2010 Seventh International Conference on Information Technology

Patent Search Decision Support Service

Aviv Segev Jussi Kantola
Department of Knowledge Service Engineering, KAIST
335 Gwahangno (373-1 Guseong-dong), Yuseong-gu,
Daejeon 305-701, Korea
{aviv, jussi} @kaist.edu

***Abstract*—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 paper describes 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 model is currently being implemented and analyzed in assisting the decision process in the Korean Patent Office.**

***Index Terms*—Knowledge Engineering, Decision Support, Patent Service, Fuzzy Logic, Ontology**

I. Introduction

The patent service is unique compared to other knowledge based services because of the requirement to identify whether similar knowledge exists as opposed to the need to locate knowledge. Contemporary knowledge based services are based on using existing information, while the patent support service is required to assist in identifying similar domains and patterns which would reject a patent request. Furthermore, patents in different countries are not classified under one classification system.

The premise of the patent system lies in its mutual benefit to both the inventor and the public. In return for full public disclosure, a patent offers certain rights to an inventor for a limited period of time, during which the inventor may exclude all others from making, using, importing or selling his or her invention. The patent is published and disseminated to the public so that others may study the invention and improve upon it. The constant evolution of science and technology, spurred by the monetary incentive the patent system offers to inventors, strengthens the economy. New inventions lead to new technologies, create new jobs, and improve our quality of life.

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 decision support system presented here aims at benefitting both the patent office decision maker who needs to decide if to grant a patent for each request and to companies and inventors which would like to inquire about existing patented technology.

The patent decision support system described in this paper presents 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. The model is based on two types of inputs: the patent request document which is written in free text and the service user, the patent officer, queries which can be either structured or free text. The service assists in determining the likelihood that the patent request is not covered by previous patents or existing knowledge. The service 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 model is described in Figure 1 and includes the following main modules: Domain Representation Process, Patent Domain Representation, User Domain Representation, Fuzzy Logic Knowledge Interface, and Fuzzy Logic Decision Support Service.

The Domain Representation Process is based on extracting information from the free text based documents. The extraction process includes identifying keywords which describe the context of the patent request. The Domain representation forwards the information to the patent domain and user domain representation modules.

The Patent Domain Representation is based on creating an ontology which allows all existing patents to be mapped according to the predefined concepts. 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 User Domain Representation involves a directed process by the patent officer of classifying the patent domain according to user perspective of knowledge. The knowledge is usually defined according to the patent officer domain of expertise. Consequently, a specific patent can be classified both by the general concepts and by an existing structure which defines the patent office workers expertise.

The problem of patent search is that the inquirer cannot always find those documents that have the maximum relevance to the inquirer. The reason for this is the crisp approach of searching the relevance. Fuzzy set theory [23] and Fuzzy Logic [24] provide a robust and tractable way to move away from precise search approach. An imprecise fuzzy patent search can find related documents that otherwise cannot be found. This is possible when we introduce the degree of relevance to the patent search. Thus, the knowledge interface becomes fuzzy - like it is in the real world!

