Patent Threat Analysis Search Engine

Yung Chang Chi
Department of Industrial and Information Management and
Institute of Information Management,
National Cheng Kung University,
Tainan City, Taiwan ROC
e-mail:charles.y.c.chi@gmail.com

Hei Chia Wang
Department of Industrial and Information Management and
Institute of Information Management,
National Cheng Kung University,
Tainan City, Taiwan ROC
e-mail:hcwang@mail.ncku.edu.tw

***Abstract*-This paper proposes a framework for a patent threat analysis search engine. The framework employs text mining based on the patent map approach to identify which particular patent is similar to the one in dispute. The patent map is a visual representation that uses technological proximities among patents. The patent threat analysis method will analyze patent infringement issues from judgements in the USA and Europe. Having examined and compared the patent map and patent infringement analysis, we can identify what kind of product and technology are subject to the threat of patent infringement with the help of solution integration and analysis of the two different databases.**

Keywords-patent;patent threat;patent map;content analysis.

I. INTRODUCTION

Patents are important knowledge sources for industrial research and product development because of their innovation and practicability. In recent years, patent analysis increased in importance for high-technology management as the process of innovation became more complex, the cycle of innovation became shorter and the market demand more volatile [22].

Patent analysis technologies include patent bibliometric data analysis [21], patent citation analysis [5], patent statistical analysis [26] and patent classification. Bibliometric analysis of patents provides information on the growth of the inventive activity and technological trends [22].

Patent mining is an emerging research topic that grew in recent years. So far, only few researches have been done on the topic. Patent mining consists of patent retrieval, patent categorization and patent clustering [25].

Retrieving patent documents can be done through the cluster-based approach [6]. Distributed information retrieval for patent can be done by generating ranking lists for the query by CORI (The collection retrieval inference network) or KL (Kernighan–Lin) algorithms [14]. Categorizing patent documents can be done automatically using the k-Nearest Neighbor classifiers and Bayesian classifiers [12][13], or by using a variety of machine learning algorithms [1], the k-Nearest Neighbor on the basis of patent's semantic structure [7], the classifier built through back-propagation network [19]. Patent documents can be clustered through the k-Means algorithm and represented in a visualized patent map [8], and the structured SOM (Self-Organizing Map) clustering algorithm [3]. Clustering algorithms can also be adopted to form a topic map for presenting patent analysis and summarization results [19], to create a system interface for retrieving patent documents [3].

Content analysis is indigenous to communication research and is potentially one of the important research techniques in social sciences. It seeks to analyze data within a specific context in view of the meaning someone –a group or a culture-attributes to them. Communications, messages, and symbols differ from observable events, things, properties, or people in that they inform about something other than themselves; they reveal some properties of their distant producers or carriers, and they have cognitive consequences for their senders, their receivers, and the institutions in which their exchange is embedded [9].

Content analysis is a research technique for making replicable and valid inferences from texts to the contexts of their use. As a technique, content analysis involves specialized procedure. It provides new insights, increases a researcher's understanding of particular phenomena, or informs practical actions. Content analysis is a scientific tool [10].

The judgements of patent infringement, unlike the patent documents, can be mined using text mining techniques, since the judgements are legal documents. The judgements can be transformed into patterns by content analysis, and readers can easily access them the same way as reading newspapers to understand the key points and issues in dispute.

The rest of the paper is structured as follows. Section II presents the research background. Section III states our objective. In Section IV, we describe our proposed research method. The paper concludes with expected results and future work considerations.

II. RESEARCH BACKGROUND

So far, patent analysis technologies include patent bibliometric data analysis [21], patent citation analysis [5], patent statistical analysis [26], and patent classification. Patent mining consists of patent retrieval, patent categorization and patent clustering [25] that focuses just on the patent documents analysis and patent mining. However, patent infringement constitutes the biggest threat in patents use. Through patent analysis and mining, one can just discover newly developed products and their similarity with

 ISBN: 978-1-61208-408-4

the claims of other patents, but no one can foresee where potential patent threats are and the likelihood of patent infringement.

This framework of patent database is based on the United States Patent and Trademark Office (USPTO) patent database and the European Patent Office (EPO) patent database. The patent infringement judgements are based on the case judgements in United States and European Union.

III. RESEARCH OBJECTIVE

The purposes of this study is to provide the patent threat analysis and reference regarding patent infringement as well as technology trends for new product designers and technology research engineers at the stage before and after developing a new product or technology. This will also provide the information needed so that management can make strategic decisions.

Because of a lack in legal background, it is difficult for ordinary readers to fully grasp the judgements rendered by professional judges. With the implementation of content analysis, ordinary people will be able to use content analysis technology to analyze the patent infringement verdict contents, and try to use big data concepts across different databases to discover any relation.

