J. Soc. Korea Ind. Syst. Eng Vol. 38, No. 2 : 17-30, June 2015
http://dx.doi.org/10.11627/jkise.2015.38.2.17
ISSN : 2005-0461(print)
ISSN : 2287-7975(online)

Quantifying the Process of Patent Right Quality Evaluation : Combined Application of AHP, Text Mining and Regression Analysis

Janghyeok Yoon* · Jaeguk Song** · Tae-Kyu Ryu***†

*Department of Industrial Engineering, Konkuk University
**Department of Economics and Statistics, Korea University
***Korea Institute of Intellectual Property

특허권리성의 정량적 평가방법에 대한 연구 : AHP, 텍스트 마이닝, 회귀분석의 활용

윤장혁* · 송재국** · 류태규***†

*건국대학교 산업공학과
**고려대학교 경제통계학과
***한국지식재산연구원

Technology-oriented national R&D programs produce intellectual property as their final result. Patents, as typical industrial intellectual property, are therefore considered an important factor when evaluating the outcome of R&D programs. Among the main components of patent evaluation, in particular, the patent right quality is a key component constituting patent value, together with marketability and usability. Current approaches for patent right quality evaluation rely mostly on intrinsic knowledge of patent attorneys, and the recent rapid increase of national R&D patents is making expert-based evaluation costly and time-consuming. Therefore, this study defines a hierarchy of patent right quality and then proposes how to quantify the evaluation process of patent right quality by combining text mining and regression analysis. This study will contribute to understanding of the systemic view of the patent right quality evaluation, as well as be an efficient aid for evaluating patents in R&D program assessment processes.

Keywords : Patent Right Quality, Program Evaluation, Analytic Hierarchy Process, Text Mining, Regression Analysis

1. Introduction

The ability to create, diffuse and accumulate intellectual property is becoming the core activity in modern knowledge-based economies, because IP is a key source for the sustainable development and future competitiveness of nations [29]. Regarding this, social ecologists have declared the future to be a knowledge-based society and said intellectual property, such as patents, will become the most important asset in economic activity [11]. In practice, leading companies and nations are increasingly formulating intellectual property strategies to secure their place in today's competitive technological environment [22].

In Korea, patents produced by government-supported R&D

Received 23 March 2015; Finally Revised 7 April 2015;
Accepted 7 April 2015
† Corresponding Author : tkryu@kiip.re.kr

<Table 1> Technology Transfer Rates and Royalties of National R&D Patents in 2008 [3]

	Korea	United States	Canada	EU
Technology transfer rate (%)	22.7	25.6	34.4	33.5
Royalty (million Korean Won/patent)	9.7	748.8	96.6	140.5

programs, called national R&D programs, have rapidly increased from 7672 in 2006 to 17969 in 2010 [35]. Despite the quantitative increase of patents, the Korean government has recognized that national R&D programs need to be further improved in light of the quality of their patents, such as technology transfer rates and royalties, to make them comparable with patents of the United States, Canada, and the EU nations (<Table 1>). For this reason, the Korean government is now putting great emphasis on the quality of national R&D programs. Several studies have stated the importance of evaluating R&D programs [15, 37], and thus it is expected that the quality evaluation of an R&D program could promote the affected research institutes and researchers to create quality patents [23].

Patents are considered to be a proxy of technological advancement, and accordingly evaluating the quality of patents produced by national R&D programs becomes an important process for qualitative evaluation of R&D programs. In particular, the patent right quality, which indicates excellence in the legal protection of a patent, has been used as a key factor for patent value assessment, together with marketability and utility [16]. The patent right quality has two main components : the patent right scope and patent right strength [23]. As the quality measures of a patent, the patent right scope indicates the legal and technological breadth of the patent right, and the patent right strength means the technological solidity of the patent right. In addition, in recent national R&D policies, evaluating the patent right quality of patents by R&D program could be a useful source by which to formulate a quality strategy for R&D programs at the government level.

However, the current approach for evaluating the patent right quality of national R&D programs has relied mostly on the human experts, who may be expensive or unavailable. In addition, the number of national R&D patents in Korea is rapidly increasing by various national R&D programs and this is making the expert-based evaluation process costly and time-consuming. In this regard, quantifying the process of patent right quality evaluation could be an efficient aid for technology-oriented R&D policy making processes.

Therefore, we suggest an approach to develop measures for patent right quality that exploit the analytic hierarchy process (AHP), text mining and multiple linear regression analysis (MLRA). The proposed approach 1) defines a systemic hierarchy of patent right quality and uses patent attorneys to evaluate the patent right quality according to the hierarchical structure using AHP, 2) extract values for the variables affecting patent right quality using text mining of patent data, and 3) applies MLRA to develop the models for patent right quality by incorporating the evaluation results by patent attorneys and the extracted values of explanatory variables. The proposed approach is illustrated using information technology (IT)-related patents from the Korean national R&D patent database. This study contributes to a systemic view of patent right quality and in addition, its models will become an efficient aid to assist human experts in evaluating patents for R&D program assessment.

2. Groundwork

The proposed approach is based on AHP, text mining, and MLRA, so this section provides overviews for those theoretical backgrounds.

2.1 Analytic Hierarchy Process in R&D Management

AHP is a method for multi-criteria decision making [39] and a structured technique composed of 1) decomposing a decision problem into a hierarchy of more easily comprehended sub-problems that can be analyzed independently, 2) evaluating various elements of the hierarchy by comparing them with respect to their effect on an element above them in the hierarchy, and 3) converting those evaluations to numerical values that can be processed and compared over the entire range of the problem [38]. Using AHP, decision-makers can identify a numerical weight or priority for each element in the hierarchy, which distinguishes AHP from other decision-making techniques [17].

By considering problems in R&D management as decision-making problems and priority identification problems, AHP-based studies have addressed issues on selecting government-sponsored R&D projects (using fuzzy AHP) [18], building technology roadmaps [13], identifying key factors to R&D program evaluation [19, 26, 44], estimating the value of several technologies [6], analyzing national competitiveness in the area of hydrogen energy technology [24], and measuring the research performance of national R&D organizations [2, 43].

