Centre for Technology Management working paper series
ISSN 2058-8887

No. 2

November 2017

A literature review on the state-of-the-art on intellectual property analytics

doi.org/10.17863/CAM.13928

Leonidas Aristodemou (CTM, University of Cambridge) *
Frank Tietze (CTM, University of Cambridge)

* Please contact the corresponding author for feedback:
la324@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

IfM Centre for Technology Management

Centre for Technology Management working paper series
ISSN 2058-8887

No. 2

November 2017

A literature review on the state-of-the-art on intellectual property analytics

doi.org/10.17863/CAM.13928

Leonidas Aristodemou (CTM, University of Cambridge) *
Frank Tietze (CTM, University of Cambridge)

* Please contact the corresponding author for feedback:
la324@cam.ac.uk

A Literature Review
on the state-of-the-art on
Intellectual Property Analytics (IPA)

Leonidas Aristodemou (la324@cam.ac.uk)
Dr. Frank Tietze (frank.tietze@eng.cam.ac.uk)

Centre for Technology Management (CTM), Institute for Manufacturing (IfM),
Department of Engineering, University of Cambridge

20th January 2018

Abstract

Big data is increasingly available in all areas of manufacturing and operations, which presents an opportunity for better decision making, to introduce the next generation of innovative technologies. Recently, there has been a large development in the field of patent analytics, which describes the science of analysing large amounts of patent information to discover relationships and trends (Trippe, 2003). Similarly, we define Intellectual Property Analytics (IPA) as the data science of analysing large amount of IP information, to discover relationships, trends and patterns. It is a multidisciplinary approach to gain valuable knowledge from data and to drive decision making, rooted in the business context. With the rise of artificial intelligence, there are a number of techniques available for analysing big data; however, while these techniques have widely been applied in other fields to complement management processes, they have hardly been applied in the IP field (Lupu, 2017). In this paper, we contribute in solving the problem of incomplete adoption of IPA within firms, by producing a literature review on the state-of-the-art on IPA methods and techniques.

Research Background

Big data is increasingly available in all areas of manufacturing and operations (OECD, 2016). Data as such presents value for enabling a competitive data-driven economy (EPSRC's Delivery Plan 2016, theme of "Connectedness" (EPSRC, 2016b)), which is at the heart of the Internet of things and Industry 4.0 (EPO, 2016). Increased data availability presents an opportunity for better decision making and strategy development (EPSRC, 2016a), to introduce the next generation of innovative and disruptive technologies and drive business innovation through digital transformation (EPSRC's Area of "Connected Nation" (EPSRC, 2016c)).

Over the last two decades, there has been a large development in the field of patent analytics. Patent analytics describes the science of analysing large amounts of intellectual property information, in relation to other data sources, to discover relationships and trends (Abbas et al., 2014; Baglieri and Cesaroni, 2013; Moehrle et al., 2010; Trippe, 2003). With the digitization of patent data, the world's largest repository of technical information has become accessible for rapidly decreasing costs. Patent data has long been considered the world's largest repository of technological information, and only with its digitization since the BACON project in 1984 (Dintzner and Van Thieleny, 1991) and gradual improvements of analytics over the last decades, patent data has become increasingly accessible to a non-specialist audience (Aristodemou et al., 2017; Raturi et al., 2010). With the rise of artificial intelligence (AI), machine learning (ML), deep learning (DL) and artificial neural networks (ANN), there are a number of methods and techniques for analysing IP data (Abbas et al., 2014; Lupu, 2017; Oldham and Fried, 2016; Trippe, 2015). However, while machine learning and deep learning algorithms have widely been applied in other fields to analyse large amounts of data and complement management processes, they have hardly been applied in the IP field (Lupu, 2017).

In particular, in a study we run, we have used the technology roadmapping approach (Phaal et al., 2012) to explore the future of patent analytics (Aristodemou and Tietze, 2017). We identify 11 priority technologies, such as artificial intelligence and artificial neural networks, that are important to be adopted in the patent analytics domain (Lupu, 2017). We also identify 21 enablers for potential breakthrough progress of the field that cluster around four themes: technology development cycles and methodologies; legislation and standardisation for patent data quality; continuous professional development; and cooperation between industry and academia. From these, we concentrate on understanding analytic techniques further, and in specific, we identify the need of adoption of these computer science techniques techniques, to complement decision processes and provide decision support (Aristodemou et al., 2017; Lupu, 2017). This is very much in line with the propositions by Agrawal et al. (2017), which suggest that AI can improve prediction capabilities, which complements human judgement in making decisions (Ciccatelli, 2017; Simmer, 2001; Stading, 2017a,b). They argue that machine prediction is a complement to human judgement, and can provide a form of decision support (Turban et al., 2005).

