Expert Systems with Applications 40 (2013) 7010–7023

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Analyzing multilingual knowledge innovation in patents

Aviv Segev [a,*], Jussi Kantola [b], Chihoon Jung [a], Jaehwa Lee [a]

[a] Department of Knowledge Service Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, South Korea
[b] Department of Production, University of Vaasa, P.O. Box 700, Vaasa FI-65101, Finland

ARTICLE INFO

Keywords:
Conceptual modeling
Ontologies
Knowledge management applications
Database semantics

ABSTRACT

In the process of analyzing knowledge innovation, it is necessary to identify the existing boundaries of knowledge so as to determine whether knowledge is new – outside these boundaries. For a patent to be granted, all aspects of the patent request must be studied to determine the patent innovation. Knowledge innovation for patent requests depends on analyzing current state of the art in multiple languages. Currently the process is usually limited to the languages and search terms the patent seeker knows. The paper describes a model for representing the patent request by a set of concepts related to a multilingual knowledge ontology. The search for patent knowledge is based on Fuzzy Logic Decision Support and allows a multilingual search. The model was analyzed using a twofold approach: a total of 104,296 patents from the United States Patent and Trademark Office were used to analyze the patent extraction process, and patents from the Korean, US, and Chinese patent offices were used in the analysis of the multilingual decision process. The results display high recall and precision and suggest that increasing the number of languages used only has minor effects on the model results.

1. Introduction

In the analysis of the boundary of knowledge, such as in the process of granting patents, there is a difference between the need to locate knowledge and the need to identify whether similar knowledge exists. The search of the boundary of knowledge examines whether given concepts exist, while regular knowledge search looks for instances of existing concepts. Contemporary knowledge-based services depend on using existing knowledge, while Patent Knowledge Extraction is required to assist in identifying similar domains and patterns that will facilitate the decision whether to grant the patent request (Cong & Tong, 2008). Furthermore, another difficulty is that patents in different countries are not classified under one classification system and employ multiple languages.

Conversely, to invalidate a patent, relevant documents must be identified as "prior art", open to the public before the patent was filed. Analysis of patents involves searching for relevant patents and documents that could invalidate a claim within the patent or for a set of patents that could invalidate a claim when used together.

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, low recall, then a patent could be granted to an already existing work. Conversely, retrieving an irrelevant document, low precision, 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 main advantage of the model presented here is that it decreases the time required to review a patent request by supplying a semi-automatic guided search. The model 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.

In the growing number of open markets, the identification of patent knowledge is a challenging task due to the language barrier. Analyzing knowledge innovation for a patent request usually involves identifying the main concepts of the invention and searching for existing documents relating to the innovation. The process of knowledge analysis is usually limited to the languages of the patent seeker.

The Patent Knowledge Extraction method described in this paper presents a model based on ontology for the domain representation of the patent request combined with Fuzzy Logic for the decision support. The Patent Knowledge Extraction method has two main advantages: the knowledge is represented using the ontology modeling technique and the user is presented with powerful reasoning in knowledge extraction using the Fuzzy Logic methods.

The Patent Knowledge Extraction method is based on free text input in the language of the patent. An example of a sample patent

* Corresponding author. Tel.: +82 10 3402 1881.
E-mail addresses: aviv@kaist.edu (A. Segev), jussi.kantola@uwasa.fi (J. Kantola), chihoon.jung@kaist.ac.kr (C. Jung), tkod119@kaist.ac.kr (J. Lee).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.013