Although AHP has been widely adopted for R&D program selection and evaluation, little attention has been paid to its application for patent right quality. This study uses AHP to define the importance of elements in the hierarchy of patent right quality.

2.2 Text Mining in R&D Management

Text mining, also called text data mining, is a process to analyze high-quality information from textual data [7, 41]. Generally, text mining extracts and then interprets patterns and features from textual information [20, 46]. Text mining has been widely used for studies related to clustering, classification, and retrieval because it is an efficient tool to extract significant patterns from massive textual information [4].

In patent analysis, many researchers have adopted text mining. Studies have been conducted to identify similarities between patent documents and scientific publications [27], to extract terms and vocabulary patterns in specific patent domains [30], and to identify significant keywords for efficient patent search and analysis [25]. Some text mining-based studies have suggested visualization tools for R&D planning by forecasting the patent landscape [47], identifying new business areas exploiting text mining and data envelopment analysis [40], identifying technological trends constructing subject-action-object based networks of patents [8, 49], and evaluating the risk of patent infringement [3, 33].

As made apparent by the prior studies, text mining is an effective tool to quantify content analyses of massive patent data. Therefore, it is incorporated into our approach to develop measures to evaluate a patent's legal protection capabilities.

2.3 Multiple Linear Regression Analysis in R&D Management

MLRA is an approach to model the linear relationship between a dependent variable and one or more explanatory variables [1] MLRA has been used in much research to analyze the economic effects of R&D and intellectual property. Such MLRA-based studies describe the incentive effects of R&D investment in research productivity and firm growth [28], identify R&D effects for firms' market entry condition and new product development [42], analyze the relationship among R&D investment, innovation, and economic growth in the EU [5, 10], and measure the effects of R&D investment and specialization for labor productivity growth [34]. Other studies using MLRA have identified the effect of intellectual property by analyzing the determinants of knowledge production and their effects on regional economic growth [32], analyzing collaboration patterns among regions, nations, and international partners by patent transfer [31], identifying innovation trends in NAFTA nations by exploiting an econometric analysis of patent applications [36], developing a patent alert system based on linear regression analysis of patent information [12], and forecasting patent trends of biotechnology by comparing linear regression and Poisson analysis [21].

MLRA has been widely used to identify various effects of R&D and intellectual property from macroscopic and microscopic views. In this paper, we combine MLRA with text mining to develop quantitative measures for patent right scope and strength, which are the main components of patent right quality.

3. Expert Group

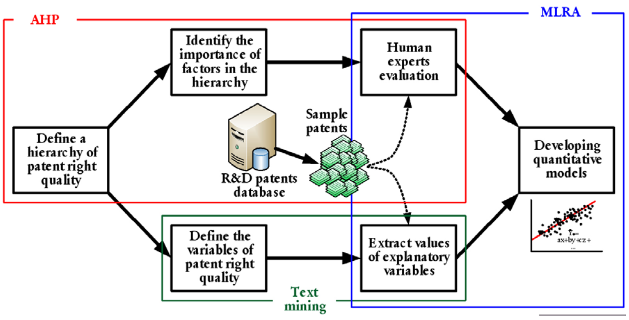
For the hierarchy definition and evaluation of patent right quality, we organized an expert group of IT-related patent attorneys (<Table 2>) with experience (avg. 7.9 years) in the IT fields of communication and network technology, robot technology, electronic device technology, and display technology. They come from various organizations, including government institutes, private companies, universities, and patent and law firms. With the expertise and knowledge of this group, we constructed a hierarchy of evaluation factors constituting patent right quality and identified the relative importance of each factor.

<Table 2> Patent Attorneys for the Expert Group

Patent attorney	Organization	Major fields in information technology	Experience
PA1	Patent and law firm	Robot technology, Internet-based technology, Communications technology	8 years
PA2	Patent and law firm	Communications technology, Broadcasting, Network technology	7 years
PA3	Government institute	Communications technology, RFID technology	8 years
PA4	Patent and law firm	Communications technology, Robot technology	10 years
PA5	Patent and law firm	Communications technology, Antenna technology	9 years
PA6	Government institute	Communications technology, Electronic devices, Display technology	7 years
PA7	Government institute	Communications technology, Electronic devices, Display technology	10 years
PA8	University	Communications technology, Semiconductor technology	6 years
PA9	Private corporation	Robot technology, Communications technology	7 years
PA10	Patent and law firm	Sensor technology, Communications technology	7 years

4. Research Framework

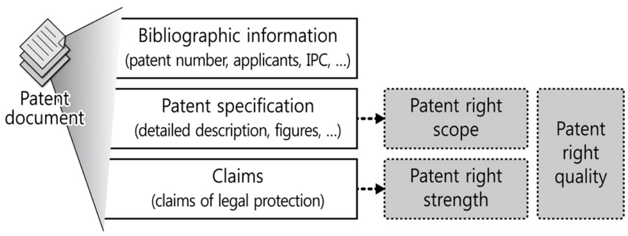
The proposed procedure to develop the measures for patent right scope and strength (which together form patent right quality) is composed of 1) defining a hierarchical structure of patent right quality, 2) identifying the importance of each element in the hierarchy, 3) defining explanatory variables for patent right scope and strength, 4) extracting values for the explanatory variables, and 5) developing quantitative models for patent right scope and strength by incorporating human experts' evaluation results and the explanatory variables' values (<Figure 1>).



<Figure 1> Overall Research Procedure

4.1 Defining a Hierarchy of Patent Right Quality

Each patent document includes all the information related to an invention and contains bibliographic information, patent specifications, and claims (<Figure 2>). The bibliographic information includes patent numbers, application dates, applicants, international patent classification (IPC), and so on. The patent specification presents a detailed description, figures, and implementation examples of the invention. The claims explicitly describe the main points for which an inventor claims legal protection [48]. According to our expert group, patent right scope is strongly related to the patent claims because the claims can be legally protected. Patent right strength is mainly related to the patent specification section, which describes technological details and supports patent right scope with detailed description, including background, implementation examples, and figures.



