Centre for Technology Management working paper series

ISSN 2058-8887

No. 5

November 2017

Exploring the Future of Patent Analytics

A Technology Roadmapping approach

doi:10.17863/CAM.13967

Leonidas Aristodemou (CTM, University of Cambridge) *
Frank Tietze (CIM, University of Cambridge)
Nikoletta Athanassopoulou (IfM ECS)
Tim Minshall (CTM, University of Cambridge)

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

This paper has been accepted to the R&D Management Conference in Leuven, Belgium (2017)

UNIVERSITY OF CAMBRIDGE

IfM Centre for Technology Management

Exploring the Future of Patent Analytics: A Technology Roadmapping approach

Leonidas Aristodemou[1], Frank Tietze[2], Nikoletta Athanassopoulou[3] and Tim Minshall[4]

[1, 2, 4] Centre for Technology Management (CTM), Institute for Manufacturing (IfM), Department of Engineering, University of Cambridge –[1]la324@cam.ac.uk, [2]frank.tietze@eng.cam.ac.uk, [4]thwm100@cam.ac.uk
[3] Institute for Manufacturing Education and Consultancy Services (IfM ECS), Cambridge – [3]naa14@cam.ac.uk

In a connected world, where successful technological development depends increasingly on collaboration between different partners, effectively utilizing patent data analytics has huge, yet unexploited potential. Given suitable analytics solutions, this high-quality data can be used for decision making on a strategic level in all kind of organizations. The paper contributes to expanding the field of patent analytics for more effective exploitation of the worldwide largest repository of technological information. We do this by developing a domain level technology roadmap following a three-stage technology roadmapping and problem-solving approach. Firstly, from desk research and expert discussions, we identified five main problem themes in the patent analytics field (patent data, database interconnectedness, data analysis, data visualisation, and patent quality). Secondly, we verified and expanded these problem themes through an online survey with 70 respondents. Thirdly, we explored the future direction of the field through a workshop, with inputs from the preparatory stages above, with 28 leading experts. The approach served to develop a technology roadmap to facilitate collaboration and coordinated action within the patent analytics community. We identify thirteen priority technologies, such as artificial intelligence and neural networks, fifteen complementary technologies, such as block chain, and five new technologies, such as technologies for linking databases, to be adopted in the field and are important in overcoming the problems. We also identify twenty-one 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. Key next actions include the generation of use cases for different users, the standardization and harmonization of patent ontologies and the implementation of reporting standards.

Key words: Patent Analytics, technology roadmapping, patent data, visualisation, data quality, database interconnectedness

1. Introduction

In a connected world, where successful technological development depends increasingly on collaboration of different partners (Tietze & Lauritzen 2016), effectively utilizing patent data has huge, yet only partially exploited potential (Lee et al. 2011). Patent data has long been considered the world's largest repository of technological information. With the digitization of patent data since the BACON project (Dintzner & Van Thieleny 1991) and gradual improvements of analytics over the last decades, patent data has become increasingly accessible to a non-specialist audience. While the quality of patent data has increased substantially over the last decades and gradually better software tools for analysing the data are being developed, still today large potential of utilizing patent data remains undeveloped (Lupu et al. 2011; Tietze & Probert 2015).

Trippe (2003) defined patent informatics as the science of analysing large amounts of patent information to discover relationships and trends. Patent analytics is part of this field. Abbas et al. (2014) provides an overview of a set of tools and approaches, with key features and weaknesses, for analysing patent documents for the purpose of forecasting future technological trends, conducting strategic technology planning and identifying technological hotspots and patent vacuums. Moehrle et al. (2010) apply a business process model, which maps the main tasks in patent analytics to the available tools and techniques.

** Paper presented at the R&D Conference 2017, Leuven, Belgium*

Given the lack of a specific definition for patent analytics in this research, we propose to define it as the science of analysing large amounts of patent information to derive meaningful insights to support decision making, which constitutes of the deployment of different technologies, techniques and approaches.

The recent advancements of data technologies, such as machine learning, deep learning and artificial intelligence, seem to potentially deliver breakthrough progress to enable completely new use cases for patent data with substantial economic benefits. While these technologies already impact several areas, their impact on patent analytics remains to be understood. These technologies which are either well established in other fields, or emerging, have been used in a limited way to explore and exploit the patent data repository. At the same time, in patent analytics, there exists a large number of problems that remain unsolved today (Lupu et al. 2011; Raturi et al. 2010; Trippe 2003).

