Ontology-based Knowledge Management of Chinese Historical Official Title: an Overview of the HOTKB[1] Project

Zhu Junwu[1,2], Jiang Yi[2]

[1] Joint LAB of Intelligent Computing, Jiangsu University of Science and Technology, 212003 Jiangsu Zhenjiang, China

[2] Department of computer science and technology, Yangzhou University, 225009, Jiangsu Yangzhou, China
zjwjy@hotmail.com

Abstract. This paper describes the results of the HOTKB (Knowledge Base of Historical Official Title) project, in particular the method of concepts analysis in HOT (Historical Official Title) domain. The goal of the project was to build a knowledge base which can provide a sharable and reusable knowledge resource about Chinese historical official title. Consider the complexity of historical domain and the ill-structured representation of historical knowledge, this paper introduces an ontology-based method of building knowledge base of HOT firstly, and then reports the process of constructing the HOT ontology which is a fundament of HOTKB.

Keywords: Historical Official Title; Analysis of Concepts; Knowledge Base; Ontology; Conceptualization.

1 Introduction

The history of China, as documented in ancient writings, dates back about 3,300 years. Centuries of migration, amalgamation, and development brought about a distinctive system of writing, philosophy, art, and political organization that came to be recognizable as Chinese civilization. For us, the past is a valuable legacy, but how to conserve and promulgate it is inevitable responsibility.

Unfortunately, most of the historical knowledge, including knowledge of Chinese historical official titles, is stored in paper or in electronic image, so knowledge retrieval and dissemination is difficult. Meanwhile, the representations are based on nature language, which is more ambiguous than formal one. That is to say, owing to the lack of precise and formal definitions of terms and concepts, communication gap between two agents is an undoubted fact. In order to construct a shareable and interoperable knowledge base of historical official titles, more attention should be paid to

[1] The HOTKB project was supported by a grant from the Ministry of Science and Technology (No. #2001CCA03000).

specification ontology and domain ontology which can provide an explicit specification of a conceptualization.

This paper is organized as following. Section 1 is a brief introduction. Section 2 analyses the aims and the system structure of HOTKB. Section 3 gives the method and the process of construction of HOT ontology, which is a crucial step of HOTKB project. Section 4 concludes the paper.

2 The System Structure of HOTKB

The HOTKB project aims to construct a knowledge base which can provide sharable and interoperable domain knowledge. As shown in figure 1, the system structure of HOTKB consists of three layers, and they are ontologies, search engine, and user interface. As a representation, ontologies play an important role in knowledge management, while search engine depends on users' models and the results of searching are raw materials of knowledge design.

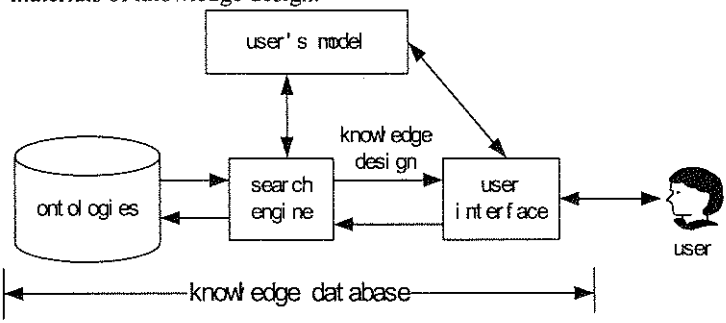

Figure 1: The system structure of HOTKB

2.1 Ontologies

In recent years, research on ontology has become a hotspot in the computer science community. Ontology studies the existence of all kinds of entities (abstract or concrete), and the specification of conceptualization. Ontology provides a set of vocabulary through which existing things can be described. Although there has been no unequivocal definition of ontology so far, Gruber has provided a widely accepted definition: "An ontology is a specification of a conceptualization". That is to say, ontology is related to language and domain. The language is the basis of specification, while the conceptualization is abstraction of reality. We advocate that the abstraction of reality and concept analysis in certain domain are jumping-off points of the HOTKB project.

In knowledge representation community, ontology has something to do with sorting of the concepts in the world, but ontology is not equal to catalogue. In the past decades, cognitive scientists have put forward synthetic accounts of relations between concepts and categories based on the premise that, although representations are di-

verse, their contents are all properties of certain object cognized by agents. Two viewpoints should be clarified, before we perform the description of concepts of certain domain and the construction of HOT ontology. On one hand, no properties, no concepts. When we talk about concepts, properties (including relations) are involved. Concept is a set of properties, and the number of properties is enough to distinguish one concept from others. On the other hand, sorting depends on properties. According to certain property, different concepts perhaps have different value of certain property, and the instances that have same value of this property can be put in one category. Then we can get many categories, and instances in different categories have different value of certain property.

