Patent Analysis for Organization based on Patent Evolution Model

Yunji Jang
KISTI, UST
Korea Institute of Science and Technology Information, KISTI
University of Science and Technology, UST
Daejeon, South Korea
yunji@kisti.re.kr

Jangwon Gim
KISTI
Korea Institute of Science and Technology Information, KISTI
Daejeon, South Korea
jangwon@kisti.re.kr

Jinpyo Lee
KAIST
Korea Advanced Institute of Science and Technology, KISTI
Daejeon, South Korea
jinpyo.lee@cs.kaist.ac.kr

Do-Heon Jung
KISTI
Korea Institute of Science and Technology Information, KISTI
Daejeon, South Korea
heon@kisti.re.kr

Hanmin Jung
KISTI
Korea Institute of Science and Technology Information, KISTI
Daejeon, South Korea
jhm@kisti.re.kr

ABSTRACT
With the rapid progress in science and technology in recent times, new research fields are being discovered and studied each day and numerous research findings are being published and presented. Considering this, organizations from various countries have been investing considerable effort to bring about internal and external changes in their organizations. In fact, at many organizations, studies are being carried out to derive meaningful results by analyzing research outcomes. Thus, in this paper we propose an evolution model through analysis of the patent titles from one specific institution. First, we classified the keyword of title according to properties of keyword and then defined the relation case of patent. After, we suggest the evolution model of relation based on timeline and applied to the actual data. It can predict keyword of future patent by applying to actual data.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Data Mining
I.6: [Simulation and Modeling]: Model Validation and Analysis

General Terms
Theory

Keywords
Patent Data, Patent Analysis, Patent Keyword, Patent Evolution, Organization Patent

1. INTRODUCTION
With the rapid progress in science and technology in recent times, new research fields are being discovered and studied each day and numerous research findings are being published and presented. In several countries including South Korea, USA, and Japan, policies for strengthening the protection of intellectual property are being implemented. Moreover, the number of patent applications for research results has also increased [1]. Considering this, organizations from various countries have been investing considerable effort to bring about internal and external changes in their organizations. In other words, organizations believe that they may not be able to survive in today's competitive market without innovation, and therefore, much effort is being invested in that direction [2]. In fact, at many organizations, studies are being carried out to derive meaningful results by analyzing research outcomes [3,4]. In general, among the research outcomes, patent data is a type of research outcome that can be used as an indicator for measuring the technological and innovative competency of an organization. Patents are not only essential from the standpoint of copyrighting and publishing of research and development, but also for aiding future research and development plans [5].

Patent data consist of title, technology implementation details, technology category code, citation information, and owner information. Analyzing such patent information is very important because it can help in interpreting the changes in the technology, trends, level, and commercial value. Patent analysis includes various kinds of analysis such as frequency analysis, share analysis, time-series analysis, citation analysis, and rights analysis. Time-series analysis and two-dimensional analysis, in particular, are more common [6].

Table 1. Types of patent analysis

Theme	Details
Frequency Analysis	Filed country, Inventor, Applicant, Technology Classification ..
Share Analysis	Filed country, Inventor, Applicant, Technology analysis, Detailed technical classification ..
Time Series Analysis	Application rate analytics , National analysis, Inventor analysis, Applicant analysis, Technical analysis, New applications analysis, New inventor analysis ..
Correlations	Inventor correlation map, Applicant correlation map, Technical correlation map ..
Citation Analysis	Citation relation analysis , Core patent analysis ..
Rights Analysis	Patent family map ..

Several studies have attempted to analyze the characteristics of companies based on the citation relationship of patents, for example, the number of research projects that attempt to determine the technological strategy of a competing company through citation relationship has been on the rise [7-10].

These studies analyze a company based on the information of the patents the companies have applied for; however, to the best of our knowledge, no study has performed a time-series analysis of patent data yet. Therefore, in this paper, we propose a patent evolution model based on time-series analysis. Furthermore, unlike previous papers, in this paper, we only analyze the title, among the many patent data, as a patent title can adequately serve as representative information about a patent.