<Figure 2> Patent Sections Related to Patent Right Quality

Thus, we identified the evaluation factors for patent right quality and defined their hierarchical structure in meetings with our expert group. The group concluded that the hierarchy of patent right quality can be represented as three levels : Level 1 (2 factors), Level 2 (6 factors), Level 3 (12 factors) (<Table 3>). An interesting point made by the expert group was that the sub-evaluation factors and their hierarchical relationships are not restricted to technology domains; that is, the hierarchy is feasible for other domains. However, the relative importance of each sub-evaluation factor could vary by technology domain, including nanotechnology and biotechnology, so we collected experts' pairwise comparisons among the evaluation factors.

4.2 Identifying the Importance of Evaluation Factors in the Patent Right Quality Hierarchy

Using the hierarchy of patent right quality (<Table 3>), this step collected pairwise ratings of the evaluation factors at each level. To this end, we provided each patent attorney in the expert group with a pairwise comparison survey form. We designed the form so individual patent attorneys could evaluate the relative importance of all pairs of evaluation factors at each level in the hierarchy.

With respect to the pairwise comparison on each level of the decision-making hierarchy, the consistency index (CI) of each decision-maker's pairwise comparison matrix should be less than the threshold value 0.1 to ensure that each decision-maker is consistent in assigning paired comparisons [39]. Otherwise, the comparison results for a decision-maker should be reconsidered or excluded in calculating the importance of evaluation factors [45]. This paper identified all experts' CI values with respect to Level 2 of the hierarchy; Level 1 and Level 3 have only two evaluation factors with respect to the factor above them, so they do not require a consistency analysis. All of 10 patent attorneys were adequately consistent in assigning pairwise comparisons because their CI values were all less than 0.1 (<Table 4>).

Thus, this step identified the importance of each evaluation factor in the hierarchy using the pairwise comparison matrices of individual experts and AHP. Generally, the pairwise comparison matrices of human experts at each level can be integrated into a pairwise comparison matrix for AHP using the geometric mean [45]. We identified the importance of the evaluation factors from a hierarchical perspective (<Table 5>). Patent right scope (0.725) was significantly more important than patent right strength (0.275) in Level 1. In Level 2, breadth of claims (0.379), ease of substantiation (0.215) and sufficiency of examples (0.133) were found crucial. In Level 3, the most important evaluation factor was appropriateness of claim contents (0.275). Thus the critical factor in patenting inventive knowledge is how broadly and properly the patent describes the claim points. On the other hand, the sufficiency of drawings (0.016) and sufficiency of description (0.030) were found to be less important. Overall, evaluation factors that pertain to patent right scope were found to be most important.

<Table 3> The Hierarchical Structure of Patent Right Quality

	Level 1	Level 2	Level 3	Definition
Patent right quality	Patent right scope	[L2-1] Breadth of claims	[L3-1] Appropriateness of claim contents	How broadly claims describe the legal protection scope
			[L3-2] Clarity of claim contents	How clearly claims describe the legal protection scope
		[L2-2] Adequacy of claims	[L3-3] Differentiation of claim contents	How well claims are differentiated from one another
			[L3-4] Reflection of examples	How well claims reflect examples in the specification
		[L2-3] Ease of substantiation	[L3-5] Appropriateness of claim composition	How proper the structure of claims is
			[L3-6] Concreteness of claim contents	How specific and concrete claims are
	Patent right strength	[L2-4] Sufficiency of examples	[L3-7] Diversity of examples	How various examples are in the specification
			[L3-8] Appropriateness of examples	How appropriate examples are in the specification
		[L2-5] Adequacy of description	[L3-9] Clarity of description	How clearly specification describes inventive knowledge
			[L3-10] Support for claims	How strongly technological description supports claims
		[L2-6] Ease of implementation	[L3-11] Sufficiency of description	How sufficiently specification describes inventive knowledge
			[L3-12] Sufficiency of drawings	How sufficiently drawings describe inventive knowledge

<Table 4> Consistency Indexes of Patent Attorneys

	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9	PA10
CI (patnt right scope)	0.043	0.054	0.037	0.051	0.008	0.002	0.069	0.016	0.026	0.038
CI (patent right strength)	0.063	0.026	0.037	0.009	0.047	0.007	0.069	0.037	0.037	0.004

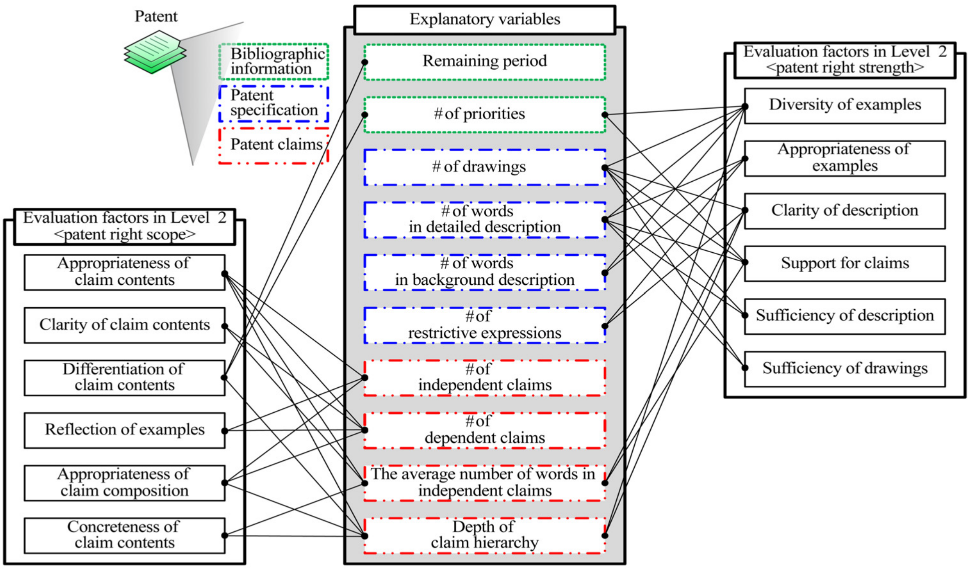
<Table 5> Importance of Evaluation Factors of Patent Right Quality

