Expert Systems with Applications 39 (2012) 9129–9140

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# A fact-oriented ontological approach to SAO-based function modeling of patents for implementing Function-based Technology Database

Sungchul Choi, Dongwoo Kang, Joohyung Lim, Kwangsoo Kim *

*Department of Industrial and Management Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang, Kyungbuk 790-784, Republic of Korea*

ARTICLE INFO

*Keywords:*
Function based Technology Database
Function Oriented Search
TRIZ
Fact-oriented ontology
Patent analysis

ABSTRACT

Function-Oriented Search (FOS) has been proposed as a tool for use in searching patent databases to find existing solutions to new problems. To implement FOS effectively, a well-structured Function-based Technology Database (FTDB) is required. An FTDB is a data repository of technology information represented as "function". To implement an FTDB, four features should be addressed: continual data updating, limited area searching, function generalization, and semantics handling. In this paper, we consider these features to suggest a fact-oriented ontological approach to implementing an FTDB by Subject–Action–Object (SAO)-based function modeling of patents. The proposed approach uses fact-oriented ontology modeling of SAO structures extracted from patent documents, and implements an FTDB, which is an SAO-based patent retrieval system to support FOS. We also verify the feasibility of the proposed approach to by using it to conduct case studies of patent retrieval.



## 1. Introduction

The Theory of Inventive Problem Solving (Russian acronym: TRIZ) was developed by generalizing technologies after extensive analysis of 40,000 patents (Altshuller, 1984). Altshuller noted that technology has followed certain patterns and rules in creating new and inventive patentable ideas (Salamatov & Souchkov, 1999). TRIZ is a useful tool for analyzing technology; it has been applied in a variety of areas, and has been validated by many researchers (Crotti, Ghitti, Regazzoni, & Rizzi, 2007; Fey & Rivin, 1999; Mann, 2003; Savransky, 2000a; Zhang & Xu, 2006; Zhang, Yang, Tian, & Tan, 2007).

Recently, a new TRIZ tool, called Function-Oriented Search (FOS) (Litvin, 2005a, 2005b) has been proposed. The main objective of FOS is to find an existing technology and apply it as a solution to the target problem. The problem solving process of FOS is based on that of TRIZ (Fig. 1), which consists of four steps: (1) identify a specific problem; (2) abstract the problem as a generic problem; (3) find a generic solution for the generic problem; and (4) apply the solution to the specific problem. However, TRIZ cannot be easily applied in industry because traditional TRIZ tools use a generic solution, and so are too abstract and limited to apply directly to industrial problems. To solve this limitation, FOS searches existing technologies to find the appropriate generic solution FOS also consists of four steps: (1) identify a target problem; (2) generalize the problem; (3) find the existing solution by using a Function-based Technology Database (FTDB); and (4) apply the existing solution to the target problem. FOS looks similar to the traditional TRIZ problem-solving process. However, by using existing technology, FOS makes TRIZ more acceptable to users in industry. Existing technology can be more easily understood than traditional TRIZ tools which use generic solutions developed by engineers and technology experts who seek a solution to a problem. Because of this advantage, FOS has been applied in a variety of areas to solve technological problems (http://www.gen3partners.com).

To implement FOS effectively, a well-structured FTDB is essential (Litvin, 2005a). An FTDB is a data repository of technology information represented as "function". The concept of function can be defined as "The action changing a feature of any object" (Savransky, 2000b). This concept provides information on the uses and purposes of a technology. To implement an FTDB, four major features should be addressed. (1) The FTDB should be continually updated, because new technology appears continuously and replaces old technology. A new technology may be useful as a new solution to an old problem. (2) The FTDB should support search for a specific technology area. If a researcher tries to find a solution by searching in all remote engineering areas, the search field is almost infinite (Litvin, 2005a). This unlimited search uncovers huge amounts of unnecessary information, which a researcher must then remove. (3) The FTDB should support function generalization. Because a direct technology search is very ineffective (Litvin, 2005a), implementing an FTDB requires that technology function be generalized. By generalizing target technology function, the researcher can use existing technologies from various

* Corresponding author. Tel.: +82 54 279 8234.
*E-mail addresses:* blissray@postech.ac.kr (S. Choi), hyunil@postech.ac.kr (D. Kang), yanzi@postech.ac.kr (J. Lim), kskim@postech.ac.kr (K. Kim).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.041

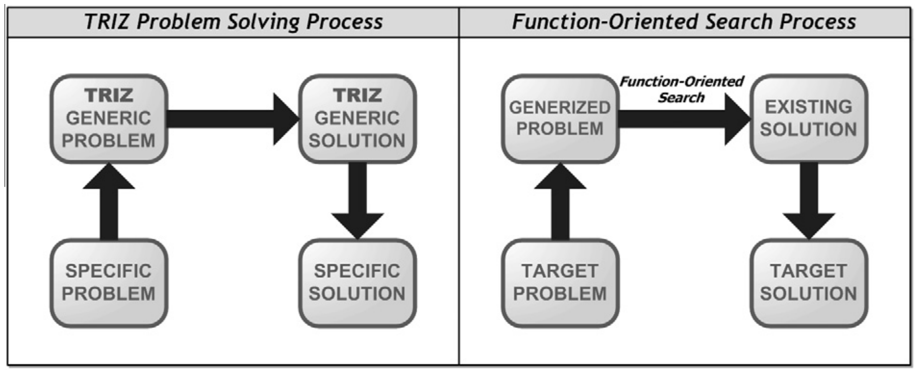


**Fig. 1.** Comparison of TRIZ and FOS process.

areas of engineering as a solution to the problem. (4) The FTDB should support semantic search. In technology, terms are represented in various formats (synonyms). For example, the terms "solar cell" and "photovoltaic cell" have the same meaning. Because of these kinds of words, users of FTDBs face the semantic confusion problem (Borenstein & Fox, 2003), which limits the effectiveness of FOS.

In this paper, we suggest a fact-oriented ontological approach to Subject–Action–Object (SAO)-based function modeling of patents for implementing FTDB. The proposed approach considers the four major features of FTDBs. For this purpose, we collect technology function information from patent information, SAO structure, and a fact-oriented ontological approach for modeling technology function information. Patent information can be used to satisfy the first and second features of an FTDB. Patent information includes valuable up-to-date technological information and is continuously updated. Patents also include bibliometric information (e.g., International Patent Classification (IPC) code), which can be utilized to search limited information. SAO structure and a fact-oriented ontological approach can be used to satisfy the third and fourth features of an FTDB. Each patent document has specific function information, which can be represented using an SAO structure. An SAO structure is commonly used to represent a function of technology. By modeling SAO structures using a fact-oriented ontological approach, the relationships among technological terms are defined. These relationships can support function generalization and remove the semantic confusion problem. A fact-oriented ontology is represented as similar structure of SAO structure and it is modeled as a simple sentence, which technology experts who have no knowledge of information modeling can use to model technology information.

The proposed approach suggests fact-oriented ontology modeling using SAO structures extracted from patent documents, and implements an FTDB, which is an SAO-based patent retrieval system to support FOS. We also verify the feasibility of the proposed approach by using it to conduct case studies.

The rest of paper is organized as follows. In Section 2, we compare the state of the art of FTDB with proposed framework. In Section 3, we describe the related work of the proposed framework. In Section 4, we suggest detailed description for our framework. In Section 5, we illustrate the proposed framework by conducting case studies of patent retrieval of function information. In Section 6, we present concluding remarks and directions for further study.

## 2. The state of the art of FTDB

In this section, we describe the state of the art of FTDBs and analyze them with respect to the features mentioned (Table 1). Current work related to FTDBs can be divided into two types: tools for constructing FTDBs, and FTDBs already constructed. Tools of these two types have been developed by companies that use TRIZ – Invention Machine (http://inventionmachine.com) and CREAX (http://www.creax.com).

The first category includes FTDB development tools. Knowledgist 2.5, developed by Invention Machine, supports construction of SAO-based FTDBs to analyze patent databases and technology documents. Knowledgist 2.5 directly supports analysis of patent databases such as that of the US patent and Trademark Office, and users can access the continually updated database and find information in specific areas. However, Knowledgist 2.5 cannot support generalization of retrieved information; it can only handle verb synonyms, and so cannot fully solve semantic problems.

