

Innovations in intellectual property rights management

Their potential benefits and limitations

Innovations
in IPRM

189

Dolores Modic

Nord University Business School, Bodø, Norway

Ana Hafner

Faculty of Information Studies in Novo Mesto, Novo Mesto, Slovenia

Nadja Damij

Newcastle Business School, Northumbria University, Newcastle upon Tyne, UK, and

Luka Cehovin Zajc

Faculty of Computer and Information Science, Univerza v Ljubljani, Ljubljana, Slovenia

Received 13 December 2018
Accepted 8 January 2019

Abstract

Purpose – The purpose of this paper is to evaluate innovations in intellectual property rights (IPR) databases, techniques and software tools, with an emphasis on selected new developments and their contribution towards achieving advantages for IPR management (IPRM) and wider social benefits. Several industry buzzwords are addressed, such as IPR-linked open data (IPR LOD) databases, blockchain and IPR-related techniques, acknowledged for their contribution in moving towards artificial intelligence (AI) in IPRM.

Design/methodology/approach – The evaluation, following an original framework developed by the authors, is based on a literature review, web analysis and interviews carried out with some of the top experts from IPR-savvy multinational companies.

Findings – The paper presents the patent databases landscape, classifying patent offices according to the format of data provided and depicting the state-of-art in the IPR LOD. An examination of existing IPR tools shows that they are not yet fully developed, with limited usability for IPRM. After reviewing the techniques, it is clear that the current state-of-the-art is insufficient to fully address AI in IPR. Uses of blockchain in IPR show that they are yet to be fully exploited on a larger scale.

Originality/value – A critical analysis of IPR tools, techniques and blockchain allows for the state-of-art to be assessed, and for their current and potential value with regard to the development of the economy and wider society to be considered. The paper also provides a novel classification of patent offices and an original IPR-linked open data landscape.

Keywords Artificial intelligence, Software tools, Big data, Social benefits, Intellectual property rights management, Linked open databases

Paper type Research paper

1. Introduction

The world today seems to be characterised by the effects of information and communication technology (ICT) on every aspect of our lives, including that of intellectual property rights (IPR) (Modic, 2017). Freeman and Louca (2002, p. 301) wrote that “even those who have disputed the revolutionary character of earlier waves of technological change, have little



© Dolores Modic, Ana Hafner, Nadja Damij and Luka Cehovin Zajc. Published in *European Journal of Management and Business Economics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

European Journal of Management
and Business Economics
Vol. 28 No. 2, 2019
pp. 189-203
Emerald Publishing Limited
2444-8494
DOI 10.1108/EJMBE-12-2018-0139

difficulty accepting that a vast technological revolution is now taking place". The surge of intellectual property is mirrored in rising IPR numbers with dissemination efforts dependent upon the available data, channels and skills. IPR data are big data, as its characteristics are high volume, high variety and high velocity of changes (Ciccatelli, 2017). Consequently, merging different types of IPR data from various databases presents a challenge (Stading, 2017; Abbas *et al.*, 2014).

When huge amounts of IPR data are connected, a new ecosystem for (open) innovation emerges. It is important to examine the best available IPR data sources, and their merge-readiness, in order to extract the maximum value. Furthermore, it is important to ensure the availability of appropriate IPR techniques and tools if we are to harness the benefits for IPR management (IPRM) and the wider social benefits of this new open IPR landscape and move towards knowledge creation assisted by artificial intelligence (AI). Examining the latest trends in technological solutions and their potential is the foci of our paper.

Figure 1 presents two dimensions: the benefits and the technology. Looking at the technology dimension, all three layers represent issues companies face. IPR software tools and techniques should better respond to business requirements, and as such support changes in databases when dealing with IPR big data, such as the implementation of blockchain technology and linked open databases.

The benefits dimension is also facing several gaps. One refers to the identification of the accessibility of employees' knowledge both in SMEs and IPR-savvy companies. In addition, there are inefficiencies when trying to transform tacit to explicit knowledge in order to further knowledge creation.

Both the technology and benefits dimensions are linked, as the technology aims to, largely unsuccessfully at the present time, to support the requirements of the IPRM, thus increasing the IPRM-derived benefits. These would consequently be translated, especially through the use of blockchain technology and IPR-linked open data (IPR LOD) databases, into increased social benefits. The question as to when, and if, the technology will become smart enough to create IPR software tools and techniques that will function in an intelligent manner remains open to debate, as we are faced with increasing transparency and inherently imbued trust.

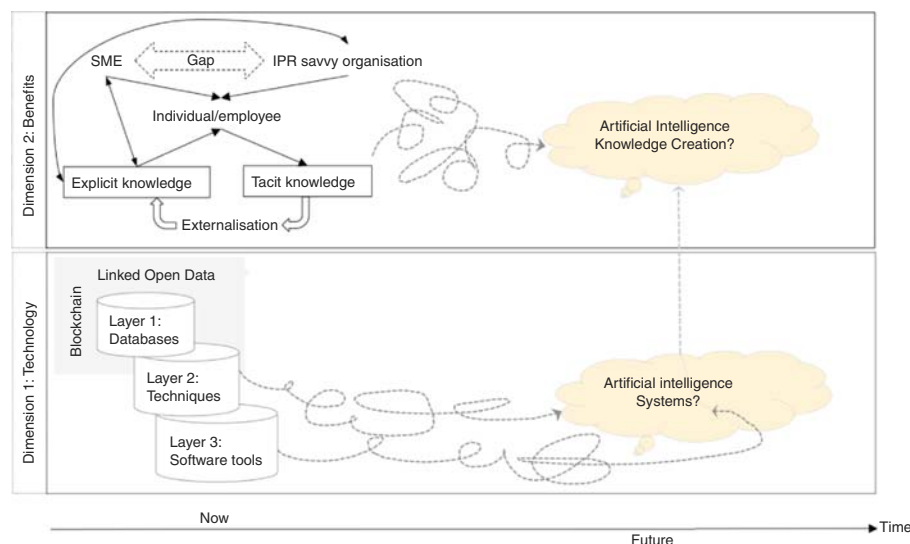


Figure 1.
Technology and
benefits in IPR
landscapes

If AI systems provide the best possible answer to every IPR-related business requirement, in order to maximise business potential, does this mean that the employees' knowledge creation will become obsolete and AI systems will be able to effectively create new knowledge?