In this paper, we contribute in solving the problem of *incomplete adoption of intellectual property analytics within firms (Aristodemou and Tietze, 2017), by producing a literature review on the state-of-the-art on IPA methods and techniques.*

Methodology

The paper aims to summarise the existing work in the field, especially when it comes to the application of machine learning, artificial neural networks and artificial intelligence in the intellectual property domain (Abbas et al., 2014; Lupu, 2017). To carry out the literature review, the narrative literature review approach has been adopted (Cronin et al., 2008), and a research search strategy has been developed (Creswell, 2013; Robson, 2011). The articles on intellectual property analaytics and patent analytics were searched through web searches.

The research follows a three phase process. We use a problem-solving approach (Alvesson and Sandberg, 2011), to identify the need for the literature review. Firstly, we identify the problem opportunity through the following: a study of the future of patent analytics (Aristodemou and Tietze, 2017; Aristodemou et al., 2017), and the problem identification (Agrawal et al., 2017; Cooper, 2007; Lupu, 2017). Secondly, we search the most relevant research work on intellectual property analytics and patent analytics, through Scopus and Google Scholar. We narrow the papers using the search strategy of key terms below:

- (TITLE-ABS-KEY("Patent" OR "patent data" OR "IP" OR "IP data" OR "Intellectual property" OR "intellectual property data") AND ("analysis" OR "analytics" OR "informatics" OR "analytic methods" OR "information retrieval")) AND ("Patent analysis" OR "Patent data" OR "Patinformatics" OR "Patent informatics" OR "Patent analytics" OR "IP analytics" OR "IP analysis" OR "IP informatics" OR "intellectual property analysis" OR "intellectual property analytics" OR "intellectual property informatics" OR "IP information retrieval" OR "Patent information retrieval")

Thirdly, the reference lists of the published research articles are scanned. Then, we select the highest cited papers from these, such as the papers by Abbas et al. (2014); Bonino et al. (2010); Lupu (2013); Moehrle et al. (2010); Trippe (2003), and we build on them by reviewing these and the ones citing. The articles published in more recent years, with a focus on the development of tools, techniques and algorithms for analyzing intellectual property data based on machine learning and deep learning, were selected for presenting a discussion on intellectual property analytics methods and techniques. The purpose of presenting the articles in detail is to provide the readers with the latest research on intellectual property analytics in a unified form.

Intellectual Property Analytics

Intellectual Property Analytics (IPA) is the data science of analysing large amount of intellectual property information, to discover relationships, trends and patterns in the data (Fig.1). It is a multidisciplinary approach that makes use of mathematics, statistics, computer programming, and operations research to gain valuable knowledge from data, to drive decision making rooted in the business context. We make use of this definition, as there is no widely accepted definition of IPA; however, this is very much in line with the definition of *Patinformatics* (Moehrle et al., 2010; Trippe, 2003).

