KNOWLEDGE SEMANTIC SEARCH IN CLOUD MANUFACTURING

HU Xiaohang[(a)], ZHANG Lin[(b)], HU Anrui[(a)], ZHAO Dongming[(c)], TAO Fei[(b)]

[(a,b)]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191
Engineering Research Center of Advanced Manufacturing System of Complex Product, Ministry of Education, Beihang University, Beijing 100191, China
[(c)]Department of Electrical and Computer Engineering, University of Michigan-Dearborn, Michigan 48128, USA

[(a)]hxh1989626@163.com [(b)] zhanglin@buaa.edu.cn

ABSTRACT
In cloud manufacturing system, the distributed stored knowledge is in multiple forms and structure, and its contents are in multiple fields. In this paper, a semantic search engine method based on shared ontology is presented. The application status of ontology in semantic search is studied. In order to enhance the recall rate and precision rate, this search engine computes the semantic matching degree between user requirements and knowledge by semantic similarity computing and logical reasoning.

Keywords: cloud manufacturing, knowledge, semantic search, shared ontology, matching degree

1. INTRODUCTION

Cloud manufacturing is a kind of networked manufacturing mode (Bo Hu Li and ZHANG 2010, ZHANG and LUO 2011, TAO and ZHANG 2011), one of the key thought of it is the servitization of manufacturing resources in the whole life cycle of manufacturing (Bo Hu Li, ZHANG, and CHAI 2010). Users visit cloud manufacturing center through cloud terminal equipment and invoke appropriate services for specific business needs. When a user invokes services through cloud manufacturing center, he has to pay fees to the owners of the resources. In the cloud manufacturing mode, resources are divided into manufacturing resources and manufacturing abilities (Zhang and Luo 2010). Manufacturing resources are resources physically exist, while the configuration and integration abilities in specific production activities are called manufacturing abilities(LUO 2012), for instance, designing and simulation abilities. Manufacturing resources consist of hard resources that include manufacturing devices, computing devices and material and soft resources including data, information, knowledge, etc. As a kind of resources, knowledge plays more and more important roles in the whole life cycle of manufacturing. Cloud manufacturing is a manufacturing mode with the purpose of building a public manufacturing environment in which users can take part in every process of product designing and manufacturing (Tao, Zhang, and Luo 2011; Tao and Guo 2012; Tao and Cheng 2012). In practical application, knowledge from resources providers and consumers works in the management of enterprise business process as cloud services. In the search, match, combination and configuration of cloud services, knowledge in multiple fields must be taken into account. As a result, cloud manufacturing is a manufacturing mode based on knowledge (Hu and ZHANG 2012).

When performing complex tasks, users need knowledge related to multiple fields from cloud manufacturing. But facing big knowledge saved in cloud manufacturing, traditional search based on key words or themes cannot satisfy the requirement of accuracy. Efficient search model that understands users' meaning correctly and achieves intelligent search has to be established. Existing mature search modes of network information primarily include the following three:

1. Search engine based on key words. Most search engines, for example, Google, using this mode. It adopts distributed search mechanism which runs in multiple storage equipment at the same time through the local agent. Search results are feedback to users after unified sorting. This mode can ensure coverage, but cannot understand requirements of users in semantic level.
2. Search engine based on metadata. This method collects and selects information of a topic or in a subject scope according to certain standards. The selected resources are described and signed for users to search rapidly, for example, subject gateways. However, the requirements of users are related to production activities of complex product in cloud manufacturing, so it's difficult to meet the requirements in specific field.

3. Search engine based on document structure. Literatrue libraries use this mode in usual. Literatures are structured in certain form, the weight of key words can be defined according to the importance of every part. The knowledge in cloud manufacturing has no standard structure, so this mode does not suits.

To the non-applicability of search engine for knowledge in cloud manufacturing in present, this paper proposes semantic search based on shared ontology as the search mode in cloud manufacturing. The semantic relativities between cloud terminal users' requirements and knowledge are got through semantic similarity computing and logical reasoning.

