

A Function-Based Knowledge Base for Technology Intelligence

Janghyeok Yoon, Namuk Ko, Jonghwa Kim*

Department of Industrial Engineering, Konkuk University, Seoul, Korea

Jae-Min Lee, Byoung-Youl Coh, Inseok Song

Korea Institute of Science and Technology Information, Seoul, Korea

(Received: March 6, 2015 / Accepted: March 12, 2015)

ABSTRACT

The development of a practical technology intelligence system requires a knowledge base that structures the core information and its relationship distilled from large volumes of technical data. Previous studies have mainly focused on the methodological approaches for technology opportunities, while little attention has been paid to constructing a practical knowledge base. Therefore, this study proposes a procedure to construct a function-based knowledge base for technology intelligence. We define the product-function-technology relationship and subsequently present the detailed steps for the knowledge base construction. The knowledge base, which is constructed analyzing 1110582 patents between 2009 and 2013 from the United States Patent and Trademark Office database, contains the functional knowledge of products and technologies and the relationship between products and technologies. This study is the first attempt to develop a large-scale knowledge base using the concept of function and has the ability to serve as a basis not only for furthering technology opportunity analysis methods but also for developing practical technology intelligence systems.

Keywords: Function-Based Knowledge Base, Technology Intelligence, Product-Function-Technology Relationship, Patent, Text Mining

* Corresponding Author, E-mail: jhkim@konkuk.ac.kr

1. INTRODUCTION

Technology intelligence is an activity that enables firms to identify the technological opportunities or threats that could affect the future growth and survival of their businesses (Ashton and Stacey, 1995; Lichtenthaler, 2004; Kerr *et al.*, 2006). Thus, a technology intelligence system can be defined as a computational system or software that identified new technological opportunities, application areas and emerging trends and delivers this information to decision makers in a useable form (Yoon and Kim, 2012; Park *et al.*, 2013; Chang and Fan, 2014). With the transformation toward a globally competitive technological environment, technology intelligence is

attracting much attention (Lichtenthaler, 2004), because it is expected to play a crucial role in identifying new, promising and valuable technological opportunities and developing them earlier than competitors in order to be technologically competitive (Yoon, 2008).

Advancement of information technology over the last two decades has enabled technical data to be accumulated systematically and retrieved and utilized efficiently. Patents, which can be accessed publicly via patent database services, are a typical type of large-scale technical data (Kasravi and Risov, 2007). Therefore, technology intelligence studies have proposed systematic technology opportunity discovery processes of collecting and processing specific patent data. In particular,

regarding the textual analysis of patents, some studies have visualized technological objects, such as patents and technological key concepts related to specific products and technologies, in the forms of maps and networks to identify the competition trends of research and development (R&D) (Yoon *et al.*, 2013; Lee and Hong, 2014), technology packages (Yoon *et al.*, 2002), and the possibility of patent infringement (Bergmann *et al.*, 2008; Cascini and Zini, 2008; Park *et al.*, 2012). Some have attempted to forecast promising technological applications and future R&D directions by applying novelty detection algorithms and network analysis indicators (Yoon and Kim, 2012; Geum *et al.*, 2013), and to find new technological items through idea generation approaches, including morphological analysis and technology vacuum analysis (Yoon and Park, 2005; Yoon and Park, 2007; Lee *et al.*, 2009).

The approaches for technology intelligence have been conducted developmentally so far and their main interests have been to identify opportunities and threats from within specific technological fields; that is, the methods have focused on analyzing restricted areas using relatively small sets of patents. However, technical data in each field are increasing explosively and moreover, the technical domains are actively converging into a new paradigm of technological evolution. Therefore, identifying the core information about technology opportunities from a large volume of technical data and across various technology fields is becoming a more important issue for the practicality of technology intelligence (Chen *et al.*, 2012). In this regard, a knowledge base could be a solution that facilitates technology intelligence methods for identifying opportunities in the large-scale and wide range of patent data.

A knowledge base is generally defined as a set of heuristically or automatically defined pieces of factual information about the world, including their hierarchical and associative relationships (Hayes-Roth *et al.*, 1983). Therefore, a knowledge base for technology intelligence should consist of factual and relational information about technological entities, such as products and technologies, which are extracted and inferred from a large amount of technical data. Such a knowledge base can contribute much to both widening the identification range of technological opportunities and improving the efficiency of technology intelligence processes. First, the knowledge base allows for identification of technological opportunities without restriction within specific technology areas, because it contains the core facts and their relationships which are distilled from a large volume and wide range of technical data. Second, it is a self-contained database that structures the pre-processed core information, ensuring that it is not necessary to collect and preprocess the technical data for specific target technology fields every time the processes for technology intelligence are executed.

The relationship between products and technologies need to be defined as loosely-coupled, because if

these relationship is rigidly fixed by word-level definitions, it would be difficult or impossible to identify new technical opportunities derived from across the domains (Yoon *et al.*, 2011). In addition, the information format for representing the products and technologies should not only be structured for usability of the knowledge base but also be able to express the specific technological principles and effects of the products and technologies.

The concept of function can be a key element to fulfill these requirements. Considering that products and technologies fundamentally exist to provide desired functions, the function is a powerful concept to explain the characteristics of a product or technology and to link among them flexibly (Park *et al.*, 2013). The function information, which comprises the actions that affect objects or change their attributes, represents what a system does or is for (Savransky, 2002). The function information appears in technical documents, including patents, is the combinational form of a verb phrase and a noun phrase, for example, ‘verb + object’ (Yoon *et al.*, 2011). In particular, in design studies, the function has been one of the fundamental building blocks used to represent what a product and its component accomplish (Stone and Wood, 2000). Moreover, the functions in patent documents have been considered represent the key concepts of the inventions and the know-how and expertise of the inventors (Moehrle *et al.*, 2005).

Therefore, this study proposes a function-based knowledge base, as part of developing a technology intelligence system. To this end, we define the product-function-technology relationship and then suggest detailed steps for constructing the knowledge base, using computational textual analysis of large-scale patent documents. This study is a novel first attempt to develop a knowledge base that contains the functional knowledge of products and technologies from patents in a wide range of technical fields. The knowledge base developed in this study can serve as a basis for furthering the processes of technology intelligence and for implementing technology intelligence systems.

2. RELATED WORK

The proposed procedure is built on the function approach. Therefore, in this section, we review the basic concept of function and present the function-based studies of technology intelligence.

2.1 The Concept of Function

The function is a representative element that describes “what a system does.” The function has a long history in design studies, and in particular, functional analysis has been widely adopted in both product design and information system design. This approach helps designers break complex systems into more manageable parts at an abstract level; for example, product design

studies have used the functional structures of products (Gero, 1990; Umeda *et al.*, 1996; Kitamura and Mizoguchi, 2004) and developed the functional basis for effective definition and reuse of product components (Hirtz *et al.*, 2002). Information system design studies have used use-case diagrams and activity diagrams to describe what an information system and its subcomponents achieve (Bennett *et al.*, 2006). In these studies, functions represent the desired output of a system (Suh, 1990) and have strong relationships with the objectives that should be achieved by the system or its components (Yoon *et al.*, 2011). In TRIZ (Russian acronym of the theory of inventive problem solving), the function is a useful concept to identify problems and generate innovative solutions, because it represents not only useful functional relationships between system components, but also harmful, ineffective, excessive, and missing functional relationships (Mann, 2002; Souili and Cavallucci, 2013).

Using textual analysis of technical documents, the function can be identified from the canonical form of the Subject-Action-Object (SAO) structure (Choi *et al.*, 2012; Choi *et al.*, 2013). An SAO structure is a syntactically ordered sentence comprising a subject (noun phrase), an action (verb phrase), and an object (noun phrase), and it represents the clear relationship between a subject and an object using an action (Yoon and Kim, 2011). In an SAO structure, S and O may refer to components of the system, and A may refer to functions performed by and on the components (Cascini *et al.*, 2004). For example, given an SAO structure of “soap cleans hands,” then “soap” is the S, “clean” is the A, and “hand” is the O. In this example, the objective of the system “soap” is to “clean” the “hand,” so the function of “soap” is to “clean hand.” In this way, an SAO structure contains and represents the key concept of a system. In addition, it can express the means-end relationship in a problem-solving format; the AO states the problem to be done, and the S forms the solution (Moehrle *et al.*, 2005). For this reason, the SAO structures found in the specification and claims of patents have been considered to represent the know-how and expertise of the inventor of a patent and may even reflect the key concepts of the invention (Bergmann *et al.*, 2008; Yoon *et al.*, 2013).

