Augmented intelligence for state-of-the-art patent search

Ana Hafner
Faculty of Information Studies in Novo mesto and DOBA Business School
Novo mesto, Maribor, Slovenia
ana.hafner@fis.unm.si

Nadja Damij
Newcastle Business School
Northumbria University
Newcastle upon Tyne, UK
nadja.damij@northumbria.ac.uk

Dolores Modic
Nord University Business School
Nord University
Bodø, Norway
dolores.modic@nord.no

***Abstract* — The volume of patent data is increasing, which is a big challenge to patent examiners as well as to all inventive companies and individuals. In this paper we take the view of individual inventors who believe they invented something new. Artificial intelligence brings a promise to support their prior art search for existing (similar) inventions with machine learning and deep learning algorithms. We discuss the potential of artificial intelligence in prior art searching. We present an experiment, based on a real-life invention, comparing relevant patents we got from Boolean keyword searching with those from the semantic search supported by artificial intelligence. We can confirm that artificial intelligence has great potential in this field. However, presently it is not yet able to make traditional patent search engines obsolete, hence it still fits better with the notions of augmented intelligence or expertise.**

Keywords—patents, prior art search, artificial intelligence, augmented intelligence, machine learning

I. INTRODUCTION

The volume of patent related data is on the rise. According to World Intellectual Property Organization's report [1], 3,276,700 new patent applications were submitted in 2020, almost a 1,6% increase in comparison to 2019. To file a new patent application, a state-of-the-art analysis, also called a prior art search, needs to be carried out to ensure that the same invention has not been previously discovered. A prior art search –due to such a high volume of data – is a time-consuming task if we want to ensure the quality and accuracy of the results. This task, also referred to as a needle-in-a-haystack challenge [2], is usually entrusted to patent attorneys who possess competences to navigate efficiently through patent databases, yet in practice other patent information users, such as individual inventors, will also engage in this challenge.

In the context of intellectual property rights (IPR) several future technology trends have been identified [3], [4], [5], [6], [2] such as: 1. merging of private and public data to increase business intelligence; 2. advanced analytic techniques integrated into IPR workflow; 3. the foundation for artificial intelligence (AI) to improve data quality and IPR data exploitation, 4. when handling IPR, resources need advanced big data skills and competences (advanced skills, talent management); 5. newly appearing formats of data, linked to "open innovation", to create benefits for research and development (R&D) processes; and finally, 6. artificial (or augmented) intelligence applied to support the retrieval of relevant IPR data.

Setchi et al. [2] claim that AI has the potential to assist patent examiners in the future as part of the prior art searching process. The aim of this paper is to test a specific real-world invention (and its future patent application) to evaluate how efficiently AI can support the patent searching task and thus contribute to saving time, effort, and costs. The identified research question is: if a Boolean keyword search within the largest open patent database is compared with a semantic search within AI supported commercial databases – how accurate are their results and what are, if any, the time efficiencies?

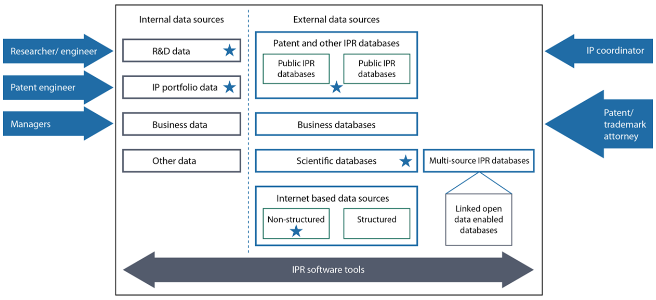
II. BACKGROUND

A. IPR Data and Sources

IPR data is suited to be analyzed with big data tools and techniques due to its high volume, variety, and velocity of changes as well as high veracity and high value (i.e. referring to the extent to which big data generates worthwhile insights and benefits) [7].