2. PATENT ANALYSIS MODEL

In this study, we analyze the evolution process of a patent title in four stages. Figure 1 shows procedure of patent analysis. First, after removing the useless, meaningless words from a patent title, the remaining words, keywords, which represent the purpose of the patent and the patent's core technology, are extracted. Next, while drawing a relation network to map the relation type between the extracted keywords, the evolution of relation types is examined by applying time-series analysis.

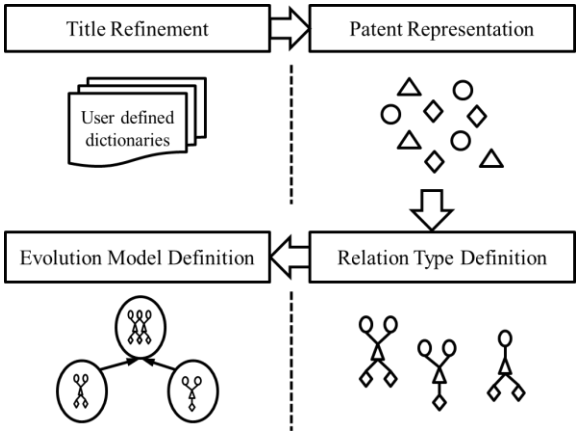

Figure 1. Patent analytics model Procedure

2.1. Title Refinement

The refining patent title stage is performed prior to the extraction of meaningful keywords such as goal and approach. In this stage, unnecessary words are removed from the patent title. On the basis of blank spaces, the patent title is divided. Considering the statistical numbers of divided words, they are removed step by step.

2.2. Patent Representation

To observe the keyword concept-based evolution process for a certain organization, the words from the refined titles are classified into Approach, Goal Object, and Goal Predicate. Goal represents keywords that indicate the purpose of the patent's invention. Since the title of a patent is the name of the invention, a goal keyword, which is the target technology, is always present. Based on the type of Patent, Approach keywords are sometimes present, which are keywords that describe the core technology used to develop the goal technology. In this study, a dictionary was built for classifying keywords into Approach and Goal. The classified Approach and Goal keywords are tagged as object and predicate through a prebuilt morpheme analyzer. When the morpheme analysis is completed, a relation network is drawn to map the relation types between Approach, Goal Object, and Goal Predicate.

2.3. Relation Type Definition

All types were defined for the types of relations for patent expressions between two or more patent titles. The relations defined in this paper refer to cases in which one or more keywords overlap among three keywords that are separated into Approach, Goal Object, and Goal Predicate.

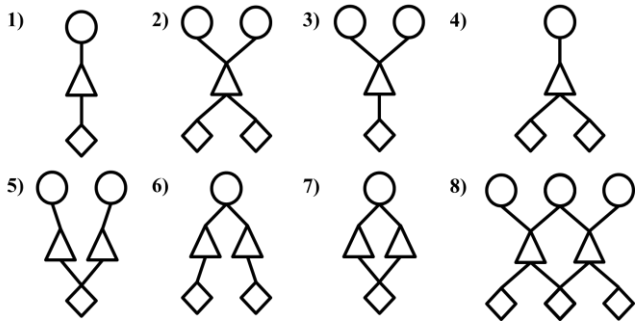

Figure 2. Patent relation types

Figure 2 shows the relation types that can be derived on the basis of Approach, Goal Object, and Goal Predicate. A circle indicates an Approach, a triangle indicates a Goal Object, and a diamond indicates a Goal Predicate. Figure 2-1) indicates one patent type. Figure 2-2) relation type is X-type. X- type is a case involving several Approaches and several Goal Predicates mapped to one Goal Object. Figure 2-3) relation type is Y- type. The Y- type is a case involving several Approaches mapped to a Goal Object and Goal Predicate pair. Figure 2-4) relation type is inverted Y- type. The inverted Y-type is a case of several Goal Predicates mapped to an Approach and Goal Object pair. Figure 2-5) relation type is V- type. In the V-type, several Goal Objects and several Approaches are mapped to one Goal Predicate. Figure 2-6) relation type is inverted V-type. In the fifth inverted V-type,

several Goal Objects and Predicates are mapped to one Approach. The ◇ type is a case involving several Goal Objects mapped to an Approach and Goal Predicate pair. Finally Figure 2-8) refers to the Double X type having the several approaches, Goal Object, and Goal Predicate. X type of relationship types can also have resulted in multiple forms, but we studied only Double X type of relationship type.