Although approaches exist that employ keyword-based techniques mainly based on statistics, i.e., counting term frequencies and identifying specific terms, their main limitation is that they cannot distinguish the role of a component in a technical system (Cascini *et al.*, 2004). In contrast to those keyword approaches, functions can provide more technical information than keyword can, as mentioned above. Therefore, this study adopts the concept of function to construct a knowledge base.

2.2 Function-Based Studies for Technology Intelligence

Research using functional analysis has recently been receiving much attention with respect to technol-

ogy intelligence. Many researchers have applied the concept of function to technology monitoring, technology opportunity identification, functional retrieval, and technology roadmapping.

Functional analysis originated from TRIZ, which describes patterns for solving technical problems (Altshuller, 1984). Genrich Altshuller, the originator of TRIZ, utilized functions to develop TRIZ by generalizing and abstracting technologies based on extensive analysis of about 200,000 patents, then used the functional approach to suggest the patterned technical solutions. An SAO structure, which is based on the functional approach used in TRIZ, was used to decompose a complex system into functional relationships among the system's components, and these relationships functionally organized among the components become the building blocks for the problem-solving processes of TRIZ (Mann, 2002).

The theoretical foundation was formulated by TRIZ, whereas the practical and computational analysis of technical documents has been advanced by Italian researchers. In order to represent the structure of a system in the composition of SAO structures, Cascini *et al.* (2004) developed a system called PAT Analyzer, which applied natural language processing (NLP) to technical documents in order to extract SAO structures. PAT Analyzer enables a functional search for “key concepts” instead of “keywords.” Using the functional analysis by PAT Analyzer, studies identified the TRIZ contradiction underlying a given technical system (Cascini and Russo, 2007) and measured similarity between functional trees of patents for plagiarism detection (Cascini and Zini, 2008).

Other studies used SAO structures to measure semantic patent similarities and construct patent maps and networks, thereby applying them to identifying technological implications for decision making in R&D management processes. Moehrle *et al.* (2005) suggested a method that profiles the technological capability of inventors for inventor assessment, human resource development, and R&D team-building. In the study, they extracted SAO structures from patents and mapped the patents onto a two-dimensional space by comparing similarities among patent SAO structures. Others proposed systematic methods to identify the risk of patent infringement (Bergmann *et al.*, 2008; Park *et al.*, 2012). The underlying concept behind the methods was that a high similarity between a pair of patents is likely to involve a high possibility of infringement. Patent maps based on SAO structures have been applied for the identification of dynamic technological trends over time and among competitors (Yoon *et al.*, 2013). In addition, Yoon *et al.* (2011) suggested a semantic patent network using SAO structures. They computed patent similarities using WordNet 3.0 (wordnet.princeton.edu), a semantic English thesaurus, in an automated way, and constructed the patent network based on the textual similarity of patents' SAO structures. It was a novel approach that overcame the citation-lag limitation of citing-cited rela-

tionship-based patent network analysis by suggesting several indices to identify the technological importance of patents, the characteristics of patent clusters, and the technological capabilities of patent applicants.

Some studies have focused on the use of functional information to analyze technological contents. Considering Ss and Os as nodes, and As as links, studies constructed an SAO-based network and applied network analysis techniques to analyze technological implications of the network by using patents of polymer electrolyte membrane technology (Choi *et al.*, 2011). In addition, using SAO structures of patents, studies proposed computational methods to develop technology trees (Choi *et al.*, 2012) and developed technology roadmaps (Choi *et al.*, 2013). In these studies, they grouped functions (AOs) into three types, such as effect function, attribute function, and structural function, and identified products, functions, and technologies by taking into consideration the types of SAO structures.

The previous studies described in this section have developed computational approaches for technology intelligence processes, but they have some limitations despite their contributions. The limitations arise from the experts' interventions and the range and volume of technical data. The previous approaches heavily depend on frequent interventions by technology experts to process technical data, and do not cover a large-scale technical data set over various technical fields. Therefore, construction of a function-based knowledge base remains a required task for the practical development of patent intelligence systems.

3. DESIGNING A FUNCTION-BASED KNOWLEDGE BASE

This section defines the conceptual relationship among products, technologies and functions and then presents the structure of the proposed function-based knowledge base.

3.1 Defining the Conceptual Relationship of Product-function-technology

The core building blocks for the proposed knowledge base are the functions. This is because the functions can represent what products and technologies do, that is, their objectives or effects, and conversely, products and technologies contain their functions. By general definition, a product means anything that can be offered to a market that might satisfy a want or need. Extending this meaning of product from technological perspective, in this study we consider a product as a tangible or intangible artifact that is composed of one or more core functions, which are achieved by technologies. Next, a technology lexically refers to the making, modification, usage, and knowledge of tools, machines, techniques,

systems, or methods to solve a problem, improve a pre-existing solution to a problem, or achieve a goal. In this research, we define it a method, process or technique that achieve a specific objective using other objectives, effects and components. That is, a technology has a core function and other sub-functions and structural components to achieve the core function.

Then, building on the concepts of products, technologies and functions, we can define the loosely-coupled relationship between products and technologies by engaging them via functions (Figure 1). A product provides various functions and it is produced by combinations of one or more technologies that perform those functions. Consider that product B has functions c, d, e and f; currently, product B is developed using technology 4, 5, 6 and 7. An interesting feature in this product-function-technology relationship is that a function can be achieved by different technologies with the same function. For example, function c can be achieved by technologies 3 and 4, function e can be achieved by technologies 6 and 8 and function f can be achieved by technologies 7 and 9. In other words, function c of product B is currently achieved by technology 4, but technology 4 can be potentially replaced into technology c to achieve its same function. Therefore, from a functional perspective, we can think that product B could be developed using technology 5 of product B, technology 3 of product A and technology 8 and 9 of product C.

3.2 Structure of the Function-Based Knowledge Base

Previous studies have modeled products and technologies as having a tight relationship, while our approach engaging products and technologies via functions makes the relationship between them loosely-coupled. According to function-based studies, function information is identified from SAO structures in the textual information of patents. Technically, if a product name is given, then we can extract the functions of a product from the SAO structures in the patent text, each of which contains the product name in its S, and the function information as its AO.

To construct a function-based knowledge base that concretizes the product-function-technology relationship, this study designed an entity-relationship diagram of the knowledge base (Figure 2). In the diagram, major entities, such as products, patents, saos, atomic_aos, gen_aos, technologies, interrelate with one another. The following are the detailed descriptions of the diagram entities.

- Patents (*patents*) include bibliographic and textual data, and each of patents is related to one or more SAO structures (*saos*). Each product can be variously represented in the patent text, so a product (*products*) includes a representative name of the product and it is related to one or more alternative names (*keywords*) of the product.