2. KNOWLEDGE SEARCH IN CLOUD MANUFACTURING

2.1. Characteristics of knowledge in Cloud Manufacturing

The knowledge in cloud manufacturing has the following characteristics:

1. Multiple forms and structures. According to the knowledge level principle proposed by Alan Newell in 1982, knowledge modeling is on the level of concept. In other words, one of the basic rules of knowledge modeling is to focus on the concept of knowledge structure and expression ability while the representation of knowledge is ignored. As a result, knowledge is very different in forms, structures, storage modes, data structures and application interface, the interoperability of knowledge is pretty low.
2. Complex content and fields. Except for the real-time information from manufacturing site, most knowledge is edited and maintained by engineers, managers or experts. As a result, the knowledge is complex in content and fields. In cloud manufacturing, reasonable knowledge system is needed to join the complex knowledge up in to achieve semantic search.
3. Distributed storage. The knowledge in cloud manufacturing is stored in distributed devices. Considering the varied forms and complex content of knowledge, it's difficult to find suited search mode to achieve efficient distributed parallel search when the manufacturing resources are unstable in cloud manufacturing.

2.2. Knowledge Search Requirements in Cloud Manufacturing System

The concept of semantic search is proposed on the basis of semantic Web, which is aimed to find the implied relation of information through semantic computing. Semantic computing includes the following three aspects.

The first is computing the semantic distance (ZHONG, ZHU, and LI 1995) between concepts or terms using concept system to realize human-machine interaction in semantic. Second, logical relation can be applied to the reasoning process of search engine (LIU and LI 2005) in order to improve search quality. The last is to learn the user's habits and search intentions by mining historical data. In this way the personalization of search results is formed.

In view of the above characteristics of knowledge in cloud manufacturing, in order to understand users' intentions and find out knowledge highly related to requirement in semantic from big data, the semantic search engine must have the following features:

1. Reasonable resource description mechanism. A reasonable description mechanism could be set up to achieve the universality of semantic search in cloud manufacturing. The mechanism formalizes description of knowledge in different forms and structures. The search for knowledge is turned into that for knowledge description information.
2. Complete knowledge system. Because of the complexity of the users' requirements and knowledge in different fields, the semantic search engine must set up complete knowledge system including concept, instance, relation and definition, etc. The system should be expressed in language and grammar that can be identified by computer. Knowledge system is the foundation for computer to understand users' intentions.
3. High recall rate. The recall rate of search engine is the percentage of searched related information in all related information. The goal of semantic search engine is to find knowledge that related to users' requirements impliedly by semantic relativities computing. This process enhances call rate to a great extent.
4. High precision rate. The precision rate of search engine is the percentage of related information in all results. Precision rate is the fundamental measure of search performance.

3. KNOWLEDGE SEARCH MECHANISM

The operating principle of knowledge semantic search engine is shown in figure 1.

The semantic search engine in cloud manufacturing includes two main function modules, *information extraction* and *semantic computing based on ontology*. This chapter will elaborate the technical realization of these two function modules.

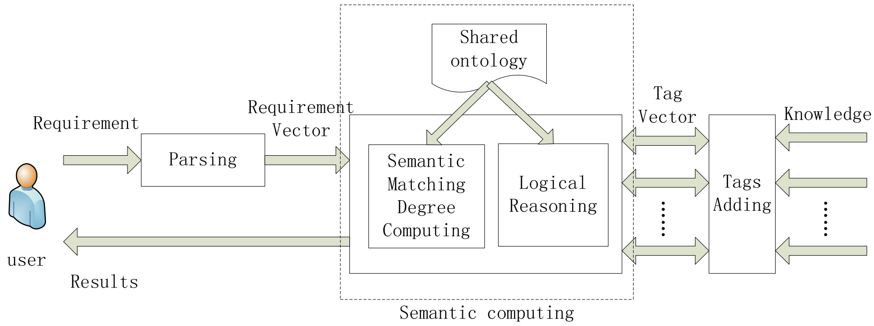

Figure 1: Search Operating Principle

3.1. Information Extraction

3.1.1. Vector Space Model (VSM)

VSM is a model that universally used in natural language processing. This model regards every document as a vector composed of characteristics. And the characteristic are endowed with a certain weight by algorithm computing. The mapping relations between documents and vectors in VSM are as table 1.

Table 1: mapping relations in VSM

documents	vectors
word	characteristic
weight	coordinate value
document	vector
document collection	vector set

Documents are represented as vectors of characteristics and the weight of them in VSM. Thus natural language processing becomes vector mathematic operation. The semantic computing between requirement and knowledge in cloud manufacturing is realized in vector space model. Thus requirement and knowledge should be in the form of vectors.

