KNOWLEDGE DISCOVERY SYSTEM FOR PATENTS

AVIV SEGEV JUSSI KANTOLA
Department of Knowledge Service Engineering
KAIST
335 Gwahangno, Yuseong-gu, Daejeon 305-701, Korea
{aviv, jussi}@kaist.edu

Patent search requires identifying the boundaries of existing knowledge. Every patent request requires a decision maker to study all the aspects of the request. The system implements a model for representing the patent request by a set of concepts related to existing knowledge ontology. The search for patent information is based on Fuzzy Logic decision support, allowing a more comprehensive search. The system is currently being analyzed in assisting the decision process in the Korean Patent Office.

1. Introduction

Patent knowledge discovery is unique compared to other knowledge based systems because of the requirement to identify whether similar knowledge exists as opposed to the need to locate knowledge. Contemporary knowledge based systems are based on using existing information, while the patent support system is required to assist in identifying similar domains and patterns that would facilitate the decision whether to grant the patent request (Cong and Tong, 2008). Furthermore, another difficulty is that patents in different countries are not classified under one classification system.

The main problem encountered when searching for existing patents is verifying that all relevant documents related to the current invention were retrieved. If a relevant document is missed, then a patent could be granted to an already existing work. On the other hand, retrieving an irrelevant document would only lead to minor additional work from the patent inquirer or decision maker. The current decision process for granting patents averages 3-4 years depending on the specific field of technology. The system presented here aims at benefitting both the patent office decision maker who needs to decide whether to grant a patent for each request and inventors and companies that would like to inquire about existing patented technology.

The patent knowledge discovery method described here is based on a model for designing a service based on ontology for the domain representation of the patent request combined with Fuzzy Logic for the decision support (Figure 1). The model inputs are the patent request document, which is written in free text, and the user, the patent officer, queries, which can be either structured or free text. The service assists in extracting relevant knowledge for determining the likelihood that the patent request is covered by previous patents or existing knowledge. The service allows the decision maker an option to identify the reasoning and to modify the requirements or the decision qualifications for each patent request.

2. Related Work

Ontologies have been defined and used in various research areas, including philosophy (where it was coined), artificial intelligence, information sciences, knowledge

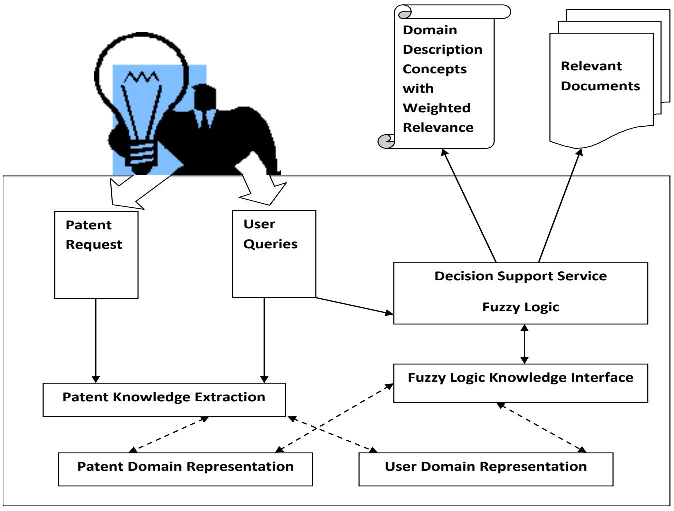

Figure 1 - Patent service model

representation, object modeling, and most recently, eCommerce applications. Bunge (1979) defines Ontology as a world of systems and provides a basic formalism for ontologies. Typically, ontologies are represented using Description Logic (Borgida, 1993), where subsumption typifies the semantic relationship between terms, or Frame Logic (Kifer *et al.*, 1995) where a deductive inference system provides access to semi-structured data.

The realm of information science has produced an extensive body of literature and practice in ontology construction, such as DOGMA (Spyns *et al.*, 2002), which provides an engineering approach to ontology management. Work has been done in ontology learning, such as Text-To-Onto (Maedche and Staab, 2001) and Mapping Context to Ontology (Segev and Gal, 2007) to name a few. Finally, researchers in the field of knowledge representation have studied ontology interoperability, resulting in systems such as Protégé (Noy and Musen, 2000).

Vagueness in linguistics can be captured mathematically by applying fuzzy sets (Lin and Lee, 1996). Fuzzy sets represent objects and concepts better than do crisp sets. There are two reasons for this. First, the predicates in propositions representing a system do not have crisp denotations. Second, explicit and implicit quantifiers are fuzzy (Zadeh, 1983). A fuzzy set can be defined mathematically by assigning to each possible individual in the universe of discourse a value representing its grade of membership in the fuzzy set. This grade corresponds to the degree to which that individual is similar to or compatible with the concept represented by the fuzzy set (Klir, 1995).