IV. RESEARCH METHOD

The patent documents can be collected from United States Patent and Trademark Office (USPTO) patent database and the European Patent Office (EPO) patent database.

A. Patent documents analysis

Base on the collected patent documents and the subject-action-object (SAO) structures extracted by using Natural Language Processing (NLP), the study uses a content analysis approach to generate the patent map.

NLP is a text mining technique that can conduct syntactic analysis of natural language; NLP tools include Stanford parser (Stanford2013)[27], Minipar (Lin2003)[28] and KnowledgistTM2.5[29].

NLP tools will be used for build a set of SAO structures from the collected patents.

Multidimensional scaling (MDS) is a statistical technique used to visualize similarities in data [11][16]. Patent documents in different fields have different key issues that trigger different Multidimensional scaling, so the paper will design a new algorithm to identify which particular patent field shall correspond to what extent of scaling.

B. Patent infringement verdict content analysis

The most obvious source of data appropriate for content analysis is text to which meanings are conventionally attributed: verbal discourse, written documents, and visual representations. The text in the patent infringement judgements is important because that is where the meanings are. For this reason, it is essential for the content analysis technology to analyze the patent infringement text in order to develop strategies and preventive measures in patent litigation.

Content analyses commonly contain six steps that define the technique procedurally, as follows:

Design. Design is a conceptual phase during which analysts define their context, what they wish to know and are unable to observe directly; explore the source of relevant data that either are or may become available; and adopt an analytical construct that formalizes the knowledge available about the data-context relationship thereby justifying the inferential step involved in going from one to the other.

Unitizing. Unitizing is the phase of defining and ultimately identifying units of analysis in the volume of available data. Sampling units makes possible the drawing of a statistically representative sample from a population of potentially available data, such as issues of a newspaper, whole books, television episodes, fictional characters, essays, advertisements.

Sampling. While the process of drawing representative samples is not indigenous to content analysis, there is the need to (1) undo the statistical biases inherent in much of the symbolic material analyzed and (2) ensure that the often conditional hierarchy of chosen sampling units become representative of the organization of the symbolic phenomena under investigation.

Coding. Coding is the step of describing the recording units or classifying them in terms of the categories of the analytical constructs chosen. This step replicates an elementary notion of meaning and can be accomplished either by explicit instructions to trained human coders or by computer coding. The two evaluative criteria, reliability as measured by inter coder agreement and relevance or meaningfulness, are often at odds.

Drawing inferences. Drawing inferences is the most important phase in a content analysis. It applies the stable knowledge about how the variable accounts of coded data are related to the phenomena the researcher wants to know about.

Validation. Validation is the desideratum of any research effort. However, validation of content analysis results is limited by the intention of the technique to infer what cannot be observed directly and for which validation evidence is not readily available.

C. Search engine

Our proposed search engine is a program that has three parts: (1) The first part searches patent documents for specified keywords and returns a list of the documents where the keywords were found. Then, the engine will use data and text mining technology to design a specified algorithm (first algorithm) in order to analyze the legal documents and try to find out the most similar patents or patent group. (2) Next, the engine searches the patent infringement judgements for specific keywords and returns a list of the documents as above patent documents by introducing the content analysis technology into specified design algorithm (second algorithm) in order to analyze the infringement cases/precedents. It also finds the nearest infringement judgements/precedents. (3) Finally, the engine uses different analysis technologies in two different

 ISBN: 978-1-61208-408-4

databases to render a cross comparison to generate a possible result algorithm (third algorithm) with the introduction of big data concepts.

A search engine is really a general class of programs. However, the term is often used to specifically describe systems. Our proposed search engine core technologies are used to analyze patent infringement content and to use algorithms and the comparative analysis between two databases in order to generate accurate result.

D. The framework of patent threat analysis search engine

The framework of patent threat analysis search engine is depicted in Figure 1. The top part of Figure 1 represents the content analysis research process [9]. We implement the patent infringement judgement is this process. The process framework needs to be designed as algorithm.

The middle of Figure 1 is the framework of patent documents analysis process. The process includes SAO structure extraction (NLP) and patent characteristic measurement and visualization (MDS). Here, we attempt to generate the patent map. In this phase, the study has generated some results based on past research.

The lower part in Figure 1 represents the proposed search engine core technology. The study will construct the knowledge and technology database in order to support the findings of particular products that are likely to be sued, technology trends, and the threat of patent infringement. The cross patent comparison and analysis will also utilize big data concepts to construct the algorithm.

V. EXPECTED RESULT AND FUTURE WORK

This study aims to develop a search engine similar to Google for patent analysis. When the user enters a keyword, the engine does an analysis and will inform on the related patents as well as potential patent threats. It can also provide the technology trend analysis.

The study aims to employ different analysis methods to analyze different databases and further use the analysis results by cross-comparison. An accurate algorithm in different fields can be constructed and achieved in patent threat analysis.