In our opinion, the analysis of concepts is more important than the specification of conceptualizing. We should paid more attention to properties, analysis of concepts and conceptualizing of certain domain, because they are intension of cognitive objects and results of cognitive actions. In ontology community, without efficient cognitions, the cognitive agents can not discovery the properties of concepts. As a consequence, the representation will lose its object. So we argue that the more important steps of knowledge engineering and ontology engineering are cognition, analysis of concepts and conceptualizing, but not the representation of conceptualization.

2.2 Search Engine

Search engine is also an important component of HOTKB, and it is mechanism of knowledge selection. In fact, ontology of HOTKB is a system of knowledge management, and affluent semantic meanings are depicted. Contrary to keyword matching, search engine of HOTKB is semantic-based.

2.3 User Interface and Knowledge Design

Knowledge design aims to offer the right knowledge to the right person in the right manner at the right point of time. It is a science of selecting, organizing and communicating knowledge in a huge knowledge space in a proper way so that it can be sensed, digested and utilized by human beings efficiently and effectively.

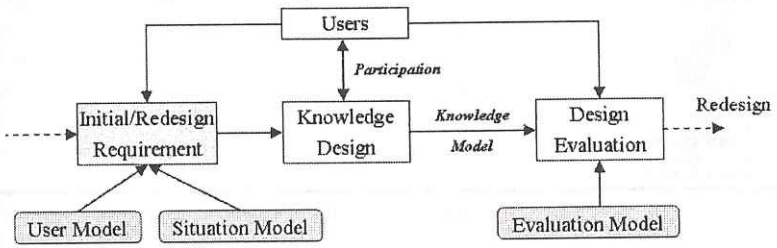

Figure 2: the process of knowledge design

As shown in figure 2, knowledge design is an iterative process, and bases on users' models Knowledge design is actually a useful tool for many research areas, such as computer-aided design, knowledge engineering, information design, and educational technology. With well-designed knowledge, people can learn, master and apply knowledge more efficiently. In HOTKB project, knowledge design, which bases on users' models, can accelerate search engine and enhance the interaction of user interface.

3 Construction of HOT Ontology

Conceptualization of HOT domain and construction of HOT ontology are two main task of HOTKB project. The detail process and methodology are following.

3.1 Analyses of Characteristics of Concepts in HOT Domain

The concepts of HOT are complex, because it is difficult to confirm the values of properties and the relations among different concepts. In order to get an explicit exhibition of domain knowledge, a tool of knowledge management shown in figure 3 had been developed in Chinese. This tool is still a result of pre-treatment, representation of knowledge is also textual and informal.

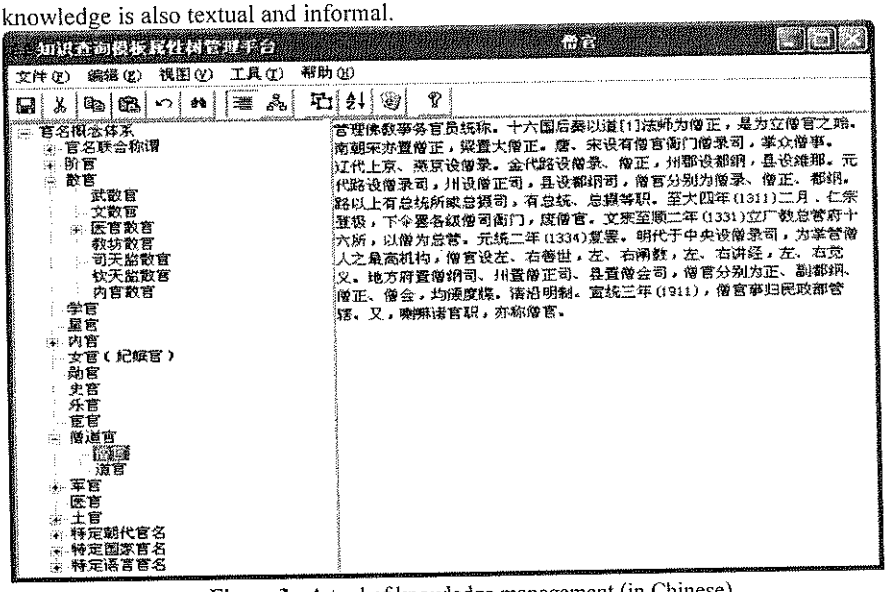

Figure 3: A tool of knowledge management (in Chinese)