3.1.2. Parsing

As requirements are sentences or phrases generally. They should be transformed into vectors by stem segmented and stop removal in VSM before semantic computing.

One word can show a variety of forms because of inflection or derivation. Stem segmentation (WU and QIAN 2012) is the process of simplifying the forms into a common stem. Stem segmentation can be divided into two classes, algorithm-based and dictionary-based. In this paper, porter stemmer is used, which is a kind of mature algorithm-based stem segmentation. Stop words are functional words in documents, the engine will skip and process the next one when meets these words. The stop removal is often done by querying the stop words dictionary. Results by parsing are elements that compose the requirement vectors.

3.1.3. Tags Adding

As knowledge is in multiple forms in cloud manufacturing, it's difficult to find a mechanism suits all knowledge. So it's necessary for the engine to add formalized tags to knowledge. The relativities between requirement vectors and tag vectors are achieved by semantic computing.

Tags can be added manually or automatically.

1. The tags of knowledge in forms of image, audio or formula can be added manually by engineers in the process of servitization. The tags are published by maintainers of cloud manufacturing after being audited.
2. For the knowledge in forms of document, tags can be added automatically through TF*IDF framework which is known as feature extraction. The computing process is as formula 1.

$$w_k = \frac{W_{Tf}\left(k,\vec{d}\right)\times IDF_k}{\sqrt{\sum_{k\in\vec{d}}\left[W_{Tf}\left(k,\vec{d}\right)\times IDF\right]^2}} \quad (1)$$

TF is the term frequency, which means the times of the key word appearing in files. This paper takes the logarithm to it to smooth the influence of word frequency jumping, as formula 2.

$$W_{Tf}\left(k,\vec{d}\right) = \log\left[1 + Tf\left(k,\vec{d}\right)\right] \quad (2)$$

IDF is inverse document frequency, which reflects the distribution of characteristics in the full set of documents. The value of IDF means the quantity of information a characteristic brings, as formula 3.

$$IDF_k = \log\left(\frac{N}{n_k} + 0.01\right) \quad (3)$$

The product of IF and IDF can be used to get the weight matrix of document collection. Take the characteristics with maximum weight as the tags of each document.

3.2. Semantic Computing based on Ontology

3.2.1. Shared Ontology Construction

Ontology is selected to be the form of knowledge system in semantic search as it's the core of semantic Web. Ontology is a kind of shared conceptual model can be recognized by computer, which abstracts objective reality in certain fields. It is the basis of semantic interoperation.

Ontology is different in modeling depth. Guarino (Guarino and Nicola 1997) regards the detail level of description as the basis of classification of ontology. The ontology with high detail level of description is named reference ontology. It takes the comprehensive concepts of object on the World Wide Web as metadata to establish fine-grained ontology model. Shared ontology is the ontology with low detail level of description, which mainly consists of concepts, relations and the hierarchical structure in certain fields. Shared ontology is independent of the specific application environment and task instances. It describes the constraints of relations instead of expressing the detail content of knowledge. The ontology example edited by protégé 4.1 is as the following figure 2.

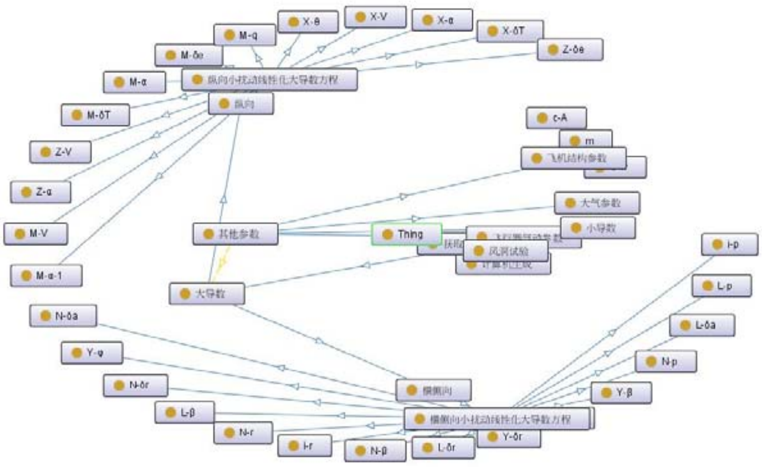

Figure 2: Ontology Example

Shared ontology plays the role of field concepts dictionary to some degree. A shared ontology has realized decoupling with manufacturing resources as soon as it is constructed. Thus it can be applied to different configuration environment and application requirements. Shared ontology makes it possible for computer to understand resources information and users' requirements when it provides the standard terminology to computer. In search engine, the shared ontology with reasonable structure and accurate content is the premise to realize efficient semantic search.