	Level 1	Level 2	Level 3
Patent right quality	Patent right scope (0.725)	[L2-1] Breadth of claims (0.379)	[L3-1] Appropriateness of claim contents (0.275)
			[L3-2] Clarity of claim contents (0.104)
		[L2-2] Adequacy of claims (0.132)	[L3-3] Differentiation of claim contents (0.061)
			[L3-4] Reflection of examples (0.070)
		[L2-3] Ease of substantiation (0.215)	[L3-5] Appropriateness of claim composition (0.106)
			[L3-6] Concreteness of claim contents (0.108)
	Patent right strength (0.275)	[L2-4] Sufficiency of examples (0.133)	[L3-7] Diversity of examples (0.088)
			[L3-8] Appropriateness of examples (0.045)
		[L2-5] Adequacy of description (0.096)	[L3-9] Clarity of description (0.040)
			[L3-10] Support for claims (0.055)
		[L2-6] Ease of implementation (0.046)	[L3-11] Sufficiency of description (0.030)
			[L3-12] Sufficiency of drawings (0.016)

4.3 Defining the Variables of Patent Right Scope and Strength

Defining dependent variables and their explanatory variables is a prerequisite to constructing our quantitative measures. In this step, we defined the relevant explanatory variables for patent right scope and patent right strength. Explanatory variables need to be derived carefully by analyzing their relationship with the dependent variables. To this end, we defined the explanatory variables using the hierarchy of patent right quality.

Related to the evaluation factors in Level 2 of the hierarchy, the expert group identified 10 explanatory variables (<Table 6>). The explanatory variables, including "the number of drawings," "the number of independent claims," "the number of dependent claims," and "depth of the claim hierarchy tree," are all extractable by computational analysis of patent documents. Some variables, for example "the number of independent claims" and "depth of claim hierarchy tree," can be identified by textual analysis, and some, such as "remaining period of a patent" and "the number of priorities," can be directly identified from patent bibliographic information. The expert group found that each explanatory variable relates to one or more evaluation factor in the hierarchy (<Figure 3>). Some variables, such as "the number of independent claims," directly affect both patent right scope and strength, and some, such as "the number of drawings," relate only to patent right strength. Although "the number of citations" has been widely used for patent analysis, we do not use citations because their use is not mandatory in Korean patents.



<Figure 3> Links between Evaluation Factors and Explanatory Variables

<Table 6> Explanatory Variables

Explanatory variables	Description and calculation
Remaining period (exp_dur)	Definition : A patent right can be maintained for a certain period of time. In most patent laws, the period is 20 years from its application date. A granted patent with a long remaining period implies more chance of using the patent right. Calculation : The remaining period (days) of a granted patent from the present to its expiration date.
The number of priorities (pri_cnt)	Description : Patent owners often apply for more advanced inventions from their previous patents. The number of priorities of a patent is the number of previous patents by an applicant whose later patent directly succeeds its previous patents. A patent with many priorities is likely to be an advanced invention being steadily developed by the applicant. Calculation : The number of priorities stated in the patent bibliographic information.
The number of drawings (img_cnt)	Description : Inventors use drawings in a patent to describe inventive peculiarities, so the number of drawings can be a good measure for the sufficiency of examples. Calculation : The number of drawings numbered in the patent specification
The number of words in detailed description (wrd_cnt_of_inv)	Description : The detailed description in the patent specification is textual information that explicitly describes the inventive knowledge. A patent with sufficient explanation is likely to have more examples and provide solid support for the relevant claims. Calculation : The number of words in the section 'detailed description of the invention' except for irrelevant words such as 'a,' 'the,' 'this,' and 'above.'
The number of words in background description (wrd_cnt_of_bgr)	Description : In general, a patent's 'background' section describes relevant previous studies or patents. A lengthy background indicates that the invention relies on much previous work, which might indicate a weak patent right strength. Calculation : The number of words in paragraphs related to background technology and previous patents in the 'background' section.
The number of restrictive expressions (wrd_cnt_of_lim)	Description : Frequent use of restrictive expressions implies that the invention does not work under specific conditions. Therefore, restrictive expressions in the patent specification suggest weak patent right strength. Calculation : The number of restrictive expressions, including 'always,' 'essential,' 'advisable,' 'only,' and 'have to,' in the 'detailed description of invention' section.
The number of independent claims (clm_ind_cnt)	Description : Patent claims are a unique system to protect an invention. Independent claims, which stand on their own, are typically written with very broad terms to prevent competitors from circumventing the claim. Therefore, an invention with many independent claims has a strong possibility of being broadly protected. Calculation : The number of claims that do not include dependency-related expressions such as 'according to claim 1,' or 'according to claims 1 and 2.'
The number of dependent claims (clm_dep_cnt)	Description : An applicant can protect his or her invention using patent claims. Dependent claims, which depend on a single claim or on several claims, generally express particular embodiments. Although the legal scope of dependent claims is generally narrower than that of the independent claims on which they depend, an invention with many dependent claims has a strong possibility of being legally protected in various specific cases. Calculation : The number of claims in 'claims' section that include dependency-related expressions.
The average number of words in independent claims (wrd_cnt_of_clm_ind)	Description : To construct the legal protection scope of an invention, independent claims are typically written using broad language and the fewest possible terms. Because the doctrine of equivalents must be applied to individual elements of the claim under the "all elements" rule [9], it is advisable that the descriptions of independent claims be concise and non-restrictive. Calculation : The average number of words in independent claims except irrelevant terms such as 'a,' 'the,' 'this,' and 'said.'
Depth of claim hierarchy (clm_depth)	Description : Generally, a dependent claim depends on one or more other claims; therefore patent claims form a hierarchical structure. If the hierarchy of claims disclosed is deep and wide, the relevant invention has a well-designed claim structure and may have more implementation examples for patentability. Calculation : Depth of the hierarchy that consists of independent claims and dependent claims.

4.4 Extracting Values for the Explanatory Variables

This step extracted values for the explanatory variables by analyzing patent information such as bibliographic information, patent specification, and claims. Generally, the information in patent documents available online can be stored in an electronic format, such as a text file or a Microsoft Excel file, from the patent database. Some explanatory variables, including "remaining period" and "the number of priorities," could easily be identified by simple computation of numerical information, whereas other variables, such as "the number of detailed descriptions," and "depth of claim hierarchy," could be identified by the analysis of textual information (<Table 7>).