Due to improvements in technology, IPR big data can be used by different types of experts in need of IPR information as well as is supporting the efforts of IPR management (IPRM) to involve diverse human resources. IPRM has three types of goals, three technological layers [6] and two components. The goals are connected to R&D processes, IPR itself and achieving business (strategic) goals. The three layers are connected to the databases, the techniques and the software tools. From this, several consequences for handling IPR data and IPR tools arise in relation to human and other resources, such as dealing with different types of users, diverse users' entry points and IPR analytic skills, different types of IPR activities to be carried out by users, and finally, the extraction of information from various IPR databases and data sources. The components consist of a strategic and an operational layer. The first requires individuals interested in exploiting IPR assets to achieve business goals (keeping competitive advantage, building strategic alliances, understanding future trends, maximizing IPR portfolios etc.), whereas the second is focused on the individual level (e.g., the contribution of specific IPR data to further development). The first component requires aggregate IPR data, the second in-depth IPR data. Both utilize the opportunities of big data analytics and tools.

Fig. 1 presents the data sources, actors and entry points related to IPR (big) data. The sources of external data are patent and other IPR databases, business databases, scientific literature databases, Internet broad search information and multi-source IPR databases with Linked Open Databases (LOD-enabled) as a related format. Free public national and international databases (e.g. patent offices' databases) allowing raw data retrieval and some metadata; and commercial, usually more sophisticated, IPR databases, constitute the first group.



Legend: stars= entry points

Fig. 1. Data sources, actors and entry points

Next, business databases e.g., Orbis, can be useful to gather information on IPR holders (their portfolios, strengths, and IPR-connected behavior). Scientific databases offer scientific papers, conference contributions and some aggregate metadata. Miscellaneous online data sources include less structured sources such as business news, wiki or blog based IPR related texts, LinkedIn etc., and more structured sources (e.g. IP Nexus providing information on IP experts). Multi-source IPR databases provide broader information including citation or IPR quality and business connected data. A special format is also the above mentioned IPR LOD-enabled databases (or LOD databases in short).

The four general types of internal data sources include R&D data (non-structured and often a mixture of codified and tacit knowledge), IP portfolio (comprised of IPR data, IPR analytics, know-how data), business data (financial data, data on strategic alliances, etc.), and other operational (day-to-day) data (court decisions, personnel files, etc.). Internal data increases the usefulness and validity of external data. The knowledge contained within internal IP-related resources, expertise and goals, and data derived from external data depicting the knowledge, strengths, weaknesses and skills of other organizations, together provide a more complete picture.

Fig. 1 also highlights diverse entry points for different IPR data users. e.g. for a manager, the initial entry point may be a business-news article from an unstructured data source; a researcher can consult an IPR/scientific databases or internal R&D data to gain ideas for inventions, or can use IPR databases providing data on state-of-the-art in order to prevent duplication and identify opportunities. Similarly, an individual inventor's first entry point might be their own R&D data, then moving on to consulting IPR and scientific databases (typically due to resource scarceness, using those that are not behind a paywall).

Employed human resources will thus have diverse competences, goals and initial information, but a more level playing field can be achieved by appropriate IPR tool solutions, allowing easier access, visualization, broader access to diverse data and better understanding of IPR data. Timely integration of new data and information sources; permitting open-ended functionality and tailor-made solutions or functionalities; allowing the companies not only to integrate their own data, but also their own ontologies semantic based functionalities, and integrating different precision-to-recall solutions for diverse users inside IPR tools all mitigate this.

B. IPR Techniques and Tools

Modic et al. [6] already pointed out the relevance of not only examining the best available IPR data sources and their attributes, but also ensuring the availability of appropriate IPR techniques and tools, for harnessing the benefits of IPR data, and to move toward knowledge creation assisted by AI.