The second category includes FTDB systems currently available. CREAX developed Function DB (http://function.creax.com) and Attribute DB (http://attributes.creax.com). The CREAX FTDB provides a web-based interface so that users can easily access the databases (DBs). Moreover, the DBs support search for well-defined generalization information composed of four generalizations (Solid, Liquid, Gas, and Field). However, the DBs have limitations to their use as FTDBs. First, the DBs are static, so updating information is not provided. Second, the DBs cannot support specific search because they do not consider bibliographic information. Lastly, the amount of information provided by the DBs is absolutely insufficient and the interface to access the DB is inconvenient. Invention Machine developed Goldfire Innovator by adding new features to Knowledgist 2.5. The 'Scientific Effects DB' module in Goldfire provides function-based technology information as well as does the FTDB of CREAX. Compared with the CREAX DB, this module can provide more generalized information, but like the CREAX DB cannot support the specific search. But with new modules, Goldfire can overcome the limitations of the Scientific Effects DB. The patent collection and knowledge base modules enable continual update of function information and allow search for information in specified areas. These two modules support search of patent documents and construction of technology knowledge bases. The main limitation of Goldfire is that it cannot generalize function information. Although Goldfire can handle ontology and synonym information, this feature cannot suggest related generalized terms. For example, if users search the word 'ceramic', Goldfire retrieves only documents that contain 'ceramic'; it does not retrieve documents that include 'brick' which is a specialized type of 'ceramic'. This retrieval makes users have difficulty in finding all documents related to ceramics. In addition, Goldfire can only solve the semantic confusion problem for verbs, not for nouns or noun phrases. If users want to find a document that includes acronyms such as 'PEMFC', Goldfire cannot interpret the acronyms as an original word (in this case, Proton exchange membrane fuel cell), and cannot find documents that include it.

**Table 1**
Comparison of tools for FTDB with features of FTDB.

| | Products | The important features of FTDB | | | |
|---|---|---|---|---|---|
| | | Continual data updating | Specific searching | Function generalization | Semantics handling |
| FTDB Development Tools | Knowledgist 2.5 | Supported | Supported | Not supported | Not fully supported (verb synonyms only) |
| FTDB Systems | CREAX (Function DB, Attribute DB) | Not supported | Not supported | Supported | Not supported |
| | Goldfire (Scientific Effects, Patent Collections, and Knowledge base) | Supported | Supported | Not fully supported | Not fully supported |

## 3. Related work

### 3.1. SAO structure

For constructing the FTDB, we use SAO structures extracted from patent text; these structures are composed of subject (noun phrase), action (verb phrase) and object (noun phrase). This simple sentence explicitly describes a relationship between components that appear in the patent text. For example, an SAO structure can simply represent the function of a battery as ''Battery Energizes Bulb''. In this example, ''Battery'' is the subject, ''Energizes'' is the action, and ''Bulb'' is the object. The technology purpose of ''Battery'' is to energize ''Bulb'' and the function of ''Battery'' is ''To energize Bulb''.

SAO structures can represent the variety of technology information. Subjects and objects may refer to components of a system and actions may refer to functions performed by and on components (Cascini, Fantechi, & Spinicci, 2004a). Moreover, an SAO structure can be organized in a problem–solution format if the action–object (AO) states the problem and the subject (S) forms the solution (Moehrle, Walter, Geritz, & Müller, 2005). SAO structure also states partative relationships between products or technologies (Cascini, Lueehesi, & Rissone, 2001). If the action word is a partative verb such as ''have'', ''composed of'', or ''be made up of'', the component of a subject may include the component of an object.

To computationally extract SAO structure from patent textual descriptions, use of natural language processing (NLP) is necessary (Cascini, Fantechi, & Spinicci, 2004b) in the Part-Of-Speech tagging process. Because NLP studies have been already conducted to extract SAO structures from sentences (Charniak et al., 1996; Choi, Lim, Yoon, & Kim, 2010; Liu, 2004; Liu & Singh, 2004; Tsourikov, Batchilo, & Sovpel, 2000), we extract SAO structures using methods described in those studies.

### 3.2. Fact-oriented ontological approach

To solve the semantic confusion problem, we use the ontological approach. An ontology is an explicit specification of a conceptualization (Gruber, 1993). The ontology includes definitions of concepts and an indication of how concepts are inter-related; these definitions collectively impose a structure on the domain and constrain the possible interpretations of terms (Uschold, 1998). The ontology is used to assist in communication between human agents, to achieve interoperability among computer systems, or to improve the process, quality, or both, of engineering software systems (Uschold & Jasper, 1999).

To use ontology to construct an FTDB, we use fact-oriented ontology modeling, which is a conceptual modeling method that enables one to model and query specific domains to find facts of interest, where all facts and rules may be verbalized in language readily understandable by non-technical users of those specific domains (Halpin, 2007). Essential elements of the fact-oriented approach are facts. Facts build on fact types, and fact types build on concepts, fact types and facts. The fact-oriented approach uses facts presented as sentences. Fact-oriented ontologies can be expressed with Structured English which uses a small number of English structures and common words to provide a simple and straightforward mapping to its concepts, fact types and facts (Object Management Group, 2005). This feature of the approach has several advantages including human-readability, semantic stability, expressiveness, extensibility, and changeability (Halpin, 2007; Kang, Lee, Choi, & Kim, 2010).

These advantages of the fact-oriented approach are suitable for representing SAO structures, for three reasons. Firstly, Structured English helps technology domain experts generate an SAO structure. Although these experts must participate in the SAO modeling process, they may not be familiar with information modeling methods such as Unified modeling language (UML) or Object-Oriented Modeling. However, they can easily understand the modeling expression by using the Structured English. Because the expressions are described based on simple English grammar, using them improves the accuracy of generating SAO structures and decreases their uncertainty. Secondly, the fact-oriented approach is suitable for managing changeability of a concept. Technology concepts frequently change or are newly generated. To manage this feature of technology words, the ontology can support extensibility. Finally, the approach can easily express SAO structures. The fact-oriented ontology uses sentences to describe ontology. This format is suitable for representing SAO models without information loss.

## 4. Fact-oriented ontology modeling for SAO-based FTDB

To construct an SAO-based FTDB, we propose fact-oriented ontology modeling (Fig. 2) which consists of fact-oriented SAO modeling and fact-oriented patent information modeling. We also suggest Resource Description Framework (RDF) modeling to implement the proposed fact-oriented modeling as a computerized system.

The SAO structure extracted from patent documents represents a specific technology function, not a general technology function. Because one technology function can be used in various technology areas, we must abstract the specific technology function. Technology abstraction is a basic approach of TRIZ.

To support abstraction of technology, we propose an abstract SAO ontology model, which is composed of noun concept and fact type (also called verb concept). First, the noun concept is used to describe the information of subject concept or object concept. The noun concept has one of four concept state types (solid, liquid, gas, and field), which were proposed by CREAX (http://www.creax.com). CREAX uses the states to search for various scientific effects to solve specific problems (CREAX, 2005) and to represent the abstract