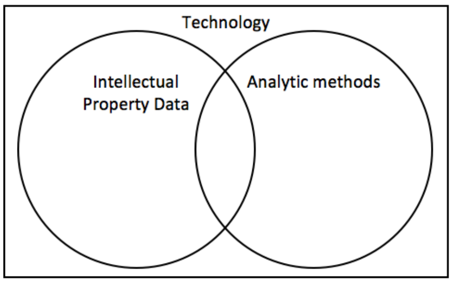
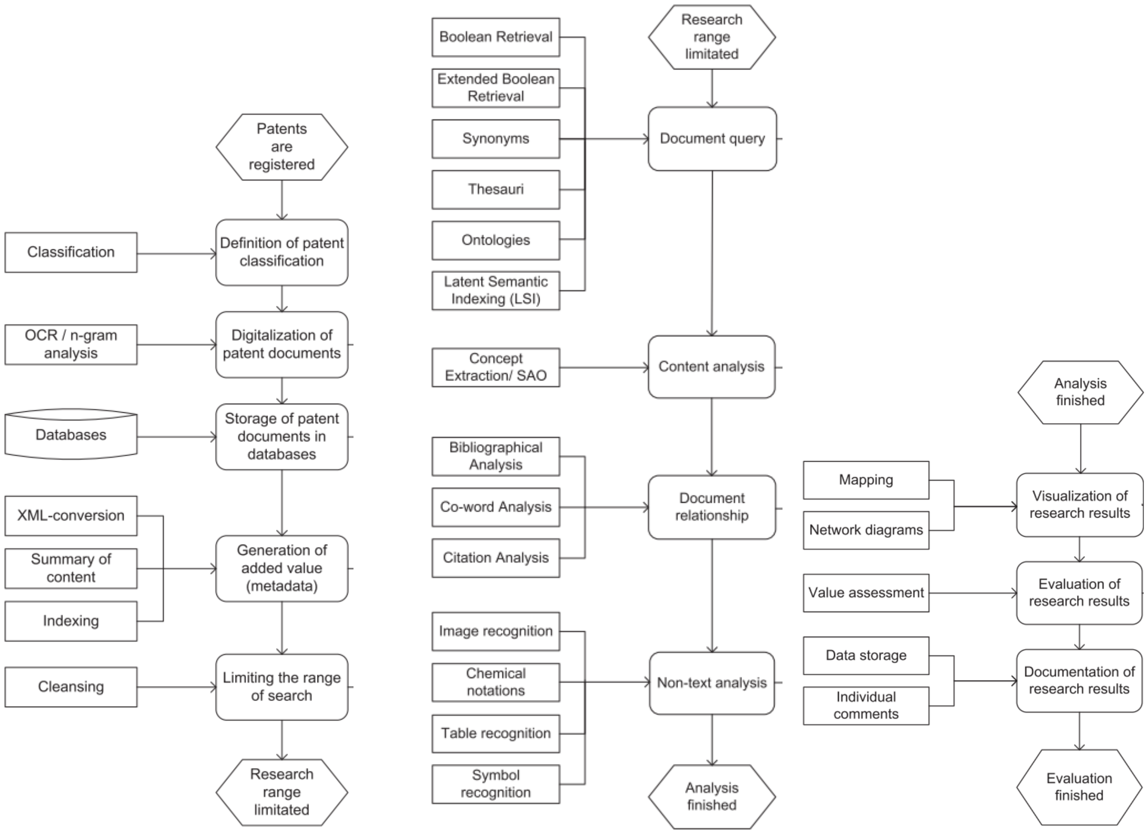


Fig. 1 Positional Venn diagram on the definition of intellectual property analytics

Intellectual Property Analytics Process

The process of intellectual property analytics has been discussed in the literature in different context, mainly around the analysis of patents, due to the nature of the structured and unstructured data they contain within. Mainly, it has been viewed as the process with which one can analyse patent data with different methods to arrive at meaningful conclusions. Trippe (2015) has created a WIPO guide, which identifies and explains a large number of concepts on patent analysis and the methodology on how to run the different types of analysis. With the recent advancements of artificial intelligence, there has been a positive amount of activity around the different methodologies involved that could be applied to intellectual property data (Aristodemou et al., 2017; Lupu, 2017).

Most of the literature makes use of the process as defined by Moehrle et al. (2010) in a business context (Fig.2), and consists of three main stages: the pre-processing stage, the processing stage and the post processing stage. In the pre-processing stage, the data are collected, after information extraction, cleaned and well prepared, with the purpose in providing these data in high quality, correctness and completeness. In the processing stage, the analysis of the data extracted in the pre-processing stage takes place using different methods to classify,



(a) Pre-processing stage (b) Processing stage (c) Post processing stage

Fig. 2 Patent analytics process, source: Moehrle et al. (2010)

cluster, and identify meaningful insights from the information. In the post processing stage, also known as discovered knowledge, the results and information from the processing stage are visualized and evaluated to support strategic decision making.

This is similar to Abbas et al. (2014), who presents a generic patent analysis work-flow, with the distinction that every analysis made has a specific purpose (Fig.4). Raturi et al. (2010) argues that this process is a complementary process to the innovation cycle, and that the analysis of intellectual property data has many application in many fields. Bonino et al. (2010) links the patent life cycle to the patent related information sources and the different tasks along the patent analytics tasks. They argue that a patent analytics process is a purpose-driven process, which consists of search tasks (patent ability, validity, infringement, portfolio survey, technology survey), analysis tasks (micro and macro assessment of business value, technical assessment and technology suggestions), and monitoring tasks (early sign monitoring, technology monitoring, portfolio monitoring, single patent monitoring).