978-0-7695-3984-3/10 $26.00 © 2010 IEEE
DOI 10.1109/ITNG.2010.99

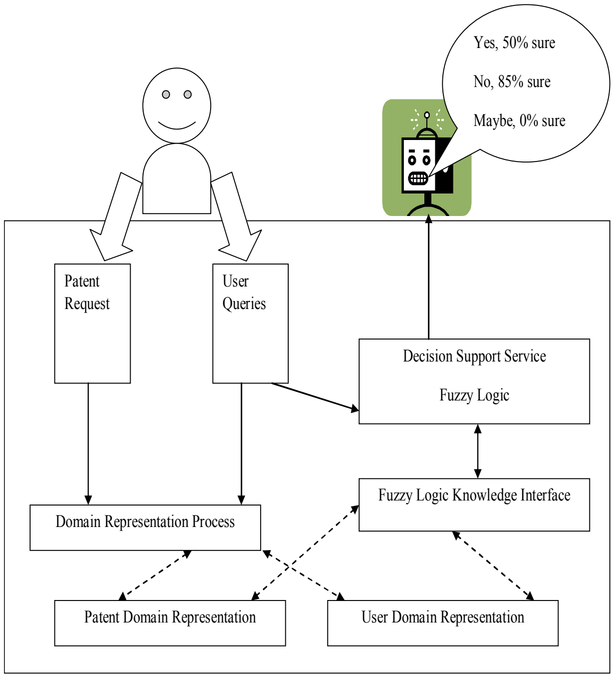

Figure 1 - Patent Service Model

II. Related Work

Ontologies have been defined and used in various research areas, including philosophy (where it was coined), artificial intelligence, information sciences, knowledge representation, object modeling, and most recently, eCommerce applications. In his seminal work, Bunge defines Ontology as a world of systems and provides a basic formalism for ontologies [3]. Typically, ontologies are represented using Description Logic [2][6], where subsumption typifies the semantic relationship between terms, or Frame Logic [9], where a deductive inference system provides access to semi-structured data.

Recent work has focused on ontology creation and evolution and in particular on schema matching. Many heuristics were proposed for the automatic matching of schemata (e.g., Cupid [13], GLUE [5], and OntoBuilder [8]), and several theoretical models were proposed to represent various aspects of the matching process [12][16].

The realm of information science has produced an extensive body of literature and practice in ontology construction, e.g., [21]. Other undertakings, such as the DOGMA project [20], provide an engineering approach to ontology management. Work has been done in ontology learning, such as Text-To-Onto [14], Mapping Context to Ontology [18], and OntoMiner [4], to name a few. Finally, researchers in the field of knowledge representation have studied ontology interoperability, resulting in systems such as Chimaera [15] and Protégé [17].

Vagueness in linguistics can be captured mathematically by applying fuzzy sets [11]. 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 [25]. 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 [10].

Fuzzy logic is reasoning with imprecise things. Fuzzy logic has two principle components. The first is a translation system for representing the meaning of propositions and other semantic entities. The second component is an inferential system for arriving at an answer to a question that relates to the information resident in a knowledge base [25]. Fuzzy logic provides Decision Support Systems with powerful reasoning capabilities.

In an ongoing work in the European Union called PATexpert [22] several areas of patent services are targeted. The goal of the project is to bring patent services to a new level by applying several new approaches and methods to various areas in patent services. The search method proposed in this article is different from the approach described in PATexpert. 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 in this article, the fuzzy refers to fuzzy logic for the reasoning and decision making process.

Recent research on representing patent content with the help of ontologies exists, for example [22], but in this proposed research the patent content is matched to the context as well. There are many publications about fuzzy information or document retrieval from the early 1970's till today, see for example [1], but we could not find any work about fuzzy concept search, as described in this article. We believe that the value of this research in comparison to existing research lies in the joined application of ontology matching and fuzzy sets that enables a searcher-friendly service.

III. Patent Service Experiences

The implementation of the model is currently being tested on 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.

The quality of a patent has two different meanings. From an economic perspective, it refers to the patent's technological value or profitability. From a legal perspective, it refers to the soundness of the decision to grant a patent and exclusion of any reasons for invalidation.

Customers have recently shown a preference for high-quality patent examinations over speedy examinations. There is also a new international grouping of major Intellectual Property (IP) offices. The trilateral cooperation among the US, Japan, and Europe has been expanded to include Korea and China. These five major offices, known as IP5, are undertaking ten foundation projects that are designed to improve the quality of examinations and promote the creation of high-quality patents. The IP5 offices handle an aggregate of approximately 1.35