In an ongoing work called PATexpert (Wanner *et al.*, 2008) several areas of patent services are targeted. PATexpert approach is different from the search method proposed in this system. First, in PATexpert the classification process is manual. In our method the classification/search is a semi-automatic process. Second, the meaning of fuzzy in PATexpert is in the morphological and spelling sense. In the method proposed, the fuzzy refers to Fuzzy Sets and Fuzzy Logic for the reasoning and decision making process. We present a joint application of ontology matching and Fuzzy Sets that enables a searcher-friendly service.

3. Patent Service Model

The implementation of the model begins when the patent office user initializes the process of evaluating the patent request. A simple syntactic search might look for documents relating to a term, such as *Length,* which appears in the text. However, the described model expands the search results to include documents related to additional concepts not mentioned in the text.

3.1. *Patent Knowledge Extraction*

Each claim is analyzed separately through the Domain Representation process. To analyze the claims, a context extraction algorithm is used. For each claim the context, a set of descriptors, is extracted by the Patent Knowledge Extraction and then compared with the ontology concept by the Patent Domain Representation.

The Patent Knowledge Extraction process uses the World Wide Web as a knowledge base to extract multiple contexts for the textual information. The algorithm input is defined as a set of textual propositions representing the claim information description. The result of the algorithm is a set of contexts – textual descriptors that are related to the propositions. The algorithm was adapted from (Segev et. al, 2007).

3.2. *Patent Domain Representation*

Patent Domain Representation performs the ontology matching process that directs the claim to the relevant ontological concepts. One of the difficult tasks is matching each information datum with the correct concepts without the usual training process required in ontology adjustment and usually performed over a long period of time.

A simplified representation of an ontology is $O \equiv <C,R>$, where $C=\{c_1,c_2,\ldots,c_n\}$ is a set of concepts with their associated relation R. A concept can consist of multiple context descriptors and can be viewed as a meta-representation of the patent domain. The added value of having such a meta-representation is that each context descriptor can belong to several ontology concepts simultaneously. For example, a context descriptor <Length, 20> can be shared by many ontology concepts that have interest in length analysis (such as *Distance* or *Wave*) although it is not in their main role.

The relevance of the information to each concept is evaluated according to the weight attributed to each concept. The weight can be calculated according to the number of references of the concept in the Web, number of references to the concept in the patent, or both combined. For example, a patent can be associated with concept *Distance* with weight 4 and concept *Wave* with weight 3. To evaluate the matching of the concepts with the information and its context, a simple string-matching function is used.

3.3. *User Domain Representation*

When a new patent request is processed, the first step involves the ontology matching process. Once the patent request is classified, the following relations with existing patents can occur: If the patent is related to concepts that are associated with existing patents, the decision process requires reviewing the existing patents and comparing them to the request. If the patent is not related to concepts similar to existing patents, the decision maker can extend the search according to related concepts until related patents are identified with overlapping concepts associated with the patent request. If the second option is encountered, the decision maker faces a dilemma of whether to grant the patent based on the relation of existing patents to the current patent. To assist in the process of decision making in these instances, a fuzzy logic process is presented.

3.4. *Fuzzy Logic Knowledge Interface*

In fuzzy information retrieval the relevance of the index terms is expressed by a fuzzy relation: $R:X^xY \rightarrow [0,1]$ where the membership value R(x,y) for each x_i and y_i represents the grade of relevance of index term x_i to document y_i (Aliev & Aliev, 2001). The basic scheme of fuzzy information retrieval is where U1 is a fuzzy set representing a particular inquiry. When U1 is composed with Thesaurus (T), then U2 becomes an inquiry augmented by associated index terms: $U2=U1\circ T$. U2 can be expressed as follows: $U2(x_i)=\max[\min[U2(x_i),T(x_i,x_j)]]$. Then a relevant document search can be expressed by: $D=U2\circ R$. The role of Fuzzy Thesaurus T can be carried out by a set of ontologies that are further linked to the lexical database Wordnet (Fellbaum, 1998). In the proposed approach the role of the fuzzy thesaurus (T) is carried out by the ontology matching process (O). The basic scheme of fuzzy information retrieval U2 becomes an inquiry augmented by associated index terms from ontology matching: $U2=U1\circ O$. The inquirer can inspect all the documents that have support D, or she can filter the inspection to those supported by some α-cuts.

3.5. *Decision Support Service Fuzzy Logic*

The user examining patent claims can expand the search to other possibly related concepts as well by selecting a mode for extended search by choosing “Strict” mode or “Vague” mode. The user enters a patent into the Web based ontology matching process. A list of related concepts, together with the degrees of relevance, is presented. The degree of relevance is calculated based on the concept weight in searched documents provided by the ontology matching algorithm and fuzzy membership functions.