Due to the rapidly rising volume of IPRs, obtaining a patent without a clear understanding of the empty space is getting more difficult. One of reasons for the growing importance of IPR big data-enabled analytics is, the more transparent the IPR, the easier it is for the company or individual to understand its value. Hence, software tools and their functionalities, which help companies in creating added value from their IPR assets, are on the rise. To be truly able to understand IPR processes as business processes [8] and efficiently utilize all available resources, IPR processes need to be supported by appropriate technology. Companies that maintain manual search processes to obtain relevant IP information are not only using their resources less efficiently, they are also missing out on the critical business insight that comes from connecting disparate (big) datasets [3]. The Vs of IPR Big data, their formats, sources, and the requirement to effectively manage the IPR, demand the use of big data analysis solutions and software tools if companies want to maximize the value of IPRs, especially via cross-genre or cross-database retrieval of information [9].

Turning now to the techniques. In their seminal paper Abbas et al. [5] made a taxonomy of proposed computer-assisted patent analysis techniques where they distinguish between text mining and visualization approaches. Modic et al. [6] also present a typical computer-assisted document analysis pipeline as an IPR techniques classification framework, distinguishing between document preprocessing related techniques, feature extraction techniques and feature analysis.

On the other hand, e.g., Baudour and van de Kuilen [10] or Satchi et al. [2] focus in particular on the patent information retrieval part of the pipeline. Baudour and van de Kuilen [10] already pointed out in 2015 that the "field of patent information searches has dramatically evolved" (p. 4), and have predicted two directions of development of techniques and related tools. Firstly, there are solutions focusing on the "simple" search of raw, unformatted, unstructured, non-cleaned data. They [10] postulated that this will consist mostly in the development of new tools able to digest large amount of information, and search the content intelligently. In this line, e.g. for the European Patent Offices' EP Linked open data, a simple solution iplod.io with a beta interface was developed which in particular focuses on the advanced data disambiguation techniques [11] supported by advanced algorithms, but which is focused only on a particular type of entity. Secondly, they predicted the development of new solutions for powerful searches in enhanced content databases, where the data have been cleaned and formatted beforehand.

Setchi et al. [2] also provide a review of relevant AI techniques focused on data retrieval and, in particular, prior art searches. They have investigated the feasibility of AI for prior art search, whilst delineating the related techniques through a series of technical requirements. They posit that e.g. the natural language processing (NLPs) techniques (such as; text segmentation, normalization, lemmatization, stemming, co-occurrences) are relevant for all five technical requirements (query expansion, document classification, document similarity, ranking and visualization), with e.g. supervised learning techniques (support vector machine, naive Bayesian learning, decision tree induction, random forest, neural networks) limited to only one technical requirement (i.e. that of document classification). Other classes of techniques (unsupervised learning and semantic technologies) are in their opinion usable for only some of the technical requirements.

Furthermore, several authors have engaged in an evaluation of IPR tools. More than a decade ago Bonino et al. [9] were optimistic with regard to semantic-based solutions, however, some of the tools they describe are now in poor condition or unavailable. Modic et al. [6] have analyzed the websites of more than 10 IPR tools providers as identified by interviewees and/or the Hyperion MarketView™ Report and Capterra's review, with the focus on their functionalities, and have estimated that the tools have not yet fully integrated all the possibilities. More recently, Satchi et al. [2] have developed a protocol to investigate the feasibility of AI for prior art search.

AI for IPR in many articles remains relatively open in terms of its conceptualization, often simply described with techniques that can be enveloped under the umbrella of artificial intelligence for IPR (e.g., [2], [13]). Similarly, in recent conferences we could hear that ad minimum machine learning is included under this umbrella, which is sometimes also its downfall, as often artificial intelligence and machine learning are considered almost synonyms. A notable exception in attempting to further delineate the concept was made by Lupu [12] where he attempts also a semantic analysis, yet he also stops short of providing a comprehensive AI definition.

On the other hand, in the IPR communities the term of augmented intelligence is also present [6]. This stems from the fact that currently there is a consensus that patent experts will not be replaced by automated and intelligent systems, but rather these will remain supportive technologies. The hybrid approach, with AI and machine learning augmenting human intelligence, when paired with human knowledge and intuition is what is commonly called augmented intelligence [14]. Hence we can speak of augmented intelligence for IPR, i.e. AI and machine learning augmenting human intelligence, paired with human knowledge and intuition on IPR. This is also the angle, albeit only implicit, to several evaluations of AI potentials for IPR and their assessment (e.g. [2]). However, Satchi et al. [2] focus on the view of patent examiners, but the profiles dealing with IPR and attempting to retrieve relevant information are much wider. We focus on an average user (inventor, entrepreneur), needing this information for practical reasons.