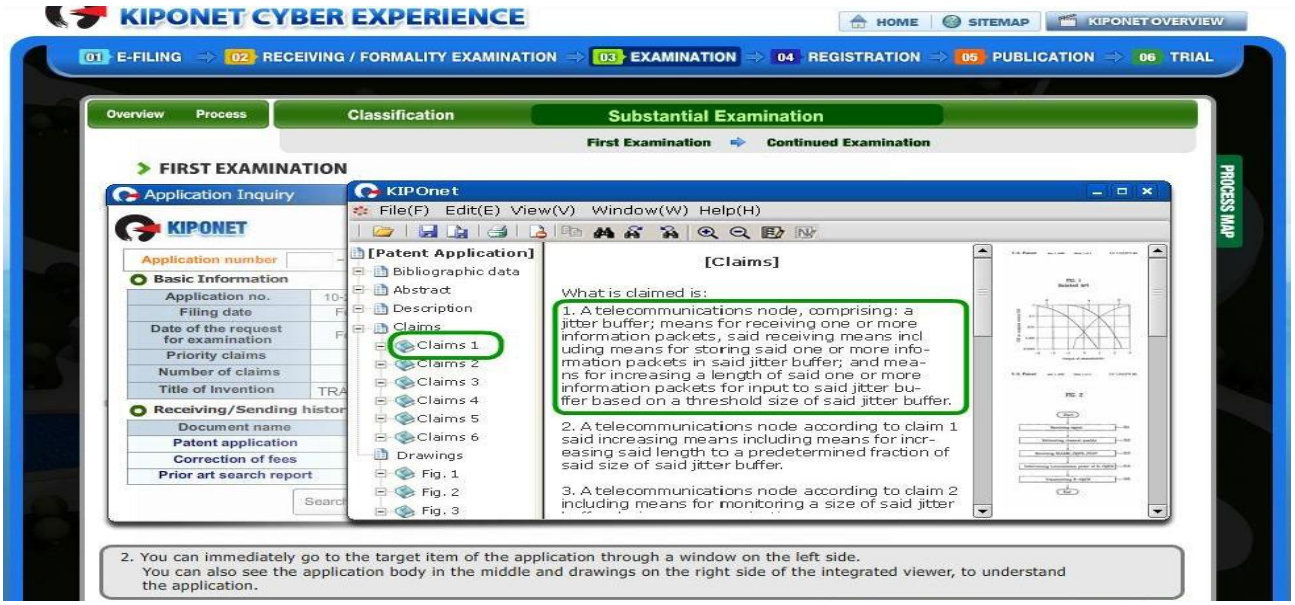

Figure 2 – Patent Service Model Input

million patent applications, which represent 76 percent of all the patent applications filed throughout the world.

KIPO has operated the IP search database since 1999 and, according to the patent technology information sharing policy, has uploaded a total of 85 patent technology databases from 21 countries and five IP offices and has continuously updated them. KIPO has also been offering them online at http://www.kipris.or.kr/ since 2000. There are about 173 million pieces of patent information on the database as of 2008 and the quantity of information is increasing, up by 14 million pieces from 2007 to 2008.

IV. Ontology Matching

The implementation of the model begins when the patent office user initializes the process of evaluating the patent request. Figure 2 displays the input of the application formatted according to the patent claims and accompanying drawings. 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 which are not mentioned in the text. Each claim is analyzed separately through the Domain Representation Process.

The ontology matching process 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.

To process the ontology for optimal information flow, the following method is proposed. Let $O_1, O_2, \ldots, O_n$ be a set of ontologies, representing either a single ontology or each representing different domain knowledge. A simplified representation of an ontology is $O \equiv \langle C, R \rangle$, where $C = \{c_1, c_2, \ldots, c_n\}$ is a set of concepts with their associated relation R.

To analyze the claims, a context extraction algorithm can be used. To handle the different vocabularies used by different information sources, a comparison based on context will be used in addition to simple string matching. The context will be extracted for each document and then compared with the ontology concept.

The 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 - terms which are related to the propositions. The context recognition algorithm was adapted from [19] and consists of the following three steps:

1. Context retrieval: Submitting each token to a Web-based search engine. The contexts are extracted and clustered from the results.
2. Context ranking: Ranking the results according to the number of references to the keyword, the number of Web sites that refer to the keyword, and the ranking of the Web sites.
3. Context selection: Last, the set of contexts for the textual proposition, defined as the outer context, is assembled.