Involving numerous key stakeholders, such as technical experts, lead users of patent analytics solutions, patent specialists and decision makers, this study contributes results from a technology roadmapping exercise for the future of patent analytics (similar to Ferrari et al. 2014). The roadmap contributes to identifying breakthroughs and further enhancing academic and industrial development of the patent analytics field for more effective exploitation of the worldwide largest repository of technological information.

2. Methodology

This study deployed a technology roadmapping approach (Gerdsri 2013; Jeong et al. 2015; Phaal 2015; Phaal 2004; Phaal et al. 2001; Probert et al. 2003) consisting of three stages, with the first two preparatory stages providing inputs to the third stage, a workshop run in March 2017, as a core element of this approach for developing a patent analytics domain roadmap. The research is guided by principles commonly used to establish the quality of a research: validity and reliability (Bryman 2012; Creswell 2013; Flick 2009), increasing quality and robustness of the research design. Figure 1 illustrates the research process.

Firstly, in the identification stage, we conducted desk and literature reviews (Creswell 2013; Cronin et al. 2008) as well as expert consultations, to identify problem themes and technologies that could have a substantial impact in the patent analytics domain.

Secondly, in the verification stage we reached out to relevant stakeholder communities using an online survey (Bryman 2012; Flick 2009). 70 respondents provided input to further identify, prioritise and eliminate technologies and problem themes from stage 1.

In the third exploration stage we ran a workshop with 28 carefully selected experts covering a variety of stakeholder perspectives both from academia and industry. The workshop had three main phases; in the first phase participants followed a problem-solving approach to develop five mini-technology roadmaps in groups. Secondly, they extracted information on technologies, from the technology layer of the mini-technology roadmap, which can enable the field. In the third phase, the technology roadmap was synthesised by combining the key elements from initial mini-technology roadmaps (phase 1 and 2) created for each of the five patent analytics domain problems, the information from stage 1 and 2 of the research design, and the examination of the three layers (problem milestone, technology and enablers).