C. *Needle-in-a-Haystack Challenge*

An average user lacks knowledge of, for example, international patent classifications (IPCs), which in turn prevents them from exploiting all possibilities of open patent searching platforms, such as Espacenet or United States Patent and Trademark Office (USPTO). Simple keyword search using Boolean operators such as AND, OR and NOT may be the user's only choice to either limit or extend the search. Since these platforms commonly return thousands of hits, limiting the search is a crucial task, while at the same time it is important to ensure that the search minimizes the risk of missing any important results. Consequently, retrieving relevant results is difficult and time-consuming.

AI research particularly linked with natural language processing, deep learning techniques and machine learning algorithms to extract the basic knowledge of patent documents has already given some encouraging results [23], [24], [25]. A semantic search is a search technique, supported by machine or deep learning algorithms, that relies on the ability of the algorithm to consider the contextual meaning of search phrases. Semantic searching does not merely count the repetition of keywords and measure the proximity of terms as keyword searches do, but rather uses AI to predict and understand the contextual meaning of query phrases [26]. This allows the search parameters to include whole sentences of invention description instead of just keywords. Due to this feature, much hope is placed in semantic search, which could simplify retrieval of relevant patent documents.

III. METHOD

The paper uses the experiment as a research approach to compare the keyword and full-text search results to evaluate their accuracy and time efficiency. The experimental design approach has been used, for example, by Roberts et al. [22] to "introduce elements of the experimental design that inform document collection within a generally applicable framework".

In this paper we design the experiment of AI efficiency in prior art search in a way that allows its replication on the same or different (other proposed) patents and tools and their included datasets. Specifically, the paper compares the results of the search of relevant patent documents connected to a focal invention (described in detail below) by applying Boolean search within the Espacenet patent database as a benchmark, and applying AI supported semantic search within IP tools of private providers. The protocol schematics are presented in Fig. 2.

The search is based on an invention for which the patent application is already drafted, but not yet filed at the time of writing this paper. The invention refers to chocolate packaging (typically of a chocolate bar) that includes a new integral part enabling more hygienic consumption of a bar (or pieces of chocolate or candies in the chocolate box) so that the user's fingers do not come into direct contact with the chocolate. Possible IPCs for this invention are B65B, B65D, A47G 21/10, A23G 1/50, A23G 3/50. Patent includes 22 figures and 9 patent claims.

The experiment enables testing of AI prior art search efficiency by comparison of the relevant patent documents for described invention retrieved from Espacenet with those retrieved from AI supported searching tools.

Espacenet is the largest patent database in the world with free Web-based access to over 130 million patent documents. It was developed by the European Patent Office (EPO) together with the member states of the European Patent Organization. With 25,000 daily users, Espacenet is one of the most frequently accessed patent information services. The EPO also continually works to improve its offerings [15].

For the AI supported patent search tool providers, the paper only included the tools as indicated by the results that appeared in the first page of internet search engine (Google) when inputting keywords "artificial intelligence" and "patent search" at the same time. These were three (at the end of January 2017): Patentfield [16], Dennemeyer's Octimine [17], and Integrator [18]; however, after contacting these providers we discover that we would be unable to test the latter one (demo version will only be available in second quartile of 2022), hence we tested in its stead IPRally [19], which was suggested to us by one of the professional associations. Free access was granted to us to the demo version of Octimine and IPRally, with Patenfield already being openly accessible.