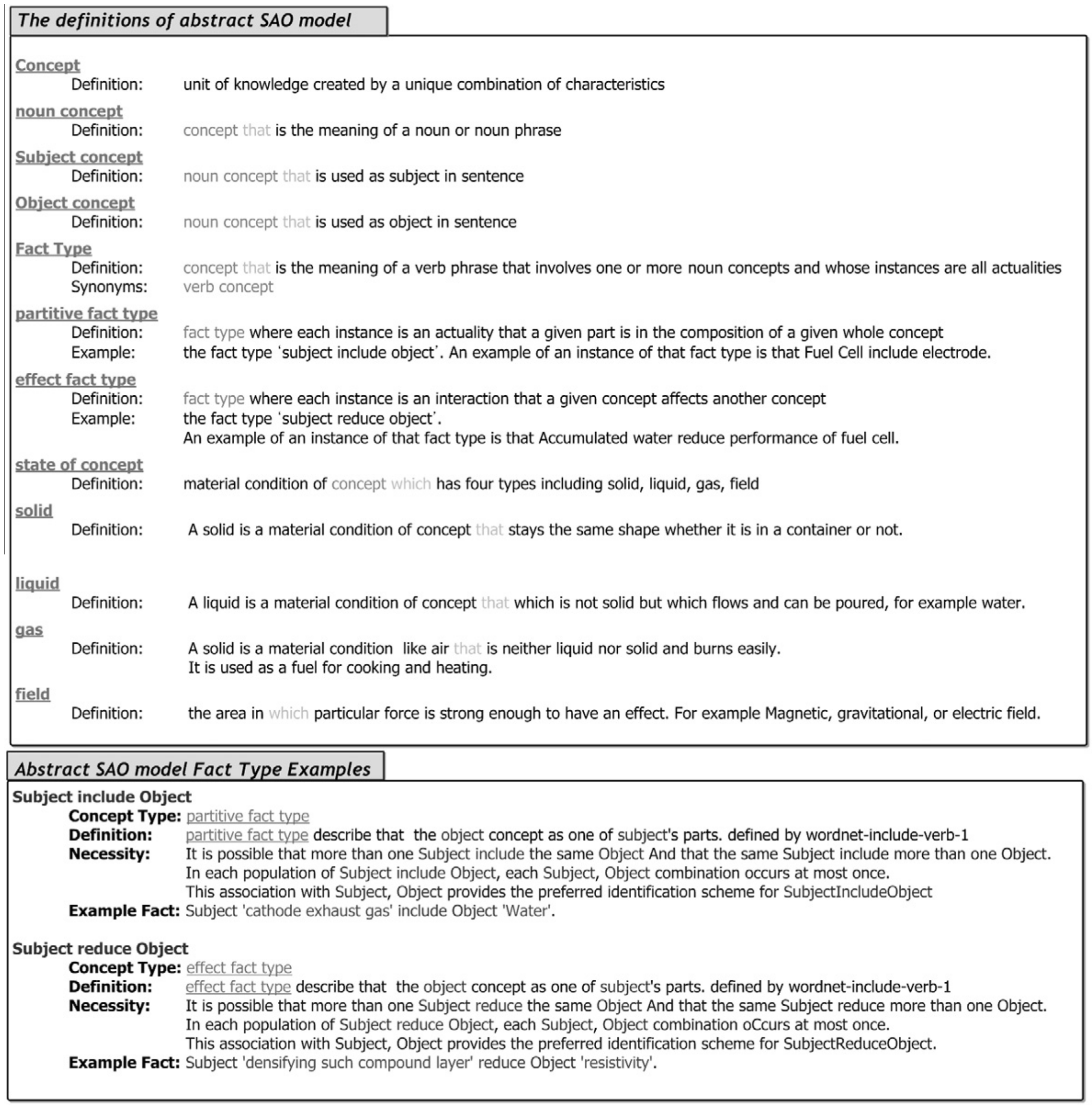


**Fig. 2.** A fact-oriented ontology for representing an abstract SAO model.

information of the noun concept. Then the fact type is used to represent a relationship between subject noun concept and object noun concept. The fact type of an abstract SAO model is composed of partative and effect fact types. Partative facts describe the inclusion relationship between two noun concepts. For example, verbs such as 'include', 'consist', or 'be composed of' represent partative facts. The effect fact type describes how the subject concept affects the object concept. For example, verbs such as 'reduce', 'improve', or 'increase' represent effect facts. Each fact type may have various inherited fact types. For instance, the 'include fact type' is inherited from the partative fact type and 'reduce fact type' is inherited from the effect fact type.

The relationship between an abstract SAO model and an extracted SAO structure (also called an SAO instance) is expressed as a relationship between fact-type and fact (Fig. 3). For example, if we extract an SAO model expressed as 'Activated reactive evaporation process-employ-electron beam' from a patent document, this SAO model is a fact of 'employ fact type'. The fact type of an abstract SAO model is generated based on the extracted SAO model and the extracted SAO models mapped to the fact type. All concepts used in SAO models are expressed as a pre-defined word stored in WordNet. For example, 'amorphous silicon' and 'c-Si' both represent a type of silicon. By using WordNet, we define the two words as 'wordnet-silicon-noun-1'.

Based on an abstract SAO model, we model fact-oriented SAO models which describe patent information (Fig. 4). With the patent information described as an SAO model, patents can be retrieved. The model is modeled as follows. Each patent has several SAO models: 'PatentKey' as patent identifier and 'ModelID' as an SAO model identifier. Each SAO model is an instance of an abstract SAO model and has information about subject concept and object concept. For example, Patent US20070065708 has an SAO model having 'ModelID' SAO_KB_113. The SAO model is an instance of an abstract SAO model called 'SubjectAbsorbObject'. This model has 'amorphous silicon' as subject concept and 'UV light' as object concept. 'amorphous silicon' and 'UV light' are generalized by WordNet definition 'wordnet-silicon-noun-1' and 'wordnet-light-noun-1' respectively.

The proposed ontology modeling supports all features of the FTDB. First, the abstract SAO model and WordNet definition define the concepts used in SAO structures. This approach supports function generalization and semantics handling. Second, by using patent information, the proposed modeling method allows continual update and specific search. The patent information is continu-