The relevance of the information to each concept is evaluated according to the weight attributed to each concept. The weight is calculated according to the number of references of the concept in both the Web and the document. For example, the document in Figure 2 can be associated with concept *Distance* with weight 0.4 and concept *Wave* with weight 0.3.

To evaluate the matching of the concepts with the information and its context, a simple string-matching function is used, denoted by $\boldsymbol{match_{str}}$, which returns 1 if two strings match and 0 otherwise. I is defined as the information, and D^I is the information descriptor. Also, n is defined as the size of D^I. The match between the concept and the information is defined as the sum of the concept matching values:

$$match(I, c_j) = \sum_{t_i \in D^I} match_{str}(t_i, c_j)$$

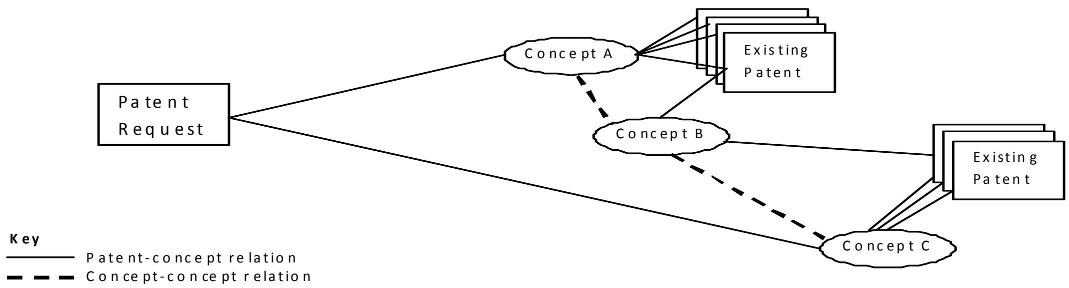

Figure 3- Patent Decision Dilemma

The overall match between the ontology and the information is defined as a normalized sum of the concept matching values:

$$match(I, O_i) = \frac{1}{n} \sum_{c_j \in O_i} \sum_{t_i \in D^I} match_{str}(t_i, c_j)$$

V. Classifying New Patent Requests

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 that are 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 (Figure 3).

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.

In fuzzy information retrieval the relevance of the index terms is expressed by a fuzzy relation: R: X x Y → [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 [1]. The basic scheme of fuzzy information retrieval is shown in Figure 4 where U1 is a fuzzy set representing a particular inquiry. When U1 is composed with T, then U2 becomes an augmented inquiry by associated index terms: U2 = U1 ∘ 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 ∘ R. The inquirer can inspect all the documents that have the support D, or she can filter the inspection to those supported by some α-cuts [1]. In patent search α-cut level 1.0 represents the current precise document search. The search index must have full relevance to the document index. The inquirer can "expand" her patent inquiry by setting α-cut to lower level. For example, α-cut level 0.5 would also bring up those documents that are meaningful to a specific search but not to a full degree. Setting α-cut to a very low level would bring up those documents that are vaguely related to a given inquiry

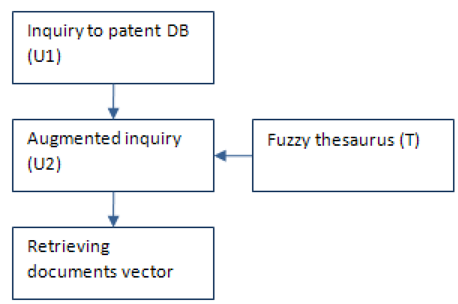

Figure 4 - Fuzzy information retrieval scheme [1]

The role of Fuzzy Thesaurus T can be carried out by a set of ontologies that are further linked to the lexical database Wordnet [7], [c.f. 23]. In the proposed approach the role of the fuzzy thesaurus is carried out by the ontology matching process. The basic scheme of fuzzy information retrieval U2 becomes an augmented inquiry by associated index terms from ontology matching: U2 = U1 ∘ S.

