are approached and solved at the European Patent Office.

## **AI in and for Patent Analytics: A hype or an efficient support tool for patent analysts?**

**Irene Kitsara**

World Intellectual Property Organization  
irene.kitsara@wipo.int

### **Abstract**

Over the years, different automation tools for patent analytics tasks were proposed to management and patent information professionals, promising efficiency, reduction of time and necessary human resources. Patent information professionals have often been skeptical, raising concerns about quality, precision, transparency and control of the process and the outcome. With the AI advancements and related trend, governments, businesses, and individuals are eager to leverage the potential of AI and deploy them in their workflow. While AI or “AI-powered” tools start appearing, and AI is explored by IP offices and academia, two questions arise: is it working and is it worth it? The future of patent analytics is expected to include AI, even if the exact form and extent are not yet clear. In this talk we will share some thoughts and observations about the status of AI tools for patent analytics, related benefits and challenges. We will use as basis for these thoughts a. WIPO’s exploratory work (2016 and ongoing work) on the use of open source tools and machine learning for patent analytics tasks in the framework of preparation of related methodological resources; and b. USPTO’s report (2020) comparing the performance of a patent professional team using traditional search and analysis approaches for the WIPO Technology Trends report on AI (2019) with the results of an AI model to retrieve and group AI-related patent documents, using WIPO’s patent dataset as benchmark.

## **Patentability Search: University’s Perspective**

**Tanja Sovic**

Head of Patent & Licence Management  
Technische Universität Wien  
Tanja.sovic@tuwien.ac.at

### **Abstract**

Prior art search is crucial for the university’s research. Being aware of relevant literature and patents related to the research topics can proof that our work is unique. Accelerated technological development and increasing number of interdisciplinary collaborations between different scientific areas lead to the expanding complexity in the prior art search. How can AI support these trends?

# Industry Demos: Integrating linguistic knowledge and Deep Learning into patent search tools

## WIPO Pearl - Insights into the Concept Map Search and Linguistic Search

Geoffrey Westgate and Cristina Valentini

World Intellectual Property Organization  
{name.surname}@wipo.int

### Abstract

In this workshop we shall present WIPO Pearl, the multilingual terminology portal of the World Intellectual Property Organization (WIPO), a specialized agency of the United Nations <sup>1</sup>. The nature of the linguistic dataset made available in WIPO Pearl will be described and we shall show how multilingual knowledge representation is achieved and graphically displayed. Secondly, we shall demonstrate how such data is exploited to facilitate search of prior art for patent filing or patent examination purposes, by leveraging the validated linguistic content as well as the validated conceptual relations that are presented in “concept maps”. We shall discuss how, in addition to humanly validated concept maps, “concept clouds” are generated by means of machine learning algorithms which automatically cluster concepts in the database by exploiting textual data embedded in the terminology repository. Finally, we shall present opportunities for collaborations with WIPO in the field of terminology. WIPO Pearl was launched in September 2014. The portal gives free access to the contents of the terminology database of WIPO’s Patent Cooperation Treaty (PCT) Translation Division (PCT Termbase), a repository of scientific and technical terms extracted from patents in ten languages. Its aim is to promote accurate and consistent use of terms across different languages, and to make it easier to search and share scientific and technical knowledge.

WIPO Pearl contains multilingual language data and semantic data, all fully validated by language experts, and constitutes an innovative project amongst terminology databases freely available on the Web today. The design of WIPO Pearl seeks to offer users flexible and distinct yet complementary ways of searching the terminology dataset: a traditional search by term, called Linguistic Search, and a search by concept, called Concept Map Search, which allows users to browse the conceptual system organized by subject field / subfield and by language. Moreover, WIPO Pearl allows users to exploit synergies between the terminology database and other WIPO patent-related resources, notably PATENTSCOPE, WIPO’s database of patent applications, and machine translation services embedded in the latter such as PATENTSCOPE CLIR. Redirection to PATENTSCOPE, in particular, allows users to look for prior art for

---

<sup>1</sup><http://www.wipo.int/reference/en/wipopearl>

patent filing purposes by using the validated terms of the PCT Termbase as “seed terms” or keywords.

Since its launch, new versions of WIPO Pearl have been released, with enhancements targeted at improving the user’s experience by facilitating the navigation and filtering of results, localizing the user interface (currently available in ten languages), and offering additional features such as a quick term-list view, image search, and a “concept path” search option within Concept Map Search that allows users to find the path between two concepts, showing all the related concepts in between. The concept path search function also allows users to launch a combined keyword search in PATENTSCOPE after having selected a concept path, thus allowing users to exploit validated semantic relations existing in the terminology records (partitive, generic, associative, as well as synonyms) to enhance patent search and search of prior art. Finally, an innovative recent feature involves the generation of “concept clouds” in Concept Map Search to display relationships between as yet unlinked concepts (i.e. concepts that are not yet part of the validated concept maps), as suggested by a machine learning algorithm trained on the corpus of validated contexts and relationships existing in WIPO Pearl.