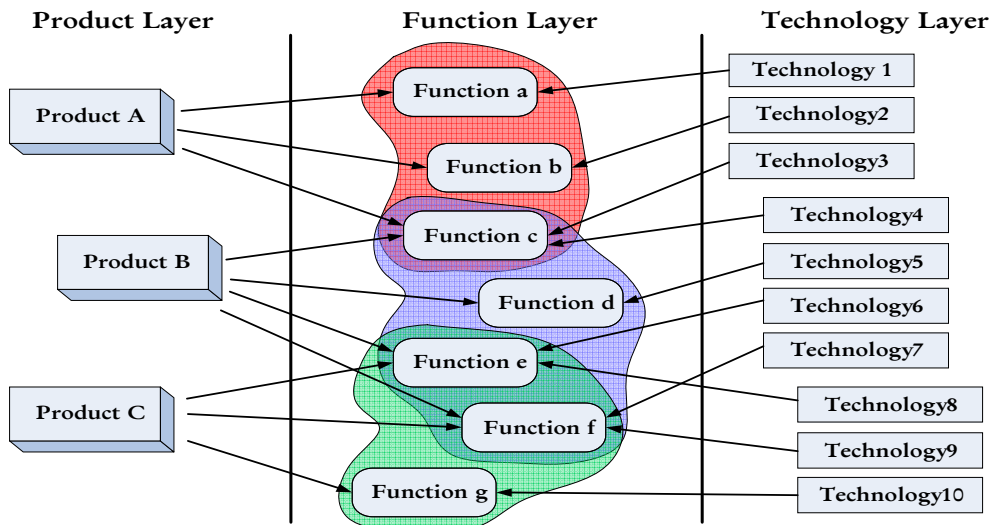


Figure 1. The product-function-technology relationship.

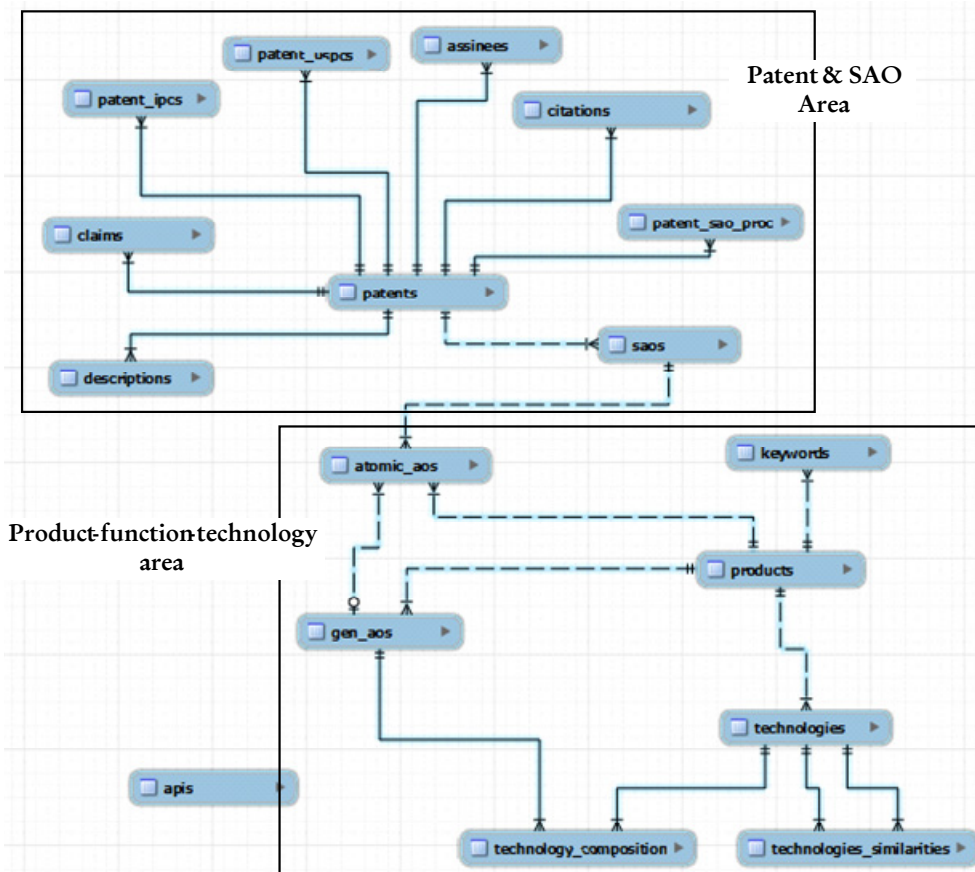


Figure 2. The structure of the function-based knowledge base.

- Some of the SAO structures extracted from patents can have product names in their S. When this is the case, atomic AOs (*atomic_aos*), which are composed of a verb and a noun, are created from their

SAO structures. An atomic AO can include the function information and structural information of its product. In addition, because there may be atomic AOs that are identical semantically with one another,

generic AOs (*gen_aos*) is created to incorporate such atomic AOs.

- Then, a product can be represented into a set of generic AOs and some generic AOs, which are considered to be relatively important by occurrence frequency, represent the functions of the product. A product can have one or more technologies (*technologies*) and the functions of a product are achieved by the technologies, each of which has a core function and is represented into the composition of various functions and structural components (*technology_composition*).

Constructing the function-based knowledge base involves storing all of the information related to these major entity types. In the next section, we will describe the steps for the knowledge base construction in more detail.

4. CONSTRUCTION PROCEDURE OF A FUNCTION-BASED KNOWLEDGE BASE

This section presents a procedure for construction of the proposed knowledge base that is composed of 7 steps (Figure 3): 1) constructing product database, 2) constructing patent database, 3) extracting patent SAO structures, 4) extracting products' atomic AOs, 5) generating products' generic AOs, 6) identifying products' core objective functions, and 7) identifying products' technologies.

4.1 Product Database Construction

A database for product name is required prior to identifying the SAO structures related to each product, because we are only interested in the SAO structures in which the Ss contain the product name. However, the product names can be variously represented in natural text; for example, an 'ac-dc converter' can be called an 'ac-to-dc converter,' ac to dc converter,' or 'ac/dc converter.' Thus, 'ac-dc converter' could be used as the representative product name and the other forms could be alternative names that indicate the same product. In this regard, it is necessary to organize such divergent product names.

To related the alternative names of product to its representative name, we utilized the results of a previous study (KISTI 2013). This study extracted product names, which include single words and compound noun phrases, from the product dictionary of the United States Patent and Trademark Office (USPTO). Then, it compared the product names with all of the noun phrases extracted from patent text. The final output of the study was a set of unique and valid product names (168048), a set of representative product names (139426), and the relationships among the unique product names and the representative names.

4.2 Patent Database Construction

Because the data source for the proposed function-based knowledge base is patent documents, it is a prerequisite to construct a patent database. This study in-

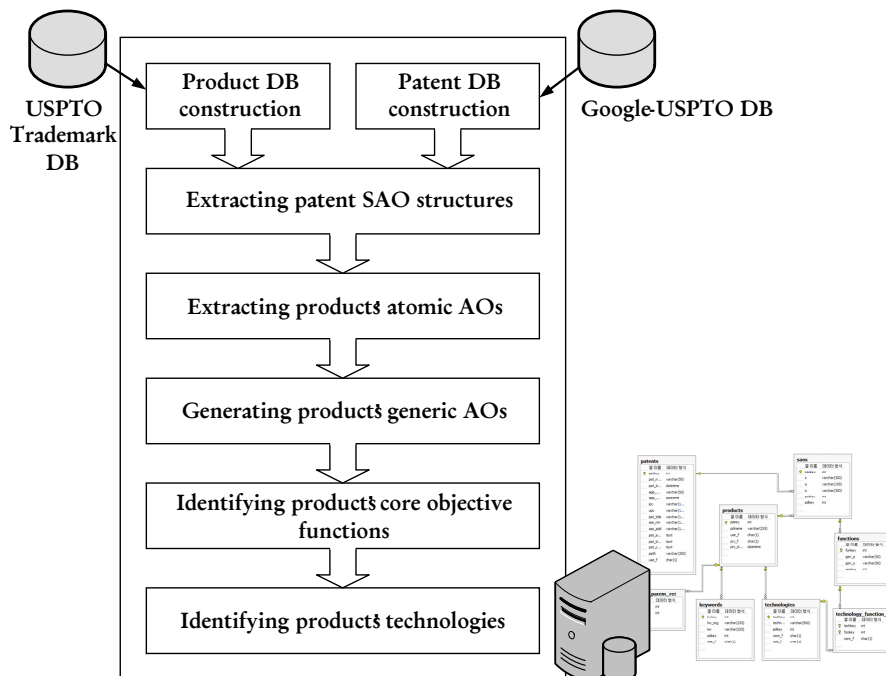


Figure 3. Overall procedure.

cluded all of the patents granted in the USPTO database over the period from 2009 to 2013; the total number of patents was 1110582.