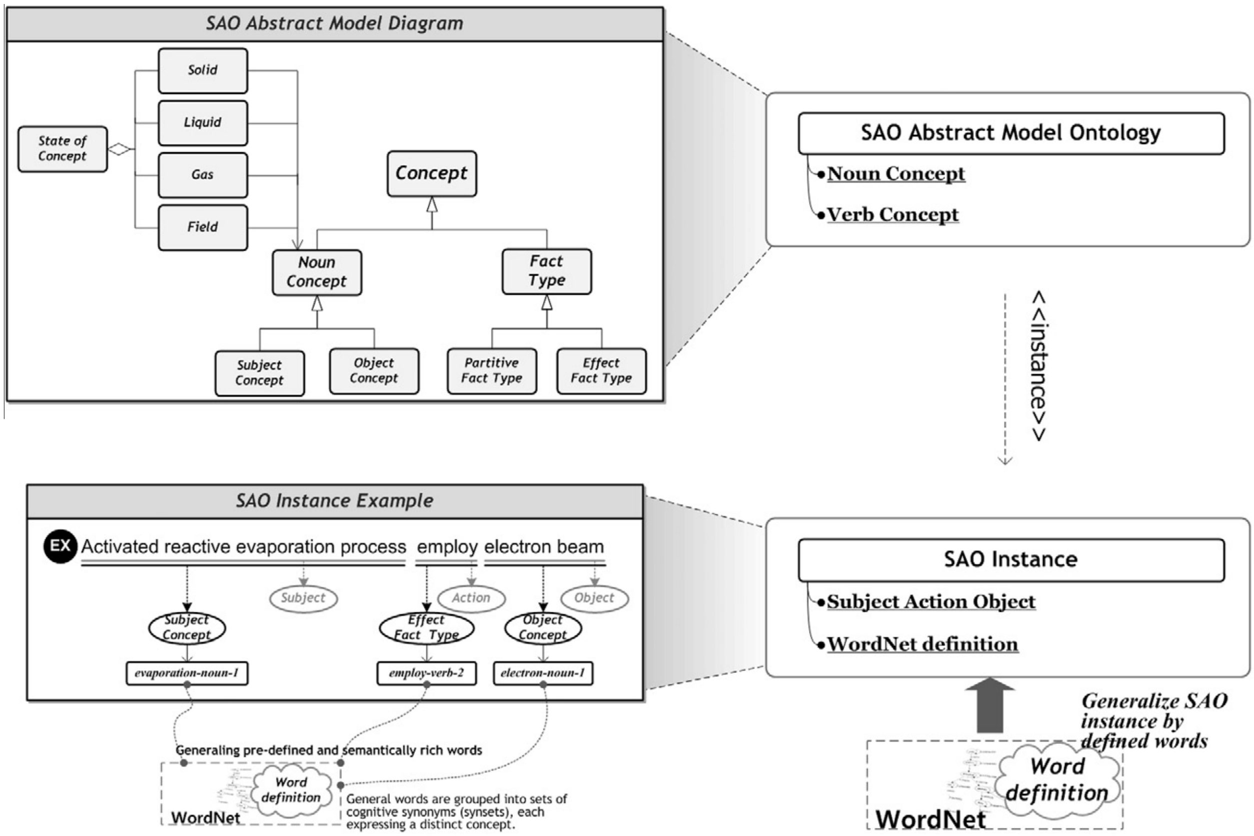


**Fig. 3.** Relationship between an abstract SAO model and an SAO instance.

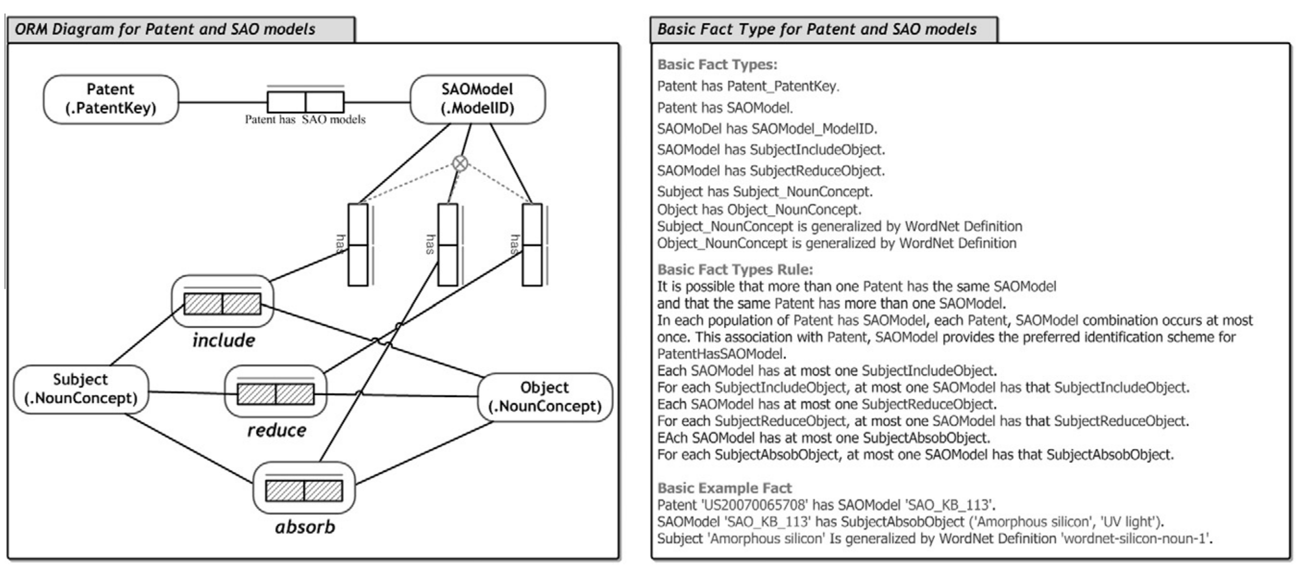


**Fig. 4.** Fact-oriented patent information modeling with SAO models.

ally updated and managed with 'PatentKey' information. 'PatentKey' information can be used as a patent identifier and this information provides the bibliographic information of patent documents. By using bibliographic information such as IPC code or assignees, the user can search the technology information in specified area.

The last task for modeling fact-oriented ontology in this paper is to transform the ontology to RDF. Because the current fact-oriented modeling is not machine-processable, we need to generate RDF format to make it so. The RDF was originally designed as a metadata data model by the World Wide Web Consortium. It is used as a general method for conceptual description or modeling of information that is implemented in web resources. Because the basic structure of RDF consists of subject, predicate, and object (Fig. 5), this structure can easily be mapped onto the SAO model and onto fact-oriented modeling. To map the modeling to RDF,

| Concepts | RDFS |
|---|---|
| Concept | <rdfs:Class rdf:about="&patent;Concept" rdfs:label="Concept"><br><rdfs:subClassOf rdf:resource="&rdfs;Resource"/><br></rdfs:Class> |
| noun concept | <rdfs:Class rdf:about="&patent;NounConcept" rdfs:label="NounConcept"><br><rdfs:subClassOf rdf:resource="&rdfs;Concept"/><br></rdfs:Class> |
| subject concept | <rdfs:Class rdf:about="&patent;SubjectConcept" rdfs:label="SubjectConcept"><br><rdfs:subClassOf rdf:resource="&rdfs;NounConcept"/><br></rdfs:Class> |
| Object concept | <rdfs:Class rdf:about="&patent;ObjectConcept" rdfs:label="ObjectConcept"><br><rdfs:subClassOf rdf:resource="&rdfs;NounConcept"/><br></rdfs:Class> |
| Fact Type | <rdf:Property rdf:ID="FactType" rdfs:comment="···"><br><rdfs:domain rdf:resource="&patent;NounConcept"/><br><rdfs:range rdf:resource="&patent;NounConcept"/><br></rdf:Property> |
| partitive fact type | </rdf:Property><br><rdf:Property rdf:ID="PartitiveFactType" rdfs:comment="···><br><subPropertyOf rdf:resource="#FactType"/><br><rdfs:domain rdf:resource="&patent;SubjectConcept"/><br><rdfs:range rdf:resource="&patent;ObjectConcept"/><br></rdf:Property> |
| effect fact type | <rdf:Property rdf:ID="EffectFactType" rdfs:comment="···"><br><subPropertyOf rdf:resource="#FactType"/><br><rdfs:domain rdf:resource="&patent;SubjectConcept"/><br><rdfs:range rdf:resource="&patent;ObjectConcept"/><br></rdf:Property> |

| Fact Type | RDFS |
|---|---|
| Subject include Object | <rdf:Property rdf:ID="include" rdfs:comment=" ···><br><subPropertyOf rdf:resource="#PartitiveFactType"/><br><rdfs:isDefinedBy rdf:resource="&wn20instances;wordnet-include-verb-1"/><br><rdfs:domain rdf:resource="&patent;SubjectConcept"/><br><rdfs:range rdf:resource="&patent;ObjectConcept"/><br></rdf:Property> |
| Subject reduceObject | <rdf:Property rdf:ID="absorb" rdfs:comment="···"><br><subPropertyOf rdf:resource="#EffectFactType"/><br><rdfs:isDefinedBy rdf:resource="&wn20instances;wordnet-absorb-verb-1"/><br><rdfs:domain rdf:resource="&patent;SubjectConcept"/><br><rdfs:range rdf:resource="&patent;ObjectConcept"/><br></rdf:Property> |

| Fact | RDF |
|---|---|
| SAOModel 'SAO_KB_113' has SubjectAbsobObject ('Amorphous silicon', 'UV light'). | <patent:SAOModel rdf:ID="SAO_KB_113"><br><patent:hasModel><br><patent:SubjectConcept rdf:ID="Amorphous silicon"><br><rdfs:isDefinedBy rdf:resource="&wn20instances;wordnet-silicon-noun-1"/><br><patent:absorb><br><patent:ObjectConcept rdf:ID="UV light"><br><rdfs:isDefinedBy rdf:resource="&wn20instances;wordnet-include-verb-1"/><br></patent:ObjectConcept><br></patent:absorb><br></patent:SubjectConcept><br></patent:hasModel><br></patent:SAOModel> |
| Patent 'US20070065708' has SAOModel 'SAO_KB_113'. | <patent:Patent rdf:ID="US20070065708"><br><patent:hasSAOModel><br><rdf:bag rdf:ID="SAOList"><br><rdf:li rdf:resource="#SAO_KB_113" /><br><rdf:li rdf:resource="#SAO_KB_114" /><br><rdf:li rdf:resource="#SAO_KB_115" /><br></rdf:bag><br><patent:hasSAOModel><br></patent:Patent> |

**Fig. 5.** Mapping fact-oriented modeling to RDFS and RDF.

the concepts and fact types are transformed to the RDF Scheme (RDFS), and facts are transformed to RDF.

## 5. Implementation of an SAO-based patent retrieval system using FTDB

We have implemented a preliminary prototype (Fig. 6) of an SAO-based patent retrieval system that uses the proposed fact-oriented ontology. We used using JAVA 1.6.0 to implement the system, MySQL Server 5.1 to manage the FTDB, Sesame Server to store and manage RDF data, JWI 2.1.5 API to manage the WordNet information, JENA API to handle RDF data, and Knowledgist 2.5™ to extract SAO structures.

The proposed system consists of four components – Patent bibliography analyzer, Patent content analyzer, Patent content ontology manager, and Patent retrieval manager.

### 5.1. Patent bibliography analyzer

The patent bibliography analyzer uses the patent bibliography crawler module to extract patent information from a legacy patent database, and then uses the patent bibliography manager module to store extracted information. Seventeen bibliographic fields are managed: Country code, Kind code, Application number, Application date, Title, Main IPC, Applicant, All IPCs, Applicant country, Main applicant, Inventor, Inventor country, Publication date, Firm number, Firm date, Patent number, and Issue date. The stored bibliographic information is used with SAO models when users request patent retrieval. This bibliography analyzer is used as the component for supporting specific search.

### 5.2. Patent content analyzer

The patent content analyzer collects the full descriptions of patents and extracts SAO models from them. The patent content analyzer has two modules for these purposes. The patent full description crawler module collects the full description of patents from FreePatentOnline (http://www.freepatentsonline.com), which provides full descriptions of patents in HTML format. The crawler module uses an HTML parser to collect the description and store it in the FTDB. Then, the SAO extractor module (Knowledgist 2.5™) uses the NLP parser and tagger to extract SAO models from the description (Fig. 7).

### 5.3. Patent content ontology manager

The patent content ontology manager transforms stored SAO models to RDF format. The fact type ontology manager module generates fact type modeling and displays modeling information. Then, the ontology and domain expert checks the validity of the suggested ontology. Together, the two experts select a patent and, if necessary, modify the WordNet definition by checking existing modeling and WordNet information (Fig. 8). The SAO ontology manager module stores the completely revised ontology in the FTDB in RDF format. This ontology manager module is to support function generalization and semantics handling.