Similarly, Baglieri and Cesaroni (2013) argue that patent analysis is a form of patent intelligence to support decision making. They argue that there are two meanings to the term of patent analysis, the process that considers all of the above, and the actual analysis of the patent data. They use the research by Bonino et al. (2010) to define the three patent analysis tasks, patent searching, patent analysing and patent monitoring, and link value of information from these to the open innovation funnel (Fig.3).

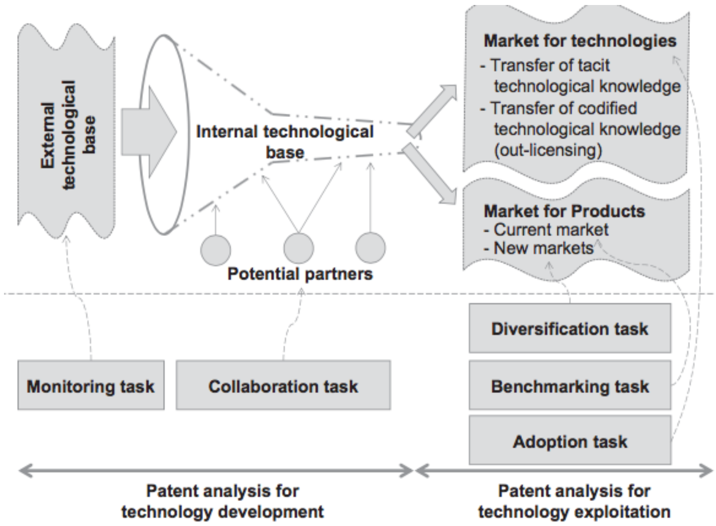


Fig. 3 Patent analysis within the innovation context, source: Baglieri and Cesaroni (2013)

Intellectual Property Analytics Methods

There are several analytic methods that have been used with intellectual property data, and specifically patent data (Abbas et al., 2014; Trippe, 2015). Fig.5 shows the methods or approaches that have been used to analyse IP data.

One form of IP data analysed are patent data, which contain a series of structured and unstructured data.Abbas et al. (2014) provide a comprehensive literature review on the patent analytics techniques, where they distinguish between text mining and visualization approaches and the applicability to structured and unstructured data. Mainly the approaches shown in Fig. 5 are concentrated around text mining techniques, due to the nature of the patent data itself (Abbas et al., 2014; Bonino et al., 2010). However, visualization approaches also exist that translate the patterns and information from the analysis to meaningful insights, to aid decision making (Moehrle et al., 2010), as shown in Fig.2.

NLP is concerned with the interactions between computers and human (natural) languages, and, in particular, the processing of large natural language corpora. It uses computational

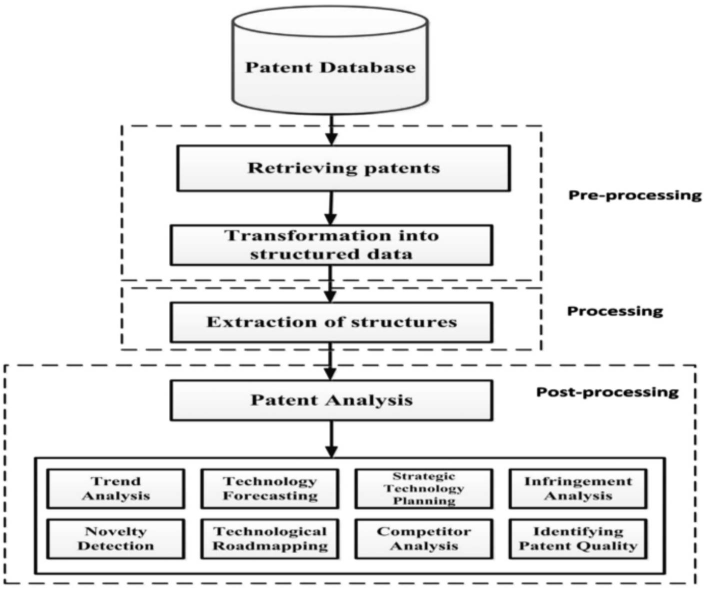


Fig. 4 Generic patent analysis work flow, source: Abbas et al. (2014)