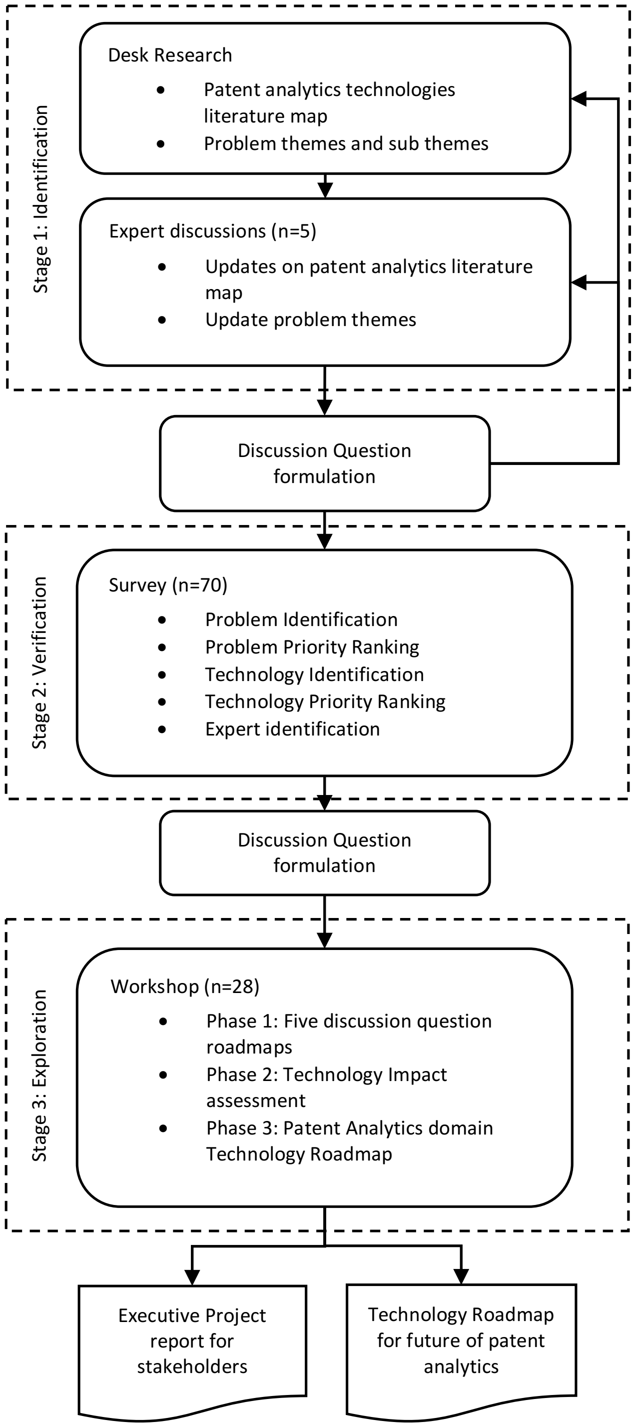


Figure 1. Research design

3. Results and Discussion

The developed technology roadmap provides a glimpse into the future of patent analytics, identifying key milestones/ breakthroughs and enabling factors for fundamental problems in the field. The technology roadmap aims to contribute to coordinating further activities in the field of patent analytics by helping research and the industry to explore potential breakthroughs and by increasing collaborations.

3.1 Patent analytics domain problems

Over the last decade, there has been a large push to improve areas of the patent analytic field and expand the capabilities of the field (Baudour & van de Kuilen 2015; Bonino et al. 2010). However, even today, there is a very large number of problematic areas (Lupu et al. 2011). Overcoming these issues should enable to improve and expand the boundaries of the patent field. The problem themes have been identified through desk research and expert discussions in stage 1, verified through the survey in stage 2, and formulated into discussion questions used in the stage 3.

3.1.1 Problem theme A - Patent data: This concentrates around the patent data itself. It tackles issues during the pre-processing stage of patent analytics (Bonino et al. 2010; Moehrle et al. 2010) in relation to data management, data preparation, data cleaning and data quality. Firstly, a sub-theme emerged with the existence of several patent family un-harmonised definition (Martinez 2010; Martínez 2011). Secondly, there are no common standards for data preparation or a current best approach. In addition, often the data are inconsistent and not accurate (Baudour & van de Kuilen 2015), and there is no global standard for patent numbering across different patent offices. Furthermore, patent taxonomies need improvement and ontologies are largely absent. Discussion question A formulated is: *How can patent data be improved?*

3.1.2 Problem B – Patent database interconnectedness: This focuses on database interconnectedness, and tackles the issue, where different types of data, such as intellectual property data, financial data, litigation data, market data etc., can be combined for more comprehensive analysis. Currently, patent data are linked primarily to legal data. Discussion question B formulated is***:*** *How to enable interconnectedness of patent databases with other data sources?*

3.1.3 Problem C – Patent data analysis: This theme concentrates on data analysis effectiveness (Brügmann et al. 2015; Gassmann et al. 2012; Lupu et al. 2011), and tackles the problem, of understanding and deciding what type of analysis is more suitable for a certain dataset, and why. Several sub-themes have emerged for this problem, such as the type of analytic techniques available (Abbas et al. 2014; Raturi et al. 2010), how to deploy them, how to measure their effectiveness, and which of these are more suitable for which decisions. In addition, sub-themes included the building of a "corporate memory" of past analysis for future users to start utilizing deep learning and machine learning capabilities, saving timing and resourcing, and changing the analytic perspective to a prospective/ adaptive framework, to enable a future-oriented approach of patent analytics. Discussion question C is formulated as: *How to make better use of the valuable information contained in the patent data?*

3.1.4 Problem D – Patent information visualisation: This theme focuses on the problem of information visualisation and its effectiveness (Masiakowski & Wang 2013), and tackles issues, where one needs to decide and understand visualizations arising from patent analysis. Sub-themes concentrate on the types of visualisations available, how these can be improved, and their effectiveness for different decisions. Discussion question D is: *How to visualize results from patent analysis more effectively for better decision making?*

3.15 Problem E – Patent quality: This concentrates on the problem of patent quality (Squicciarini et al. 2013; Trappey et al. 2012) and invalidity. Sub-themes include the definition of patent quality, how it is measured, how can we make judgements about it, and how can we identify invalid patents. Discussion question E is: *How to determine patent quality and patent invalidity?*

3.2 Technology as a key enabling factor

Technology is regarded as a key enabling factor to help resolve many of the key problem themes in the patent analytics domain. The technology layer from each mini-technology roadmap has been carefully analysed to extract and identify future technology developments. In addition, the current state (literature map) of technologies and techniques in the patent analytics domain, was also assessed (figure 2).

Priority technologies for the patent analytics domain (figure 3), from the technology roadmapping, have emerged as priorities across different problem groups from a scoring expert exercise during the workshop, in priority order. The matrix can also be read from the problem perspective, and what is the collective and individual level of impact for each technology on the specific problem.