4.5 Developing Quantitative Models

In this step, we collected values for the dependent variables, patent right scope and strength, through expert evaluation of the sample patents. To this end, the 10 experienced patent attorneys of our expert group were asked to evaluate the patent right scope and strength for the sample patents on a scale from 1 to 9. Then we calculated the average evaluation scores for the sample patents (<Table 8>). We then used those scores as the dependent variable to develop the measures for patent right scope and strength.

We then used MLRA to construct the quantitative models for patent right scope and strength using the values of the independent variables and the experts' evaluations. For MLRA of the models, we summarized the distribution of the variables (<Table 9>) and conducted a correlation matrix examination (<Table 10>). The matrix suggested that the explanatory variables are not highly inter-correlated; generally, correlation of 0.7 to 0.8 indicates a strong positive correlation, 0.5 to 0.6 is normal positive correlation, and below 0.4 is weak correlation [14].

<Table 7> Part of the Extracted Values of Explanatory Variables by Patent

No.	img_cnt	pri_cnt	wrd_cnt_of_inv	wrd_cnt_of_bgr	wrd_cnt_of_lim	exp_dur	clm_ind_cnt	clm_dep_cnt	wrd_cnt_of_clm_ind	clm_depth
1	3	0	798	1128	1	5489	2	4	73	5
2	0	0	898	469	16	5489	1	6	86	4
3	6	0	1614	1025	8	5495	1	8	63	7
4	0	0	0	233	1	5496	2	8	75	6
5	4	0	735	666	4	5496	2	2	119	2
6	10	1	588	261	2	5555	4	11	59	2
7	4	1	1024	168	3	5555	3	11	97	4
8	7	1	1709	119	20	5560	3	7	80	5
9	10	0	2723	119	8	5576	2	14	106	4
10	3	0	570	319	2	5590	2	0	92	1
11	7	1	1179	195	10	5595	3	3	105	2
12	6	1	2074	277	16	5609	2	11	72	4
13	5	1	891	286	10	5623	2	1	125	2
14	4	0	2138	124	14	5629	2	10	131	7
15	2	0	282	583	2	5714	2	4	87	4
16	5	0	1548	62	7	5726	3	7	63	4
17	9	0	2119	138	7	5836	2	12	92	5
18	11	2	2512	240	10	5836	3	5	69	2
19	4	0	1292	108	4	5842	2	5	113	4
20	5	1	1886	55	4	5850	3	7	69	4
21	6	0	1344	101	6	5850	1	8	44	6
22	6	0	1349	182	3	5852	2	7	67	3
23	4	0	351	181	1	5857	4	8	67	2
24	13	2	2013	125	13	5860	6	2	59	3
25	3	0	437	362	0	5866	2	5	82	2
26	4	0	1762	247	8	5866	2	5	61	4
27	3	0	940	251	2	5867	2	7	58	5
28	9	1	1199	133	2	5903	1	3	129	2
29	9	1	1197	567	5	6009	4	5	87	3
30	6	1	925	96	12	6009	2	8	74	2

<Table 8> Part of Patent Right Scope and Strength Scores Rated by Experts

Patent No.	Patent right scope	Patent right strength
1	5.233777	4.549417
2	5.452298	4.747161
3	5.280572	5.37653
4	5.822792	5.570177
5	4.885788	4.377724
6	6.939445	7.282836
7	6.889643	6.080939
8	6.794416	6.632871
9	6.343191	6.799507
10	5.380303	5.657249
11	4.668036	5.467177
12	6.213554	6.005187
13	3.840363	5.384569
14	5.201337	5.473421
15	5.563765	4.699029
16	6.379359	6.427544
17	6.006129	7.218328
18	6.710558	7.495839
19	5.145406	5.827548
20	6.787952	5.901394
21	6.170734	5.594226
22	5.627114	5.967372
23	6.185864	5.448515
24	7.802952	7.512878
25	6.401863	5.883161
26	6.262363	5.961182
27	6.479258	5.845572
28	6.040666	6.560814
29	6.494853	6.165679
30	5.812291	5.759182

<Table 9> Means and Standard Deviations

Variable codes	Mean	Std. Dev.	Min.	Max.
y_strength	5.33	0.95	3.14	7.51
y_scope	5.18	1.15	1.37	7.80
img_cnt	6.66	3.27	0	15
pri_cnt	0.14	0.40	0	2
wrd_cnt_of_inv	1581.16	929.78	0	4303
wrd_cnt_of_bgr	333.43	200.19	0	1128
wrd_cnt_of_lim	8.84	6.63	0	34
exp_dur	5621.52	272.27	4785	6086
clm_ind_cnt	2.28	1.16	1	7
clm_dep_cnt	7.58	4.80	0	30
wrd_cnt_of_clm_ind	94.44	60.91	21	452
clm_depth	3.4	1.40	1	7

The regressions for the patent right scope and strength turned out to be proper : an adjusted R^2 of 60.9% for patent right scope (<Table 11>) and an adjusted R^2 of 54.4% for patent right strength (<Table 12>). Next, the general F-test for the model significance rejected the null hypothesis on the effect of all explanatory variables in each MLRA model, so the two models were found significant.

In the analysis of the patent right scope, we found that all 6 explanatory variables were significant (<Table 11>). The only variable with a negative effect on patent right scope was "the average number of words in independent claims" : when that variable increases 1, the patent right scope decreases by 0.007. Some of the explanatory variables had strong positive effects on patent right scope : "the number of priorities" (0.8661), "the number of independent claims" (0.2798), and "depth of claim hierarchy tree" (0.2227). Of particular note, "depth of claim hierarchy tree" was newly introduced as an explanatory variable in this research, and its effect suggests that a deep and wide claim hierarchy is significantly positive for patent right scope.