2.4. Evolution Model Definition

In this section, a definition is provided for the evolution model. The evolution model is made using the characteristics of relation types based on the time-series data of a certain organization. Figure 3 shows a model that can evolve according to the relation type. As an example of an evolvable model, "A type can evolve to B type" refers to a case where the condition of B type is satisfied when A type and B type are combined. In other words, when A type and B type are combined , it should be B type.

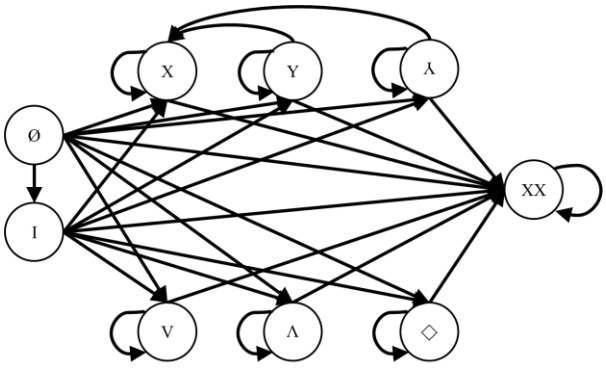

Figure 3. Evolution model

All relation types can evolve into the type of each one. Ø and I relation types can be evolved into all relative types. In addition, all relation types can evolve into all relative types. However, Y and ⅄ can evolve into the X relation types.

3. EXPERIMENT

3.1. Data Set

The data set was composed of 82 patents of the Computer Intelligence Lab of Korea Institute of Science and Technology Information (KISTI) from 2005 to 2013. At first we started 99 data but 17 titles having parallel structures were discarded.

3.2. User Defined Dictionaries

As explained in the overall process stage, a process involving analysis and tagging of a sentence structure was performed. Based on the patent title set, several word dictionaries were built. Modeling was performed to identify similar results through cognitive analysis. Tables 2 and 3 is a dictionary to translate Korean to English. Table 2 shows the part of dictionary used for refining useless words.

Table 2. Semantic stopword dictionaries

Semantic Stopword	General Stopword
System and Method	Optimum
Apparatus and method	Effective
Method and System	For
System and Apparatus	About
Framework	Included
System	At
:	:

Table 3. Distinguishing word dictionaries

A-G Distinguishing word	A-G Distinguishing word
Based on	Applied
Based	Through
Using	By
Centered	According to

Table 3 used is a terminology dictionary for classifying Approach and Goal. After classifying Approach and Goal, the remaining words are processed as useless words. In fact, the words of Korean are more than English word. So Table 3 is smaller than Korean dictionaries.

3.3. Statistics

The statistics produced when applying the unnecessary word dictionary for 82 cases are shown. Goal Object and Goal Predicate were extracted in all 82 cases, and Approach was extracted in only 44 cases. We made the statistics about the relation type and evolution model. Table 4 shows the cumulative statistics of relation cases from 2005 to 2013. The X-type type appeared most, followed by V-type.