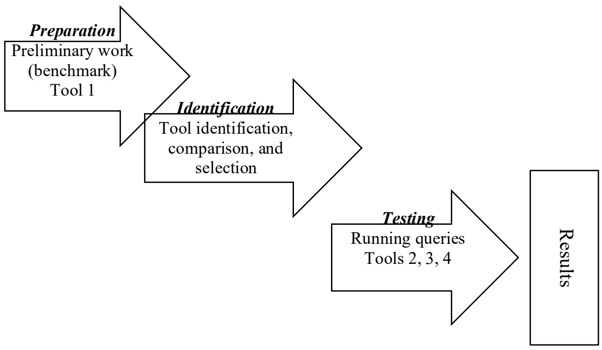


Fig. 2. Protocol schematics

Patentfield is an open patent search and analysis platform which also includes AI machine learning semantic search and AI classifications options [16]. The application IPRally is supported by deep learning neural networks trained with millions of real-world patent cases and, as they claim, "has learned to mimic the work of patent examiners" [19]. IPRally promised a unique AI knowledge graph technology for patent searching. Dennemeyer's Octimine supports AI powered patent search by Octimine's SaaS application using machine-learning algorithms that are tailored to patent data in combination with a clear and easy-to-use interface which was developed by researchers from the Max Planck Institute and LMU Munich [17].

Patentfield, which is open access, includes Japan (JP), US, European (EP), and PCT (Patent Cooperation Treaty) data. Octimine, besides already named offices also includes data from several EU offices, and other countries such as China (CN), Canada (CA), South Korea (KR), Australia (AU) and Russia (RU). IPRally includes data from 45 major patent offices. Data coverage has to be taken into account when comparing different patent searching tools.

For the purpose of this research, two basic metrics that measure the performance of such tools, called precision (quality ratio of the retrieved hits or documents) and recall (the proportion of correct answers found) [20] will be applied for the analysis of the search results, which are also widely used for the assessment of data retrieval [11].

IV. RESULTS

A. *Backround and preliminary work*

The initial search regarding the invention was completed in September 2021 before the patent was drafted. The state-of-the-art search was performed with Tool 1 by using keywords and their combinations such as chocolate packaging, chocolate AND packaging, chocolate AND hygiene, candies OR truffles OR biscuits AND packaging, etc. to identify any similar solutions that were already patented. The search identified over 200 patents and the 14 most relevant were chosen. Table I presents the results of our benchmarking exercise (i.e. preparation step). All 14 most relevant patents are, according to the EPO's Guidelines for Examination, considered Type "A" documents as they are connected to the invention, but not prejudicial to the novelty or inventive step of the claimed invention [21].

TABLE I

IP tool	Tool 1
Type of searching	Boolean
Time spent on searching and examination of relevant patents	8 hours
List of relevant patent publications	KR20190134405** WO2010121731* DE102009014245* RU2520009 EP2132110* CH701734 USD594323 US2010044263* US2015136642* CN104843329 WO2010040592* EP3380404 CN203410834 (U) AU201716183

B. *Testing*

In February 2022, the AI algorithms in patent search were applied by firstly inputting in the search tool only the abstract, then only the claims and, lastly, the full patent text, i.e., complete description including abstract and claims. Results are presented in Table II.

All three IP tools calculated similarity scores of patent documents in their databases with the entered text and then sorted the patent documents from the most to the least relevant. Table II presents 15 of the most relevant patents in each category (abstract/ claims/ full description text) and only those connected with the invention (i.e., relevant patents), are marked with *. This means that after our in-depth investigation the majority of the search results proved to be irrelevant. AI offered results such as a method and device for making marshmallows, chocolate calendar, packaging for fast foods, even a fragrance testing strip combined with a nose clearing devise appeared. Time to recall these patents was a matter of seconds, but examination of their relevance was time consuming (longer in the case of Tools 3 and 4 due to patent titles indicating more relevant results). Namely, in the case of Tool 2 some patents were very quickly rejected based on their title and the drawing only – without the need to read the patent description.

Based on the search results, several highlights can be made.