linguistic mechanisms to represent the text found in any document. NLP has been used in hybrid structure with bibliographic coupling and text mining to discover patterns in a patent retrieval and analysis platform (Liu et al., 2011). Yoon and Lim (2013) construct patent maps dynamically by analysing the Subject-Action-Object (SAO) extracted structures, to identify the technological competition trends.Park et al. (2013b) utilize SAO structures, from extracted NLP language structures in patent documents, to compare them again TRIZ evolution trends to evaluate technological evolution. This method can be expanded to SAO based intelligent patent analysis, where the semantic similarities between patents can be visualized in patent maps and patent networks (Park et al., 2013a). Choi et al. (2012) develop a technology tree using NLP to extract SAO structures and perform similarity detection between patents. Park et al. (2013b) detect infringement by using SAO structures to express the relationships that exist between technological components, calculating semantic similarity.

In addition, integrating SAO structures and technology roadmapping approaches improves decision making by utilised the product-function-technology maps. Gerken and Moehrle (2012) utilize NLP through syntactic analysis to create semantic SAO structures to identify novelty among patents. Some authors use the property-function analysis, which uses grammatical analysis to extract properties and functions from patent documents, to create patent networks (Dewulf, 2013; Yoon and Kim, 2012).

Rule based approaches make use of inference and association rules. Shih et al. (2010) propose patent trend change mining, which has a change detection module, where keywords searching within patent documents allows for structure extraction, associating the structures and