<Table 10> Pearson's Coefficient Matrix

	img_cnt	pri_cnt	wrd_cnt_of_inv	wrd_cnt_of_bgr	wrd_cnt_of_lim	exp_dur	clm_ind_cnt	clm_dep_cnt	wrd_cnt_of_clm_ind	clm_depth
img_cnt	1									
pri_cnt	0.1745	1								
wrd_cnt_of_inv	0.3808***	-0.0112	1							
wrd_cnt_of_bgr	0.0826	-0.2233	-0.0605	1						
wrd_cnt_of_lim	0.2867***	0.0236	0.6026	-0.0719	1					
exp_dur	-0.0141	0.1805	-0.0707	-0.2446**	-0.1343	1				
clm_ind_cnt	0.1857	0.2841***	0.2006**	-0.0891	0.0772	0.2387*	1			
clm_dep_cnt	0.3185***	-0.1312	0.4423	-0.099	0.3748***	-0.0657	0.0942	1		
wrd_cnt_of_clm_ind	0.0269	-0.0697	0.327***	0.0965	0.1901	-0.1595	-0.2146**	-0.1637	1	
clm_depth	0.0675	-0.1362	0.1859	0.1421	0.093	-0.1311	-0.1761	0.5074	-0.2108**	1

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

<Table 11> Regression Result for Patent Right Scope

F(6, 93) : 26.700, Prob > F : 0.000, R-squared : 0.633, Adj R-squared : 0.609.

Definition	Dependent variable	Coef.	Std.Err.	t	P > t
	y_scope				
The number of priorities	pri_cnt	0.8661	0.191601	4.52	0.000***
Remaining period of patent	exp_dur	0.0005	0.00028	1.95	0.055*
The number of independent claims	clm_ind_cnt	0.2798	0.070996	3.94	0.000***
The number of dependent claims	clm_dep_cnt	0.0430	0.018126	2.37	0.020**
The average number of words in independent claims	wrd_cnt_of_ind	-0.0070	0.001275	-5.53	0.000***
Depth of claim hierarchy tree	clm_depth	0.2227	0.06396	3.48	0.001***
	constant	0.9342	1.62148	0.58	0.566

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

<Table 12> Regression Result for Patent Right Strength

F(7, 92) : 17.900, Prob > F : 0.000, R-squared : 0.577, Adj R-squared : 0.544

Definition	Dependent variable	Coef.	Std.Err.	t	P > t
	y_strength				
The number of drawings	img_cnt	0.1001	.0222265	4.50	0.000***
The number of priorities	pri_cnt	0.8404	.1695954	4.96	0.000***
The number of words in detailed description	wrd_cnt_of_inv	0.0003	.0000982	3.45	0.001***
The number of words in background description	wrd_cnt_of_bgr	-0.0007	.0003438	-2.12	0.037**
The number of restrictive expressions	wrd_cnt_of_lim	-0.0293	.0122268	-2.40	0.019**
The number of words in independent claims	wrd_cnt_of_ind	-0.0049	.0012023	-4.09	0.000***
Depth of claim hierarchy tree	clm_depth	0.1291	.0504159	2.56	0.012**
	constant	4.5408	.2591309	17.52	0.000

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

In the analysis of patent right strength (<Table 12>), we found that all 7 explanatory variables were significant. "The number of drawings," "the number of words in detailed description," and "depth of claim hierarchy tree" all had positive relations with patent right strength. "The number of priorities" and "depth of claim hierarchy tree" were important factors in both patent right scope and patent right strength. "The number of words in background description," "the number of restrictive expressions," and "the average number of words in independent claims" had a negative effect on patent right strength. Among the variables, "the number of priorities" had the strongest positive relation (0.8404) and "the number of restrictive expressions" had the strongest negative relation (-0.0293).

The linear models for patent right scope and strength are used as the measures to estimate the legal excellence of national R&D patents.

5. Evaluating the Patent Right Quality of IT-Related National R&D Programs

The developed models for patent right scope and patent right strength can be used to evaluate the intellectual property dimension of national R&D programs. To introduce an application of the developed models, we chose 10 national R&D programs that produced many patents. Then we used textual processing of patent descriptions and claims to extract the values for our explanatory variables for patent right scope and strength from 203 patents. By putting the values into the measures, we calculated the patent right scope and strength for each patent and compared the evaluation results among the R&D programs; the average patent right scope score was 5.121, and the average patent right strength score was 5.253 (<Table 13>).

<Table 13> Quantified Evaluation of the Selected National R&D Programs

R&D program title	# of patents	Patent right scope		Patent right strength	
		Average	Rank	Average	Rank
Fourth generation mobile communication technology	35	5.582	2	5.648	1
Mobile WiMAX	25	5.276	3	5.002	9
E-3G based multi-media convergence technology	22	5.852	1	5.637	2
Interoperable ubiquitous computer development	14	5.011	6	5.074	6
Wearable interface technology for disaster relief	21	4.729	8	5.347	4
Data capturing mobile communication technology	18	4.630	9	5.046	8
Hybrid antenna technology for mobile broadcasting services	12	5.213	5	5.056	7
Interoperable humanoid technology based on distributed network	27	5.268	4	5.363	3
Intelligent robot sensors	14	4.961	7	5.252	5
Wearable mobile interface technology for the physically handicapped	15	3.692	10	4.542	10
SUM	203	5.115		5.261	

Among the programs, the legal protection scope of patents in the R&D program "E-3G based multimedia convergence technology" was found to be the widest. In fact, according to the in-depth examination of patent attorneys, the program's patents were overall evaluated as well maximizing their legal protection scope by making full use of broad language and the fewest possible terms in their claims. Furthermore, the patent right scope-related numerical factors of the program's patents were definitely superior to those of other programs: independent claims = 2.56 (average = 2.25), dependent claims = 10.31 (average = 7.59), and depth of claim hierarchy = 2.63 (average = 3.46). From an aspect of patent right strength, the R&D program "the fourth generation mobile communication technology was found to be the most solid. Numerically, the average number of its patents' drawings (5.82) was similar to the average number of drawings (6.06) of all R&D programs, but patents of the program had a large number of priorities of 0.20 (average = 0.15). Interestingly, the program "Mobile WiMAX" was relatively high-ranked in patent right scope but low-ranked in patent right strength. This suggests that, in an overall sense, patents of that program are somewhat weak in supporting their claims by concretizing the relevant inventions with textual descriptions and drawings; in that program's patents, the number of words in detailed description was 1674.7 (average = 1809.6), the number of drawings was 5.02 (average = 6.06), and the depth of claim hierarchy tree was 2.2 (average 3.46).