### 5.4. Patent retrieval manager

The patent retrieval manager (Fig. 9) retrieves function-based technology information (Fig. 10) in response to a user's request.

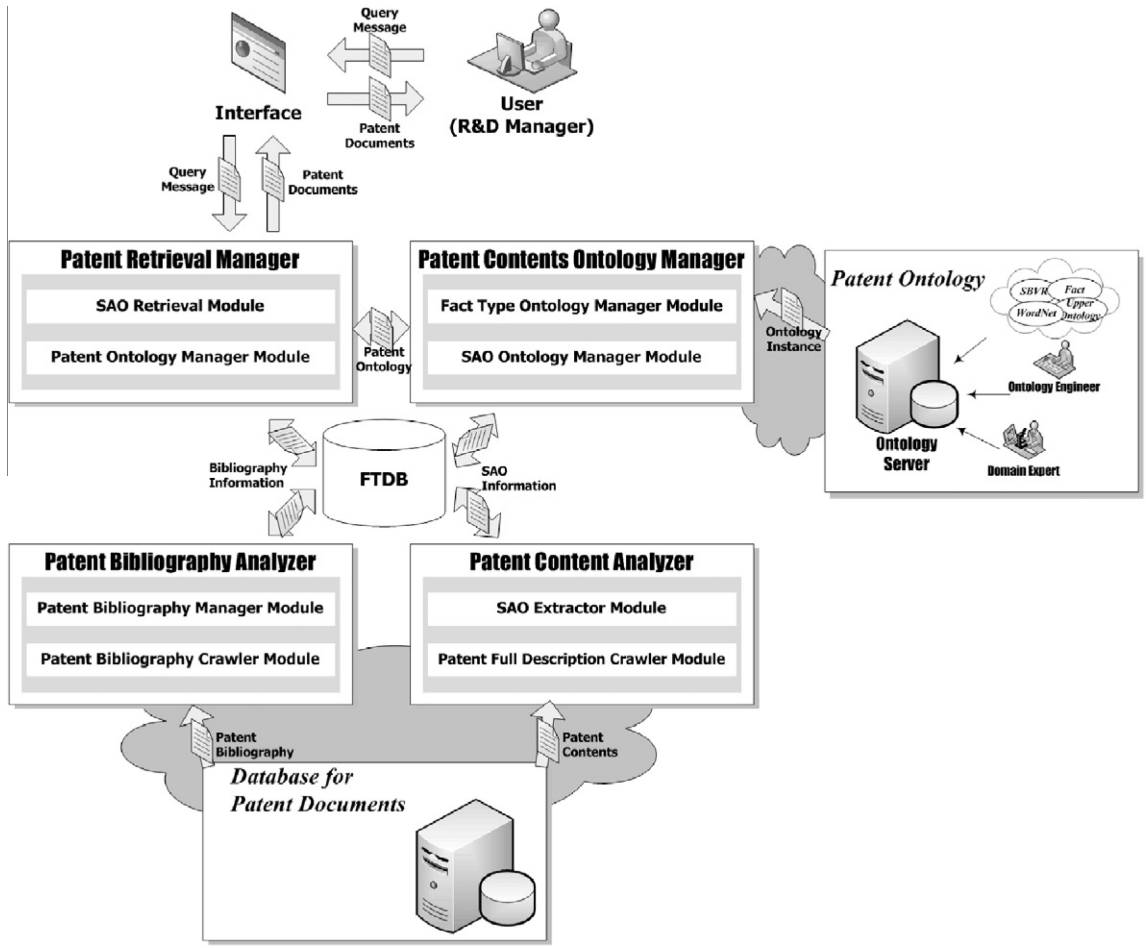


**Fig. 6.** Overall system architecture for the proposed approach.

*The SAO Extraction Example(from US5360745)*

Original Sentence:
**The MG-Si powder is melted by the high-temperature plasma and sprayed onto the base substrate 11.**

Extracted SAO Model:
**high-temperature plasma - melt - MG-Si powder**

**Fig. 7.** Example of SAO extraction.

The proposed approach uses bibliographic information and an SAO structure to retrieve function-based technology information. This retrieval system allows a user to search patents by using the 17 bibliographic fields (Section 5.1). By using bibliographic information, users can search the function-based technology information in a specified area. Users choose the bibliographic information for their search and choose the function information that they want to find. The bibliographic information uses function information to limit the results of search. In an SAO structure, this system allows a user to search patents by using Subject–Action–Object information. Based on the ontology modeling, this retrieval consists of abstract SAO model retrieval, WordNet retrieval, and SAO instance retrieval. Abstract SAO model retrieval uses the abstract level of the SAO model to support retrieval. In an abstract SAO model, the subject concept and object concept each has one of four concept state models (solid, gas, liquid, field), and the action is either partative or effect. In this abstract SAO model retrieval a user can choose one of the four state models in the subject and object, and one of the two verb concepts. WordNet retrieval supports retrieval based on WordNet information. All concepts of SAO models are generalized as WordNet definitions. Therefore, a user can choose a specific WordNet definition for SAO based patent retrieval. By using two retrieval types, users can exploit function generalization and semantics when searching the function-based technology information. Finally, SAO instance retrieval supports retrieval of extracted SAO models directly. Sometimes, users want to search patents using SAO models without abstraction. This retrieval allows users to search for patents that include specific SAO models. As an example of SAO-based patent retrieval, we can conduct SAO instance retrieval by inserting the SAO model composed of [Accumulated water, reduce, performance of fuel cell] to search for patents that include a similar SAO model. As an example of a problem–solution relationship search, we can find patents that include solution information for ‘fixing something solid’, by combining WordNet retrieval and abstract SAO model retrieval as follows: ‘wordsense-adhere-verb-6’ and ‘solid’.

When searching for function-based technology information, users can use the ‘‘Search within Results’’ function. For example, after searching for patents that include a specific IPC code, they can then search those patents for a specific SAO model. Similarly, they can first conduct SAO-based retrieval, and then use bibliographic data to search within the results.

Because function-based technology information is stored in RDF format, the patent retrieval manager component transforms a user’s request to RDF query language. For this process, we use the Sesame Server to manage ontology information stored as RDF format and Sesame RDF Query Language (SeRQL) to query the