The paper offers a review and an interview-based analysis of the requirements and expectations of some of the top IPR experts from IPR-savvy multinationals, as well as a consideration of the potential social benefits. This is followed by a web-based analysis and data retrieval-based evaluation of the current evolution of IPR (LOD) databases. Furthermore, the practical solutions available have been critically evaluated with respect to IPR databases and IPR software tools. The results of the analysis of the state-of-the-art with the available techniques are presented. Finally, a debate-style conclusion is presented.

2. Background and prepositions

This paper investigates IPRM and IPR social benefits by answering what are the potential social and IPRM benefits of adopting new ICT solutions when dealing with IPR, and especially what is the current state of all three technological layers? The research is based on the following prepositions constructed following the literature review and the evidence-based approach.

The first preposition is linked to the availability of data and its connectivity, hence to state-of-the-art of Layer 1 in our framework in Table I. One of the newer directions for the field of intellectual property is IPR data in linked open data (LOD) format. This is following two trends: the linked open data idea, introduced as a vision more than a decade ago by Berners-Lee (2006) – envisioning the web as a web of data rather than a web of linked documents; the second is based on the notion of open government. European countries are developing policies to release data as open data and putting in place the “systemic” prerequisites for effective use and re-use of them (European Data Portal, 2017; Bauer and Kaltenböck, 2012). The road to effective open data systems are nonetheless long, hence we investigate where on the Berners-Lee Five Star Open Data Plan individual patent offices are at the moment, before presenting the LOD IPR map:

P1. The IPR-linked open data (IPR LOD) map is still in its infancy, thus the full potential of their social benefits are still not realized.

There is a certain hype created in the IPR community either under the moniker of augmented intelligence (Fleischman, 2018) or augmented expertise (White, 2018). In connection to the

Type	Data available in	Patent offices
***** (and *****)	LOD (and RDF)	EPO; USPTO ^a ; Korean Patent Office ^b ; IPO UK ^b ; IP Australia ^c
***	CSV ^d	French Patent Office; Norwegian Patent Office; German Patent Office
**	XLS	Hungarian Patent Office; Austrian Patent Office; Polish Patent Office; Swedish Patent Office
*	PDF ^d	Italian Patent Office; Bulgarian Patent Office; Benelux Patent Office; Spanish Patent Office; Estonian Patent Office; Finish Patent Office; Irish Patent Office; Czech Republic Patent Office; Lithuanian Patent Office; Slovak Patent Office; Portuguese Patent Office; HR Office; Slovenian IP Office; Swiss Patent Office; Hellenic Patent Office; Romanian Patent Office; Danish Patent Office

Notes: ^aTaking into account the AKSW database (different provider); ^bthe patent offices have done additional steps non-related to the format to make merging of data easier; ^cthe database can be described as providing linking data, yet it is not an LOD database in classical sense; ^dif taking into account the bibliographical export in .csv by Espacenet on its web-pages designed in cooperation with national patent offices (e.g. <https://sk.espacenet.com/>), there are such data provided for most, however, the end document exports remain .pdf

Table I.
Classification of
patent offices
according to the
Berners-Lee Open
Data Plan

second preposition, our presumption is that the current state of techniques do not support a sufficient level of semantic understanding that would contribute to successful automation of retrieval and comparison scenarios. Techniques like deep learning show some promise to significantly contribute in this case, yet these approaches are still in active development. Through the investigation of this preposition, we focus on Layer 2 in the begin framework:

P2. AI is a term used very broadly when connected to IPR techniques, to oversell various information retrieval (IR) and machine learning (ML) methods.

The third preposition is connected to Layer 3. In this part – in contrast to the first preposition – we move from the public sector to the private sector, and to especially IPRM benefits. We theorise there is a lack of accessibility and transfer of employees' knowledge, and a low level of transformation of tacit to explicit knowledge, as we believe the current tools function more as visualisation, project management and docking tools. The holistic IPRM-supporting tools, which would allow for internal/external merge of data, as well as support back office (in particular also information, technology and knowledge transfer) as well as front-office IPR activities (Modic and Damij, 2018) are lacking. This part deepens the work started in the study of Modic and Damij (2018) and the evidence is both interview-based as well as a results of IPR tools testing and web searches:

P3. The tools do not correspond to the needs of users as expressed by top IPR managers.

Amongst the several IPRM and social benefits that the paper investigates, due consideration is given to blockchain potential IPR-connected benefits (P4). Several private companies as well as governmental and intergovernmental organisations are currently researching the possibilities of blockchain use in many different fields, including record keeping and smart contracts (Morabito, 2017) which are crucial for the IPR issues:

P4. Blockchain has the potential to produce both IPRM and IPR-connected social benefits if some issues are solved.

The outputs of this paper are the classifications of IPR databases and patent offices according to Berners-Lee Open Data Plan, and IPR LOD map as connected to patents as well as classification of tools and techniques. A mixed methods approach has been used, every part diligently designed with methodological notes.

3. Methodology

We derive our analysis of potential benefits of new solutions for IPR and the potential of IPR tools from interviews with ten prominent IP experts. First, interviews with ten prominent IP experts were conducted. Seven out of the ten IP experts were head IP managers within their respective companies. The companies selected are positioned highly in terms of patent applications and quality rankings. Furthermore, they appear on top innovation listings, such as MIT's list of the 50 Smartest companies. All respondents are executives with years of experience; and one of the interviewees appeared twice in the 50 most influential people in IP, as listed by the *Managing Intellectual Property* magazine. Views expressed inside the interviews are their own and not the views of the companies they are affiliated with. Interviews were conducted either in person, via Skype or via similar VoIP during 2016 and with follow-ups in 2017. Transcripts were analysed using MAXQDA Analytics Pro 12 software. Interview questions were divided into three sections: IPRM (1), formalization (2) and optimisation of processes and gaps reduction (3)). In particular for this paper three topics and their related questions that were included in this semi-structured interview questionnaire are harnessed upon (pertaining to either part (1) or part (3): What is the missing information and/or resources?; Which software tools do you use inside your processes? What are their pros and cons?; What kind of (big) data analysis would be particularly interesting? Who can provide them?