Currently, the online website for USPTO patent bulk download service by Google (www.google.com/googlebooks/uspto.html) provides the raw data about patents, including grants, published applications, assignments, classification information and maintenance fee events. From this website, we downloaded the full text of patent grants in the form of tagged text and we were able to extract the patent details, including bibliographic information, abstract, specification and claims, by pre-processing the tagged textual information.

4.3 Extracting Patents' SAO Structures

In this step, we extracted all SAO structures from the gathered patents, using an NLP tool. There are various NLP tools available, including Stanford parser (nlp.stanford.edu), Proxem Antelope (www.proxem.com), and AlchemyAPI (www.alchemy.com). In our study, we used AlchemyAPI, which provides various text mining services, such as entity extraction, sentiment analysis and relation extraction. Among the services provided by AlchemyAPI, the relation extraction is a service for extracting SAO structures from text. In order to exploit this service, we used the full text of a patent as its input parameter and then obtained a list of the SAO structures found within the full text, as the result of its service execution.

A total of 393503554 SAO structures were extracted from all of the patents. After eliminating unavailable SAO structures, for example, S or O is an empty string and A is 'be,' we were able to obtain a total of 173291660 valid SAO structures; the average per patent was 156.04 SAO structures.

4.4 Extracting Products' Atomic AO Structures

This step extracts atomic AO structures from the SAO structures of which Ss contain the product names of our product database, because we need to identify functions that belong to the products of our product database; an atomic AO has been used as a concise and effective format with one verb and one noun to describe

function information (Yoon *et al.*, 2011; Yoon and Kim, 2012).

According to studies (Choi *et al.*, 2012; Choi *et al.*, 2013), AO information can be classified into three types: effect type, attribute type and partative type. An effect type AO represents an objective or effect of its S, an attribute type AO represents how its S affects the technology attribute and partative type AO represents inclusion relationship between its S and O. Building on the prior studies, we defined three types of atomic AOs: objective function (OF), attribute function (AF) and structural relation (SR). First, the OF type is the functions that represents a general objective that a product provides and it is composed of VS'(VS = a verb representing structure)+OT'(OT = a noun object representing technology attribute); for example, 'heat water' and 'reduce pollution' for product 'solar heater.' Second, the AF type is the functions that represents how a product changes a technical attributes, including temperature, energy, length and density, and it is composed of VS'+OT; for example, 'control temperature' and 'receive light.' Third, the SR type is the compositional information representing partative relationship between a product and its object and it is composed of VS+OT'; for example, 'include sensor' and 'contain tank.'

To computationally allocate the type of each atomic AO, this study defined a set of VSs and a set of OTs (Table 2 and Table 3) through the examination of 5 experts who are familiar with function analysis and TRIZ. First, we defined a total of 23 classes for technology attributes, incorporating functional basis (Hirtz *et al.*, 2002) and TRIZ attribute schema (Mann, 2002). To this end, we selected the top 10000 nouns that occurred the most frequently in Os of the extracted SAOs. Then, we assigned the noun to a technology attribute class, by taking into consideration whether the noun is a technology attribute. As a result, a total of 392 nouns were found technology attributes. For example, noun words 'voltage' and 'electricity' were assigned to electrical-energy class, 'rigidity' and 'flexibility' were grouped into hardness class, and 'torque' and 'rotation' were classified into mechanical-energy class.

Second, we defined a set of verbs that represent partative relationship. To this end, we selected the top 8000 verbs that occurred the most frequently in As of

Table 1. Types of atomic AOs

Types	Definition	AO composition	Example (S = solar heater)
Objective function (OF)	Functions representing the general effect or objective that a product (S) provides	VS' + OT'	Heat (A) water (O) Reduce (A) pollution (O)
Attribute function (AF)	Functions representing how a product (S) changes a technical attribute (O)	VS' + OT	Control (A) temperature (O) Receive (A) light (O)
Structural relation (SR)	Compositional information representing partative relationship between a product (S) and its object (O)	VS + OT'	Include (A) sensor (O) Contain (A) tank (O)

Note: A = VS (verb representing structure) or VS' (not VS), O = OT (noun object representing technology attribute) or OT.'

Table 2. Technology attributes' 23 classes

Classes of nouns representing technology attributes	
Weight	Hardness
Surface state	Sound/Ultrasound
Density	Color
Speed	Size
Pressure	Mechanical-energy
Temperature	Chemical-energy
Light	Biological-energy
Homogeneity	Electrical-energy
Porosity	Electromagnetic-energy
Viscosity	Magnetic-energy
Resistant	Vibration/Frequency
Slope	

the extracted SAOs and then for the verbs with the explicit part-whole meaning, we classified them into two groups: include class and constitute class. For an SAO structure, a verb of the include class type shows that the O is a part of the S and a verb of the constitute type means that the S is a part of the O. For example, verbs of the include type were 'include,' 'comprise,' 'consist' and 'encompass' and verbs of the constitute type were 'constitute', 'compose' and 'construct.'

Using the sets of OTs and VSs and the compositional definition of atomic AO types, we were able to identify all of the atomic AOs, which each product in the product database has, and their types. As a result, a total of 63072898 atomic AOs were extracted for all of the products; total OFs = 46340540, total AFs = 2496600, total SRs = 14235758. Statistically, the average number of atomic AOs per product was approximately 749; average OFs per product = 550, average AFs per product = 30, average SRs per products = 169.

4.5 Generating Products' Generic AOs

Although atomic AOs are written differently, some of them could be identical semantically and thus need to be incorporated into common AOs. This research calls the common AO expressions the generic AOs. Beside the advantage of incorporating atomic AOs, generic AOs can act as a fundamental mechanism for searching similar functions from across different domains because they allows the concept matching at an upper level.

To incorporate atomic AOs, this research employed the concept generalization. The concept generalization is a process to transform a word to its upper level word in a concept hierarchy and it can be realized by WordNet. WordNet is a large lexical database of English (Miller, 1995; Fellbaum, 2010). Noun, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms, called synsets. By means of conceptual-semantic and lexical relations, WordNet's hierarchical and relational structure makes it a useful tool for computational linguistics

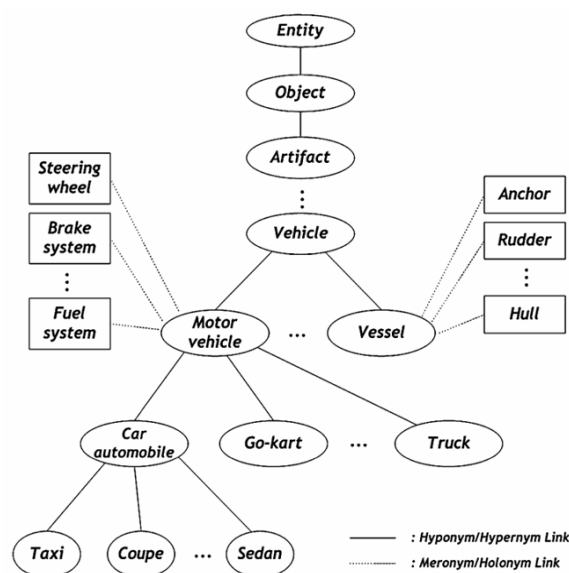


Figure 4. An example of WordNet's noun hierarchy.

and NLP (Pedersen, Patwardhan *et al.*, 2004). For example, 'motor vehicle' and 'vessel' are the sub-concepts of 'vehicle' and 'taxi,' 'coupe' and 'sedan' are the sub-concepts of 'car/automobile'; to identify common concepts, 'motor vehicle' and 'vessel' can be grouped into their upper concept 'vehicle' and 'taxi,' 'coupe' and 'sedan' can be transformed to their upper concept 'car/automobile' (Figure 4).