Firstly, patents that were extracted from four different tools, are almost completely different – there is no patent publication which would appear among the top relevant publications in all four tools (we also considered that patent families might have included different publication numbers). However, in the case of Tool 3 and Tool 4 there is a small overlap since patents highlighted with bold letters are members of the same patent family. Additionally, in case of Tool 3 we were able to extract 1000 of the most relevant patents and among them there are six publications which are marked by * in Table I. With Tool 4 we were able to extract 50 most relevant patents and among them there is one marked by ** (Table I). Therefore, these patents coincide with those extracted from Tool 1, but they were not recognized among the 15 most relevant (in Table II). For Tool 2 the same test could not be made as the extraction of the patents was not available to us.

Secondly, according to results, Tools 2, 3 and 4 do not differ only by data coverage (Tools 3 and 4 including more databases) but also by their machine learning or deep learning algorithms. It looks as if Tool 3's algorithm is more focused on the field of invention, Tool 2's on how it is made, while for Tool 4 we got the impression it is focused on the technical problem to be solved.

Thirdly, by entering the full patent text the highest number of relevant results was achieved (see precision and recall – P/R in Table II). This confirms the finding of Wretblad [20] that using a full-text document as an input improves recall considerably. However, it seems that length of inserted text is not the only factor; namely, abstract is much shorter than the patent claims section, but the worst results in case of the Tool 3 were received when entering the patent claims which are supposed to be the most exact technical description part of a patent.

Lastly, all three Tools (2-4), very precisely determined IPCs for the invention. Therefore, such tools enabling AI-supported search can be a great support to the national patent offices' patent examiners who have to determine IPCs for new patent applications.

Finally, to answer the research question: there are time savings when comparing semantic and Boolean search, but the accuracy of the results was not always better.

TABLE II

IP tool	**Tool 2**	**Tool 3**	**Tool 4**
Search type	Semantic	Semantic	Knowledge graph-powered search
Time spent	1 hour	2.5 hours	2.5 hours
Abstract similarity	US2015140180	DE20020699	CN209152220*
	WO2012104578	DE3635858	CN209152221*
	US2015245631	DE20119962	WO2011027621
	US2005095326	US3108874	BE1021271
	US2018368597*	BE1018227*	EP2845700*
	WO03086266	GR1009499	EP2982250
	WO2004086886	DE19805134	CN203897194
	US2020253222	JPS5018664	CN209518211
	US2012328742	US2558128	CN204335724
	WO8905764	DE202004005621*	CN110384158
	US2017050793	US2571516*	KR20020060510
	US2014076179	US20210259277	JP2770117
	US5649728*	US2863772	DE6921931*
	US2015079242	US2539518	RU115626
	US2006051463	FR2689730	KR20100007441
Patent claims similarity	WO9844832	FR3082733	US20170119010
	US2005095949	DE20020699	CN209152227*
	WO9737906	US3252702	US20160227812
	US6299918	US3189055	CN215270398
	WO9944482	US3422609	BE1021313*
	US2013081961	US3245681	US20100196548
	US2007034226	US3116763	CN209152221*
	EP0973410	US3332450	US8974850
	WO2008153844	US3205644	BE1021271
	EP2818052	US3092949	EP2845700*
	WO9716075	US3222851	US20160152406*
	US2018208342	US3298889	AU2013204649
	WO9830138*	US3229726	CN103879170
	WO2008118965	US3413787	WO2018065797
	WO2008137183	US3114397	DE6903169*
Full description similarity	US2010089857	BE1018227*	US20170119010
	US2010089856	DE20020699	EP2845700*
	US2005089604	ES2586598	**US20160152406***
	US2015245631	NL1043206	RU44459
	US2016120208	DE202011108574*	WO2018065797
	US2010151091	DE19605815*	US20160227812
	US2010285184	BE1019423*	CN209152227*
	WO9844832	DE202008001694*	DE1981252*
	US2009148571	FR2859979*	DE1974564*
	WO2018189562*	DE20120329	CN209152221*
	WO2009068451	EP2447187*	BE1021313*
	WO9737906	**GB201314107***	NL1043206
	WO0134003*	FR2874002*	WO2007063265*
	EP2227423	US20130240387	DE6903169*
	US2021053718	DE202011051492*	RU2275825
P/R	**2/15, 1/15, 2/15**	**3/15, 0/15, 10/15**	**4/15, 6/15, 9/15**

V. DISCUSSION AND CONCLUSIONS

The experiment tested AI supported IP tools for a particular invention, garnering mixed results.