Table 4. Statistics of relation type

	Base	X	Y	⅄	V	Λ	◇
2005	1	0	0	0	0	0	0
2006	0	1	0	0	0	0	0
2007	1	1	0	0	1	0	0
2008	3	1	0	0	0	0	0
2009	2	3	0	0	1	1	0
2010	2	4	1	1	4	2	0
2011	2	4	2	1	5	3	1
2012	3	5	3	1	5	3	1
2013	3	6	3	1	5	3	1
Sum	**3**	**6**	**3**	**1**	**5**	**3**	**1**

Table 5. Statistics of evolution model

	I → XX	I → X	X → X	X → XX	XX → XX	Y → XX	Ø → I	Ø → V	Ø → Y
2005	0	0	0	0	0	0	1	0	0
2006	0	1	0	0	0	0	0	0	0
2007	0	0	0	1	0	0	1	0	0
2008	0	0	0	0	0	0	2	0	0
2009	1	1	0	0	1	0	2	1	0
2010	0	1	0	1	3	0	1	0	1
2011	0	0	2	1	0	1	0	0	0
2012	0	1	1	0	1	0	1	0	0
2013	0	0	0	0	1	0	0	0	0
Sum	**1**	**4**	**3**	**3**	**6**	**1**	**8**	**1**	**1**

Table 5 is statistics of evolution model. It was the most abundant to evolve into the Ø types from I type. This evolution case means new patent is indicated in KISTI.

3.4. Result

In this section, we compare the actual results with the evolution model proposed in this paper.

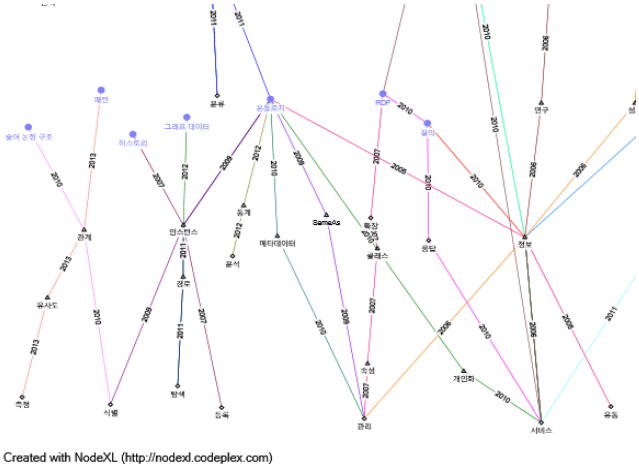

Figure 4. Relation network of KISTI patent

Figure 4 shows the final relation network of Approach, Goal Object, and Goal Predicate for 2013. The network was drawn by NodeXL [11]. The evolution of relation types was examined by drawing the relation network for patent titles keywords of 2005 to 2013, as shown in Figure 4. Figure 4 shows the evolution model produced from the KISTI patents. Figure 5 shows the evolution model produced from the KISTI patents.

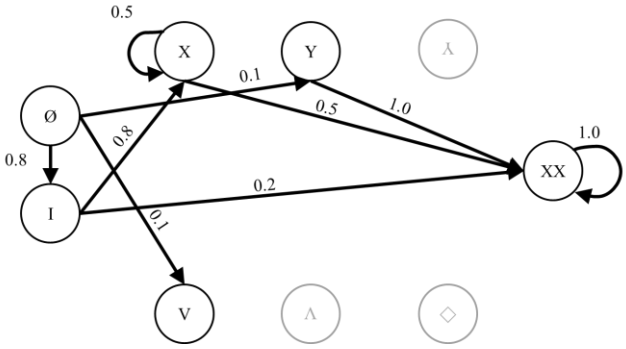

Figure 5. Evolution model in KISTI

The Evolution Model which can be discovered in KISTI patent among 30 Evolution Model is 9. The numbers in Figure 5 is the probability to go in the direction of the arrow. I type, Y type , V type are evolved from the O type. The probability of evolving into I types is 0.8, the probability of evolving into Y and V types are 0.1. The types which can be evolved from I type are X and XX type. The probability of evolving into X type is 0.8 and the probability of evolving into XX type is 0.2. The types which can be evolved from X types are X types and XX type. And each probability is 0.5. The XX type is only type which can be evolved from Y type and XX type evolves into XX type. The inversed Y type, Λ type and ◇ type of evolution model are not observed in KISTI patent. By predicting the future evolution type KISTI patent based on this result, it can be extract keywords that match the type of evolution. For example, when the patent evolves into the patent of X types from I type, we can predict that KISTI will research about Goal Object.