Some "additional" technologies (shown with bold), not included in figure 3 and are essential to be developed and used in this domain, are identified as important. These technologies such as the ones for linking databases, are shown to have the highest impact in the domain, together with artificial intelligence technology, incorporating artificial neural network analysis, deep learning analytics, and machine learning. These are followed closely by classification algorithms and concordance with data system, NLP approaches and open source. New technologies that allow linking and combining databases can potentially have a substantial impact in progressing the field. From the priority and

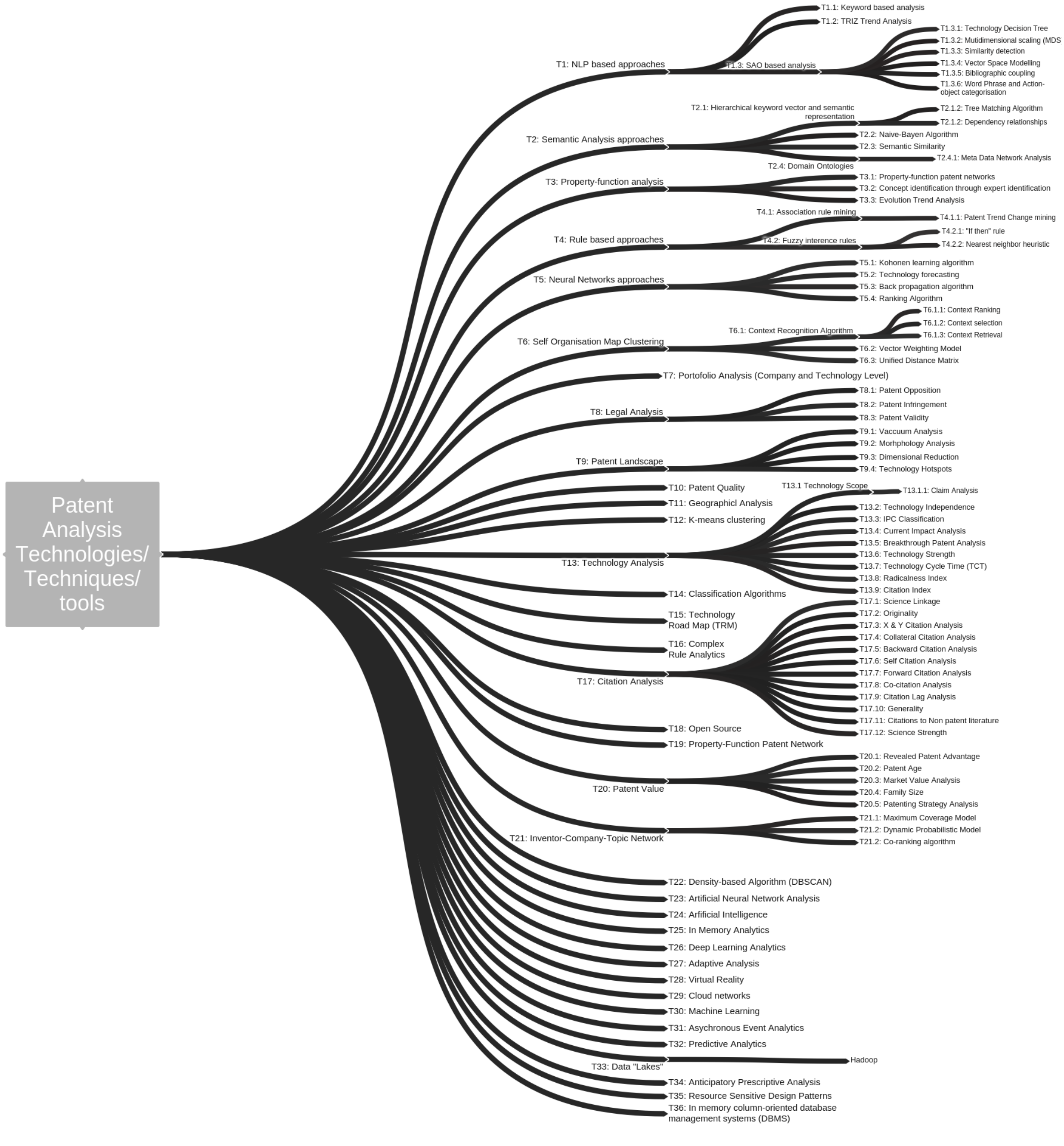


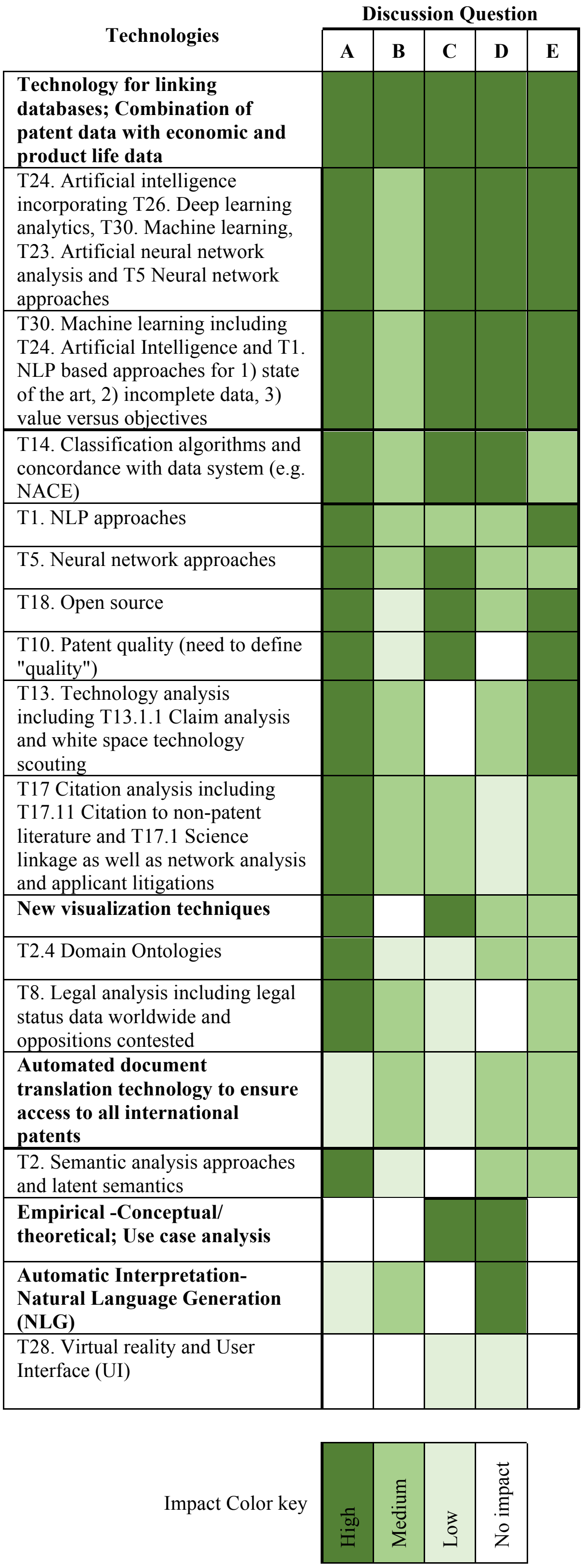
Figure 2. Patent Analytic technologies, techniques, and tools

new technologies identified, the majority complements DQ A, followed by E, D and C. It is also clear that there is a gap in the technology for database interconnectedness and thus the need for it.

During the workshop, several complementary technologies were identified that may potentially play an important enabling role in accelerating the adoption and/or integration of the priority technologies into the patent analytics domain. These have been clustered into three main categories (table 1).

3.3 Patent analytics domain technology roadmap

The patent analytics domain technology roadmap, arising from the three-stage roadmapping and problem-solving approach, has a vision of a fully adaptive, interactive, intelligent, personalized system with searching, analysis, visualizations and interpretation. The time frame that this is envisioned is about 15+ years. Figure 4 shows the overall technology roadmap with three clearly articulated layers: the problem-solving milestones layer, technology developments layer and key enablers layer that are required over time to progress the field. Different pathways are highlighted for resolving the most pressing problems in the domain, for instance, either through the further development of AI technologies and their integration with neural networks and related citation protocols of technologies, or by facilitating the implementation of key enablers (table 2) necessary for the resolution of the issues in this area.



Notes: Dark color indicates high impact, whereas blank indicates no impact. Technologies in bold are new technologies identified, whereas all the others are priority technologies. Technology numbers (T1 etc.) refer to the technology numbering shown in Figure 2.