We first develop and later adopt an easily replicable protocol, consisting of a benchmarking exercise in the preparation phase and the testing itself in the second part. The value is also added as our experiment relies on a concrete case at its heart. Furthermore, contrary to much of the literature, we focus not on patent examiners, as for example recently [2], but point out the variety of users of the patent information – and in particular highlight individual inventors.

Turning now to the generated results. On the one hand, AI did not provide within the top 15 results even one relevant document that was previously discovered during our benchmarking exercise with Tool 1. The results we got from three different AI tools were also surprisingly different, indicating that even AI can have a diversity of opinions on (di)similarity of IP. On the other hand, AI came back with a lot of new very relevant documents that were previously missed. Hence, although we need to be careful about false positives, we can diminish the problem related to false negatives – both have been pointed out as important issues [11].

Additionally, the user friendliness of Tools 2, 3 and 4 is increased in comparison to Tool 1 that requires the user to have some basic training to begin the search. Tools 2 to 4 are designed so that anyone can use them without any previous knowledge in data retrieval. As such, tools integrating AI can offer support to individual inventors and small enterprises who face barriers in affording the cost of patent attorneys to check the state of the art for their ideas, as long as the cost of the tools is not prohibitive. But there is an obstacle too: longer texts obviously give better results. However, investing in patent drafting with the sole purpose of checking the prior art may also not be a reasonable decision for those with limited resources.

To conclude, currently solely relying on AI supported search is not yet recommended, but it can represent a great support to traditional Boolean search and can lead to augmented expertize. Tools 2 to 4 not only support full-text searches but also various combinations which can significantly improve the results. For example, in our case, when we saw that AI got stuck at chocolate recipes, we could simply guide it to chocolate packaging. As one of our tool providers commented: "AI will always need a little help from people". But these search variations are not the subject of this paper since they are different in each tool, but we wanted equal conditions for all and to see what we get if we enter exactly the same text without any corrections in searching techniques.

We also point out that the data coverage plays an important role. If there is no data, not even the most sophisticated AI can find patents included in Tool 1.

Albeit conducting the experiment diligently, our work still faces the issue of generalizability, similar to other qualitative methodology based works. The external validity remains limited to the analytical generalization [27], whereby our approach with a real life case in the heart of our experiment allowed us to provide an in-depth understanding. We believe that our focus on a benchmark tool and two further tools provided enough insight to substantiate our findings. However, issues related to our sample are both a limitation of this study and an opportunity for future research. Hence, we encourage further larger n-sample analyses to complement our own.

Similarly, questions may find different answers depending on the context under scrutiny, since some aspects of a phenomenon and/or theorizing of the phenomenon may not transfer across contexts. We thus also encourage constructing further cases to be at the heart of the study, as well as those that take into consideration the graduality and feedback loops of writing a patent application, since we discover that longer texts (and presumably those already more aligned with terminology specific to patent writing) can conceivably increase the usefulness of the AI-related solutions in various IPR tools.

References

[1] WIPO (2020). World Intellectual Property Indicators Report: Worldwide Trademark Filing Soars in 2020 Despite Global Pandemic. Available at: https://www.wipo.int/pressroom/en/articles/2021/article_0011.html (accessed January 24, 2022).

[2] Setchi, R., Spasić, I., Morgan, J., Harrison, C., & Corken, R. (2021). Artificial intelligence for patent prior art searching. World Patent Information, 64, 102021.

[3] Stading, T. (2017). Using big data to make intellectual property a strategic weapon. Available at: https://dzone.com/articles/using-Big-Data-to-make-intellectual-property-a-str (accessed February 10, 2021).