The complexity of HOT knowledge does not come from the scarcity of cognitive tools, cognitive methods or cognitive symbols, but from cognitive objects (HOT). Characteristics of knowledge in HOT domain are listed as following:

- **Time-related** Historical knowledge is closely related to time and constrained by it. Without constraint of time, historical knowledge is inaccurate. Most knowledge about HOT hold only in a special period, and time is a common facet of the properties of concepts in HOT domain.
- **Time-order** Time creates history. Similarly, historical knowledge is accumulated by time. There are many special relations about time-order among the knowledge of HOT.
- **Inaccuracy** Due to the influence of politics and the limitation of science and technology of those days, the completeness and accuracy of historical documents can hardly be guaranteed. Accordingly, some historical knowledge is fuzzy and inaccurate.
- **Instability** Owing to the influence of political reforms and the development of the society, some fundamental properties of HOT are different in different period. As a result, the intension and extension of HOT concept have changed.
- **Complexity** In HOT domain, the relations among concepts are very complicated. These complicated relations not only include relations among concepts of HOT, but also involve relations between concepts of HOT and other special concepts in historical domain, e.g. historical period, historical organ, historical characters, historical events etc. Richly semantic meaning in historical title is often loaded by these complicated relations.

According to the characteristics of knowledge in HOT domain, proper method and strategy should be adopted to depict the semantic of concepts.

3.2 Analysis and Taxonomy of Concepts in the HOT Domain

Now, a common problem of ontology is that its taxonomy is often poor and confusing. Perhaps, the abuse of subsumption is one of the main reasons. Nicola Guarino and Christtopher Welty put forward a formal ontology that is based on some meta-properties, which contributes greatly to the science of taxonomy. However, our classification is mainly based on contrast and differentiation. The description and taxonomy of entities are based on the contrast of properties (including relations between entities). On one hand, if we want to distinguish one entity from others, enough properties must be described and at least one different property value can be found in the description (formal or informal) of the entity. So we can say that the contrasts, which are related to the taxonomy and determine certain entity belongs to one category or another, are more important than category itself. On the other hand, the distinctions of the property value are combined to generate a category, and the category is an interpreting template of concepts. The contrasts and the distinctions are two fundamental principles for our HOT classification.

Furthermore, based on the analysis of HOT concepts, we should also clarify some relations which appear frequently in HOT concepts. They are part-of, member-of, instance-of and is-a (Figure 4 illustrate these relations). Part-of is the basic mereological relation, and it holds only between two entities. If entity A consists of some specific entities (for example A1, A2,...Ai...An), we can say Ai is part of A. Meanwhile,

the part-of relation is reflexive, asymmetric and transitive. The Member-of relation holds only between the element and the set. In HOT, many concepts have the properties of set, such as taizizanshi, wuwu-general and mizhu etc, so these relations are all represented by member-of. Instance-of is a more complex relation, because weather a concept is an instance of certain category or not is related to granularity of knowledge base. We can't say an instance is equal to an individual. Instance and individual have different meanings in knowledge engineering community. But it is an undoubted fact that instance is the leaf node on a conceptual tree and that instance has not children nodes. Is-a is a common relation between categories of HOT and only holds between two categories. In other words, is-a holds between certain category and its super-category. Generally speaking, a category may has more than one sub-category, which has a common property name and different property values, while one category only has one super-category, except for root category (thing).

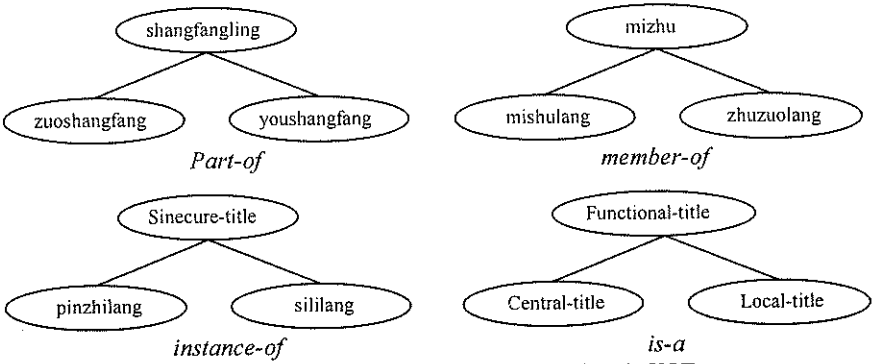

Figure 4: illustration of four common relations in HOT

In the process of constructing HOT domain Ontology, these four fundamental relations must be clearly distinguished.