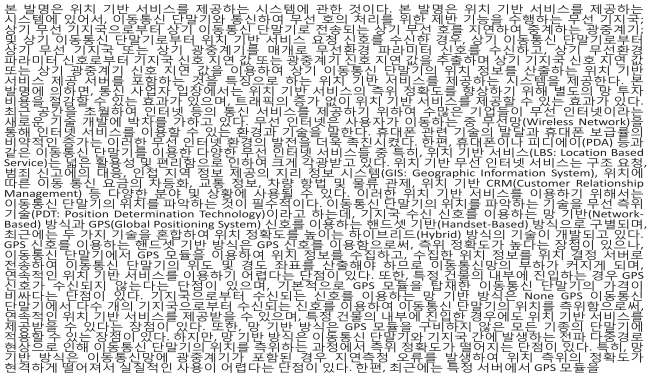
본 발명은 위치 기반 서비스를 제공하는 시스템에 관한 것이다. 본 발명은 위치 기반 서비스를 제공하는 시스템에 있어서, 이동통신 단말기의 통신하여 무선 호의 처리를 위한 기능을 수행하는 무선 기지국; 상기 무선 기지국으로부터 상기 이동통신 단말기로 전송되는 상기 무선 호를 지원하여 중계하는 광중계기; 및 상기 이동통신 단말기로부터 위치 기반 서비스 요청 신호를 수신한 경우, 상기 이동통신 단말기로부터 상기 무선 기지국 또는 상기 광중계기를 매개로 무선환경 파라미터 신호를 수신하고, 상기 무선환경 파라미터 신호로부터 기지국 신호 지연 값 또는 광중계기 신호 지연 값을 추출하며 상기 기지국 신호 지연 값 또는 상기 광중계기 신호 지연 값을 이용하여 상기 이동통신 단말기의 위치 정보를 산출하는 위치 기반 서비스 제공 서버를 포함하는 것을 특징으로 하는 위치 기반 서비스를 제공하는 시스템을 제공한다. 본 발명에 의하면, 통신 사업자 입장에서는 위치 기반 서비스의 즉위 정확도를 향상하기 위해 별도의 망 투자 비용을 절감할 수 있는 효과가 있으며, 트래픽의 증가 없이 위치 기반 서비스 제공 수 있는 효과가 있다. 최근 광대역을 초월하여 인터넷 등의 통신 서비스를 제공하기 위하여 수많은 기업들이 무선 인터넷이라는 새로운 기술 개발에 박차를 가하고 있다. 무선 인터넷은 사용자가 이동하는 중 무선망(Wireless Network)을 통해 인터넷 서비스를 이용할 수 있는 환경과 기술을 말한다. 휴대폰 관련 기술의 발달과 휴대폰 보급률의 비약적인 증가는 이러한 무선 인터넷 환경의 발전을 더욱 촉진시켰다. 하퍼, 휴대폰이나 피디에이(PDA) 등과 같은 이동통신 단말기를 이용한 다양한 무선 인터넷 서비스들 중 특히, 위치 기반 서비스(LBS: Location Based Service)는 넓은 활용성 및 편리함으로 인하여 크게 각광받고 있다. 위치 기반 무선 인터넷 서비스는 구조 요청, 범죄 신고에의 대응, 이정 지역 정보 제공의 지리 정보 시스템(GIS: Geographic Information System), 위치에 따른 이동 통신 요금의 자등화, 교통 정보, 차량 항법 및 물류 관제, 위치 기반 CRM(Customer Relationship Management) 등 다양한 분야 및 상황에 사용될 수 있다. 이러한 위치 기반 서비스를 이용하기 위해서는 이동통신 단말기의 위치를 파악하는 것이 필수적이다. 이동통신 단말기의 위치를 파악하는 기술로 무선 측위 기술(POT: Position Determination Technology)이라고 하는데, 기지국 수신 신호를 이용하는 망 기반(Network-Based) 방식과 GPS(Global Positioning System) 신호를 이용하는 핸드셋 기반(Handset-Based) 방식으로 구별되며, 최근에는 두 가지 기술을 혼합하여 위치 정확도를 높이는 하이브리드(Hybrid) 방식의 기술이 개발되고 있다. GPS 신호를 이용하는 핸드셋 기반 방식은 GPS 신호를 이용함으로써, 측위 정확도가 높다는 장점이 있으나 이동통신 단말기에서 GPS 모듈을 이용하여 위치 정보를 수집하고, 수집한 위치 정보를 위치 결정 서버로 전송하여 이동통신 단말기의 위도 및 경도 좌표를 산출해야 하므로 이동통신망의 부하가 커지게 되며, 연속적인 위치 기반 서비스를 이용하기 어렵다는 단점이 있다. 또한, 특정 건물의 내부에 진입하는 경우 GPS 신호가 수신되지 않는다는 단점이 있으며, 기본적으로 GPS 모듈을 탑재한 이동통신 단말기의 가격이 비싸다는 단점이 있다. 기지국으로부터 수신되는 신호를 이용하는 망 기반 방식은 None GPS 이동통신 단말기에서 다수 개의 기지국으로부터 수신되는 신호를 이용하여 이동통신 단말기의 위치를 측위함으로써, 연속적인 위치 기반 서비스를 제공받을 수 있으며, 특정 건물의 내부에 진입한 경우에도 위치 기반 서비스를 제공받을 수 있다는 장점이 있다. 또한, 망 기반 방식은 GPS 모듈을 구비하지 않은 모든 기종의 단말기에 적용할 수 있는 장점이 있다. 하지만, 망 기반 방식은 이동통신 단말기와 기지국 간에 발생하는 전파 다중경로 현상으로 인해 이동통신 단말기의 위치를 측위하는 과정에서 측위 정확도가 떨어지는 단점이 있다. 특히, 망 기반 방식은 이동통신망에 광중계기가 포함된 경우 지연측정 오류를 발생하여 위치 측위의 정확도가 현격하게 떨어져서 실질적인 사용이 어렵다는 단점이 있다. 한편, 최근에는 특정 서버에서 GPS 모듈을

Fig. 1. Sample free text input – patent in Korean.

input in Korean is displayed in Fig. 1. Current methods require translation of the patent or identification of the main related issues manually before searching for similar patents in multiple languages. The proposed solution is based on the automatic identification of related concepts represented in multiple languages and on the automatic extraction of relevant documents in different languages.