Figure 3. Impact of priority and new technologies on the patent analytics problems

Table 1. Complementary technologies (accelerating the adoption of priority technologies)

Technology Categories	Complementary technologies
Tools and Methods	• Block chain • Automated effectiveness evaluation • Automated patent document translation • Automated drafting of patent applications, taking into account analytics while drafting • Quantum Computing • Tools to facilitate NPL search • Technology forecasting • Computer aided design
Databases	• Building concordance between existing taxonomies • OECD database of standardized names • OROPO ownership database • Better open source database software • Technologies for loading databases
Integration of existing technologies	• Integration of machine learning with other techniques • Inexpensive cloud computing and enabling platforms to harness cloud analysis

The four main problem-solving milestones for the patent analytics domain are: firstly, automating patent classification; secondly, transparent and consistent clarification and clustering of information; thirdly, having cleaner, standardized and interlinked patent data with other data; and fourthly, the creation of appropriate use cases for user groups, for understanding decision needs. The required technology developments, are further integration and validation artificial intelligence, neural networks and citation protocols. This is complemented by the alignment of different databases to enable compatibility of data and visualizations.

The field can benefit from more emphasis being placed in key enablers, especially on cooperation of different organizations, such as WIPO, EPO, OECD. Also, incentives to applicants to write clearer abstracts that enable easier classification of patent applications can act as an enabler, followed by a standardized (harmonized) legislation. In terms of the technology cycle enablers, these can be identified as funding resources, open source development of tools and build of community, and infrastructure development for security to protect the patent data with the interconnected databases. In terms of legislation, enablers such as legislation for cooperation between intellectual property offices and internal standards are important to harmonize and converge patent data. The main gaps from these are the lack of appropriate data tagging, ontologies or taxonomies, and that the data are not well organized.

Three key insights from the process of synthesizing the patent domain technology roadmap are derived. Firstly, use cases can play an important role in progressing the field of patent analytics, as it can help to link user group needs to technology developments and decision making. Secondly, the most required technologies are already known, and some of them are in use by the patent analyst experts. A requirement to aid and guide the technology adoption, is to create a more specialized training for both developers and end users in key technologies, adopting a data science profile. The aim of this is that these analytic technologies stop being regarded as "black box" solutions and can be customized for specific needs. Finally, the biggest impact in the domain can be achieved only by cooperation of different organizations and standardized legislation, activities which normally take much longer time to initiate and implement.

Table 2. Patent analytics domain enablers

Theme	Enablers
Technology development cycle/ Methodologies	• Market (users) demand - industry, academic, Technology Transfer Office, policy and decision makers • Funders - resources - staff, premises, • Technology transfer - academia and commercialise • Producers of the technology - academics, contract research, commercial vendors • Clarify choice and definition of families • Open source tools and community. • Open data pat-information communities • Cooperation between academics and the private sector • Infrastructure to protect the linked data security standards
Legislation	• Legislation cooperation between IPOs • International Standards (e.g. WIPO) IP5 and legal changes for patents
Training/ Continuous professional development	• Changes evolution of patent scientist/analyst • Training, awareness certification. • Transparency (no black box tools) • Training of developers and end users in patent analytics and visualisation • Training for QPIP/ISBQPIP PDG
Cooperation	• "5-10" collaborations between IP tool suppliers and external visualization experts and data sciences • Increased cooperation between WIPO, EPO, USPTO, OECD • Incentives to write informative abstracts, require applicants to classify the application • Organisation(s) to run the integrated data e.g. patent offices, private intermediary firms • Concordance, collaboration with industry

4. Conclusions

The paper contributes to expanding the field of patent analytics for more effective exploitation of the worldwide largest repository of technological information to enable new use cases supporting better decision making and partnerships of R&D pursuing organizations. This is achieved by developing a public roadmap to facilitate collaboration and coordinated action of actors in the patent analytics community to further develop the capabilities for analysing patent data.

Using a technology roadmapping problem-solving approach, the research design involves 100+ experts from academia and industry in the patent analytics domain, to develop a patent analytics domain roadmap (figure 4), where a number of observations can be made.

Firstly, we identify eighteen technology families (figure 3) or clusters, which are important in overcoming the original five problem themes (section 3.1). In the top three identified, are a combination of technologies such as T24. Artificial intelligence, which incorporate T.26 Deep learning, T30. Machine learning, T23. Artificial neural network analysis, T5. Neural network approaches, T1. NLP, and T14. Classification algorithms. Secondly, out of these eighteen technology families, we identify five new technologies (figure 3 in bold), which can complement and aid this process, that are: empirical use case analysis (conceptual/ theoretical), new visualization techniques, automatic interpretation (natural language generation), technology for linking databases, automated document translation technology for international patents and combination of patent data with economic and product life data (figure 4)

Thirdly, twenty-one enablers are identified (table 2). These play an important and equal role in resolving the five problem themes in the domain, and are classified under the themes of technology development cycle/methodologies, legislation, training/continuous professional development and cooperation.