The next step will be to employ the image recognition functions to identify drawings and pictures. If the search engine has the capability to analyze drawings and pictures, the accuracy of the results will be increased in the future.

REFERENCES

[1] C. J. Fall, A. Torcsrari, K. Benzineb, and G. Karetka, Automated categorization in the international patent classification. “SIGIR Forum”. 2003, pp.10-25. 37(1).

[2] S.H. Huang, H.R. Ke, and W.P. Yang, Structure clustering for Chinese patent documents. “Expert system with application”. 2008, pp.2290-2297.34.

[3] S.H. Huang, C.C. Liu, C.W. Wang, H.R. Ke, and W.P. Yang, Knowledge annotation and discovery for patent analysis. “International Computer Symposium”. 2004, pp.15-20.

[4] H. Park, J. Yoon, and K. Kim, “Identification and evaluation of corporations for merger and acquisition strategies using patent information and text mining” Scientometrics. April 2013, pp.883–909.

[5] J. Michel, and B. Bettels, “Patent citation analysis: a closer look at the basic input data from patent search reports”, Scientometrics. 2001, pp.185-201. Vol.51. no. 1.

[6] I. S. Kang, S.H. Na, J. Kim, and J.H. Lee, Cluster-based patent retrieval. “Information Processing & Management”. 2007, pp.1173-1182.43(5).

[7] J.H. Kim, and K.S. Choi, Patent document categorization based on semantic structural information. “Information processing & Management”. 2007, pp.1200-1215.43(5).

[8] Y.G. Kim, J.H. Suh, and S.C. Park, Visualization of patent analysis for emerging technology. “Expert System with Applications”. 2008, pp.1804-1812.34(3).

[9] K. Krippendorff, Content analysis In E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, and L. Gross (Eds.), International encyclopedia of communication New York, NY: Oxford University Press. 1989, pp.403-407.Vol. 1.

[10] K. Krippendorff, “Content Analysis An Introduction to Its Methodology” second Edition, Sage Publications, Inc. 2004.

[11] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 1964, pp.1-27.29(1).

[12] L.S. Larkey, Some issues in the automatic classification of U.S. patents. In: Working notes for the AAAI-98 workshop on learning for text categorization. 1998, pp.87-90.

[13] Larkey L.S. A patent search and classification system. In: Proceedings of the fourth ACM conference on digital libraries. 1999, pp.79-87.

[14] Larkey L.S., Connell, M.E., and Callan, J. Collection selection and results merging with topically organized US patents and TREC data. In Proceedings of ninth international conference on informaiton knowledge and management. 2000, pp.282-289.

[15] D. Lin, Dependency-based evaluation of MINIPAR. In A. Abeille(Ed.), Treebanks: Building and using parsed corpora. Dordrecht: Kluwer, 2003, pp.317-332.

[16] U. Schmoch, Evaluation of technological strategies of companies by means of MDS maps. “International Journal of Technology Management”, 1995, pp.4-5.10(4-5).

[17] Stanford. The Stanford parser: A statistical parser. from http://nlp.standford.edu/software/lex-parser.shtml. Retrieved March 2013.

[18] T. Joachims “Text Categorization with Support Vector Machines: Learning with Many Relevant Features” University Dortmund Informatik LS8, Baroper Str. 301 44221 Dortmund, Germany.

[19] A.J.C. Trappey, F.C. Hsu, C.V. Trappy, C.I. Lin, Development of a patent document classification and search platform using a back-propagation network. “Expert Systems with Applications”. 2006, pp.755-765.31(4).

[20] Y.H Tseng, Y.M. Wang, Y.I. Lin, C.J. Lin, and D.W. Juang, Patent surrogate extraction and evaluation in the context of patent mapping. “Journal of Information Science”. 2007, pp.718-736.33(6).

[21] V. K. Gupta, and N. B. Pangannaya, Carbon nanotubes; bibliometric analysis of patents, “World Patent Information”. Sep. 2000, pp.185-189.Vol.22,issue 3.

[22] Y. Liang, R. Tan, and J. Ma, “Patent Analysis with Text Mining for TRIZ” IEEE ICMIT. 2008, pp.1147-1151.

[23] Y.L. Chen, and Y.C. Chang, A three-phase method for patent classification ” Information Processing and Management”. 2012, pp.1017-1030.48.

 ISBN: 978-1-61208-408-4

[24] Y.L. Chen, and Y.T. Chiu, Vector space model for patent documents with hierarchical class labels ”Journal of Information Science”. 2012, pp.222-233.38(3)

[25] Y.L. Chen, and Y.T. Chiu, An IPC-based vector space model for patent retrieval “Information Processing and Management”. 2011, pp.309-322.47.