3.3 The Structure of HOT Ontology

In the system of HOT, concepts of HOT that appear on all dynasties are our research objects. We have acquired 6,250 concepts of HOT from Chinese History Dictionary. If we adopt an isolated method to describe concepts of HOT, we will not be able to clearly explain the semantics contained in these concepts, and the significances of constructing HOT domain ontology will be lost. So, during the course of HOT-concept description, not only inherent attributes such as HOT-setting-periods, HOT-

names in different periods, and various relations between HOT concepts, but also relations between HOT concepts and other concepts in other domains of history are described. Just like the features of HOT knowledge, these relations are also complicated. But there are universal relations, e.g. relations between HOT and historical period, relations between HOT and historical organs. Only through complete description of related relations, richly semantic meaning contained in knowledge can be clearly expressed in formal structure. Meanwhile, classifying HOT concepts also depend on those relations.

Firstly, we should acquire an acknowledged HOT system in historical domain. The system is the precondition to describe HOT concepts and to organize these concepts. When we cognise and describe HOT concepts, we must depend on attributes and relations about these concepts. According to cognitive science, only through adequate description of those attributes and relations, which belong to this kind of concepts, we can describe these concepts accurately.

To cognise HOT concepts, we consider the problem by two steps, namely, concept capture and concept description. At the initial stage, we don't attempt to obtain an entire HOT system. Our main object is to acquire initial cognition about HOT. Figure 5 shows the concept HOT system.

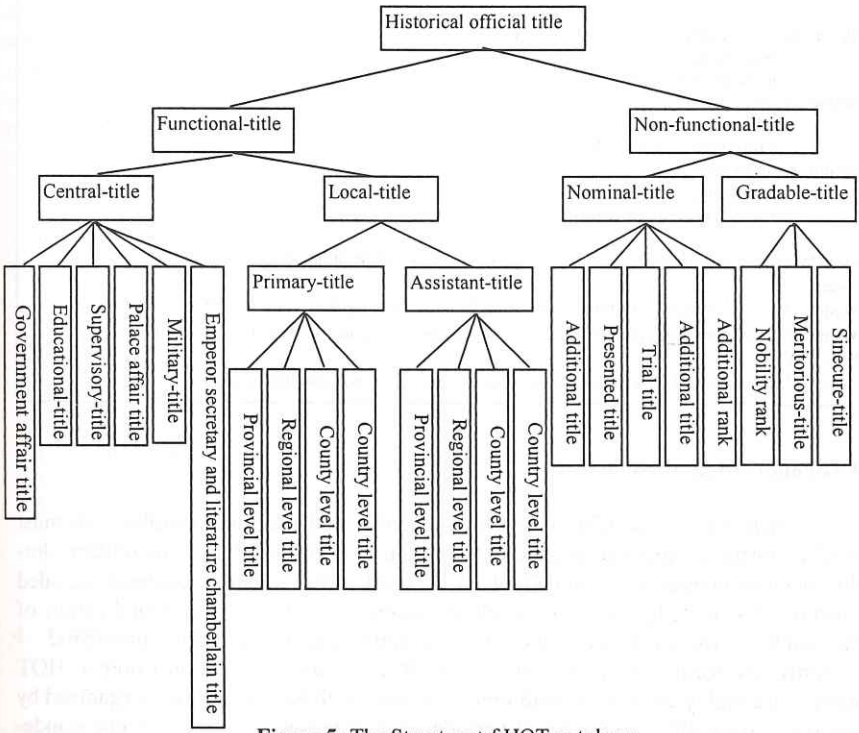

Figure 5: The Structure of HOT ontology

According to the structure of HOT ontology shown in figure 5, HOT concept can be divided into related concept categories. Through describing the relative special properties and relations, the concepts can be classified. To describe concepts, class and instance must be involved. Category is the abstraction of instance, which has common value of certain property. Only if an instance belongs to a special category, the instance can own the properties and the relations what have been defined in the category.

To describe the HOT concepts, we have defined 482 properties (including relations). Properties reflect the inherent features in concepts and relations reflect the dependence between concepts. Based on some special concepts related to the HOT relations, we divide these relations into time-relation, organ-relation, document-relation, function-relation, status-relation, lead-relation, etc. Concept properties and concept relations are generally named properties.