Four key messages are derived from this work. Firstly, better data quality is important, and there is an urgent need for more structured and cleaner standardized data. This can include a standardized definition of patent families. In addition, open data increases data quality and data repair. Secondly, the identification of different user group needs is important in extracting the information needs and use cases. Thirdly, there is need for training in using different technologies, and of transparency and traceability of using different analytic technologies, techniques and tools. Finally, legislation and standardisation can aide transparency and adoption of technologies in the patent domain.

The research design includes limitations. Firstly, the method used was not in large scale. One workshop was conducted to generate the technology roadmap, whereas an iterative process would be more suitable. Secondly, there were not any people involved to reflect on the technology roadmap from outside the patent domain field, which made the process to lack critical reflections.

The next key actions from the technology roadmap, should be to generate use cases for different users and/or user groups (these could possibly by created by technology vendors) and the standardization and harmonization of patent ontologies by WIPO and member states. The final action would be to implement standards of reporting that are disclosed. To overcome the limitations above, another round of feedback will be generated, capturing next actions, and action owners.

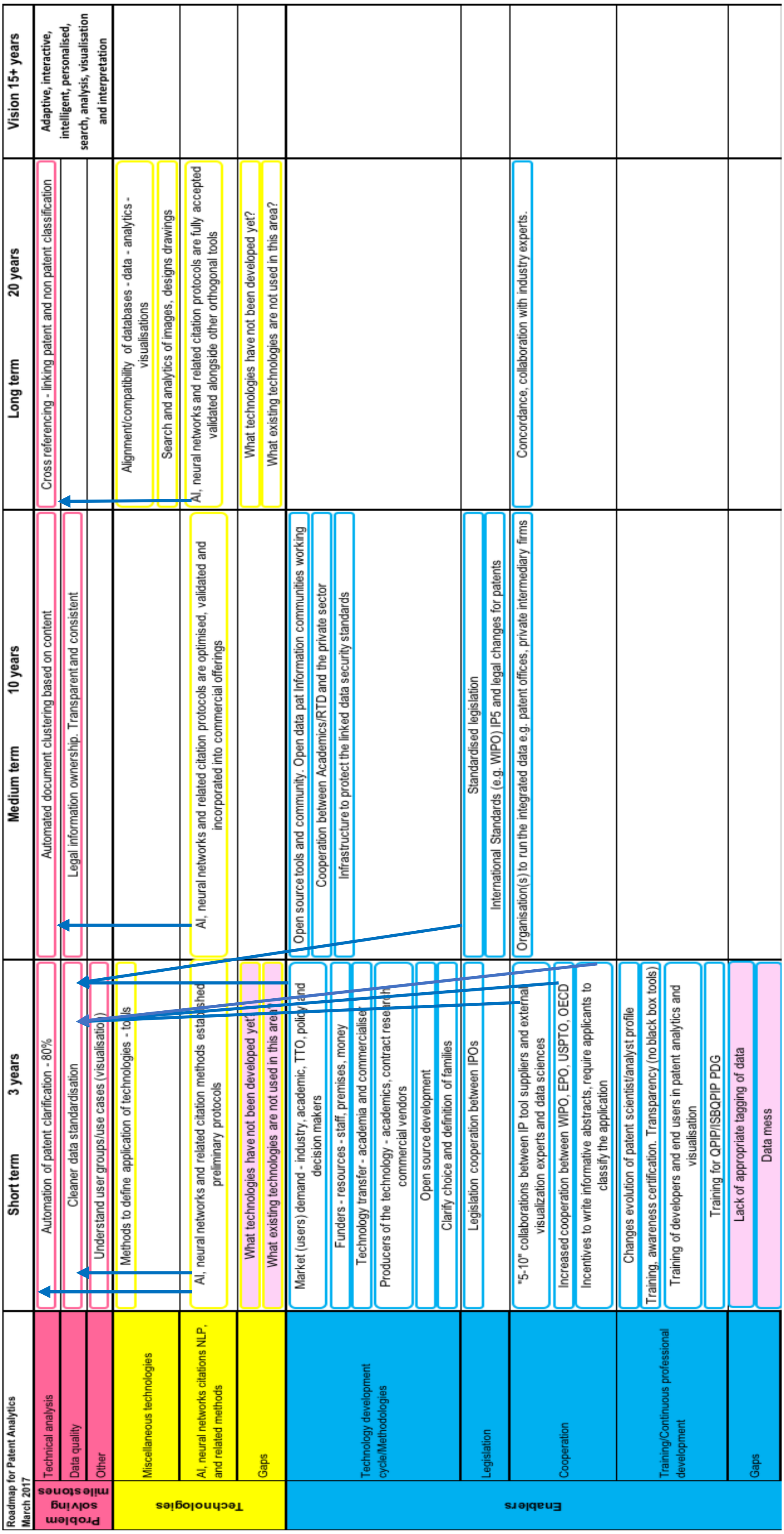


Figure 4. Patent Analytics domain technology roadmap

5. Acknowledgements

We would like to thank the UK Engineering Physical Science Research Council (EPSRC) for funding this research through the Cambridge Big Data initiative with Aistemos Ltd (Nigel Swycher and Steve Harris) supporting this research. We would also like to thank IfM ECS for running the workshop and all experts and participants who contributed to this research

6. References

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

Baudour, F. & van de Kuilen, A., 2015. Evolution of the Patent Information World - Challenges of yesterday, today and tomorrow. *World Patent Information*, 40, pp.4–9.

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*, 32(1), pp.30–38.

Brügmann, S. et al., 2015. Towards content-oriented patent document processing: Intelligent patent analysis and summarization. *World Patent Information*, 40, pp.30–42.

Bryman, A., 2012. *Social Research Methods*, Oxford University Press.

Creswell, J.W., 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* 4th Editio.,

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

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