[25] Y. H. Tseng, C. J. Lin, and Y. I. Lin, “Text mining for patent mapanalysis”. Information Processing & Mangement”. Sep. 2007, pp.1216-1247. vol.43, issue 5.

[26] Y.H. Tseng, C.J. Lin, and Y.I. Lin, “Text mining techniques for patent analysis ” Information Processing and Managemnet”. 2007, pp.1216-1247.43.

[27] The Stanford Natural Language Processing Group, The Stanford Parser: A statistical parser, http://nlp.stanford.edu/software/lex-parser.shtml

[28] MINIPAR is a broad-coverage parser for the English language. http://webdocs.cs.ualberta.ca/~lindek/minipar.htm

[29] Knowledgist retrieves, analyzes, and organizes information into a meaningful, robust, personal knowledge base. https://invention-machine.com/

 ISBN: 978-1-61208-408-4

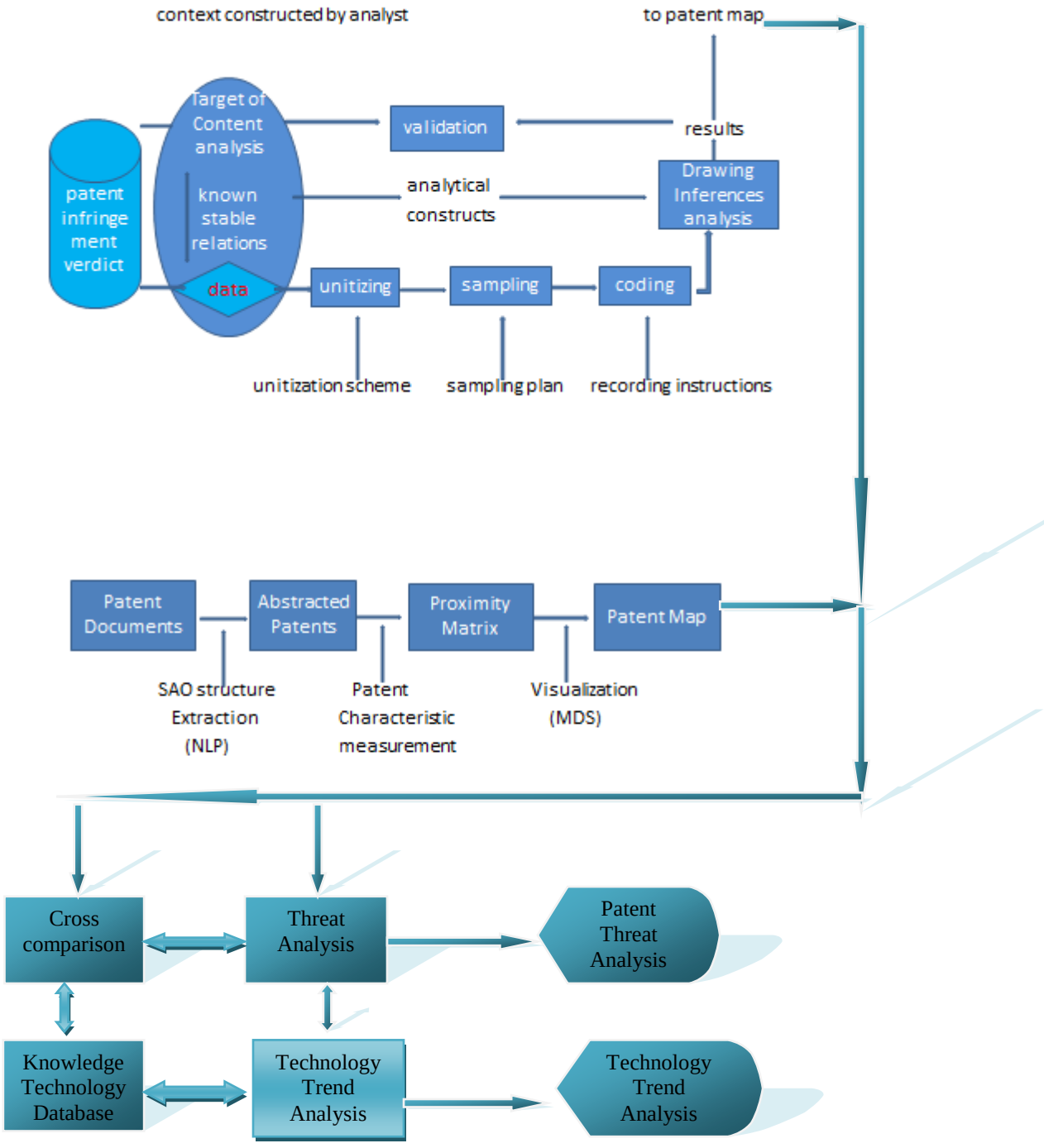

Figure 1. The framework of patent threat analysis search engine

 ISBN: 978-1-61208-408-4