Alongside these technical improvements, the contents have been regularly enhanced by adding collections of new terms and concepts, many arising from collaborations with external partners, including universities worldwide. Currently WIPO Pearl contains 205,000 validated terms and 21,000 validated concept relations. The workshop will conclude by describing opportunities for collaboration with WIPO in the field of terminology, whether for university students of terminology, or scientific and technical experts whose assistance is sought to complement the work of WIPO’s language experts in validating the contents of WIPO Pearl.

## The next generation AI-based Prior Art Search tools can be sustainable and transparent.

Linda Andersson, Peter Pollak, Tobias Fink, Florina Piroi

Artificial Researcher IT GmbH  
{name.surname}@artificialresearcher.com

### Abstract

In the workshop PatentSemTech'21, we will in our demo talk introduce software and services developed by the start-up Artificial Researcher IT GmbH (AR). The start-up was founded in 2019, and the company's text mining technology is based upon a two-time award-winning PhD research result by Linda Andersson at TU Wien. The focus of the demo will be the technology behind the **AR Data Pipeline solution**, which is a production flow that process any type of machine-readable text data and create sustainable ready to use indices and ontologies, as well as enhanced data formats, which can be integrated into client's own dataflow system. The ontology generated by the AR Data Pipeline is implemented as docker images containing the software **AR Ontology Service**<sup>2</sup> and the **AR Passage Retrieval Service**<sup>3</sup>. We have developed a modular software architecture which allows for continuous improvement releases, technology quality, and transparency to our clients.

To develop scientific and patent text mining tools for students, researchers, and patent experts, we need to understand their daily work, as well as the linguistic characteristics of the text genres. By integrating domain-specific ontologies into the information retrieval system, the AR technology provides automatic query expansions with understandable semantic information to provide Transparent Artificial Intelligence (AI). The key component to create sustainable search solutions is to provide users with several search alternatives and not limit the users to just one alternative. Different types of information needs require different search solutions such as meta-data search, text-box search, and graph search. With the novel **AR Graph Search Service** composed of domain-specific ontologies extracted from the collections, and the direct links to text paragraphs provide easy knowledge and terminology discovery. Transparent AI provides humans with explainable and thoroughly tested models, the models answer questions why terms were extracted, and how the related concepts are linked. The retrieval model is also transparent with how the query formulation was constructed.

To integrate linguistic knowledge into algorithms is essential for domain-specific text mining tools. To this day, many frequently used algorithms still postulate a *single word* can capture the entire scope of a semantic concept. For many text genres and languages, this is a valid premise, however this is not true for text genres and languages characterized by frequent multi-word term

---

<sup>2</sup><https://graph.artificialresearcher.com/>

<sup>3</sup><https://passageretrieval.artificialresearcher.com/>

(MWT) occurrences used to describe domain-specific concepts. Consequently, many of the state-of-the-art text mining techniques, as well as Natural Language Processing (NLP) tools have significant lower performance when applied on patent and scientific literature.

In the patent domain all types of issues, from very specific search requirements to the linguistic characteristics of the text domain, are accentuated. In writing processing of technical English texts, a MWT method is often deployed as a word formation strategy in order to expand the working vocabulary, i.e. introducing a new concept without the invention of an entirely new word. This productive word formation is a well-known challenge for traditional NLP tools utilizing supervised machine learning algorithms due to the limited amount of domain-specific training data (labelled data). The out-of-domain data issue, increases the unseen events and out-of-vocabulary term occurrences, which negatively affect the performance of the text mining tools. In comparison, Deep Learning (DL) algorithms do not require large amount of manually labelled training data, since the algorithms derive knowledge out of unlabelled data (hence unsupervised methods). However, using an unsupervised method does not completely exclude labelled data, since labelled data may be required to initiate the learning process.

In conclusion, DL algorithms have several advantages compared to the supervised NLP methods. However, the unsupervised algorithms need a significant amount of data to achieve implicit learning from the data. Meanwhile, supervised algorithms do explicit learning, but will only learn from the labelled data they are trained on. The unsupervised methods also require a representative data set in order to reflect the implicit learning that should take place. If the data is unbalanced (natural biases), the unsupervised algorithms will still end up with issues regarding unseen events and out-of-vocabulary term occurrences, due to the fact that implicit knowledge could not be derived from the given data. With our technology, we aim to provide text mining solutions with Transparent AI by focusing on addressing the limitation and reducing the natural biases in DL models. For the domain-specific ontology population method, the AR technology extracts single words and phrases by combining NLP and gazetteers with a domain-specific trained Bidirectional Encoder Representations from Transformers (BERT) model, part of the **AR NLP-toolkit Services**.<sup>4</sup>. The AR technology makes use of a domain-specific modified NLP module, as well as an assembly module composed of several similarity values. The assembly module targets the semantic functions, syntagmatic (i.e. MWT relations) and paradigmatic (i.e. lexical-semantic relations e.g. hyponymy, synonym). To summarize, the next generation technology needs to incorporate linguistic information and provide users with several search alternatives (meta-data search, text-box search and graph search) to give users the option to utilize the most suitable technology for a given information need. We believe scientific literature, technology and data should be findable to everyone and not just to those who know where to look and how to search.

---

<sup>4</sup><https://swagger.artificialresearcher.com/?urls.primaryName=unified-nlp-server>