In the case study, we compared only the average excellence among some national R&D programs from the views of patent right scope and patent right strength because the number of patents produced by each R&D program varied. However, the measures of this research have the potential to quantify the evaluation of the legal protection capability of patents, so they contribute by assisting experts, including patent attorneys and technology experts, as they deal with massive patents in the R&D program evaluation process.

6. Conclusions and Challenges Remaining

The patent right quality of patents produced by national R&D programs is obviously considered a leading indicator to assess the quality of program outcomes because the patents can be the only legal mechanism to secure the economic and technological value of R&D programs' final results. Through this research, we learned some lessons in designing and evaluating patents for better R&D outcomes.

First, scientists and engineers should take a slightly different approach to protecting their inventive knowledge. According to our analysis, patent right scope was much more important than patent right strength. In IT-related patents, more specifically, the broadness of patent rights was 2.64 times more important than their technological solidity. This result provides inventors with a significant implication in patenting their inventive knowledge. Many researchers in science and engineering fields tend to be technologically detailed, but our results suggest that they should change their approach. For example, under the "all elements rule," each claim is anticipated only if a single reference discloses each and every claimed element. According to the rule, claims must be written well with broad language and an appropriate minimum of technological terms to broaden the patent right scope.

Second, the hierarchical structure of evaluation factors for patent right quality holds, but their importance may vary by technological field. Despite the need to evaluate patent right quality of national R&D patents, little attention has been paid to defining the evaluation factors constituting the hierarchy of patent right quality and identifying the importance of the factors for better intellectual property evaluation. Because the current evaluation process has relied only on the knowledge of experts, evaluating patents has been subjective and biased. In this aspect, this research has a contribution in that it defines a detailed structure of patent right quality in a hierarchical way. However, although the hierarchical structure can apply in various technology domains, the relative importance of the sub-evaluation factors in the hierarchy may vary by field, including nanotechnology, biotechnology, environment technology, and space technology. Although identifying the importance of evaluation factors in a patent right quality hierarchy helps experts' R&D program evaluations to be balanced and nonbiased, expert-based evaluation is still time-consuming and costly. Thus, our measures for patent right scope and strength of patent right quality can be used as an efficient tool in the experts' patent evaluation process.

Quality evaluation of national R&D programs can contribute to the creation of high-quality intellectual property. Therefore, we developed measures for patent right scope and strength, which constitute patent right quality, using AHP, text mining and regression analysis. Our research contributes to a systemic view of patent right quality and our measures will become an efficient tool to assist human experts in evaluating patents for R&D program assessment. Furthermore, it holds the potential to become a basis for quantifying the process of patent evaluation.

Despite its contribution, there still exist some challenges in this research. First, the AHP results of the proposed method can be applied only to evaluating patents of IT-related R&D programs. In fact, the importance of evaluation factors in the patent right quality can vary according to technology fields, so future research will need to identify the importance of the evaluation factors in other fields. Second, the method did not use patent citation information, which has been widely adopted as an important factor for patent quality, because Korean patents do not require citing patents, and the sampled patents for our analysis were too recent to be cited by other later patents. Therefore, using U.S. patents in future research will identify the significance of the number of forward citations and further develop the measures for patent right quality. Finally, although this paper extracted values for some explanatory variables by using the number of words, future research will introduce natural language processing technology for more accurate textual analysis.

References

[1] Aiken, L.S., West, S.G., and Pitts, S.C., *Multiple linear regression*. Handbook of psychology, 2003.

[2] Banwet, D. and Deshmukh, S., Evaluating performance of national R&D organizations using integrated DEA-AHP technique. *International Journal of Productivity and Performance Management*, 2008, Vol. 57, No. 5, pp. 370-388.

[3] Bergmann et al., Evaluating the risk of patent infringement by means of semantic patent analysis : the case of DNA chips, *R&D Management*, 2008, Vol. 38, No. 5, pp. 550-562.

[4] Berry, M.W., *Survey of text mining : clustering, classification, and retrieval*, 2004, Springer-Verlag New York Inc.

[5] Bilbao-Osorio, B. and Rodriguez-Pose, A., From R&D to innovation and economic growth in the EU. *Growth and Change*, 2004, Vol. 35, No. 4, pp. 434-455.

[6] Chiu, Y.-J. and Chen, Y.-W., Using AHP in patent valuation. *Mathematical and Computer Modelling*, 2007, Vol. 46, No. 7-8, pp. 1054-1062.

[7] Choi et al., *Patent function network analysis : A function based approach for analyzing patent information*. 2010,

[8] Choi et al., SAO network analysis of patents for technology trends identification : a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, 2011, Vol. 88, No. 3, pp. 863-883.

[9] Chunlin, S. and King-Lien, L., The strategy of designing around existing patents in technology innovation : Case study of critical technology of OTFT. *Journal of Chinese Entrepreneurship*, 2010, Vol. 2, No. 3, pp. 270-281.

[10] Coccia, M., What is the optimal rate of R&D investment to maximize productivity growth?. *Technological Forecasting and Social Change*, 2009, Vol. 76, No. 3, pp. 433-446.

[11] Cohen, W.M. and Merrill, S.A., *Patents in the knowledge-based economy*, 2003, National Academies Press.

[12] Dereli, T. and Durmusoglu, A., A trend-based patent alert system for technology watch. *Journal of Scientific*

and Industrial Research, 2009, Vol. 68, No. 8, p. 674.

[13] Gerdsri, N. and Kocaoglu, D.F., Applying the analytic hierarchy process (AHP) to build a strategic framework for technology roadmapping. *Mathematical and Computer Modelling*, 2007, Vol. 46, No. 7-8, pp. 1071-1080.

[14] Green, W.H., *Econometric analysis*. International Edition, 2000.

[15] Gueorguieva et al., The Program Assessment Rating Tool and the Government Performance and Results Act. *The American Review of Public Administration*, 2009, Vol. 39, No. 3, pp. 225-245.

[16] Hirose, Y. and Hiruma, F., *Patent Valuation Model and PatVM*, 2005 : Toyo Keizai.

[17] Ho, W., Integrated analytic hierarchy process and its applications-A literature review. *European Journal of Operational Research*, 2008, Vol. 186, No. 1, pp. 211-228.