[4] Lupu, M. (2017). Information retrieval, machine learning, and natural language processing for intellectual property information. World Patent Information, Vol. 49, pp. A1-A3.

[5] Abbas, A., Zhang, L. & Khan, U.S. (2014). A literature review on the state-of-the-art in patent analysis. World Patent Information, No. 37, pp. 3-13.

[6] Modic, D., Hafner, A., Damij, N., & Zajc, L. C. (2019). Innovations in intellectual property rights management: Their potential benefits and limitations. European Journal of Management and Business Economics.

[7] de Camargo Fiorini, P., Seles, B.M.R.P., Jabbour, C.J.C., Mariano, E.B., & de Sousa Jabbour, A.B.L. (2018). Management theory and big data literature: From a review to a research agenda. International Journal of Information Management, 43, 112-129.

[8] Modic, D. and Damij, N. (2018), Towards Intellectual Property Management: Back Office and Front Office Perspectives, Palgrave Macmillan, Cham.

[9] Bonino, D., Ciaramella, A. & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Information, No. 32, pp. 30-38.

[10] Baudour, F., & van de Kuilen, A. (2015). Evolution of the Patent Information World–Challenges of yesterday, today and tomorrow. World Patent Information, 40, 4-9.

[11] Modic, D., Johnson, A., Lužar, B., Hafner A., Rožac B., Vučkovič M. & Rasmussen E. (2021). IP LodB project: Final report for EPO Academic Research Programme. Available at: https://www.epo.org/learning/materials/academic-researchprogramme/research-project-grants.html (accessed: January 29, 2022).

[12] Lupu, M. (2018). Artificial intelligence and intellectual property. World Patent Information, 53, A1-A3.

[13] Trappey, A. J., Lupu, M., & Stjepandic, J. (2020). Embrace artificial intelligence technologies for advanced analytics and management of intellectual properties. World Patent Information, 61, 101970.

[14] Hurwitz, Morris, H., Sidner, C. L., & Kirsch, D. (2019). Augmented Intelligence : The Business Power of Human-Machine Collaboration. Auerbach Publishers.

[15] EPO (2018). Espacenet: 20 years of free access to patent information in Europe. Available at: https://www.epo.org/news-events/news/2018/20181019.html (accessed February 14, 2022).

[16] Patentfield (2022). https://en.patentfield.com/ (accessed February 14, 2022).

[17] Dennemeyer (2022). https://www.dennemeyer.com/about/ (accessed February 14, 2022).

[18] Integrator (2022). https://patentsearch.intergator.cloud/ (accessed February 14, 2022).

[19] IPRally (2022). https://www.iprally.com/ (accessed February 26, 2022).

[20] Wretblad, L. (2022). AI search tools for patents: How to test & compare them? IPscreener, April, 2020.

[21] EPO (2022). Guidelines for Examination. Available at: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2_2.htm (accessed February 14, 2022).

[22] Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. Journal of the American Statistical Association, 111(515), 988-1003.

[23] Trappey, A. J., Trappey, C. V., Wu, J. L., & Wang, J. W. (2020). Intelligent compilation of patent summaries using machine learning and natural language processing techniques. Advanced Engineering Informatics, 43, 101027.

[24] Chikkamath, R., Endres, M., Bayyapu, L., & Hewel, C. (2020). An empirical study on patent novelty detection: A novel approach using machine learning and natural language processing. In 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1-7). IEEE.

[25] Trappey, A., Trappey, C. V., & Hsieh, A. (2021). An intelligent patent recommender adopting machine learning approach for natural language processing: A case study for smart machinery technology mining. Technological Forecasting and Social Change, 164, 120511.

[26] IAM (2020). How semantic search/automated patent analysis tools can save researchers time. Available at: https://www.iam-media.com/how-semantic-searchautomated-patent-analysis-tools-can-save-researchers-time (February 15, 2022).

[27] Cook T.D. & Campbell, D.T. (1979). Quasi-experimental Design: Design and Analysis Issues for Field Settings. Rand McNally, Skokie.