A product has various atomic AOs and thus it can be represented into a network that uses atomic AOs as nodes and their co-occurrence relationships as links, assuming two different atomic AOs appeared in the same patent have strong relationship. For solar heaters (Figure 5), 'heat water,' 'boil water,' 'cover film' and 'reduce pollution' are atomic AOs of the OF type, 'receive light,' 'receive ray' and 'transmit radiation' are atomic AOs of the AF type, and 'contain tank' and 'comprise tank' and 'include sensor' are atomic AOs of the SR type. However, there are many atomic AOs that can be incorporated, so we convert all of the atomic AOs into generic AOs by searching upper level concepts of As and Os from the WordNet's hierarchy; the generalization for As (verbs) was performed using the upper level concept of depth 2 and the generalization for Os (nouns) was performed using the upper level concept of depth 8. For example, 'heat water' and 'boil water' could be converted into 'change_state liquid', 'receive ray' and 'receive light' could be incorporated into 'receive radioactive_energy' and 'contain tank' and 'comprise tank' could be incorporated into 'include container'.

For all of the products in the product database, we were able to generate generic AOs by taking into consideration the types of atomic AOs. As a result, 7707417 generic AOs were identified from 65072898 atomic AOs; the average per product = 91.56 generic AOs.

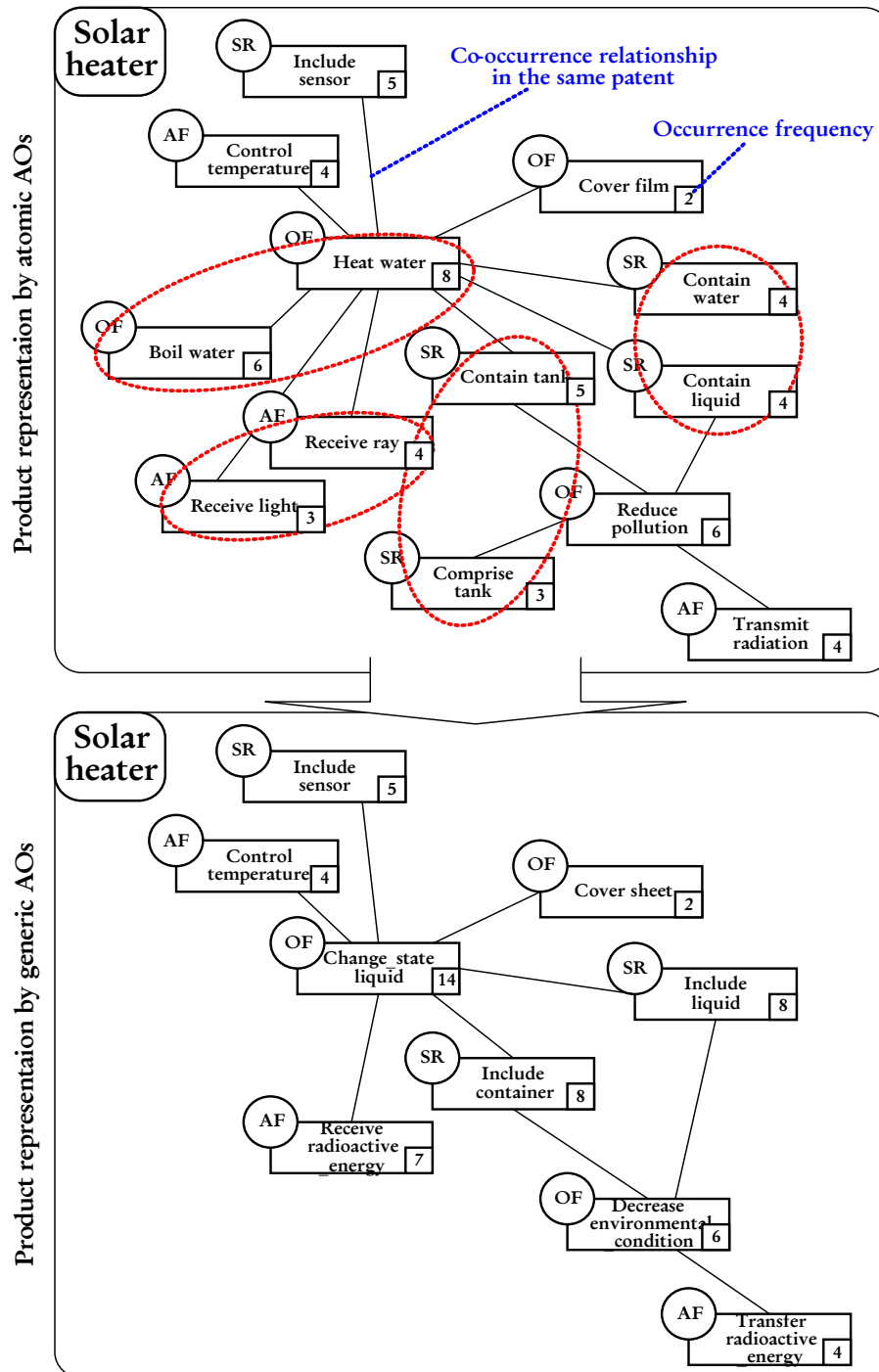


Figure 5. Product representation by atomic AOs and generic AOs.

4.6 Identifying Products' Core Objective Functions

This step identifies the core functions of each product using its generic AOs. Identifying products' core functions is a preliminary step for identifying the technologies for a product because the core functions are the starting point for constructing groups of relevant AOs;

each technology is structured by a group of a core function and its adjacent AOs. A product's core functions can be selected from its generic AOs of the OF type. This is because generic AOs of the OF type represent general effects or objectives that a product (S) provides; generic AOs of the AF type do not describe the objective of their S and generic AOs of the SR type only de-

scribe structural relations.

The idea of identifying core functions for a product is based on how often each function of the product appears. Our assumption is that generic AOs of the OF type with high occurrence frequencies have a greater likelihood of being core functions. However, because each product has a different number of generic AOs, it is difficult to use the absolute occurrence frequency of generic AOs of the OF type to identify core functions. For this reason, this study used a threshold value p , where $0 \leq p \leq 1$. By setting a proper threshold value p , we could obtain all core functions for each product. For example, $p = 0.05$ considers the core functions to be those with the top 5% of occurrences for a product. However, it is difficult to select the core functions for products with small number of generic AOs of the OF type, so we excluded products with less than 20 generic AOs of the OF type.

After trial-and-error processes, we set p to 0.02, and finally we were able to extract total 92930 core functions for 40282 product; the average per product = 2.3 core functions.

4.7 Identifying Products' Technologies

Using the core functions of each product, this step identifies technologies for the product. According to our conceptual definition, a technology has a core function and other sub-functions and structural components to achieve the core function. Building on this definition, we structure a technology as the composition of generic AOs that are adjacent to a core function.

Given two core functions OF1 and OF3 for a product, two technologies for the product can be identified

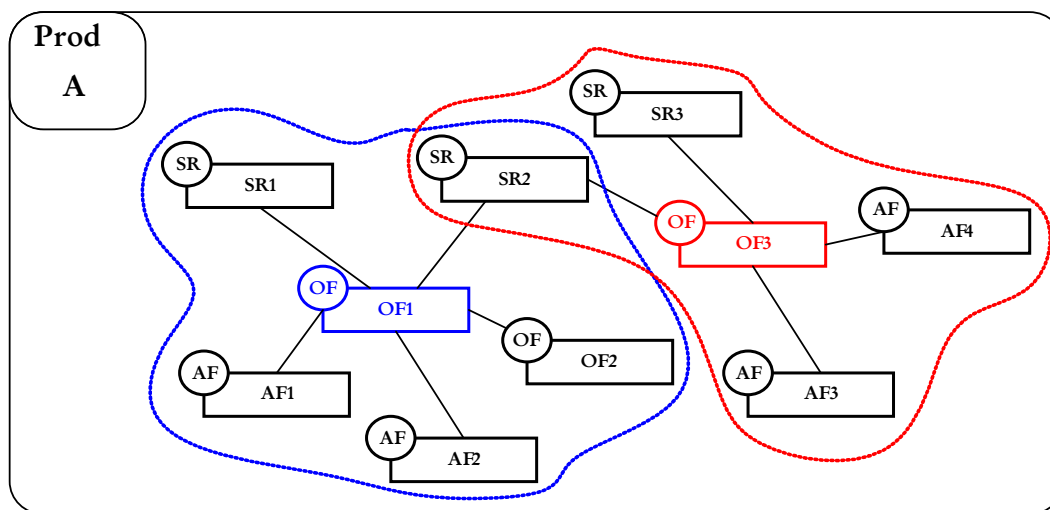
by OF1 and OF3 (Figure 6). The technology to achieve OF1 is defined as the cluster that includes OF1 as its core function and its adjacent generic AOs as sub-functions and structural components (OF2, AF1, AF2, SR1, SR2) and the technology to achieve OF3 is the cluster that is composed of OF3 and its adjacent generic AOs (AF3, AF4, SR2, SR3).