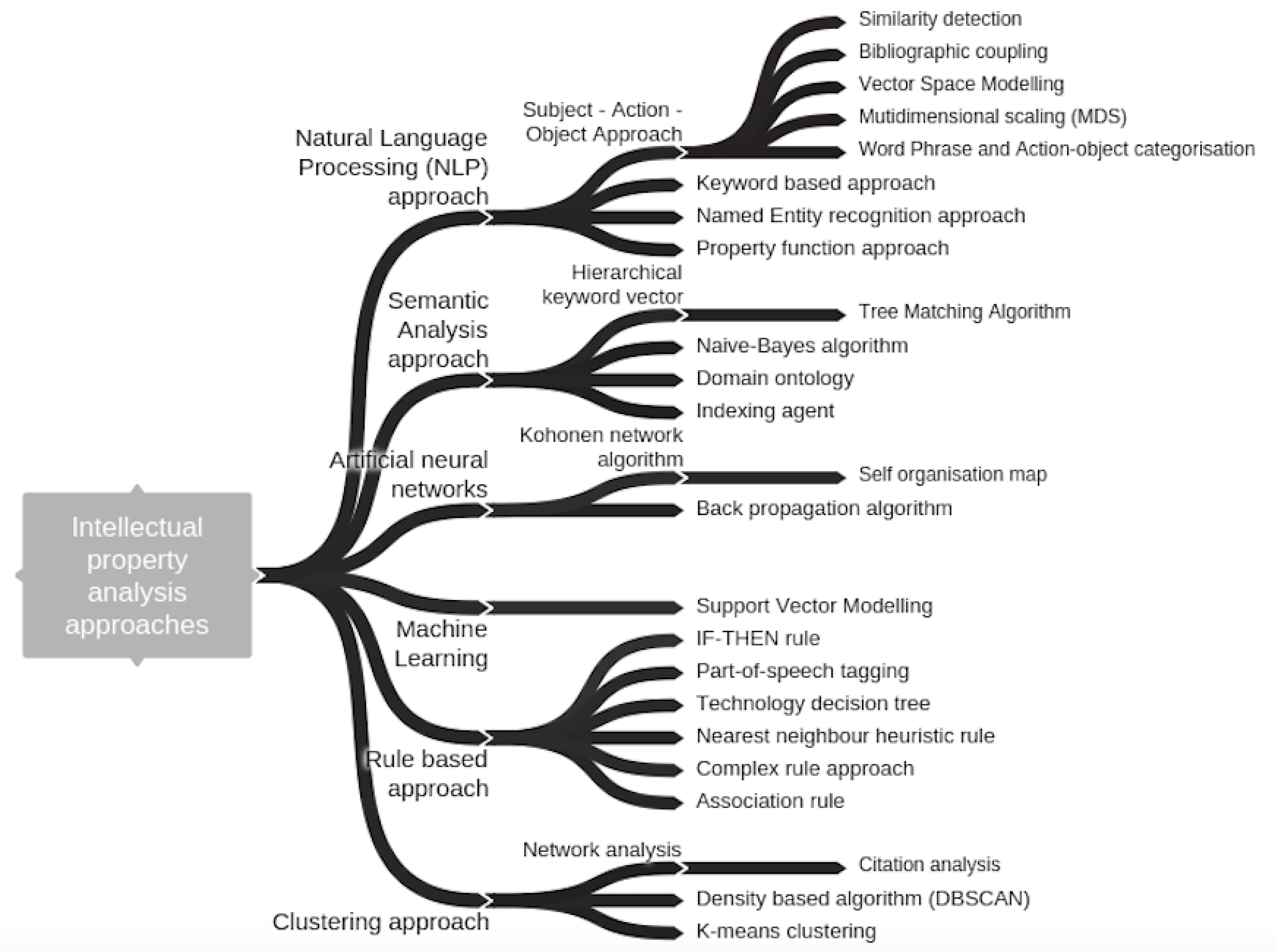


Fig. 5 Intellectual property analytics methods in the literature

extracting trends. Yu and Lo (2009) use the IF-THEN rules in conjunction with the Kohonen learning algorithm and the first nearest neighbour heuristic to plan technology strategies.

Semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings. They rely on domain knowledge and create relationships among domain specific concepts. Lee et al. (2013) proposes the detection of infringement using semantic analysis and dependency relationships. Wang and Cheung (2011) extract key concepts from patent documents to discover and use the abstracts of patent documents collected from the USTPO database to classify patent documents using the Naives-Bayes algorithm. Also, the development of ontologies for multiple domains can serve as an integration platform, to develop the knowledge base by populating the ontology classes (Taduri et al., 2011). The authors expand this to propose a knowledge based framework to facilitate retrieval of patent documents (Taduri et al., 2012).

ANN are computational algorithm that have found extensive utilization in solving many complex real-world problems. Mainly, patents data have been used for classification and technology forecasting. Lamirel et al. (2003) propose a Kohonen self organizing map to perform a viewpoint oriented analysis to classify patent documents. The back-propagation algorithm has also been used to classify patent documents and create a knowledge management technology system (Trappey et al., 2005). In order to process large numbers of explicit knowledge documents, Trappey et al. (2006) develop a document classification and search methodology based on neural network technology that helps companies manage patent documents more effectively. The authors expand this research to include ontology-based artificial neural networks to classify patent documents (Trappey et al., 2012). Lai and Che (2009) propose a valuation model for the monetary legal value of patents, based on the extension neural network incorporated with the factor analysis. Chen et al. (2013) utilizes the back-propagation algorithm to explore the non-linear influences of number of inventors, average age of patents, and age of patenting activities on patent citations and corporate performance in the US pharmaceutical industry. Karanikić et al. (2017) develop and apply the extreme learning machine (ELM) using granted patents as one of the attributes to forecast gross domestic product (GDP) growth rate. In a similar approach, Jokanović et al. (2017) use as input attribute patent applications to estimate economic development.

Conclusion

In this paper, we have reviewed the literature on intellectual property analytics methods and techniques. While there is a large amount of literature on analysing IP data and several analytic methods deployed (Abbas et al., 2014), the application of computer science techniques, machine learning and deep learning, in the IP field, has hardly been applied (Ciccatelli, 2017; Lupu, 2017; Stading, 2017a,b). We contribute by reviewing the literature on the use of machine learning and artificial neural network methods in analysing IP data. In addition, we also contribute to the theoritcal foundations of IPA, by definning the term IPA, and the technologies, techniques and tools that constitute it (Aristodemou et al., 2017); a definition which has hardly been proposed, with efforts mostly concentrating on patent analytics/ informatics (Bonino et al., 2010; Moehrle et al., 2010; Trippe, 2003). This ensures the development of the industrial foundations of IPA, and how can firms use them to increase their knowledge on IP analytics. Further research is required in this field to identify use cases of IPA methods within the innovation process and apply these methods in firms.

References

Abbas, A., Zhang, L., and Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37:3–13.