3.2.2. Semantic Matching Degree Computing

Ontology can be considered as a tree structure for semantic distance computing as its layer of structure is stable. The calculating flow is as figure 3.

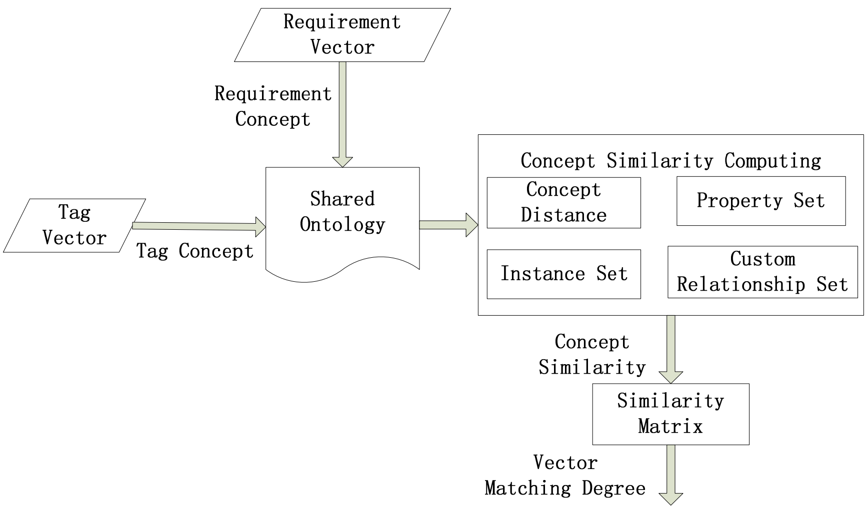

Figure 3: Semantic Relativity Calculating Flow

This paper takes shared ontology as the basis of semantic matching degree computing between requirement and knowledge tag. While concept similarity computing is the fundamental computing (WANG and MA 2007), the semantic similarity between two concepts can be got by concept similarity computing.

Concept similarity computing consists of four parts in this paper. This is a synthesis algorithm taking the distance, property and content of ontology into consideration. Concept distance can be got by analyzing the shared ontology structure. The property, instance, custom relationship set similarity can be got by computing Cartesian product of two concepts in ontology. Then we obtained the concept similarity by weighted summing the above four aspects. Every concept is one dimension of requirement vector or tag vector. The similarity matrix is generated from vectors by semantic similarity computing. Finally, semantic matching degree, which is the matching degree of two vectors, is worked out based on the matrix.

3.2.3. Logical Reasoning

With the emergence of semantic Web, a series of ontology representation language arises. In the modeling of complex knowledge, OWL has been recommended as a standard of ontology modeling by W3C since 2004. OWL has logical reasoning ability and extensibility to a certain extent. OWL is used as the description language to construct shared ontology in this paper.

The basic building blocks of OWL include class (concept), role (properties) and individual (concept or instance). The role is relation between class and class, class and individual, or individual and individual.

For instance, in figure 4, the similarity between B and C can be more than 0 if computing it by semantic distance. But A and B have the relation of *disjointWith*, according to the definition of *disjointWith*, the intersection of A and B is empty. So the similarity between C and B is 0 as C is the subclass of A.

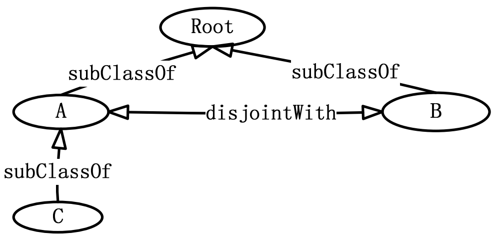

Figure 4: Ontology Logical Reasoning

This shows that logical reasoning can be used to filter the results of semantic similarity computing in ontology-based semantic search. The number of search results is decreased through logical reasoning and the results left behind suit requirements better. The precision rate of search engine is improved.