The technologies section brings further methods. The classification of patent offices was done in the period January–February 2018 by conducting web searches and experimental searches with consequent search retrievals inside patent search machines either for full patent documents or at least bibliographical exports. The classification encompasses primarily EU Patent Offices as well as a selection of other relevant patent offices[1]. The framework for the patent map relies on The Linking Open Data cloud diagram, however, it has been significantly upgraded by including material gathered via web searches guided by discussions with various patent offices' staff members. Analysis of techniques is based on critical literature review. We also reviewed websites of 11 top IPR tools providers as identified by interviewees and/or the Hyperion MarketView™ Report (2016) and Capterra's review (2017). Analysis is based on reviews of websites (November, 2017) by Anaqua for Corporations, IP One (from CPA Global), InnovationQ (from ip.com), IPfolio, PatentSight, Unycom Enterprise, Wellspring's IP management software, Patricia (form Patrix), Alt Legal, Inteum, Dennemeyer's DIAMS iQ[2].

4. The potential social and IPRM benefits of new advances in the field of IPR

One of the biggest problems of IPR data usability is the rapid growth of number of IPR, especially patents. They are written in different languages and it has become increasingly challenging to understand the state of the art, this consequently causing duplication of research and increasing the number of invalid patents granted. Once errors can be corrected, it will be easier to identify inherently invalid patents previously granted, and consequently leading to a natural rise in the quality of IPR.

Governments have a large quantity of IPR-related data, which can be of economic and social value to society. European Patent Office (EPO) sees the advantages of its new LOD patent databases, one of the outlets of the new open data trend, as increased availability of data from different sources via one channel, less “data friction” when combining different data sets, more effective linking with business information and increased trust thanks to provenance (Kracker, 2017). The Korean Patent Office (KIPO) also saw its efforts in a similar manner (KIPO, 2016).

The growing importance of IPR Open (linked) data is connected to better transparency making it easier for companies to understand their value. However, if we could not only have exploitable open databases, but if these could also be combined with IPR techniques with AI functionality, and additionally, IPR tools which supported the handling of IPR data by integrating some AI functionalities, we could be seeing a new form of tacit knowledge, the “Artificial intelligence knowledge” creation (see Figure 1). Therefore, the often problematic issue of tacit knowledge inside the IPR field embodied in individuals (note that the usual way of gaining IPRM, exploitation and other connected IPR knowledge is through apprenticeship and that the rotation of individuals presents a serious problem for especially company IPR departments, Modic and Damij, 2018)) would be transformed into a latent explicit knowledge (knowledge available on recall as opposed to explicit knowledge, always available). Solutions, like IBM Watson, seem to also be a game changer in this area. Watson identified compounds on which the patent protection has already lapsed, and the pilot results suggest that Watson can accelerate identification of novel drug candidates and novel drug targets by harnessing the potential of patent (and connected) big data (Chen *et al.*, 2016). The IBM team believes the insights provided by Watson technology are to be used as a guide, i.e., as augmented intelligence – which is capable of ingesting, digesting, understanding and analysing data and can be harnessed in various elements of IPR processes: from evidence of use, to prior art, patent landscapes and portfolio analysis (Fleischman, 2018). If the technology was widely available with all its features, this could present a significant change, as it would enable smaller entities to access knowledge that is now tacit knowledge.

When discussing traceability, blockchain is one of the frequently debated issues. Several potential social benefits, as derived from the utilisation of blockchain in the field of IPR, are present. A tool for registration of IPRs could simplify registration and lower the costs (Vella *et al.*, 2018; Morabito, 2017) or could be an alternative to IPR registration, especially patents. Thus, it has a potential particularly for small entities (independent inventors, SMEs, non-profit organisations), as well as inventors and organisations from less developed countries, who are unable to access the current world patent system simply because it is too expensive for them.

Blockchain provides a robust and trustworthy method of establishing business ownership on intangible assets, including IPR (Morabito, 2017) and thus has the potential to enhance transparency of IPR transactions (Vella *et al.*, 2018). Not only does this have positive effects for individual companies, but it can also streamline the costs of operations for patent offices, and reduced options for litigation can lower court case numbers and reduce court backlogs. Furthermore, it also has the potential to enable half open licensing, when royalties start only when IPR-based income is generated by downstream users; meaning that without income generation, the half open licenses allow for IPR-based solutions to be spread in an open environment. Moreover, it would allow tracking commons' knowledge (under open licenses or not) incorporation into corporate IPR portfolios disallowing the privatisation of gains.

With regard to potential IPRM benefits, IPRM deals with managing IPR big data efficiently, and differently (Braganza *et al.*, 2017; Davenport *et al.*, 2012). McAfee and Brynjolfsson (2012) argue that companies will not reap the full benefits of the transition made in exploiting big data, unless they are able to manage change effectively.

Analysis of the interviews showed a clear trend that IP executives are aware of the growing importance of ICT, and their role in IPRM, however, they continue to struggle with defining how to integrate IPR tools to achieve best outcome. A Senior IP Counsel at a German multinational chemical manufacturing corporation stated that, "IT developments will have a big impact in the near future on IP development, because the more transparent you make the IP, the easier it is for management to understand its value".

Utilising the ICT in IPR processes is possible, however, doing it in the most efficient way to enable companies to achieve maximum benefits, is the ideal. Some companies use a range of different software tools connected to IPR and IPRM, whilst others try to find or develop software that integrates as many features and data sources as possible and are able to connect to other business processes and databases. Generally, the more comprehensive the tool, the less information is missing, and consequently, the higher the satisfaction level. Nonetheless, some experts, such as the Head of Legal Operations and IP Management at a European multinational pharmaceutical corporation, believe that IPR tools often promise more than they deliver. He states that they, "do not think there are any particularly good IP management tools on the market /.../the whole industry still lacks are real IP management tools, helping to relate to the business value more". IPR experts are seeking a tool that would, in addition to being a comprehensive docketing system and simple interface retrieval of data from public IPR databases, also encompass supplying or channelling invention disclosures to pertinent individuals, providing functionality for IPR valuation, evaluation and analysis.