Figure 5 shows an example of the proposed approach. Say the patent officer is examining the claims shown in the Figure 2. She wants to expand the search to other possibly related concepts as well. She can select a mode for her extended search by choosing "Strict" mode or "Vague" mode. She enters a document into the Web based ontology matching process. After a couple of minutes she can see a list of related concepts together with the degrees of relevance. The degree of relevance is calculated based on concept weight in searched documents provided by the ontology matching algorithm and fuzzy membership functions. The membership function is different for the "Strict" and for the "Vague" search modes. For example, in Figure 5 we can see that the ontology matching algorithm would return weights for *Distance* (w=0.4) and *Wave* (w=0.3) concepts.

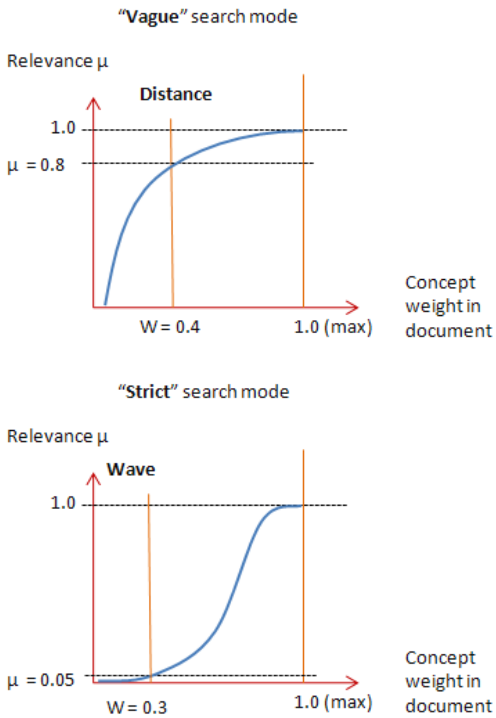

Figure 5 - The relevance of new concepts

According to the membership function in the "Vague" search mode, the *Distance* concept is highly relevant (membership value µ=0.8) as well. In contrast, the *Wave* concept is not really relevant (µ=0.05) at all in the "Strict" search mode. In this case the user will receive not only documents related to *Length* which appears in the text, but also documents related to *Distance*. She will not receive documents related to *Wave*.

Figure 6 illustrates how the α-cuts are used to filter the new expanded set of results. For example, the *Wave* concept is part of the new expanded set if the α-cut is set to a level of 0.4. However, the *Wave* concept is part of result set if there is α-cut level of 0.6.

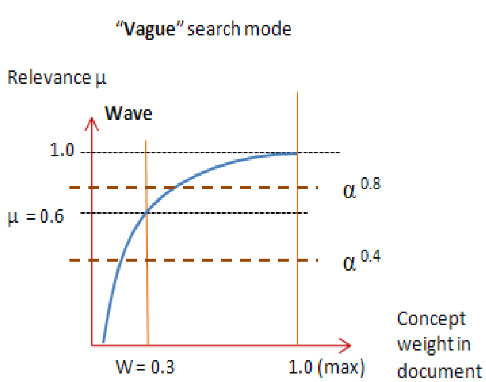

Figure 6 - The relevance of concepts

The patent officer can adjust her expanded search by selecting a "Strict" or "Vague" search mode and also by setting the α-cut level of individual concepts (or for all the concepts) to a Low, Medium or High level. According to this proposed method, the patent officer can carry out expanded searches by using her own language. Therefore, she does not need to convert meanings in her mind to some numerical scales, indexes or variables. Hopefully, this way we can get more meaningful results and at the same time provide a more human-like search approach for the users.

VI. Conclusion and Future Work

The patent search model described in the paper allows queries on the boundaries of existing knowledge to be performed. The model shows promising in extending the field of patent search. This paper describes work-in-process, and we are currently working on validating the proposed joined approach. Future work involves evaluating the patent search model against past patent request decisions. In addition, optional fuzzy set representations of the "relevance of concepts" linguistic variable will be tested. Another direction is to extend the model to multiple languages.