Semantic search engine sorts the search results based on the matching degree between requirement vectors and tag vectors in cloud manufacturing. Sorting results are feedback to users through cloud manufacturing center. What users need is knowledge but not its description information, so the results should be application interfaces or links

4. CONCLUSION

This paper studies the knowledge semantic search engine in cloud manufacturing. In view of the diversity, complexity and dispersibility characteristics of knowledge in cloud manufacturing, it establishes semantic search mode on the basis of knowledge system. This paper computes the semantic relativities between users' requirements and knowledge by concept semantic similarities computing based on shared ontology. In the process of computing, the engine takes use of the logic relations of OWL to make logical reasoning for requirement in order to improve recall rate and precision rate.

The next work consists of the algorithms improvement of tags adding and concept similarity computing. And prove systemically the correctness of them.

ACKNOWLEDGMENTS

This work is supported in part by NSFC project (No.61074144 and No.51005012) and the Fundamental Research Funds for the Central Universities in China.

REFERENCES

Bo Hu Li, ZHANG Lin, et al. Cloud manufacturing: a new service-oriented networked manufacturing model. *Computer Integrated Manufacturing Systems*. 2010,16(1):2-4.

L Zhang, YL Luo, WH Fan, F Tao, L Ren. Analyses of cloud manufacturing and related advanced manufacturing models. *Computer Integrated Manufacturing Systems*, 2011.11(3):458-468. (in Chinese)

F Tao, L Zhang, VC Venkatesh, YL Luo, Y Cheng. Cloud manufacturing: a computing and service-oriented manufacturing model. *Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture* (Proc IMechE Part B: J Eng Manufact), 2011,225(10):1969-1976. (8 Pages) Oct. 2011

Bo Hu Li, ZHANG Lin, CHAI Xudong. Introduction to Cloud Manufacturing. *ZTE COMMUNICATIONS*. 2010,16(4)

L Zhang, YL Luo, F Tao, L Ren, H Guo. Study on the key technologies for the construction of manufacturing cloud. *Computer Integrated Manufacturing Systems*, 2010, 16(11)：2510-2520

LUO Yongliang, ZHANG Lin, TAO Fei, etc. Key technology of manufacturing capability modeling in cloud manufacturing mode. *Computer*

Integrated Manufacturing Systems. 2012,18(7):1357-1361.

F Tao, L Zhang, YL Luo, L Ren. Typical characteristics of cloud manufacturing and several key issues of cloud service composition. *Computer Integrated Manufacturing Systems*, 2011.17(3):477-486. (in Chinese)

F Tao, H Guo, L Zhang, Y Cheng. Modelling of combinable relationship-based composition service network and theoretical proof of its scale-free characteristics. *Enterprise Information Systems*, 2012, 6(4) : 373-404 (32 Pages), Oct. 2012

F Tao, Y Cheng, L Zhang, D Zhao. Utility modeling, equilibrium and collaboration of resource service transaction in service-oriented manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture* (Proc. IMechE Part B: J Engineering Manufacturing) 2012, 226(6):1099-1117 (19 Pages) Jun. 2012

A Hu, L Zhang, F Tao, Y Luo. Resource Service Management of Cloud Manufacturing Based on Knowledge. *Journal of Tongji University*, 2012, 40(7):1092-1101

Klinger T. Image Processing with LabVIEW and IMAQ Vision. *Upper Saddle River, New Jersey: Prentice Hall PTR*, 2003.

ZHONG J, ZHU H, LI Y, et al. Using information content to evaluate semantic similarity in a taxonomy. *processing of the 14th International Joint Conference on Artificial Intelligence* (IJCAI-95). Washington: ACM Press, 1995.

LIU Jin, LI Bing. Research on Logical Analysis of Web Ontology Language. *Computer Engineering*. 2005,4.

WU Sizhu, QIAN Qing, et al. Comparative Analysis of Methods and Tools for Word Stemming. *Library and Information Service*. 2012,52(15).

Guarino, Nicola. Semantic Matching: Formal Ontological Distinction for information Organization, Extraction, and Integration. *Pazienza M T, eds. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Springer Verlag. 1997. 139-170

WANG Liancheng, MA Qiang.The Computation of Ontology Similarity Based on Concept Value. *Computer Technology and The Application Progress*. 2007:692-694.