The next chapter will provide more detail deal with regard to the technological dimension, providing an analysis on the current state of linked open databases, software tools for IPRM and techniques that support IPR data correction and analytics.

5. Technology

5.1 Databases and linked (open) data

Since the Venetian patent statute of 1474, IPR have retained their connection to the concept of openness and dissemination of ideas in exchange for limited time monopolies. There are various types of databases and online sources connected with IPR constituting Layer 1 in the framework in Table I. Public patent databases as the original sources allow raw data

retrieval and the use of interfaces by providing patent texts and some metadata. Related IPR databases include, for example, those related to patent disputes, patent citations. Business databases provide information on IPR owners, etc. Scientific databases provide us *inter alia* with data on inventors. Miscellaneous online data sources include less or more structured sources, e.g., business news, blogs-based IPR-related texts, information on IPR experts. Multi-source IPR databases provide broader information, e.g., on IPR quality and business connected data. Two examples of the latter are the data set linking the EPO and USPTO patent data to Amadeus business database and the Oxford Firm-Level IP Database (Thoma and Torrisi, 2007; Helmers *et al.*, 2011).

Linked open data (IPR LOD) databases are the latest evolution in IPR databases, although the concept of LOD goes back to 2006, when principles such as using uniform resource identifiers as names for things and including links were put forward (Berners-Lee, 2006). Linked data are data published on the web in a machine-readable format, which can be linked to or from external data (Bizer *et al.*, 2009). LOD is in essence a format allowing for efficient (multi-source) database utilisation as the term refers to a set of practices for publishing and interlinking structured data (Auer, 2014).

Combining this to ideas of open data, we get LOD, structured data made available for others to be reused (Mezaour *et al.*, 2014). The concept is connected to the Open Data movement to ensure public government data are accessible in non-proprietary formats (Bauer and Kaltenböck, 2012). However, LOD landscape includes databases provided by non-governmental entities. DBPedia, extracting structured knowledge from Wikipedia, is often seen as the “nucleus” of LOD (Auer *et al.*, 2007). Furthermore, patent data of individual patent offices are sometimes provided by outside providers, such as in the case of USPTO or (formally) the EPO.

Table I shows the classification of patent offices and their data according to the Berners-Lee Five Star Open Data Plan. More stars indicate data formats more conducive to open data policies, as they allow for easier export and import of data, and more streamlined merging and analysis. The category **** is redundant as there is no standalone RDF providing databases; and, we would suggest an introduction of the *****+ category, where the additional criteria is the existence of linkages with other data, signalling the real uptake of the raw data by users (see Table I). The Type indicates the most Open data friendly format, though patent offices often provide other formats simultaneously. They often also provide more than one database, and the degree of the export varies for bibliographical data (Swiss Patent Database offering up to 25 variables).

Five patent offices are leading in terms of IPR LOD; USPTO, EPO, KIPO, IPAustralia and IPO UK. Cooperation of national offices with Espacenet was also advantageous, as it produced the option of a limited bibliographic data download in .csv format (not taken into account above). However, most of the patent offices can still be categorised only as Type * or Type **, their data remaining in linkable open data unfriendly formats.

There are only a few databases that could be categorised as *****+, or that have shown other initiatives to make exporting, merging and analysing data easier. For example, KIPO has not only published the IPR LOD, but also included the owners’ corporate registration number and the Australian Patent Office IPR database includes information about companies’ size, technology and geographic location, making it easier for users to link data on patents to information on related business entities (KIPO, 2016; Man, 2014).

Currently, EPO’s Linked open data is the newest of the few IPR LOD databases at users’ disposal. It builds upon their previous work in connecting patent-related data, such as their Deep Linking service, allowing users to consult the EP document’s legal status data. However, the IPR LOD database remains as a raw data product and without additional skills and resources cannot be fully utilised, which could potentially widen the gap between SMEs and IPR-savvy companies. For example, the linkage to DBPedia has also been carried out, but since then de-installed (Kracker, 2017). This year the EPO also included in their

Research grant call explicitly the field of linked open data and solutions therein, where at least one project will start end of this year linking EPO database with the Springer database (IP LodB, 2018). The current LOD IPR landscape shown below is based on the The Linking Open Data cloud diagram and upgraded[3].

Figure 2 shows patent LOD databases[4] we could call *****+, and their inbound and outbound links, as per The Linking Open Data cloud diagram (LOD cloud, 2018) – a complex LOD ecosystem currently listing 1,164 data sets. They are also linked to the most inbound and outbound link-rich LOD databases, namely, the Comprehensive KAN and DBPedia. The new EP LOD and KIPO databases have no data on linkages, even though some attempts were made as mentioned above. There are, however, several LOD databases that this patent data could be linked to; e.g. the recently published bibliographic LOD database by Springer Nature SciGraph or the older New York Times LOD.

When considering the traceability of IPR data, some patent offices offer centralised solutions, such as i-DEPOT, which allows to trace the date of inventions' creation. However, at the forefront of these debates is blockchain as a disruptive technology, due to its transparency, decentralisation and prevention of infringements and fraud. Blockchain is a chain of blocks of chronologically linked information, replicated in a distributed database. Information can be added, but never removed, changes are registered and validated. Individual blocks can be protected by cryptography, and only those authorised can access the information (McPhee and Ljusic, 2017). Blockchain application to IPR can be either inside the registration or exploitation phases (related to issues of licensing, proving authenticity and piracy) (Vella *et al.*, 2018; Morabito, 2017) as well as distribution. In case of licensing, the topic is connected to smart contracts, open licenses and IPR-based collaboration (Pilkington, 2016; Morabito, 2017). Smart contracts are computer codes that reside in the blockchain and are implemented if certain conditions are met, which is confirmable by a number of computers to ensure truthfulness (Morabito, 2017; Szabo, 1997). There are numerous potential applications of blockchain connected to IPR. Also, the Linked Data paradigm is evolving from an academic concept for addressing one of the biggest challenges in the area of information management