Assuming solar heaters have two core objective functions 'change_state liquid' and 'decrease environmental condition', we can structure two technologies 'solar heater tech to change_state (heat, boil) liquid (water)' and 'solar heater tech to decrease (reduce) environmental condition (pollution)' (Figure7). For the technology 'solar heater tech to change_state (heat, boil) liquid (water)' has OF type AOs, including 'cover sheet,' AF type AOs, including 'receive radioactive_energy' and 'control temperature', and SR type AOs, including 'include sensor', 'include container' and 'include liquid'.

Finally, this step was able to construct the product-function-technology relationship for a solar heater (Figure 8). In this way, we identified and structured technologies for each product's core functions. As a result, we were able to define total 92930 technologies for 40282 products; the average per product = 2.3 technologies.

5. ILLUSTRATION BY SYSTEM IMPLEMENTATION

To construct the function-based knowledge base according to the proposed procedure, we developed a .NET-based system using C#, a programming language. We incorporated several external open application programming interfaces, such as SharpNLP (shar-

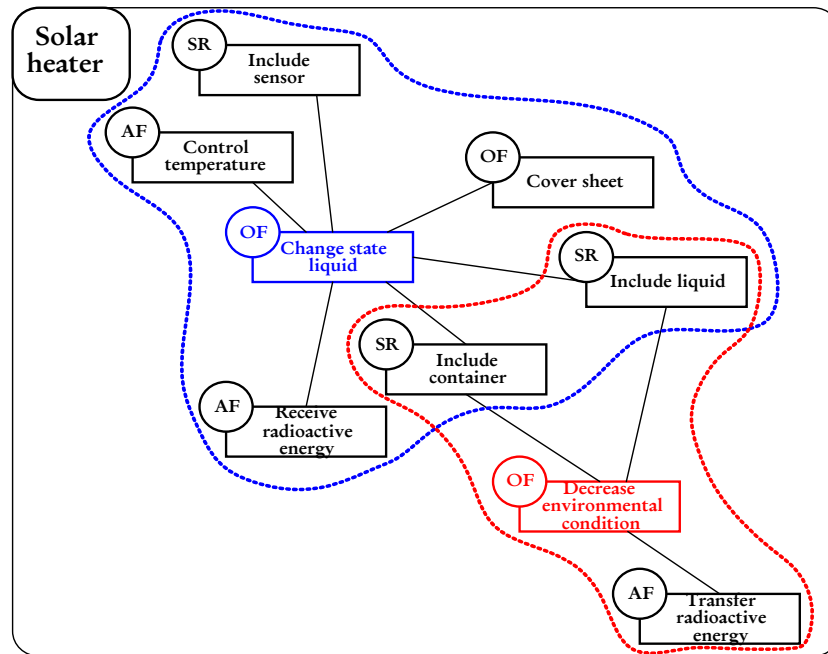


Core objective functions of product A: OF1, OF3

Technology of product A for OF1 = {OF1, OF2, AF1, AF2, SR1, SR2}

Technology of product A for OF3 = {OF3, AF3, AF4, SR2, SR3}

Figure 6. Concept of identifying technologies by core functions.



Core objective functions for solar heaters :
 → change_state(A) liquid(O) / decrease (A) environmental_condition(O)
 Two technologies by the core objective functions
 1. Solar heater tech to change_state (heat, boil) liquid (water)
 2. Solar heater tech to decrease (reduce) environmental_condition (pollution)

Figure 7. Identifying technologies of solar heaters.

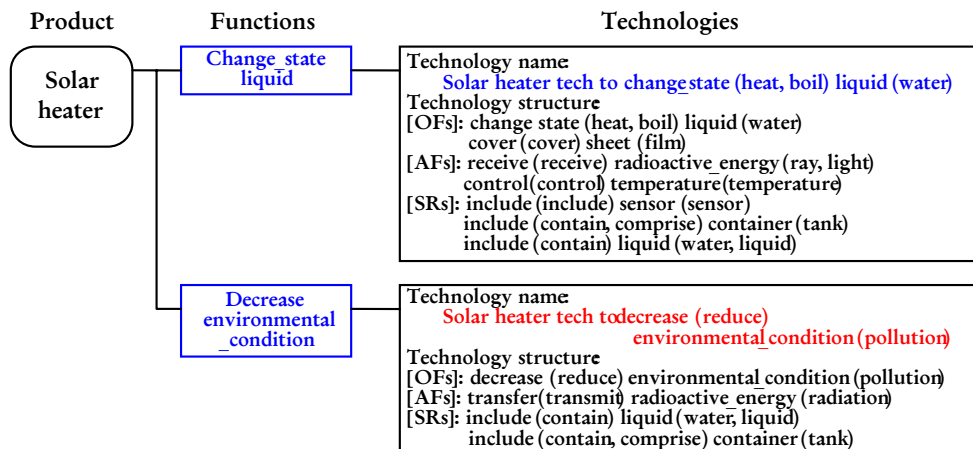


Figure 8. Product-function-technology relationship for solar heaters.

pnlp.codplex.com) and WordNet.Net (opensource.ebswift.com/WordNet.Net), in our system, and stored all relevant information in the proposed knowledge base using MySQL database server.

Using our user interface of the developed system, we were able to determine various products' functions and the technologies of these functions. For example, we searched the product 'air dryer' to examine its functions and the technologies to achieve the functions (Fig-

ure 9). We found an air dryer has a total of 8 functions, including 'give (supply, yield) air (air)', 'remove (remove) moisture (moisture, wet)', 'manage (hold) hair (hair)' and 'eat (consume) air (air)', and each of the function was found to be achieved by air dryer's technologies. Technology 'air dryer tech to remove (remove) moisture (moisture, wet)' had 'remove moisture' as its core function, 17 generic AOs of the OF type, 1 generic AOs of the AF type and 2 generic AOs of the SR type.