The Patent Knowledge Analysis model is described in Fig. 2. The model is based on two types of inputs. The first type is the patent submission request document, which is written in free text (Fig. 1). The second type is the queries performed by the service user, the patent officer, on either structured text or free text. Queries on structured text can be performed by adjusting relevant concepts weights. Queries on free text can be performed by modifying proposed concepts descriptors. The model assists in extracting relevant knowledge for determining the likelihood that the patent request is covered by previous patents or existing knowledge. The model allows the decision maker an option to drill down and identify the reasoning and to modify the requirements or the decision qualifications for each patent request. The Patent Knowledge Analysis model includes the following main modules: Patent Knowledge Extraction, Patent Domain Representation, Multilingual Domain Representation, Fuzzy Logic Knowledge Interface, and Fuzzy Logic Decision Support. The arrows represent the process flow, and the dotted arrows represent data extraction from the Patent Domain Representation, the Multilingual Domain Representation, and the storage of the Patent Ontology and the Patent Corpus.

The Patent Knowledge Extraction process is based on extracting knowledge from the free text based documents. The extraction process includes the identification of keywords that describe the context of the patent request and the association of relevant weights to each descriptor. The Patent Knowledge Extraction process forwards the knowledge to the Patent Domain Representation and Multilingual Domain Representation modules.

The Patent Domain Representation is based on using a multilingual ontology that allows all existing patents to be mapped according to the predefined concepts. Each concept is represented in multiple languages. The process allows the patent officer to create new concepts according to which existing patents can be automatically classified. The process can also be used to cluster the patents in order to seek new patent classifications.

The Multilingual Domain Representation process is directed by the patent officer who classifies the patent domain according to the user perspective of the knowledge. The knowledge is usually defined according to the domain of expertise and languages of the patent officer. Consequently, a specific patent can be classified both by the general concepts and by an existing structure that defines the patent office workers' expertise. The multilingual representation allows the user to classify the patent in one language and match it with similar patents according to the multilingual ontology.

The problem of patent search is that the inquirer cannot always find those documents that have the maximum relevance, because of the crisp approach which is defined as the exact approach of searching for relevance in database systems. Fuzzy Set theory

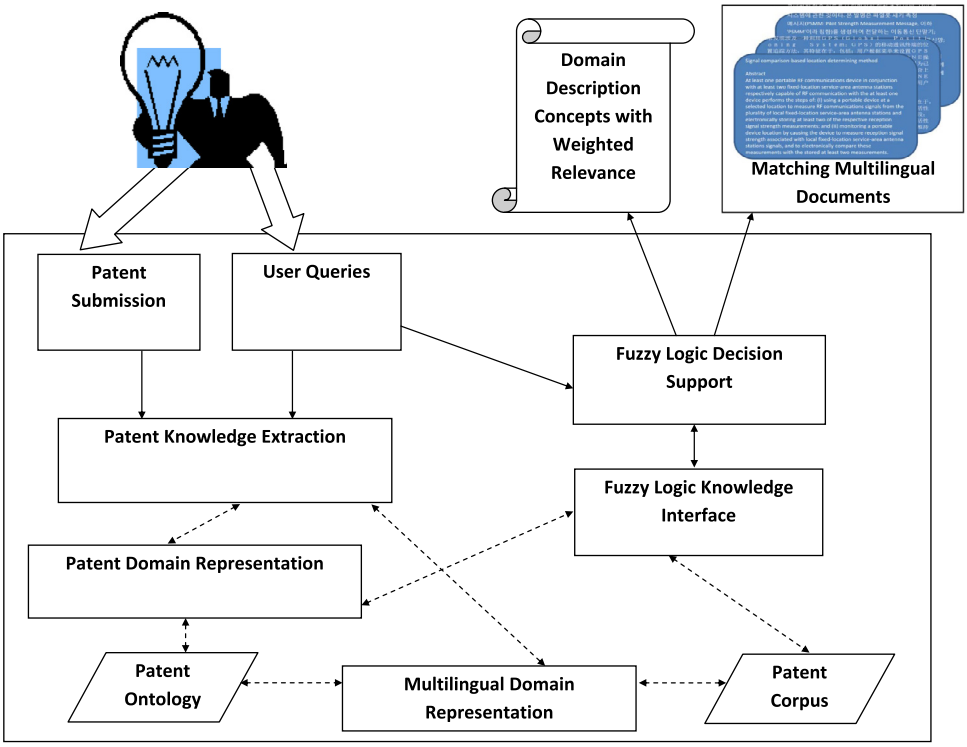

Fig. 2. Patent Knowledge Analysis model outline.