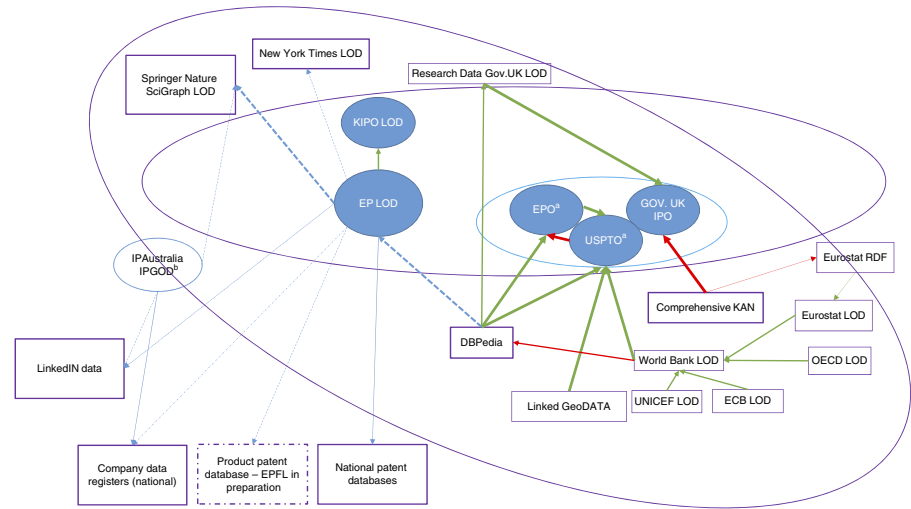


Figure 2.
Narrow IPR LOD
landscape (patent
databases)

Notes: Green arrows: ingoing links; red arrows: outgoing links; dashed blue lines: potential and planned or deconstructed (bolded). ^aNot provided by patent offices; ^blinked database, but not following LOD standards

the exploitation of the web as a platform for data and information integration; to practical applications in IPR field deriving from the transfer from the Web of Documents to a Web of Data. Yet, it is clear there is still much to be done, both in terms of the volume of IPR LOD-connected databases, as well as their functionality in linking to other LOD data sets as well as the real-life uptake of blockchain solutions.

5.2 Classification of tools and techniques

This chapter summarises the techniques and tools (technology Layers 2 and 3 as set out in Figure 1) that analyse large quantities of patent documents and other IPR data to provide useful information to various users.

The EPO's database, Espacenet, on its own, currently contains over 100m patent documents from 90 patent authorities worldwide. Whilst patent data are exceptionally important, it is also very difficult to extract some useful information from it as patents are mostly stored as images; written in different languages; countries have different patent requirements; no uniform structural requirements; some patent figures are drawn by hand, some on computer; some patent attorneys intentionally use misleading language; incomprehensible language and grammatical mistakes can be also used inadvertently. How to deal with these issues remains a challenge.

There are several possible taxonomies of IPR software. Considering their functionalities we see tools supporting different phases of the innovation cycle, those supporting financial management (record and estimate costs), archiving documents (IPR portfolio) and enabling communication between users and IPR offices. Some tools have functionality to integrate data from external databases, such as patent litigation information and patent citation indexes. In terms of intended user-base we have IPR tools for companies, for IPR experts and for technology transfer offices.

There is an upward trend in the creation of new IPRM software in recent years. However, after reviewing the websites of the 13 most important IPR tools providers by Hyperion MarketView™ Report (2016) it appears that these tools only modestly respond to the challenges raised, and largely look like any project management software. Bonino *et al.* (2010) was optimistic with regard to semantic-based solutions, however, some of the tools he describes are currently in poor condition or unavailable.

In terms of techniques utilised in semantic analysis, Abbas *et al.* (2014) made a taxonomy of proposed computer-assisted patent analysis techniques where they distinguish between text mining and visualisation approaches. These two categories are based on frequent use-cases, whilst the underlying methods are primarily inspired by IR and ML. This is not unreasonable, as patent documents are similar to other types of documents in that they contain textual and visual data as well as references to other documents.

As seen in Figure 3, a typical IR system consists of document pre-processing, feature extraction and feature analysis. Each of those steps can be based on heuristic rules or utilise machine learning methods. In the following paragraphs, we review the use of different techniques in the IPR research domain in the last decade, with a particular focus on the

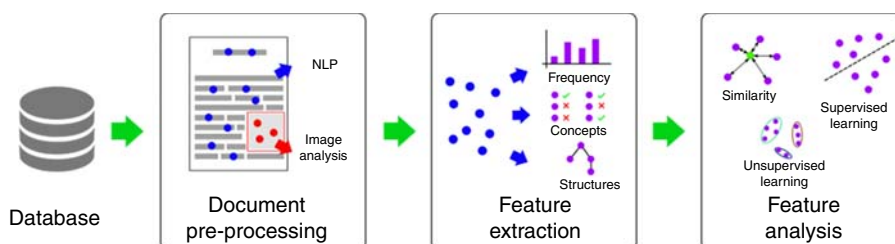


Figure 3.
A typical computer-assisted document analysis pipeline as IPR techniques classification framework

works referenced in recent literature reviews by Abbas *et al.* (2014) and Aristodemou and Tietze (2017). The list is by no means complete, it is only focussed on key examples illustrating the diversity and potential of such methods.

The patent document pre-processing step involves scanning the unstructured data (text and images) and extracting useful information from it.

Due to the nature of the patent data, the approaches mainly focus around text mining techniques; meaning using some kind of natural language processing (Wang *et al.*, 2015; Han *et al.*, 2017), such as subject–action–object analysis (Park, Kim, Choi and Yoon, 2013; Park, Ree and Kim, 2013), property–function analysis (Dewulf, 2013) or rule-based analysis to extract semantic primitives. Several authors have also proposed the utilisation of patent images and sketches in patent analysis, in order to determine similarities between patents (Bhatti and Handbury, 2013). In terms of pre-processing, image analysis challenges involve localisation of images and sub-images, categorisation of images and label recognition (Vrochidis *et al.*, 2010). The primary sources of inter-information are cross-patent citations (Altuntas *et al.*, 2015).