Agrawal, A., Gans, J., and Goldfarb, A. (2017). How AI Will Change the Way We Make Decisions. *Harvard Business Review*, July:1–7.

Alvesson, M. and Sandberg, J. (2011). Generating Research Questions Through Problematization. *Academy of Management Review*, 36(2):247–271.

Aristodemou, L. and Tietze, F. (2017). Exporling the Future of Patent Analytics. Technical report, Institute for Manufacturing, University of Cambridge, Cambridge, UK.

Aristodemou, L., Tietze, F., Athanassopoulou, N., and Minshall, T. (2017). Exploring the Future of Patent Analytics: A Technology Roadmapping Approach. In *R&D Management Conference 2017, Leuven, Belgium*, pages 1–9.

Baglieri, D. and Cesaroni, F. (2013). Capturing the real value of patent analysis for R&D strategies. *Technology Analysis & Strategic Management*, 25(8):971–986.

Basheer, I. and Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43:3–31.

Bonino, D., Ciaramella, A., and Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1):30–38.

Chen, Y. S., Tien, W. P., Chen, Y. W., Lin, C. C., and Lee, Y. I. (2013). Using artificial neural network (ANN) to explore the influences of number of inventors, average age of patents, and age of patenting activities on patent performance and corporate performance. *Proceedings - 2013 4th World Congress on Software Engineering, WCSE 2013*, pages 136–139.

Choi, S., Park, H., Kang, D., Lee, J. Y., and Kim, K. (2012). An SAO-based text mining approach to building a technology tree for technology planning. *Expert Systems with Applications*, 39(13):11443–11455.

Ciccatelli, A. (2017). The Future of Big Data and Intellectual Property.

Cooper, R. G. (2007). Managing technology development projects. *IEEE engineering management review*, 35(1):67–77.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th editio edition.

Cronin, P., Ryan, F., and Coughlan, M. (2008). Undertaking a literature review : a step-by-step approach. 17(1):38–43.

Dewulf, K. (2013). Sustainable Product Innovation: The importance of the Front End Stage in the Innovation Process. In *Advances in Industrial Design Engineering*, chapter Chapter 7.

Dintzner, J. P. and Van Thieleny, J. (1991). Image handling at the European Patent Office: BACON and first page. *World Patent Information*, 13(3):152–154.

EPO (2016). India and Europe explore the impact of Industry 4.0 on the patent system. Technical report, European Patent Office, Munich, Germany.

EPSRC (2016a). Delivery Plan 2016-2020 Top Ten Messages. Technical report.

EPSRC (2016b). EPSRC Delivery Plan 2016/17-2019/20 Science for a Successful Nation. Technical report.

EPSRC (2016c). Science Strategy.

Gerken, J. M. and Moehrle, M. G. (2012). A new instrument for technology monitoring: Novelty in patents measured by semantic patent analysis.

Goffin, K. and Mitchell, R. (2016). *Innovation Management: Effective Strategy and Implementation*. Palgrave Macmillan.

Ilevbare, I., Dusch, B., and Templeton, P. (2016). A framework and methodology for creating business tools. Technical report.

Jokanović, B., Lalic, B., Milovančević, M., Simeunović, N., and Marković, D. (2017). Economic development evaluation based on science and patents. *Physica A: Statistical Mechanics and its Applications*, 481:141–145.

Karanikić, P., Mladenović, I., Sokolov-Mladenović, S., and Alizamir, M. (2017). Prediction of economic growth by extreme learning approach based on science and technology transfer. *Quality & Quantity*, 51(3):1395–1401.

Lai, Y. H. and Che, H. C. (2009). Modeling patent legal value by Extension Neural Network. *Expert Systems with Applications*, 36(7):10520–10528.

Lamirel, J.-C., Al Shehabi, S., Hoffmann, M., and François, C. (2003). Intelligent patent analysis through the use of a neural network. *Proceedings of the ACL-2003 workshop on Patent corpus processing -*, 20:7–23.

Lee, C., Song, B., and Park, Y. (2013). How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. *Technology Analysis & Strategic Management*, 25(1):23–38.

Liu, S. H., Liao, H. L., Pi, S. M., and Hu, J. W. (2011). Development of a patent retrieval and analysis platform - A hybrid approach. *Expert Systems with Applications*, 38(6):7864–7868.

Lupu, M. (2013). *Patent Retrieval*, volume 7.