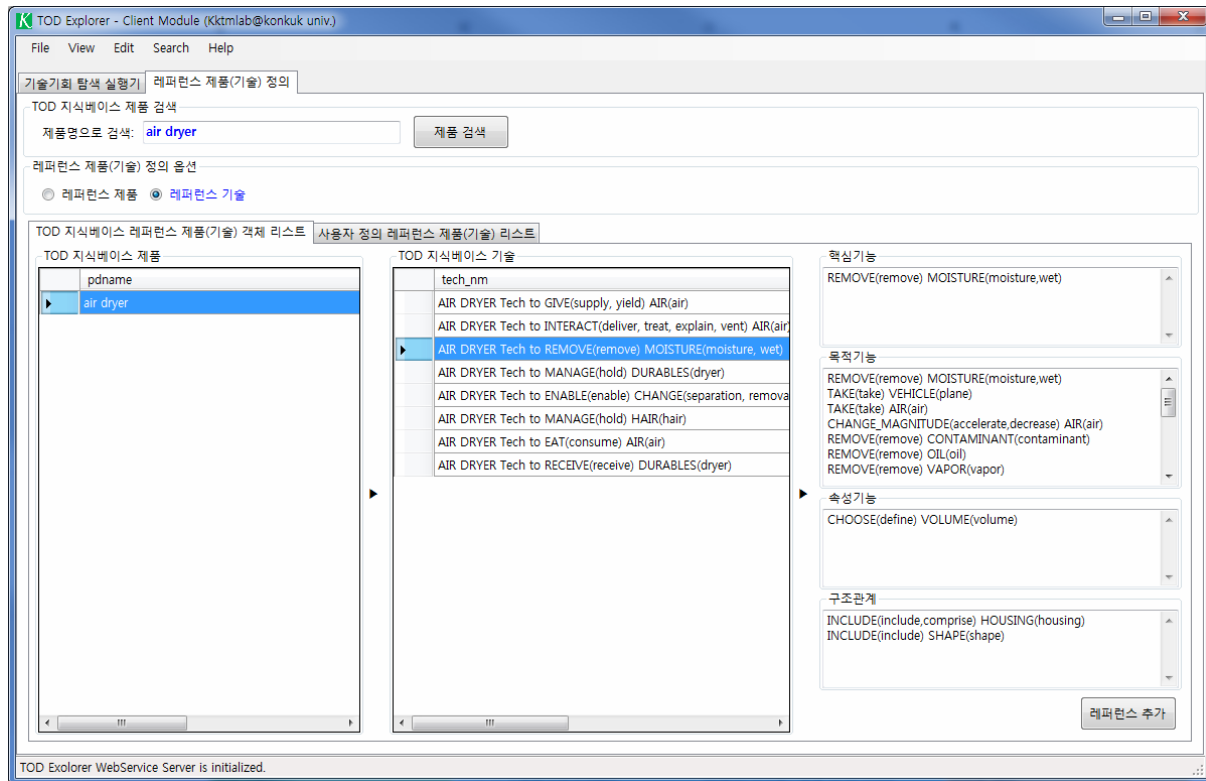


Figure 9. Snapshot of system implementation.

6. DISCUSSIONS

In this study, first, we newly presented the loosely-coupled relationship of products, functions and technologies. Typical approaches consider products and technologies to be tightly related, whereas our study models the relationship between products and technologies as coupled loosely by engaging them via functions. To this end, we redefined the concepts of products and technologies from the functional perspective. That is, a product as a tangible or intangible artifact that is composed of one or more core functions, which are achieved by technologies. Therefore, for example, if a technology has a very similar core function to another given technology, it has the potential to replace the given technology or could be modified and applied to produce the product for which the given technology is currently used. This brief example shows the potential of the product-function-technology relationship in technology opportunity analysis.

Second, to propose a function-based knowledge base, this study adopted the concept of function. In fact, the use of functions is not a new concept; it has a long history in product and information system design studies. The functional representation can explain what a system does and is for at an abstract level, so it has been considered to be the core concept for describing the operations and objectives of a system in a structured way. In

contrast to the keyword approach that only utilizes the occurrence information of the defined keywords, the functional approach can provide much more information. The function information clearly describes the facts of technological components, thereby explaining what a product or system does and provides. With regard to the ability to explain systems, there is no doubt that the function approach provides a good structural information format to abstract and represent technical information related to products and technologies. Despite this advantage of the functional representation, however, there have been surprisingly few studies that adopt functions to develop a knowledge base using textual analysis of large-scale technical data, including patents. Therefore, this study employed NLP technologies to propose a procedural approach to constructing a function-based knowledge base. In addition, this study applied the proposed approach to a function-based knowledge base by using and analyzing large-scale patent data. We expect that the knowledge base developed in the present study will provide a solid foundation that contributes much to widening the identification range of technology opportunities and improving the efficiency of processes for technology intelligence.

Third, to effectively exploit the function information, this study conducted a generalization process of AO information. According to TRIZ, a problem can be patterned and solved using conceptually similar problem models that exist elsewhere (Altshuller, 1984). As a me-

chanism for patterned problem solving, TRIZ generalizes a given problem through the problem abstraction process, and herein, the building blocks to model a problem are the function information that can be represented in SAO structures. In this study, we applied the concept of generalization to atomic AOs, i.e., functions or structural relations. To this end, we used WordNet's concept hierarchy to generate generic AOs for the atomic AOs and thereby transforming domain-specific atomic AOs into generic AOs has the potential to support searches of the functions of products and technologies. Therefore, our knowledge base will allow for identification of function analogy among different products and different technologies in diverse technical fields.

7. FURTHER RESEARCH THEMES USING THE FUNCTION-BASED KNOWLEDGE BASE

This study proposed procedural steps for a function-based knowledge base and according to the steps we constructed a knowledge base using a large number of patents. This study provides an important basis for further research that can be conducted using the proposed knowledge base.

One direct application of the function-based knowledge base is function searching. As mentioned, studies exist of function searching at the patent level, but little attention has been paid to function searching at the product and technology level. There are numerous products and technologies throughout the world and some of them may employ similar functions to achieve their own objectives. Using a large volume of patents, the function-based knowledge base structures the function information of products and technologies over a wide range of technical fields. In addition, the knowledge base contains the generalized functions, so a further research theme could be to develop a method to retrieve products and technologies that provide the functions given by search conditions.

We suggest that a second methodological theme using the knowledge base is to develop a new method for technology opportunity discovery. This theme should be supported by semantic textual analysis, i.e., semantic similarity measurement of functions. The functions provided in different products and technologies could be semantically similar or the same, but they could be represented using different terms. The similarity measurement technique between functions will allow the functional similarity measurement of technologies, and even products. Therefore, it can work as a fundamental mechanism for a technology opportunity discovery approach that starts with defining a given technology or product. Firms have their own existing technologies and products, and they could explore the potential applications of existing technologies and products, and search for external technologies to improve their existing pro-

ducts. Such an approach based on firms' existing technology and product portfolios will contrast with most previous approaches, which have focused on identifying new technologies that have not been developed and on forecasting new emerging technologies.

Third research theme could be to develop an analytical method for identifying convergence technology opportunities. Our knowledge base contains massive number of products, their functions and their technologies, and they all are structured in the form of textual information. Several data mining techniques, such as association rule mining and collaborative filtering, which extract patterns and strong relations from very large data sets, could be utilized to further the automated discovery processes of convergence technologies. Association rule mining is a technique for discovering interesting relations between variables in a large database and collaborative filtering is a recommendation technique for making automatic predictions about the interests of a user by collecting preference information. Therefore, an analytical methods could apply these data mining techniques to the information contained in the knowledge to suggest new types of potential products that can be further developed based on a given product or technology, and to provide newly deliverable functions of the potential products to R&D planners and experts.

8. CONCLUDING REMARKS

Rapid transformations toward a globally competitive technological environment are now driving firms to propose competitive products and technologies for their sustainable development. However, the activity of proposing competitive products or technologies tends to be limited by psychological inertia on one hand and by lack of knowledge on the other (Cascini *et al.*, 2004). Although, for these reasons, methodological approaches for technology intelligence processes haven suggested, few studies have dealt with the construction of a knowledge base that can act as a basis for the practical implementation of these approaches. As part of tackling a technology intelligence framework, this study proposed a function-based knowledge base.

The main contributions of this research involve both the academic and industrial perspectives. From the academic perspective, this study is the first to link between products and technologies from a functional perspective, which enables us to conduct cross-divisional technological analysis by redefining the conventional tightly-coupled and domain-specific definition of product-technology relationship to the loosely-coupled and fact-oriented relationships. In addition, the proposed knowledge base can lead to various technology intelligence methods and systems using function-based patent analysis. From the industrial perspective, the proposed knowledge base can be used in practical intelligence systems as an enabling tool that can effectively deal

with large-scale technical data in a wide range of technical fields, as well as reduce or eliminate the need for expert intervention during the information processing, which may affect objectivity and credibility of the technology intelligence system.

Despite these contributions, some aspect need to be further addressed. First, this study proposed how to construct a function-based knowledge base, but did not deal with the applications of the knowledge base. The proposed knowledge base plays a basic role for methodological applications for technology intelligence, so further research will develop technology opportunity discovery methods using data mining techniques, including collaborative filtering and association rule mining. Second, the volume of the patent database should be extended. This study, as a pilot, collected patents from a limited scope; we collected 1110582 patents granted in the USPTO database over the period from 2009 to 2013. Therefore, the patent database should contain large-scale patent documents over a wider period in order to increase the practical feasibility of the proposed knowledge base. Third, as mentioned, the fundamental limitation of the previous technology intelligence methods is their lack of practicality. The proposed knowledge base here can play a critical role in solving this shortcoming. Implementation of software systems using the proposed knowledge base will deliver practical benefits to technology intelligence processes.