References

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

[2] Borgida, A., Brachman, R.J.: Loading data into description reasoners. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 217–226. ACM Press, New York (1993)

[3] Bunge, M.: Treatise on Basic Philosophy: vol. 4: Ontology II: A World of Systems. D. Reidel Publishing Co., Inc., New York (1979)

[4] Davulcu, H., Vadrevu, S., Nagarajan, S.: Ontominer: Bootstrapping and populating ontologies from domain specific websites. In: Proceedings of the First International Workshop on Semantic Web and Databases (2003)

[5] Doan, A., Madhavan, J., Domingos, P., Halevy, A.: Learning to map between ontologies on the semantic web. In: Proceedings of the Eleventh International Conference on World Wide Web, Honolulu, Hawaii, USA, pp. 662–673. ACM Press, New York (2002)

[6] Donini, F.M., Lenzerini, M., Nardi, D., Schaerf, A.: Reasoning in description logic. In: Brewka, G. (ed.) Principles on Knowledge Representation, Studies in Logic, Languages and Information, pp. 193–238. CSLI Publications (1996)

[7] Fellbaum C. WordNet: an electronic lexical database. Cambridge (MA): MIT Press (1998)

[8] Gal, A., Modica, G., Jamil, H.M., Eyal, A.: Automatic ontology matching using application semantics. AI Magazine, 26(1) (2005)

[9] Kifer, M., Lausen, G., Wu, J.: Logical foundation of object-oriented and frame-based languages. Journal of the ACM 42 (1995)

[10] Klir, J. , G., and Yuan, B. Fuzzy Sets and Fuzzy Logic, Theory and applications, Prentice – Hall Inc. (1995)

[11] Lin C.-T., Lee C. S., Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems, Prentice-Hall, Inc., Upper Saddle River, NJ, (1996)

[12] Madhavan, J., Bernstein, P.A., Domingos, P., Halevy, A.Y.: Representing and reasoning about mappings between domain models. In: Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI), pp. 80–86 (2002)

[13] Madhavan, J., Bernstein, P.A., Rahm, E.: Generic schema matching with Cupid. In: Proceedings of the International conference on very Large Data Bases (VLDB), pp. 49–58, Rome, Italy (2001)

[14] Maedche, A., Staab, S.: Ontology learning for the semantic web. IEEE Intelligent Systems 16 (2001)

[15] McGuinness, D.L., Fikes, R., Rice, J., Wilder, S.: An environment for merging and testing large ontologies. In: Proceedings of the Seventh

International Conference on Principles of Knowledge Representation and Reasoning (KR2000) (2000)
[16] Melnik, S. (ed.): Generic Model Management: Concepts and Algorithms. Springer, Heidelberg (2004)
[17] Noy, F.N., Musen, M.A.: PROMPT: Algorithm and tool for automated ontology merging and alignment. In: Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000), pp. 450–455, Austin, TX (2000)
[18] Segev A., Gal A., Putting things in context: A topological approach to mapping contexts to ontologies. Journal on Data Semantics IX, pp. 113–140, (2007)
[19] Segev A, Leshno M., and Zviran M., Context recognition using Internet as a knowledge base, Journal of Intelligent Information Systems, 29(3) (2007)
[20] Spyns, P., Meersman, R., Jarrar, M.: Data modelling versus ontology engineering. ACM SIGMOD Record 31(4) (2002)
[21] Vickery, B.C.: Faceted classification schemes. Graduate School of Library Service, Rutgers, the State University, New Brunswick, NJ (1966)
[22] 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., Towards content-oriented patent document processing, World Patent Information, 30(1), pp. 21-33 (2008)
[23] Zadeh, L.A., Fuzzy sets, Information and control, pp. 338-353 (1965)
[24] Zadeh, L.A., Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 1, No. 1, pp. 28-44 (1973)
[25] Zadeh, L. A., Commonsense Knowledge Representation Based on Fuzzy Logic, Computer, Vol. 16, October, pp. 61-65 (1983)