[18] Huang, C.C., Chu, P.Y., and Chiang, Y.H., A fuzzy AHP application in government-sponsored R&D project selection. *Omega*, 2008, Vol. 36, No. 6, pp. 1038-1052.

[19] Imoto, S., Yabuuchi, Y., and Watada, J., Fuzzy regression model of R&D project evaluation. *Applied Soft Computing*, 2008, Vol. 8, No. 3, pp. 1266-1273.

[20] Janasik, N., Honkela, T., and Bruun, H., Text Mining in Qualitative Research. *Organizational Research Methods*, 2009, Vol. 12, No. 3, p. 436.

[21] Jun, S. and Uhm, D., Technology Forecasting Using Frequency Time Series Model : Bio-Technology Patent Analysis. *Journal of Modern Mathematics and Statistics*, 2010, Vol. 4, No. 3, pp. 101-104.

[22] KIIP, *Development and application of indicators for intellectual property competitiveness*, 2010, Korea Institute of Intellectual Property : Seoul.

[23] KIIP, *Development of Quantitative Patent Quality Measures for Government Supported Research and Development Programs*. 2010, Korea Institute of Intellectual Property : Seoul.

[24] Lee, S.K., Mogi, G., and Kim, J.W., The competitiveness of Korea as a developer of hydrogen energy technology : The AHP approach. *Energy policy*, 2008, Vol. 36, No. 4, pp. 1284-1291.

[25] Li, Y.R., Wang, L.H., and Hong, C.F., Extracting the significant-rare keywords for patent analysis. *Expert Systems with Applications*, 2009, Vol. 36, No. 3, pp. 5200-5204.

[26] Liang, W.Y., The analytic hierarchy process in project evaluation : an R&D case study in Taiwan. *Benchmarking : An International Journal*, 2003, Vol. 10, No. 5, pp. 445-456.

[27] Magerman, T., B. Van Looy, and X. Song, Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*, 2010, Vol. 82, No. 2, pp. 289-306.

[28] Merito, M., Giannangeli, S., and Bonaccorsi, A., Do incentives to industrial R&D enhance research productivity and firm growth? Evidence from the Italian case. *International Journal of Technology Management*, 2010, Vol. 49, No. 1, pp. 25-48.

[29] OECD, *The knowledge-based economy*, 1996, OECD : Paris.

[30] Oostdijk, N., Verberne, S., and Koster, C., *Constructing a broad-coverage lexicon for text mining in the patent domain*, 2010.

[31] Ortega, J.L., Collaboration patterns in patent networks and their relationship with the transfer of technology : the case study of the CSIC patents. *Scientometrics*, 2011, Vol. 87, No. 3, pp. 657-666.

[32] Parent, O. and LeSage, J. P., Determinants of knowledge production and their effects on regional economic growth. *Journal of Regional Science*, 2011.

[33] Park, H., Yoon, J., and Kim, K., Identifying patent infringement using SAO based semantic technological similarities, *Scientometrics*, 2012, Vol. 90, No. 2, pp. 515-529.

[34] Piras, G., Postiglione, P., and Aroca, P., Specialization, R&D and productivity growth: evidence from EU regions. *The Annals of Regional Science*, 2010, pp. 1-17.

[35] RIPIS, *R&D Intellectual Property Information System*. 2011 [cited 2011 Dec 21]; Available from : rndip.or.kr.

[36] Rodriguez, J.C. and Gomez, M., Innovation Trends in NAFTA Countries : An Econometric Analysis of Patent Applications. *Journal of Technology Management and Innovation*, 2011, Vol. 6, No. 3, pp. 116-125.

[37] Rogers, G.M. and Spencer, D., *Report of the Advisory Committee for GPRA Performance Assessment FY 2007*. Washington, DC : NSF-07-207, 2007.

[38] Saaty, T.L. and Vargas, L.G., *The seven pillars of the analytic hierarchy process*. Models, Methods, Concepts and Applications of the Analytic Hierarchy Process, 2001, pp. 27-46.

[39] Saaty, T.L., How to make a decision : the analytic hier-

archy process. *European journal of operational research*, 1990, Vol. 48, No. 1, pp. 9-26.

[40] Seol, H., Lee, S., and Kim, C., Identifying new business areas using patent information : A DEA and text mining approach. *Expert Systems with Applications*, 2011, Vol. 38, No. 4, pp. 2933-2941.

[41] Srivastava, A. and Sahami, M., *Text mining : Classification, clustering, and applications,* 2009, Chapman and Hall/CRC.

[42] Stam, E. and Wennberg, K., The roles of R&D in new firm growth. *Small Business Economics*, 2009, Vol. 33, No. 1, pp. 77-89.

[43] Tikoria, J., Banwet, D., and Deshmukh, S., Performance measurement of national R&D organisations using analytic hierarchy process : a case of India. *International Journal of Innovation and Regional Development*, 2009, Vol. 1, No. 3, pp. 276-300.

[44] Wang, J., Lin, W., and Huang, Y.H., A performance-oriented risk management framework for innovative R&D projects. *Technovation*, 2010, Vol. 30, No. 11-12, pp. 601-611.

[45] Wei, C.C., Chien, C.F., and Wang, M.J.J., An AHP-based approach to ERP system selection. *International Journal of Production Economics*, 2005, Vol. 96, No. 1, pp. 47-62.

[46] Yamada, Y., Kato, K., and Hirokawa, S., *Text Mining for Analysis of Interviews and Questionnaires*, 2011.

[47] Yang et al., Enhancing patent landscape analysis with visualization output. *World Patent Information*, 2010, Vol. 32, No. 3, pp. 203-220.

[48] Yoon, J. and K. Kim, An Automated Method for Identifying TRIZ Trends from Patents. *Expert Systems with Applications*, 2011, Vol. 38, No. 12, pp. 15540-15548.

[49] Yoon, J. and Kim, K., Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics*, 2011, Vol. 88, No. 1, pp. 213-228.

ORCID

Janghyeok Yoon | http://orcid.org/0000-0002-8701-0695
Jae Guk Song | http://orcid.org/0000-0001-7229-2418
Tae-Kyu Ryu | http://orcid.org/0000-0002-1928-5099