Ferrari, A.C. et al., 2014. Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems. *Nanoscale*, 7(11), pp.4598–4810.

Flick, U., 2009. An introduction to qualitative research. *Sage*, 4th, p.529.

Gassmann, O. et al., 2012. The role of IT for managing intellectual property - An empirical analysis. *World Patent Information*, 34(3), pp.216–223.

Gerdsri, N., 2013. Technology Roadmapping for Strategy and Innovation. *Journal of Chemical Information and Modeling*, 53(9), pp.1689–1699.

Jeong, Y. et al., 2015. Development of a patent roadmap through the Generative Topographic Mapping and Bass diffusion model. *Journal of Engineering and Technology Management - JET-M*, 38, pp.53–70.

Lee, C., Jeon, J. & Park, Y., 2011. Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach. *Technological Forecasting and Social Change*, 78(4), pp.690–702. Available at: http://dx.doi.org/10.1016/j.techfore.2010.11.010.

Lupu, M. et al., 2011. *Current Challenges in Patent Information Retrieval*, Springer-Verlag Berlin Heidelberg.

Martinez, C., 2010. Insight into different types of Patent Families.

Martínez, C., 2011. Patent families: When do different definitions really matter? *Scientometrics*, 86(1), pp.39–63.

Masiakowski, P. & Wang, S., 2013. Integration of software tools in patent analysis. *World Patent Information*, 35(2), pp.97–104.

Moehrle, M.G. et al., 2010. Patinformatics as a business process: A guideline through patent research tasks and tools. *World Patent Information*, 32(4), pp.291–299.

Phaal, R., 2015. Roadmapping for strategy and innovation. *Centre for Technology Management*, 47(March), pp.1–7.

Phaal, R., 2004. Technology roadmapping - A planning framework for evolution and revolution. *Technological Forecasting and Social Change*, 71(1–2), pp.5–26.

Phaal, R., Farrukh, C.J.P. & Probert, D.R., 2001. Technology Roadmapping: linking technology resources to business objectives. *International Journal of Technology Management*, 26(1), p.2.

Probert, D.R., Farrukh, C.J.P. & Phaal, R., 2003. Technology roadmapping Developing a practical. , 217, pp.1183–1195.

Raturi, M.K. et al., 2010. *Patinformatics – An Emerging Scientific Discipline*,

Squicciarini, M., Dernis, H. & Criscuolo, C., 2013. Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Science, Technology and Industry Working Papers*, (3), p.70.

Tietze, F. & Lauritzen, G.D., 2016. IP challenges in multi-partner collaboration. *Institute for Collaborative working: The Partner*, (May), pp.90–91.

Tietze, F. & Probert, D., 2015. Patent Informatics for collaboration management. *The Partner*, May, pp.109–11.

Trappey, A.J.C. et al., 2012. A patent quality analysis for innovative technology and product development. *Advanced Engineering Informatics*, 26(1), pp.26–34.

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