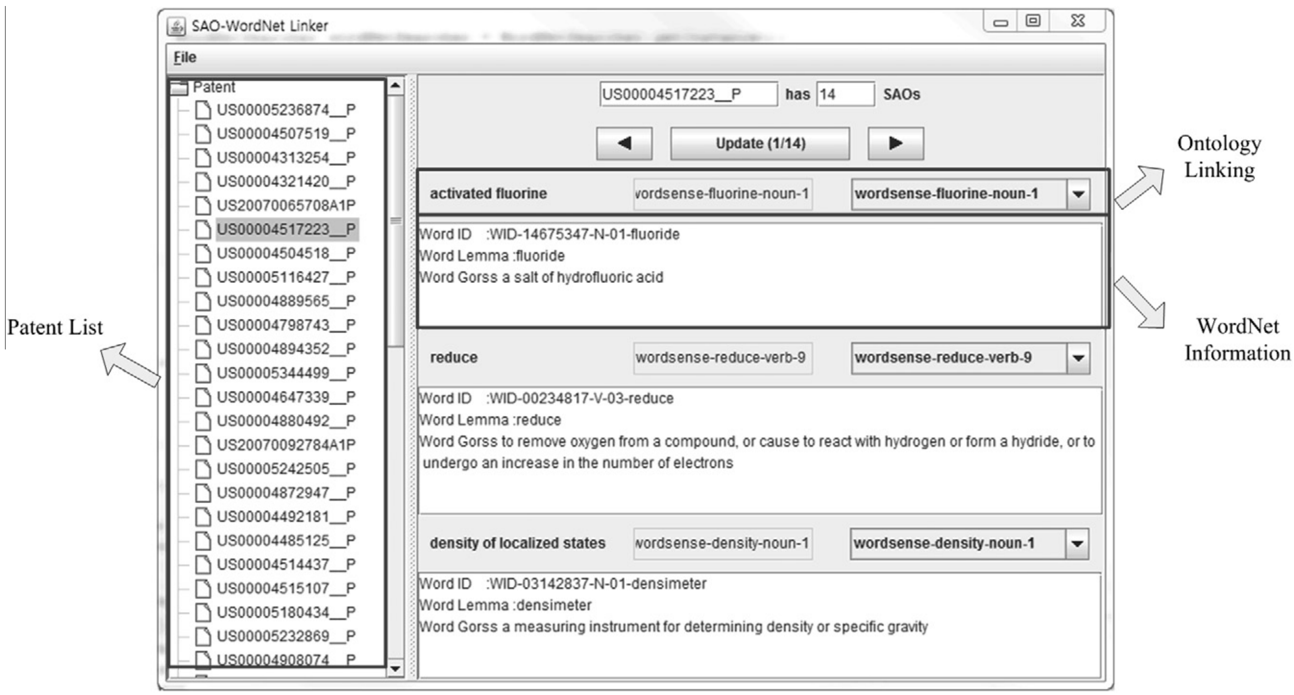


**Fig. 8.** Patent content ontology manager.

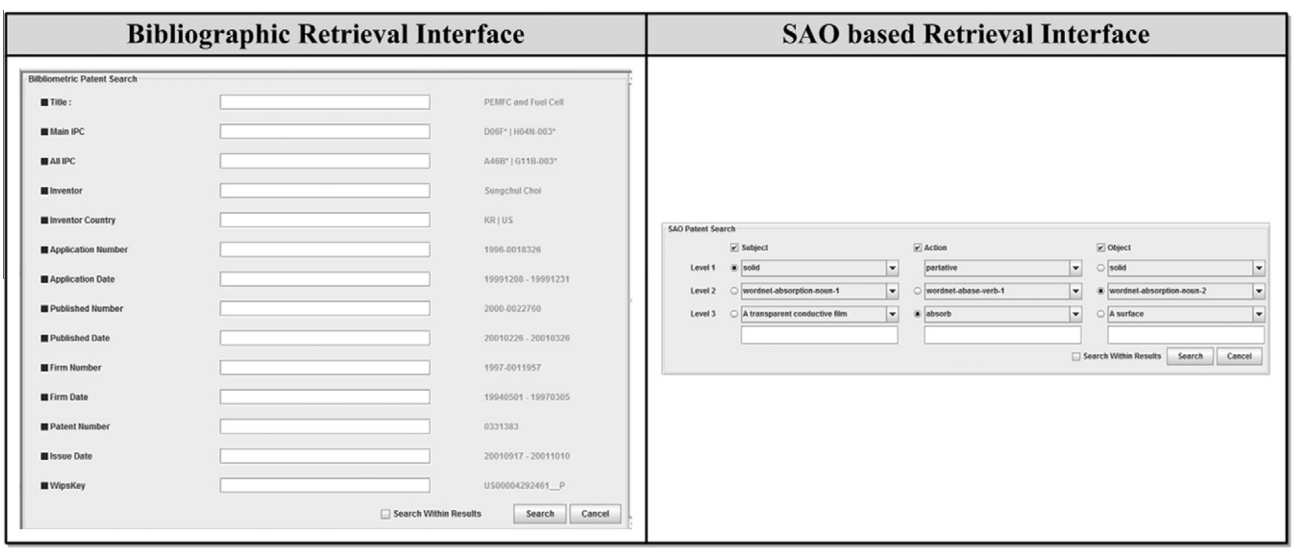


**Fig. 9.** Interfaces of patent retrieval manager.

ontology information. SeRQL uses a declarative SQL-like syntax for querying information contained in one or more RDF triples in Sesame Sever. To query SAO models, the patent retrieval manager generates queries in response to a retrieval request. We present SeRQL examples (Fig. 11) of two cases: using WordNet definition retrieval to search for an action–object model depicted as ‘reduce performance’; and combining WordNet definition retrieval in action and abstract SAO model retrieval to search for an action–object model depicted as ‘reduce field’.

## 6. Case study

To illustrate the feasibility of the approach proposed in this paper, we conducted two case studies in which we used the implemented system to search for function-based technology information. The first case is a search for patents that have the same problem–solution relationship, with the goal of supporting product development. Specifically, we analyze the problem–solution relationship of WhiteStrips™ developed by Proctor & Gamble, Inc. (P&G). The second case is a search for patents that use similar technology. Specifically, we compared Silicon Thin Film Solar Cell (TFSC) technology to Semiconductor Manufacturing technology that uses Chemical Vapor Deposition (CVD).

### *6.1. Case study 1: WhiteStrips™*

This case uses the proposed approach to analyze the development of WhiteStrips™ (GEN3-Partners, 2008). By using FOS, P&G

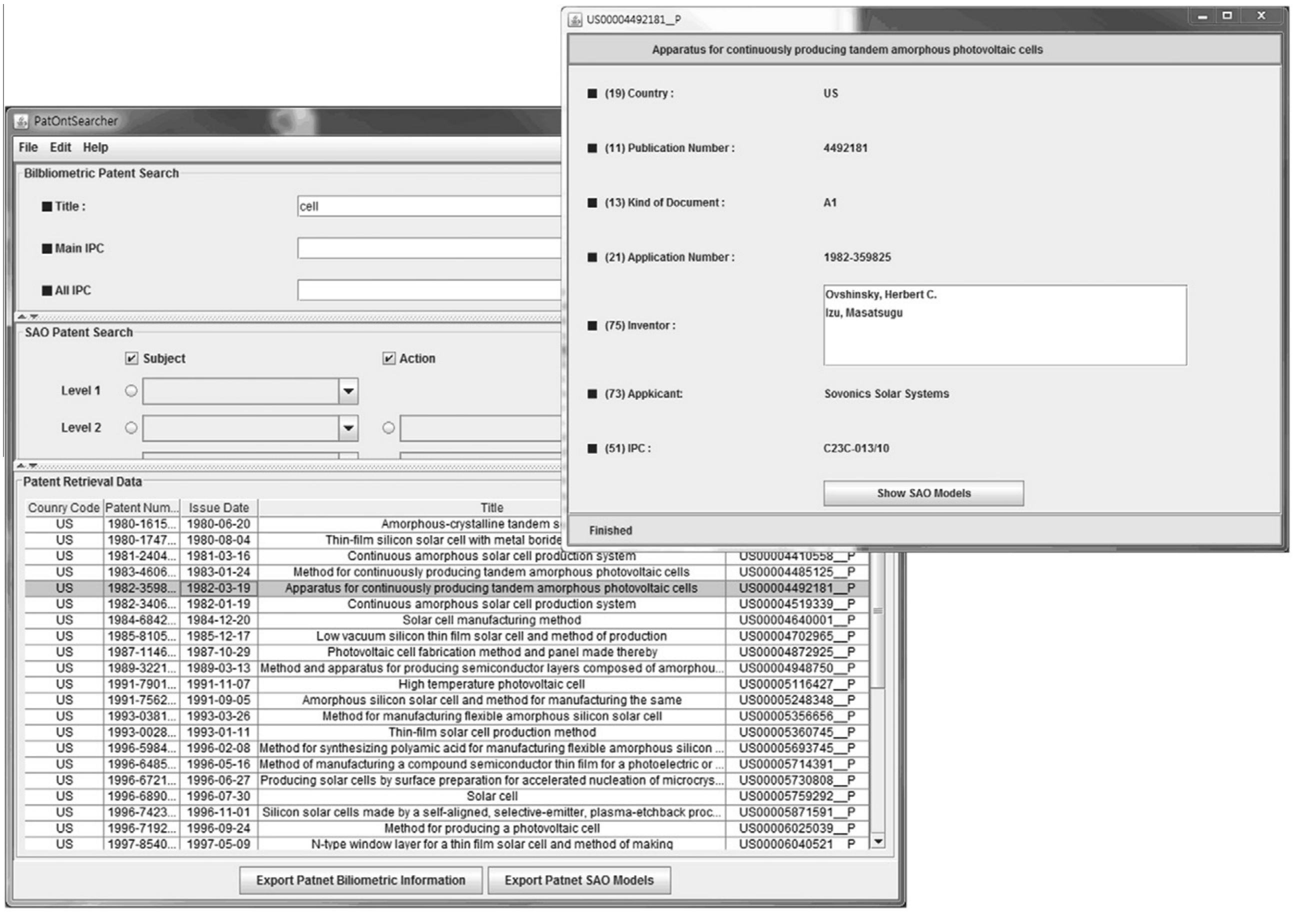


**Fig. 10.** Results of function-based technology information retrieval.

tried to develop new device for tooth whitening and contracted GEN3 Partners, a Russian TRIZ consulting company, to improve an existing P&G product that used whitening gel in a polymer tray, which users put into their mouth and bit down on. This process required considerable time and was uncomfortable to use (Fig. 12).