The feature extraction methods transform low-level semantic primitives into a document-wide representation. By involving projection of each document into a high-dimensional feature space we can determine bounds between classes or proximity of documents. When processing textual data, the semantic primitives can be frequency vectors (Chen and Yu-Ting, 2011), vectors of concepts that describe higher-level semantic information, or domain-specific hierarchical structures (Lee, 2013). In analysis of patent sketches, content is frequently encoded with shape or texture descriptors (Bhatti and Handbury, 2013) due to the line-art nature of visual information.

The method used in the feature analysis stage depends on the problem at hand, for example, retrieval of similar patents. In this case, IR techniques based on vector distances (Lee, 2013) are used to infer which documents are most similar. Another task is automatic classification of patents using ML methods. Scenarios include patent quality analysis (Wu *et al.*, 2016), patent categorisation (Vrochidis *et al.*, 2010) and determining the impact of patents on other aspect of companies (Chen *et al.*, 2013). Supervised learning methods, such as support vector machines (Wu *et al.*, 2016) or artificial neural networks (Chen *et al.*, 2013), are frequently used in such cases. In explorative analysis of the patent landscape for trend identification, people have also utilised unsupervised learning methods, like clustering (Atzmüller and Landl, 2009; Madani and Weber, 2016) and network analysis (Dotsika, 2017; Park, Kim, Choi and Yoon, 2013).

Despite the apparent contribution of IR methods in transforming access to information, they are harder to apply to semantic-sensitive fields, such as IPR analysis, with the same level of success. The crucial information in patent documents can be difficult to extract automatically because of objective (history, language) or subjective (intentional misuse of description) reasons. As noted by Lupu (2017), the level of research interest in this field has, after more than a decade of increasing optimism, decreased in the past years. This can be in part attributed to the realisation that extracting high-level semantic content from sophisticated unstructured text and images is very a challenging problem. The most successful working cognitive computing system is IBM Watson, who has already been analysing patent information in the past, with a particular emphasis in the pharmaceutical sector. However, this system is proprietary and accessible only to a limited number of influential clients.

6. Discussion

Over the last years, activities around IPR Open Data, merging of IPR data with related data, IPR Linked Data, IPR-linked open databases and the debates over utilising the Semantic Web opportunities have gained momentum. However, this should go hand in hand with organisations (both public and private) publishing structured data (complying also with linked data standards/principles), the advances in new techniques, as well as IPR tools and

their increased availability. Companies and other patent and IPR data users need to draw on those advanced technologies and tools in order to combine, query (and analyse) data as part of their business intelligence, as well as to improve their services and products.

In terms of the availability of data, the amount of IPR and IPR-connected data publicly available is increasing. Responding to P1, the new trends towards formats supporting more export-ready, merge-ready and analysis-ready data are also real, although the amount of patent data available (e.g. as LOD) is still relatively low. LOD means the data are “linkable”, not that it is already linked. This means that the uptake of these databases by the users can be slow and can even widen the gap between the IPR-savvy multinationals with sufficient resources and other smaller entities and individuals. The latter would defeat the purpose of publishing such databases, if the objective was to make IPR data more useful to more groups of users, especially also non-patent savvy users (data scientists, web developers, companies integrating IP into their products). Some steps are taken towards this, for example, IPNOVA (available at the moment as a beta version) which is the interface to the IPAustralia’s IPGOD database. Another route (contrasting somewhat with developing interfaces) is through sufficient dissemination and training workshop accompanying the releases of databases in new formats. On the other hand, the authors remain hopeful as new entities – including private and NGO entities – provide more and more LOD databases, and with growth of potential links, allowing greater potential for IPR.

In response to P2, techniques that would support IPR data correction, and IPR data analytics and software tools, which support IPRM, are still not at a sufficient stage of development for IPR managers and other users dealing with IPR. The IPR tools remain primarily visualisation tools (P3); or project management and docketing tools, applied to the field of IPR. There are few true IPRM tools that also integrate variable (external and internal) data merges and harness new advances in IPR techniques, although some solutions have been integrated. This is perhaps because the existing techniques, which are suitable for many existing retrieval and analysis tasks, are frequently branded as “AI”, a term that increases expectations about the capabilities which existing methods fail to fulfil. A complete AI system is perhaps the ultimate goal of automatic patent analysis, capable of high-level reasoning about the content of patent documents, comparing their underlying ideas and determining similarities. The current state is (far) removed from this goal. At present, it is primarily addressing very narrow domains, interpretable by data scientists and machine learning researchers. However, as also noted by Lupu (2017), recent breakthroughs in deep learning and artificial neural networks already address tasks such as machine translation and image analysis, which can be (and sometimes are) utilised in IPR analysis.

In response to P4, blockchain technology is now fairly widely discussed for its potential to change the nature of IPRs by simplifying registration, lowering costs, increasing transparency and enabling or improving licensing and other transfers of IPR. However, the technology has certain limitations and still needs significant time to develop. This is not only because of the influence that transnational companies have on policy makers, but also, the technology itself might have some weaknesses. It needs huge processing power and therefore for now requires high-volume electricity consumption. Second, field, such as the IPR field, has its own set of limitations connected to legal and judicial frameworks. Therefore, it is important to carefully determine fields where it would be used. “Despite the many interesting potential uses of blockchain technology, one of the most important skills in the developing industry is to see where it is and is not appropriate to use cryptocurrency and blockchain models” (Swan, 2015). Although there are various social and IPRM benefits of employing blockchain technologies in the field of IPR, caution must be applied.

To conclude, despite significant efforts in the last decades, in the field of information technology support to IPRs, and the more and more used buzzwords of augmented intelligence and augmented expertise also for IPR, there is more time needed before these progressive ideas will become (widespread) reality.

Acknowledgements

Dr Damij would like to acknowledge the ARRS Grant No. ARRS-P1-0383(A). Dr Hafner would like to acknowledge Operation No. C3330-17-529006 “Researchers-2.0-FIŠ-529006” supported by ERDF and Republic of Slovenia, Ministry of Education, Science and Sport. Dr Modic would like to acknowledge the JSPS International Research Grant ID No. 16774 and JSPS KAKENHI Grant No. 16F16774. Dr Hafner and Dr Modic acknowledge that this paper has been co-funded by the Academic Research Programme of the European Patent Office. The research results contained in this paper are those of the researchers only. They do not necessarily represent the views of the EPO.