4. Patent Service Model Implementation

The implementation of the model is currently being tested at the Korean Intellectual Property Office (KIPO). KIPO seeks to improve the ability to identify and classify new patents. KIPO’s goal is to optimize the examination infrastructure, improve the quality of examinations, and enhance the effectiveness of quality management.

Figure 2 shows the Fuzzy Logic Ontology Context Knowledge (FLOCK) demonstrator application that was used to test the model described. The FLOCK system for extracting concepts and relevant patent documents was evaluated by six KIPO Patent Officers who routinely process patent requests.

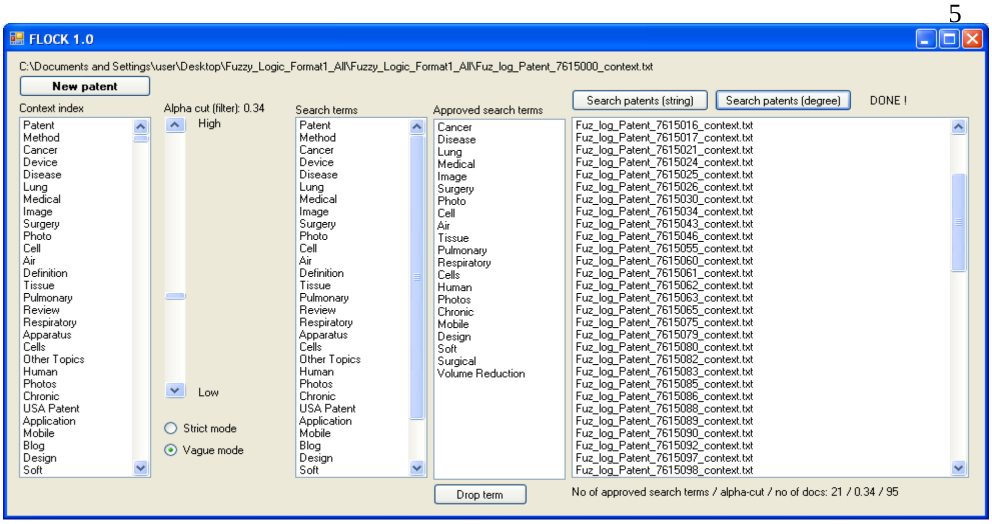

Figure 2 - The FLOCK system

References

Aliev, R.A., Aliev, R.R. (2001). *Soft Computing and Its Applications*, World Scientific, Singapore.

Borgida, A., Brachman, R.J. (1993). “Loading data into description reasoners,” *Proceedings ACM SIGMOD International Conference on Management of Data*.

Bunge, M. (1979). *Treatise on basic philosophy, vol. 4: Ontology II: A World of Systems*. D. Reidel Publishing Co., Inc.

Cong H., Tong L.H. (2008). Grouping of TRIZ Inventive Principles to facilitate automatic patent classification,” Expert *Systems with Applications*, 34: 788–795

Fellbaum C. (1998). *WordNet: An electronic lexical database*, Cambridge (MA): MIT Press.

Kifer, M., Lausen, G., Wu, J. (1995). “Logical foundation of object-oriented and frame-based languages,” Journal *of the ACM*,42.

Klir, J.G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice-Hall Inc.

Lin C.-T., Lee C.S. (1996). *Neural fuzzy systems: A neuro-fuzzy synergism to intelligent systems*, Prentice-Hall Inc.

Maedche, A., Staab, S. (2001). “Ontology learning for the semantic web,” *IEEE Intelligent Systems*, 16.

Noy, F.N., Musen, M.A. (2000). “PROMPT: Algorithm and tool for automated ontology merging and alignment,” *Proceedings of AAAI*.

Segev A., Gal A. (2007). “Putting things in context: A topological approach to mapping contexts to ontologies,” *Journal on Data Semantics*, IX: 113–140.

Segev A., Leshno M., and Zviran M. (2007). “Context recognition using Internet as a knowledge base,” *Journal of Intelligent Information Systems*, 29(3).

Spyns, P., Meersman, R., Jarrar, M. (2002). “Data modelling versus ontology engineering,” *ACM SIGMOD Record*, 31(4).

Wanner, L., Baeza-Yatesa, R., Brügmann, S., Codina, J., Diallo, B., Escorsa, E., Giereth, M., Kompatsiaris, Y., Papadopoulos, S., Pianta, E., Piella, G., Puhlmann, I., Raoe, G., Rotard, M., Schoester, P., Serafini, L. and Zervaki, V. (2008). “Towards content-oriented patent document processing,” *World Patent Information*, 30(1): 21-33.

Zadeh, L.A. (1983). “Commonsense knowledge representation based on fuzzy logic,” *Computer*, 16: 61-65.