To develop a new product, GEN3 Partners analyzed the function of the existing product and concluded that the new product should not have the bulk and expense of the molded inserts, but should maintain tooth-bleaching effectiveness. For this, they proposed that the new product be changed to a thin film that had a self-adhesion property. Based on this conclusion, they searched for related technology in the field of Medicine and Pharmacology (IPC Code: A61) and found a nicotine patch that had the necessary characteristics. P&G accepted this conclusion and four months later developed a new product called WhiteStrips™ that uses a thin flexible film (Fig. 13). In the first year after release, WhiteStrips™ generated $US 130 billion in revenue.

In this case study, the action–object combination of '*glued to the teeth*' was used to find related technology. By abstracting this action–object, we can generate an AO model as follows: the Action word is '*wordnet-glue-verb-1*' in the WordNet definition level and the object word is '*solid*' in the abstract SAO model. Because '*wordnet-glue-verb-1*' has synonyms such as '*wordnet-attach-verb-1*', '*wordnet-paste-verb-1*', and '*wordnet-adhesive-noun-1*', the system also considers other related words. By using this AO model, we also conducted patent retrieval with IPC code A61 information for limited search.

The patents found (Table 2) refer to Nicotine Patches. Regarding the AO model 'adhere to – skin', they suggest solution technologies 'active pads' or 'patch'. The results are similar to that of GEN3 Partners. As described in this case study, the proposed approach can help product development process by providing related patents that include a problem–solution relationship. The product developer uses this approach to find solutions of problem by searching function-based technology information.

To conduct the similar search with the existing tools, the user must use the word 'skin' which is not generalized. However, without background knowledge, the user has a difficulty to define 'skin' as an important concept. Compared to the existing FTDB, the proposed approach can support search of function-based technology information without defining 'skin' as an important concept. In the proposed approach, a user only uses 'Solid' which is abstracted concept. In addition, this approach uses WordNet synonym relationships when searching for information. These advantages provide better function generalization and semantics handling than existing FTDBs, and allow users to utilize FOS easily.

### *6.2. Case study 2: Silicon TFSC and CVD*

This case uses the proposed approach to compare silicon TFSC technology with CVD. Solar cells, also called photovoltaic cells, use the photovoltaic effect to convert solar energy directly into electricity. The basic structure of the solar cell is a p–n junction. The process of generating electricity is started by the absorption of photons in the p–n junction. After absorption, an electron–hole pair is generated if the photon energy is above the band gap energy of semiconductor; then, due to an electric field at the p–n junction surface, the electron–hole pair separates. The electron moves to the n-layer and the hole moves to the p-layer. The separated elec-

***Query 1***

| Purpose of Query | Looking for Patent number having AO function regarding 'reduce' 'performance' |
| --- | --- |

SeSQL

```
SELECT   Patent
FROM               {Patent}   patent:hasSAOs      {patentinstances:SAOList}
WHERE EXISTS
(
    SELECT  SAO
    FROM           {SAO}    rdf:predicate        {wn20instances:wordsense-reduce-verb-8};
                            rdf:object {wn20instances:wordsense-performance-noun-4}
    INTERSECT
    SELECT  SAO
    FROM           {patentinstances:SAOList} predicate {SAO}
)

USING NAMESPACE
rdfs = <http://www.w3.org/2000/01/rdf-schema#>,
rdf = <http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
patent = <http://phlox.postech.ac.kr:8080/patent/upo#>,
wn20instances = <http://www.w3.org/2006/03/wn/wn20/instances/>,
saoinstances = <http://phlox.postech.ac.kr:8080/patent/sao/instances#>,
fusion = <http://metadata.net/sunago/fusion/fusion.owl#>,
patentinstances = <http://phlox.postech.ac.kr:8080/patent/instances#>
```

***Query 2***

| Purpose of Query | Looking for Patent number having AO function regarding 'reduce' 'field' |
| --- | --- |

SeSQL

```
SELECT   Patent
FROM               {Patent}   patent:hasSAOs      {patentinstances:SAOList}
WHERE EXISTS
(
    SELECT  SAO
    FROM           {SAO}    rdf:predicate        {wn20instances:wordsense-reduce-verb-8};
                            rdf:object {object}   rdf:type   {patent:Field}
    INTERSECT

    SELECT  SAO
    FROM           {patentinstances:SAOList} predicate {SAO}
)


USING NAMESPACE
rdfs = <http://www.w3.org/2000/01/rdf-schema#>,
rdf = <http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
patent = <http://phlox.postech.ac.kr:8080/patent/upo#>,
wn20instances = <http://www.w3.org/2006/03/wn/wn20/instances/>,
saoinstances = <http://phlox.postech.ac.kr:8080/patent/sao/instances#>,
fusion = <http://metadata.net/sunago/fusion/fusion.owl#>,
patentinstances = <http://phlox.postech.ac.kr:8080/patent/instances#>
```

**Fig. 11.** Examples of SeRQL for SAO-based patent retrieval.

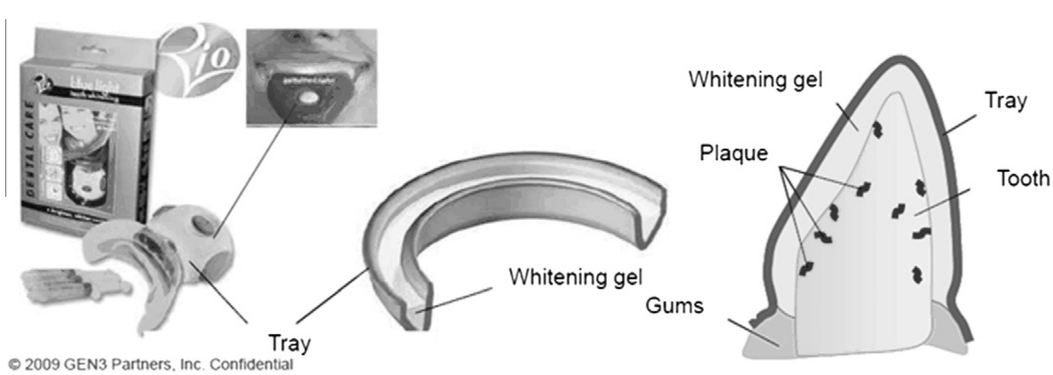


**Fig. 12.** The existing teeth whitening device (from GEN3 Partners).

tron–hole pair generates electromotive force and current flows when an electrical device is connected to the two layers.

Originally, single-crystal silicon was used for the fabrication of a solar cell. The single crystal silicon solar cell has higher conversion efficiency than cells that use other materials. However, it has high material cost and requires a complex manufacturing process, because it is produced from bulk type raw material. To overcome these problems, use of thin film solar cell deposition technology to decrease cell thickness has received considerable attention.

A representative technology for solar cell deposition is the CVD method used for semiconductor manufacturing. CVD starts by decomposing a gaseous compound. The CVD is used to form a thin film or epitaxial layer on silicon wafer by chemical reaction of the decomposed compound. The physical and chemical properties of


**Fig. 13.** WhiteStrips™ (from GEN3 Partners).

**Table 2**
The results of retrieval of AO combination of '*glued to the teeth*'.

| Patent number | Patent title | Extracted SAO structures |
|---|---|---|
| EP0916339 | Prolonged activity nicotine patch | Active pads – adhere to – skin |
| US5897522 | Flexible thin layer open electrochemical cell and applications of same | Patch – adhere to – skin |

**Table 3**
The results of retrieval: ''solar cell use''.

| Patent number | Title | Extracted SAO models |
|---|---|---|
| US6531711 | Photoelectric conversion device and process for producing photoelectric conversion device | Solar cell-use-amorphous silicon film |
| US6189485 | Plasma CVD apparatus suitable for manufacturing solar cell and the like | Solar cell-use-amorphous silicon thin film |
| US6488995 | Method of forming microcrystalline silicon film, method of fabricating photovoltaic cell using said method, and photovoltaic device fabricated thereby | Solar cell-use-microcrystalline silicon |
| US6177711 | Photoelectric conversion element | Solar cell-use-microcrystalline silicon |

**Table 4**
The results of retrieval: ''Produce film''.

| Patent number | Title | Extracted SAO models |
|---|---|---|
| US6189485 | Plasma CVD apparatus suitable for manufacturing solar cell and the like | Above mentioned plasma CVD method-make-amorphous silicon thin film |
| US6645573 | Process for forming a microcrystalline silicon series thin film and apparatus suitable for practicing said process | CVD-produce-amorphous silicon thin film |
| US6124545 | Thin film solar cell | Inventive method-produce-amorphous silicon thin film solar cells |
| US6100466 | Method of forming microcrystalline silicon film, photovoltaic element, and method of producing same | Method-produce-amorphous silicon film |
| | | Method-produce-microcrystalline silicon film |
| US5278015<br>US5258207<br>US5152833 | Amorphous silicon film, its production and photo semiconductor device utilizing such a film | PACVD process-produce-amorphous silicon film |
| US5391410 | Plasma CVD of amorphous silicon thin film | Plasma CVD method-produce-amorphous silicon thin film |

thin film produced by CVD depend on the type of substrate, the characteristics of the deposition site, and the deposition conditions. These variables affect deposition velocity, and thus determine the structure and properties of the thin film.