Notes

1. Missing from the list are the Latvian, Icelandic, Maltese and Cyprus Patent Office, as they only refer to Espacenet or there is a lack of sufficient information. The classification takes into account data that is (formally) provided by outside sources (e.g. for USPTO).
2. We have also taken into account a review of the available semantic solutions that was made by Bonino *et al.* (2010, p. 37, Table 9). However, these new technology enablers are currently in a less than ideal state (in poor condition or unavailable) and they (those which are at least available) look more like a scientific experiment than a final product that would support real patent analytics in companies. Though we sent some follow-up e-mails we did not receive much useful information so they were excluded from the paper.
3. Eito-Brun (2015) lists 31 LOD databases according to datahub.io related to patents, but they could be hardly classified as IPR databases.
4. The Linked Open Data Cloud diagram includes EPO reference, which was created and published by the research group AKSW.

References

- Abbas, A., Zhang, L. and Khan, U.S. (2014), “A literature review on the state-of-the-art in patent analysis”, *World Patent Information*, No. 37, pp. 3-13.
- Altuntas, S., Dereli, T. and Kusiak, A. (2015), “Analysis of patent documents with weighted association rules”, *Technological Forecasting and Social Change*, No. 92, pp. 249-262.
- Aristodemou, L. and Tietze, F. (2017), “A literature review on the state-of-the-art on intellectual property analytics”, Working Paper Series No. 2, Centre for Technology Management, Vol. 2017, Cambridge, pp. 1-13, available at: www.repository.cam.ac.uk/handle/1810/268007 (accessed 25 August 2018).
- Atzmüller, P. and Landl, G. (2009), “Semantic enrichment and added metadata – examples of efficient usage in an industrial environment”, *World Patent Information*, Vol. 31 No. 2, pp. 89-96.
- Auer, S. (2014), “Introduction to LOD2”, in Auer, S., Bryl, V. and Tramp, S. (Eds), *Linked Open Data – Creating Knowledge Out of Interlinked Data Results of the LOD2 Project*, Springer, Cham, Heidelberg, New York, NY, Dordrecht and London, pp. 1-20.
- Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. (2007), “DBpedia: a nucleus for a web of open data”, in Aberer, K. *et al.* (Eds), *ISWC/ASWC 2007*, Springer-Verlag, Berlin and Heidelberg, pp. 722-735.
- Bauer, F. and Kaltenböck, M. (2012), *Linked Open Data: The Essentials – A Quick Start Guide for Decision Makers*, Edition mono/monochrom, Vienna.
- Berners-Lee, T. (2006), “Linked data: a personal view note”, available at: www.w3.org/DesignIssues/LinkedData.html (accessed 30 January 2018).
- Bhatti, N. and Hanbury, A. (2013), “Image search in patents: a review”, *International Journal on Document Analysis and Recognition*, Vol. 16 No. 4, pp. 309-329.

- Bizer, C., Heath, T. and Berners-Lee, T. (2009), "Linked data – the story so far", *International Journal on Semantic Web and Information Systems*, Vol. 5 No. 3, pp. 1-22.
- 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*, No. 32, pp. 30-38.
- Braganza, A., Brooks, L., Nepelski, D., Ali, M. and Moro, R. (2017), "Resource management in big data initiatives: processes and dynamic capabilities", *Journal of Business Research*, No. 70, pp. 328-337.
- Capterra (2017), "Capterra review", available at: www.capterra.com/ (accessed 10 December 2017).
- Chen, Y. and Yu-Ting, C. (2011), "An IPC-based vector space model for patent retrieval", *Information Processing and Management*, Vol. 47 No. 3, pp. 309-322.
- Chen, Y., Argentinis, E. and Weber, G. (2016), "IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research", *Clinical Therapeutics*, Vol. 38 No. 4, pp. 688-701.
- 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*, pp. 136-139, available at: <https://ieeexplore.ieee.org/document/6754276> (accessed 25 August 2018).
- Ciccatelli, A. (2017), "The future of big data and intellectual property", available at: www.insidecounsel.com/2017/06/13/the-future-of-Big-Data-and-intellectual-property?&slreturn=1507773271 (accessed 15 May 2018).
- Davenport, T.H., Barth, P. and Bean, R. (2012), "How 'Big Data' is different", *MIT Sloan Management Review*, Vol. 54 No. 1, pp. 43-46.
- Dewulf, K. (2013), "Sustainable product innovation: the importance of the front end stage in the innovation process", in Coelho, D. (Ed), *Advances in Industrial Design Engineering*, Intech Open, Rijeka, pp. 139-166.
- Dotsika, F. (2017), "Identifying potentially disruptive trends by means of keyword network analysis", *Technological Forecasting and Social Change*, No. 119, pp. 114-127.
- Eito-Brun, R. (2015), "Patents, semantics and open innovation: the role of LOD in a business directory for knowledge intensive industries", Presentation at Dr Haxel Congress and Event Management GmbH, Nice, 20 October, available at: www.slideshare.net/Haxel/linked-open-data-in-the-world-of-patents (accessed 10 February 2018).
- European Data Portal (2017), *Open Data Maturity in Europe 2017: Open Data for a European Data Economy*, Publications Office of the European Union, Luxembourg, available at: www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf (accessed 14 February 2018).
- Fleischman, T. (2018), "Augmented intelligence, intellectual property and IBM", Presentation at AI: Hype vs Reality and the Impact on the Patent Industry Clarivate Analytics Online Webinar.
- Freeman, C. and Louca, F. (2002), *As Time Goes By: From the Industrial Revolution to the Information Revolution*, Oxford University Press, Oxford.
- Han, Q., Heimerl, F., Codina-Filba, J., Lohmann, S., Wanner, L. and Ertl, T. (2017), "Visual patent trend analysis for informed decision making in technology management", *World Patent Information*, No. 49, pp. 34-42.
- Helmets, C., Rogers, M. and Schautschick, P. (2011), "Intellectual property at the firm-level in the UK: the Oxford firm-level intellectual property database", Discussion Paper Series No. 546, Department of Economics, University of Oxford, Oxford, pp. 1-23, available at: www.economics.ox.ac.uk/materials/working_papers/paper546.pdf (accessed 14 February 2018).
- Hyperion MarketView™ Report (2016), "Anaqua enterprise as featured in the Hyperion MarketView™ research program intellectual property management systems for corporation (2016)", available at: www.anaqua.com/anaqua-hyperion-report.html (accessed 25 February 2018).