Lupu, M. (2017). Information retrieval, machine learning, and Natural Language Processing for intellectual property information. *World Patent Information*, 49:A1–A3.

Martinsuo, M. and Poskela, J. (2011). Use of evaluation criteria and innovation performance in the front end of innovation. *Journal of Product Innovation Management*, 28(6):896–914.

Moehrle, M. G., Walter, L., Bergmann, I., Bobe, S., and Skrzipale, S. (2010). Patinformatics as a business process: A guideline through patent research tasks and tools. *World Patent Information*, 32(4):291–299.

OECD (2016). Enabling the Next Production Revolution: the Future of Manufacturing and Services - Interim Report. Technical Report June, OECD.

Oldham, G. R. and Fried, Y. (2016). Job design research and theory: Past, present and future. *Organizational Behavior and Human Decision Processes*, 136:20–35.

Park, H., Kim, K., Choi, S., and Yoon, J. (2013a). A patent intelligence system for strategic technology planning. *Expert Systems with Applications*, 40(7):2373–2390.

Park, H., Ree, J. J., and Kim, K. (2013b). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert Systems with Applications*, 40(2):736–743.

Phaal, R., Routley, M., Athanassopoulou, N., and Probert, D. (2012). Charting Exploitation Strategies for Emerging Technology. *Research-Technology Management*, 55(2):34–42.

Raturi, M. K., Sahoo, P. K., Mukherjee, S., and Tiwari, A. K. (2010). *Patinformatics – An Emerging Scientific Discipline*.

Robson, C. (2011). Real world research. *Edition. Blackwell Publishing. Malden*, pages 1–608.

Shih, M. J., Liu, D. R., and Hsu, M. L. (2010). Discovering competitive intelligence by mining changes in patent trends. *Expert Systems with Applications*, 37(4):2882–2890.

Simmer, R. (2001). Using intellectual property data for competitive Intelligence. *Vancouver: University of British Columbia*, pages 1–12.

Stading, T. (2017a). The Role of Artificial Intelligence in Intellectual Property.

Stading, T. (2017b). Using Big Data to Make Intellectual Property a Strategic Weapon.

Taduri, S., Lau, G. T., Law, K. H., and Kesan, J. P. (2012). A patent system ontology for facilitating retrieval of patent related information. *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV '12*, page 146.

Taduri, S., Lau, G. T., Law, K. H., Yu, H., and Kesan, J. P. (2011). Developing an ontology for the U.S. patent system. In *Proceedings of the 12th Annual International Digital Government Research Conference on Digital Government Innovation in Challenging Times - dg.o '11*, page 157.

Trappey, A., Lin, S., and Wang, A. (2005). Using neural network categorization method to develop an innovative knowledge management technology for patent document classification. *Proceedings of the 9th International Conference on Computer Supported Cooperative Work in Design*, 2.

Trappey, A. J., Trappey, C. V., Wu, C.-Y., and Lin, C.-W. (2012). A patent quality analysis for innovative technology and product development. *Advanced Engineering Informatics*, 26(1):26–34.

Trappey, A. J. C., Hsu, F. C., Trappey, C. V., and Lin, C. I. (2006). Development of a patent document classification and search platform using a back-propagation network. *Expert Systems with Applications*, 31(4):755–765.

Trippe, A. (2015). Guidelines for Preparing Patent Landscape Reports. Technical report, World Intellectual Property Organisation.

Trippe, A. J. (2003). Patinformatics: Tasks to tools. *World Patent Information*, 25(3):211–221.

Turban, E., Aronson, J. E., and Liang, T.-P. (2005). *Decision Support Systems and Intelligent Systems*. Prentice-Hall, 7th editio edition.

Wang, W. M. and Cheung, C. F. (2011). A Semantic-based Intellectual Property Management System (SIPMS) for supporting patent analysis. *Engineering Applications of Artificial Intelligence*, 24(8):1510–1520.

Yoon, J. and Kim, K. (2012). TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents. *Expert Systems with Applications*, 39(3):2927–2938.

Yoon, J. H. and Lim, S. S. (2013). Potential trade distortion effects of state trading enterprises under the tariff-rate quota scheme. *Economics*, 7:0–20.

Yu, W. D. and Lo, S. S. (2009). Patent analysis-based fuzzy inference system for technological strategy planning. *Automation in Construction*, 18(6):770–776.