To analyze this technology using proposed approach, we used our implemented system to randomly store 100 patents regarding solar cell technology and 100 patents regarding CVD. Then we used subject and action information to search the solar cell patents. The subject was '*wordnet-solar_cell-noun-1*' and the action was '*wordnet-use-verb-1*' in WordNet definition.

The retrieval results (Table 3) indicates that in general, either amorphous silicon film or microcrystalline silicon is used as the materials of a solar cell. Based on these results, we used an AO model to search again among these patents. The Action was '*wordnet-pro-*

*duce-verb-2*' at the WordNet definition level, and the object was '*amorphous silicon*' or '*microcrystalline silicon*' at the SAO instance level. In this retrieval, the '*wordnet-produce-verb-2*' has two synonyms such as '*wordnet-make-verb-6*' and '*wordnet-create-verb-6*'. Therefore, our implemented system used a SeRQL query that considered WordNet synonyms to searched patents that contain SAO model which has '*wordnet-produce-verb-2*', '*wordnet-make-verb-6*' or '*wordnet-create-verb-6*'.

Search results (Table 4) indicate that the CVD method using plasma is mainly used to produce 'amorphous silicon thin film' or 'microcrystalline silicon film'. Plasma-Assisted Chemical Vapor Deposition (PACVD) and Plasma Enhanced Chemical Vapor Deposition (PECVD) are kinds of CVD methods that use plasma during the deposition process. This method allows deposition at lower temperatures than other CVD methods. This feature is often critical in the deposition of temperature-sensitive materials such as semiconductors, plastic or glass. The amorphous silicon film deposited by plasma-CVD is a semiconductor with excellent electrical conductivity. Because of this advantage, plasma-CVD is widely used for producing silicon films for solar cells. This case shows that the proposed approach can correctly interpret acronyms such as 'PACVD' and 'PECVD' by using WordNet to model technology concepts. In the existing FTDB, users faced difficulties when searching technology information that includes acronyms. In addition, by using a verb relationship defined in WordNet, we can search more related function information than the existing FTDB. For example, we can use verb relationships among '*wordnet-produce-verb-2*', '*wordnet-make-verb-6*', and '*wordnet-create-verb-6*' when searching the function information. This advantage improves semantics handling and simplifies utilization of FOS.

## 7. Conclusion and future work

We proposed a fact-oriented ontological approach to FTDB for supporting FOS. For this, we suggested the fact-oriented ontological modeling approach for generating FTDB and provided the implementation of an SAO-based patent retrieval system that uses the FTDB. The proposed approach can remedy the limitations of existing FTDBs. Finally, we illustrate the feasibility of the proposed approach by using the implemented system to conduct two case studies. As a tool to support FOS, the proposed approach can find new aspects of technology information which are provided incompletely by exiting FTDB systems. The proposed approach can be used as a complementary tool to help researchers or R&D managers when searching for new concepts or new technology analysis methods. From the perspective of modeling SAO structures, the fact-oriented approach helps domain and ontology experts to generate ontologies effectively. Domain experts may not be familiar with modeling representations such as UML or Object-Oriented Modeling. But because the fact-oriented approach expresses the ontologies as simple English sentences, it helps domain experts easily understand ontology modeling. This advantage allows the two types of expert to smoothly cooperate when modeling ontology.

In the future, we will research NLP techniques to improve quality of extracted SAO models. This will further minimize the participation of people in the proposed approach. By reducing the participation of experts, the cost of generating an ontology can be minimized. Another future research objective is to establish a diverse function abstraction model. To express more diverse technology idea at the abstraction level, a well-established abstraction model is needed. This topic should be studied in the TRIZ field.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2009-0088379).

## References

Altshuller, G. S. (1984). *Creativity as an exact science: The theory of the solution of inventive problems*. New York: Gordon and Breach Science Publishers.
Borenstein, J., & Fox, J. (2003). Semantic discovery for web services. *Web Services Journal*, 3.
Cascini, G., Lucchhesi, D., & Rissone, P. (2001). Automatic patents functional analysis through semantic processing. In XII ADM International Conference. Rimini, Italy. Available from: http://adm.ing.unibo.it/ADM%20Rimini/Papers/C2/1-ADM01-C2-065.pdf.
Cascini, G., Fantechi, A., & Spinicci, E. (2004b). Natural language processing of patents and technical documentation. In 11th International Conference PAM 2010 (Vol. 3163, pp. 508–520). Zurich, Switzerland.
Cascini, G., Fantechi, A., & Spinicci, E. (2004a). Natural language processing of patents and technical documentation. In S. Marinai & A. Dengel (Eds.). *Document Analysis Systems* (Vol. 3163, pp. 89–92). Berlin/Heidelberg: Springer.
Charniak, E., Carroll, G., Adcock, J., Cassandra, A., Gotoh, Y., Katz, J., et al. (1996). Taggers for parsers. *Artificial Intelligence, 85*, 45–57.
Choi, S., Lim, J., Yoon, J., & Kim, K. (2010). Patent function network analysis: a function based approach for analyzing patent information. In Y. Hosni & T. Khalil (Eds.), 19th International Conference for the International Association of Management of Technology. Cairo, Egypt.
CREAX. (2005). CREAX function database.
Crotti, A., Ghitti, M., Regazzoni, D., & Rizzi, C. (2007). Trends of evolutions and patent analysis: an application in the household appliances field. In The Future of Product Development (pp. 145–154).
Fey, V. R., & Rivin, E. I. (1999). *Guided technology evolution (TRIZ Technology Forecasting)*. TRIZ Journal.
GEN3-Partners. (2008). Client Case Study : "Teeth Whitening Strips".
Gruber, T.R. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In the International Workshop on Formal Ontology: Citeseer.
Halpin, T. (2007). Fact-oriented modeling: past, present and future. In J. Krogstie, A. L. Opdahl, & S. Brinkkemper (Eds.), *Conceptual modelling in information systems engineering* (pp. 19–38). Berlin/Heidelberg: Springer.
Kang, D., Lee, J., Choi, S., & Kim, K. (2010). An ontology-based enterprise architecture. *Expert Systems with Applications, 37*, 1456–1464.
Litvin, S. (2005a). NEW TRIZ-Based Tool — Function-Oriented Search (FOS). In TRIZCON 2005 (pp. 505–509). Florence, Italy.
Litvin, S. (2005b). TRIZ Readings – Altshuller's Tradition Continues. In ETRIA Conference TRIZ Future 2005. Graz, Austria.
Liu, H. (2004). MontyLingua: An end-to-end natural language processor with common sense. In (Vol. 2009).
Liu, H., & Singh, P. (2004). ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal, 22*, 211–226.
Mann, D. L. (2003). Better technology forecasting using systematic innovation methods. *Technological Forecasting and Social Change, 70*, 779–795.
Moehrle, M., Walter, L., Geritz, A., & Müller, S. (2005). Patent-based inventor profiles as a basis for human resource decisions in research and development. *R&D Management, 35*, 513–524.
Object Management Group (2005). Semantics of business vocabulary and business rules (sBVR) specification.
Salamatov, Y., & Souchkov, V. (1999). *TRIZ: the right solution at the right time: a guide to innovative problem solving*. Hattem: Insytec.
Savransky, S. (2000a). *Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving*. CRC Press.
Savransky, S. D. (2000b). *Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving*. London: CRC Press.
Tsourikov, V. M., Batchilo, L. S., & Sovpel, I. V. (2000). Document semantic analysis/selection with knowledge creativity capability utilizing subject–action–object (SAO) structures. In: Google Patents.
Uschold, M. (1998). Knowledge level modelling: concepts and terminology. *Knowlegdge Engineering Review., 13*, 5–29.
Uschold, M., & Jasper, R. (1999). A framework for understanding and classifying ontology applications. In 12th Workshop on Knowledge Acquisition Modeling and Management (Vol. 99, pp. 16–21): Citeseer.
Zhang, J., Yang, B., Tian, Y., & Tan, R. (2007). Technology innovation of product using CAI system based on TRIZ. In Trends in Computer Aided Innovation (pp. 97–106).
Zhang, F.-y., & Xu, Y.-s. (2006). Research on technical strategy for new product development based on TRIZ evolution theory. *International Journal of Product Development, 4*, 96–108.