- IP LodB (2018), "Granted project proposal IP linked open data: building bridges (IP LodB)", IP LodB (forthcoming), Muenich, pp. 1-96.
- KIPO (2016), "Korean Intellectual Property Office and Patent Information Expansion of providing public services (English title)", available at: www.kipo.go.kr/kpo/user.tdf?a=user.news.press1.BoardApp&board_id=press&c=1003&catmenu=m03_05_01&seq=15589 (accessed 13 February 2018).
- Kracker, M. (2017), "New PI product: linked open EP data", *Presentation at IP Statistics for Decision Makers Conference, Mexico City, 14–15 November*.
- Lee, S. (2013), "How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships", *Journal Technology Analysis & Strategic Management*, Vol. 25 No. 1, pp. 23-38.
- LOD cloud (2018), "The linked open data cloud", available at: <https://lod-cloud.net/> (accessed 20 December 2018).
- Lupu, M. (2017), "Information retrieval, machine learning, and natural language processing for intellectual property information", *World Patent Information*, Vol. 49, pp. A1-A3.
- McAfee, A. and Brynolfsson, E. (2012), "Big data: the management revolution," *Harvard Business Review*, Vol. 90 No. 10, pp. 60-68.
- McPhee, C. and Ljutic, A. (2017), "Editorial: blockchain", *Technology Innovation Management Review*, Vol. 7 No. 10, pp. 3-5.
- Madani, F. and Weber, C. (2016), "The evolution of patent mining: applying bibliometrics analysis and keyword network analysis", *World Patent Information*, No. 46, pp. 32-48.
- Man, B. (2014), "Overview of the intellectual property government open data", IP Australia Economic Research Paper No. 2, pp. 1-18, available at: www.ipaustralia.gov.au/about-us/reports/economics_research_paper02 (accessed 14 February 2018).
- Mezaour, A., van Nuffelen, B. and Blaschke, C. (2014), "Building enterprise ready applications using linked open data", in Auer, S., Bryl, V. and Tramp, S. (Eds), *Linked Open Data – Creating Knowledge Out of Interlinked Data Results of the LOD2 Project*, Springer, Cham, Heidelberg, New York, NY, Dordrecht and London, pp. 155-174.
- Modic, D. (2017), "Intellectual property rights in the information sector: sociological perspective", in Rončević, B. and Tomšić, M. (Eds), *Information Society and Its Manifestations: Economy, Politics, Culture*, Peter Lang, Frankfurt am Main, pp. 55-70.
- Modic, D. and Damij, N. (2018), *Towards Intellectual Property Management: Back Office and Front Office Perspectives*, Palgrave Macmillan, Cham.
- Morabito, V. (2017), *Business Innovation through Blockchain: The B³ Perspective*, Springer Nature, Cham.
- Park, H., Ree, J.J. and Kim, K. (2013), "Identification of promising patents for technology transfers using TRIZ evolution trends", *Expert Systems with Applications*, Vol. 40 No. 2, pp. 736-743.
- Park, H., Kim, K., Choi, S. and Yoon, J. (2013), "A patent intelligence system for strategic technology planning", *Expert Systems with Applications*, Vol. 40 No. 7, pp. 2373-2390.
- Pilkington, M. (2016), "Blockchain technology: principles and applications", in Olleros, X.F. and Zheng, M. (Eds), *Research Handbook on Digital Transformations*, Edward Elgar Publishing, Cheltenham and Northampton, pp. 225-253.
- 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 10 February, 2018).
- Swan, M. (2015), *Blockchain – Blueprint for a New Economy*, O'Reilly Media, Sebastopol, CA.
- Szabo, N. (1997), "Formalizing and securing relationships on public networks", *First Monday*, Vol. 2 No. 9, pp. 1-8, available at: <http://firstmonday.org/article/view/548/469> (accessed 29 January 2018).

- Thoma, G. and Torrisi, S. (2007), "Creating powerful indicators for innovation studies with approximate matching algorithms: a test based on PATSTAT and Amadeus databases", Working Paper No. 211, CESPRI Bocconi University, Milan, available at: <https://ideas.repec.org/p/cri/cespri/wp211.html> (accessed 14 February 2019).
- Vella, D., Falzon, M., Cassar, T. and Valenzia, A. (2018), "Blockchain's applicability to intellectual property management", *The Licensing Journal*, Vol. 38 No. 1, pp. 10-12.
- Vrochidis, S., Papadopoulos, S., Moumtzidou, A., Sidiropoulos, P., Pianta, E. and Kompatsiaris, I. (2010), "Towards content-based patent image retrieval: a framework perspective", *World Patent Information*, Vol. 32 No. 2, pp. 94-106.
- Wang, J., Lu, W.F. and Loh, H.T. (2015), "A two-level parser for patent claim parsing", *Advanced Engineering Informatics*, Vol. 29 No. 3, pp. 431-439.
- White, E. (2018), "AI in patent research", Presentation at AI: Hype vs Reality and the Impact on the Patent Industry Clarivate Analytics Online Webinar.
- Wu, J.L., Chang, P.C., Tsao, C.C. and Fand, C.Y. (2016), "A patent quality analysis and classification system using self-organizing maps with support vector machine", *Applied Soft Computing*, No. 41, pp. 305-316.

Further reading

- Lee, S., Yoon, B., Lee, C. and Park, J. (2009), "Business planning based on technological capabilities: patent analysis for technology-driven roadmapping", *Technological Forecasting and Social Change*, Vol. 76 No. 6, pp. 769-786.

Corresponding author

Dolores Modic can be contacted at: dollymod@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com