ACKNOWLEDGEMENTS

This Research has been performed as a subproject of project Development of Information Analysis System for Exploring Future Technologies and supported by the KOREA INSTITUTE of SCIENCE and TECHNOLOGY INFORMATION (KISTI). This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2012R1A1A1039303).

REFERENCES

- Altshuller, G. S. (1984), *Creativity as an exact science: The theory of the solution of inventive problems*, CRC Press.
- Ashton, W. B. and Stacey, G. S. (1995), Technical intelligence in business: understanding technology threats and opportunities, *International Journal of Technology Management*, **10**(1), 79-104.
- Bennett, S. and McRobb, S. *et al.* (2006). Object-oriented systems analysis and design using UML, McGraw-Hill Berkshire, UK.
- Bergmann, I. and Butzke, D. *et al.* (2008), Evaluating the risk of patent infringement by means of semantic patent analysis: the case of DNA chips, *R&D Management*, **38**(5), 550-562.
- Cascini, G. and Fantechi, A. *et al.* (2004), Natural language processing of patents and technical documentation, *Document analysis systems VI*, Springer, 508-520.
- Cascini, G. and Russo, D. (2007), Computer-aided analysis of patents and search for TRIZ contradictions, *International Journal of Product Development*, **4** (1), 52-67.
- Cascini, G. and Zini, M. (2008), *Measuring patent similarity by comparing inventions functional trees*, Computer-Aided Innovation (CAI), Springer, 31-42.
- Chang, S.-H. and Fan, C.-Y. (2014), Analyzing Offshore Wind Power Patent Portfolios by Using Data Clustering, *Industrial Engineering and Management Systems*, **13**(1), 107-115.
- Chen, H. and Chiang, R. H. *et al.* (2012), Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly*, **36**(4).
- Choi, S. and Kim, H. *et al.* (2013), An SAO-based text-mining approach for technology roadmapping using patent information, *R&D Management*, **43**(1), 52-74.
- Choi, S. and Park, H. *et al.* (2012), An SAO-based text mining approach to building a technology tree for technology planning, *Expert Systems with Applications*, **39**(13), 11443-11455.
- Choi, S. and Yoon, J. *et al.* (2011), SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells, *Scientometrics*, **88**(3), 863-883.
- Fellbaum, C. (2010), *Wordnet, Theory and applications of ontology: computer applications*, Springer, 231-243.
- Gero, J. S. (1990), Design prototypes: a knowledge representation schema for design, *AI Magazine*, **11**(4), 26.
- Geum, Y. and Jeon, J. *et al.* (2013), Identifying technological opportunities using the novelty detection technique: a case of laser technology in semiconductor manufacturing, *Technology Analysis and Strategic Management*, **25**(1), 1-22.
- Hayes-Roth, F. and Waterman, D. A. *et al.* (1983), Building expert systems, *Teknowledge Series in Knowledge Engineering*, Reading: Addison-Wesley, edited by Hayes-Roth, Frederick; Waterman, Donald A.; Lenat, Douglas B, **1**.
- Hirtz, J. and Stone, R. B. *et al.* (2002), A functional basis for engineering design: reconciling and evolving previous efforts, *Research in engineering Design*, **13**(2), 65-82.
- Kasravi, K. and Risov, M. (2007), *Patent Mining-Dis-*

- cover y of Business Value from Patent Repositor ies. *System Sciences*, HICSS 40th Annual Hawaii International Conference on, IEEE.
- Kerr, C. I. and Mortara, L. *et al.* (2006), A conceptual model for technology intelligence, *International Journal of Technology Intelligence and Planning*, **2**(1), 73-93.
- KISTI (2013), intelligence system development for information analysis, Daejun, *Korea Institute of Science and Technology Information*.
- Kitamura, Y. and Mizoguchi, R. (2004), Ontology-based systematization of functional knowledge, *Journal of Engineering Design*, **15**(4), 327-351.
- Lee, J. and Hong, Y. S. (2014), Business model mining: Analyzing a firm's business model with text mining of annual report, *Industrial Engineering and Management Systems*, **13**(4), 432-441.
- Lee, S. and Yoon, B. *et al.* (2009), An approach to discovering new technology opportunities: Keyword-based patent map approach, *Technovation*, **29**(6), 481-497.
- Lichtenthaler, E. (2004), Technological change and the technology intelligence process: a case study, *Journal of Engineering and technology Management*, **21**(4), 331-348.
- Mann, D. (2002), *Hands-on systematic innovation*, Creax press Belgium.
- Miller, G. A. (1995), WordNet: a lexical database for English, *Communications of the ACM*, **38**(11), 39-41.
- Moehrle, M. G. and Walter, L. *et al.* (2005), Patent-based inventor profiles as a basis for human resource decisions in research and development, *R&D Management*, **35**(5), 513-524.
- Park, H. and Kim, K. *et al.* (2013), A patent intelligence system for strategic technology planning, *Expert Systems with Applications*, **40**(7), 2373-2390.
- Park, H. and Yoon, J. *et al.* (2012), Identifying patent infringement using SAO based semantic technological similarities, *Scientometrics*, **90**(2), 515-529.
- Park, H. and Yoon, J. *et al.* (2013), Using function-based patent analysis to identify potential application areas of technology for technology transfer, *Expert Systems with Applications*.
- Pedersen, T. and Patwardhan, S. *et al.* (2004), WordNet: Similarity: measuring the relatedness of concepts. Demonstration Papers at HLT-NAACL 2004, *Association for Computational Linguistics*.
- Savransky, S. D. (2002), *Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving*, CRC Press.
- Souili, A. and Cavallucci, D. (2013), *Toward an automatic extraction of IDM concepts from patents*, CIRP Design 2012, Springer, 115-124.
- Stone, R. B. and Wood, K. L. (2000), Development of a functional basis for design, *Journal of Mechanical Design*, **122**, 359.
- Suh, N. P. (1990), *The principles of design*, Oxford University Press New York.
- Umeda, Y. and Ishii, M. *et al.* (1996), Supporting conceptual design based on the function-behavior-state modeler, *Ai Edam*, **10**(4), 275-288.
- Yoon, B. (2008), On the development of a technology intelligence tool for identifying technology opportunity, *Expert Systems with Applications*, **35**(1), 124-135.
- Yoon, B. and Park, Y. (2005), A systematic approach for identifying technology opportunities: Keyword-based morphology analysis, *Technological Forecasting and Social Change*, **72**(2), 145-160.
- Yoon, B. and Park, Y. (2007), Development of new technology forecasting algorithm: hybrid approach for morphology analysis and conjoint analysis of patent information, *Engineering Management, IEEE Transactions on*, **54**(3), 588-599.
- Yoon, B. U. and Yoon, C. B. *et al.* (2002), On the development and application of a self-organizing feature map-based patent map, *R&D Management*, **32**(4), 291-300.
- Yoon, J. and Choi, S. *et al.* (2011), Invention property-function network analysis of patents: a case of silicon-based thin film solar cells, *Scientometrics*, **86**(3), 687-703.
- Yoon, J. and Kim, K. (2011), Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks, *Scientometrics*, **88** (1), 213-228.
- Yoon, J. and Kim, K. (2012), Detecting signals of new technological opportunities using semantic patent analysis and outlier detection, *Scientometrics*, **90**(2), 445-461.
- Yoon, J. and Kim, K. (2012), TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents, *Expert Systems with Applications*, **39**(3), 2927-2938.
- Yoon, J. and Lim, J. *et al.* (2011), Ontological functional modeling of technology for reusability, *Expert Systems with Applications*, **38**(8), 10484-10492.
- Yoon, J. and Park, H. *et al.* (2013), Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis, *Scientometrics*, **94**(1), 313-331.