4. CONCLUSIONS

In this paper, we proposed the patent evolution model through the patent title analysis of the specified affiliation. We presented the new possibility by studying the Korean title which was not active in the existing research. We removed the stop words at the patent title of the specified affiliation and we separated the Approach, Goal Object, and Goal Predicate. By using separated three keywords, we drew the connection network in the patent title based on time series. It defined the types of relationships that may appear between the patent through the network connection in a given year. And then we can know the relation case of organization based on the result of test.

Patent analysis systems and related methods currently rely on basic visualization techniques and patent maps, such as bar graph, pie chart, separate table, and bubble diagram. Relation types between patents and an evolution model of relations were proposed only using the titles of patents.

In a follow-up study, we plan to perform the same analysis for a different organization, and compare it with the evolution result of this study; further, we plan to examine the expandability of the proposed model. In particular, we aim to further develop the proposed patent evolution model so that it can be applied to patent titles of other countries in addition to those of South Korea. Furthermore, we expect to use it in convergence technology prediction by predicting a relation type, in which a patent relation of certain relation type will evolve, through the evolution model.

Acknowledgments

This work was supported by the IT R&D program of MSIP/IITP. [B010-15-0353, High performance database solution development for Integrated big data monitoring and Analytics]

5. REFERENCES

[1] Sang-cheol Kil, Sung-min Kang, "The study of an Analysis on Patent Management Affecting the Company Performance, " In Proceedings of the Korea Technology Innovation Society 2008, Vol. 11, No. 2, pp. 171-193, 2008.

[2] SeongTaek Park, Youngki Kim, Taeung Kim, "A Study on Infuencing Factors of Patent Activities on Magnagement Performance," In Journal of Entrue Journal of Information Technology 2013, Vol. 12, No. 3, pp. 121-129, 2013.

[3] Hwan Lim, Hosun Rhim, Myungho Sohn, "Study for Improving the Patent Management Schema by Using Citation Index in Public-sector R&D Institute : Case Study on K Institute," In Journal of the Korea Society of IT services, Vol. 7, No. 3, pp. 115-127, 2008.

[4] Mu-hsuan Huang, Li-yun Chiang, Dar-zen Chen,. " Constructing a patent citation map using bibliographic coupling : A study of Taiwan's high-tech companies," In Proceedings of the Scientometrics , Vol. 58, No. 3, pp 489-506, 2003

[5] Sunghae Jun, Sang-Sung Park, Young-Geun Shin, Dong-Sik Jang, HoSeok Chung, "Forecasting Vacant Technology of Patent Analysis System using Self Organizing Map and Matrix Analysis," In Journal of the Korea Contents Association , Vol. 10, No. 2, pp. 462-480, 2010.

[6] Jin-Hyo Joseph Yun, Oh-Jin Kwon, Jinseo Park, Eui-seob Jeong, "A Study on the Development and Adaption of Open Innovation Analysis Model," In Proceedings of the Korea Technology Innovation Society 2010, Vol. 13, No. 1, pp. 99-123, 2010.

[7] Aidong, Peng. "The Application of Patent Citation Analysis in Enterprise Competitive Intelligence," In Proceedings of the Information Studies: Theory & Application, Vol. 27, No. 3, pp no. 16, 2004.

[8] Lingyun Sun,, Yu Song, "Research on Clustered Patent Mapping Visualization and Interaction," In Proceedings of the Computer-Aided Industrial Design and Conceptual Design, pp. 1130-1133, 2008.

[9] Jinho Choi, Junguk Kim, "Analysis of Assortativity in the Keyword-based Patent Network Evolution," In Journal of Internet Computing and Services, Vol. 14, No. 6, pp. 107-115, 2013.

[10] Xian Wu,, Li Zhang, " SIMPLE: A Strategic Information Mining Platform for Licensing and Execution," In Data Mining Workshops, pp. 270-275, 2009.

[11] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, Eric Gleave, "Analyzing (Social Media) Networks with NodeXL," In Proceedings of the fourth international conference on Communities and technologies, pp. 255-264, 2009.