Status-relation is one of the important properties categories. There are some axioms about status-relation. In this paper, we show part of these axioms.

```
def-category status-relation-category
(
relation: position-above
        :type string
        :value-type multiple-valued
relation: position-below
        :type string
        :value-type multiple-valued
relation: position-near
        :type string
        :value-type multiple-valued
    ... ...
Axiom1: (∀x,y,z: historical official title), (position-above(x,y)∧position-above(y,z)--> position-
above(x,z))
Axiom2: (∀x,y,: historical official title), (position-below(x,y)-->! position-above(y,x))
Axiom3: (∀x,y,z: historical official title), (position-below(x,y)∧position-below(y,z)--> position-
below(x,z))
Axiom4: (∀x,y: historical official title), (position-above(x,y)<--> position-below(y,x))
```

3.4 Formalization of HOT text

To achieve the abilities of machine-readable and machine-operable, we must formalize historical texts during the construction of HOT ontology. For unitary slot-value, such as integer, string and Boolean are used, richly semantic contexts included in historical knowledge can not be clearly expressed. During the formalization of historical texts, we mainly adopt the following approaches to solve these problems.

Firstly, the formalization of concepts in HOT domain is based on Context. HOT concepts are highly correlative with time. Contexts of these concepts are organized by time too. During different historical periods, knowledge of the same concept is independent respectively. So, we regard a HOT concept as components of a series of sub-

concepts: A1, A2, ..., Ai, where A1, A2, ..., Ai contain the knowledge of concept A during a special period. For every sub-concept of a special concept, there are no other semantic relations except for the same time or the same historical organ during the same period. To describe the context relations of every sub-concept, a series of properties constrained by time are defined, such as time-function, time-title-rank etc. By this kind of properties, knowledge of the same sub-concept is organized through the same time. But, these sub-concepts are not absolutely isolated. The context relations among sub-concepts must be described. Some relations about contexts, mainly including inheritance of slot-value and evolvement of slot-value, are summarized on the basis of concept analysis. In HOT domain ontology, we have defined properties to describe these relations, e.g. Time-inheritance (time1; time2). It means that the sub-concept of time2 inherits the slot-value of every properties described in the sub-concept of time1.

Secondly, the description of concepts involves affluent semantics. For every sub-concept, the main problem during the text formalization is how to clearly express rich semantic meaning contained in the knowledge. One HOT concept often involves other concepts, including other HOT concepts and other historical concepts. Namely, existence of HOT knowledge is restricted by some relative concepts. To clearly express the knowledge, relations contained in the concept must be described. So, we defined a series of multi-properties to illustrate the relations between one concept and other concepts, e.g. Time-organ-position (time; organ; string). It means, in this time, the position in the organ, described as the string, is occupied by the title.

The following textbox is an example of knowledge formalization of a HOT concept, named Siwei.

```
def-frame siwei
{
 is-instance: HOT
 firstly-erecting-time: Suiyangdi period
 time-is: (Suiyangdi period; female HOT of royal court
 time-is-member-of: (Suiyangdi period; twenty-four bureau
 time-rated-number-of-official: (Suiyangdi period; two)
 time-title-rank: (Suiyangdi period; hypo-sixth grade)
 time-belong-to-organ: (Suiyangdi period; Shanggong bureau)
 time-function: (Suiyangdi period; control key and manage eunuch)
 successive-time: Tang Dynasty
 time-changed-title-rank: (Tang Dynasty; sixth grade)
 time-with-same-system: (Song Dynasty; Tang Dynasty,
                         and (Jin Dynasty; Tang Dynasty),
                         and (Ming Dynasty; Tangdynasty)
 time-function-is-transfered-to: (Ming Dynasty; eunuch)
}
```

4 Conclusions

The requirements of knowledge base of Chinese historical official titles, which can be sharable and interoperable, are increasingly apparent and urgent. We have con-

structed the HOT ontology, and this ontology provided an efficient approach to manage knowledge in HOT domain. The method we shown can be generalized and applied to other areas, because the process of concepts analysis is common.

The benefits of the proposed approach are manifold: (□)the construction of certain ontology can clarify the properties and relations.(□)the method of HOTKB project provided a methodology to build historical knowledge base, such as historical events, historical characters, and historical organ.(□)the HOT ontology and vocabularies defined in this project is a fundament of semantic WEB.

5 Acknowledgements

During studying and working in intelligent processing laboratory institute of Computing Technology Chinese Academy of Sciences, Professor Cao Cungen and Sui Yuefei had provided great help. We are indebted to them. We would also appreciate our colleagues and other graduate students for their contributions to this project.

References

1. Feng Donghui, Cao Cungen. Building Communication Between User and Knowledge. In ICYCS'2001 (Proc 6th Int. Conf), Vol.(3), pp.1204-1208, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001.
2. T.R.Gruber, A Translation Approach to Portable Ontology Specification. Knowledge Acquisition, vol.5, no.2, pp.199-220.
3. P.Clark, J.Thompson, B.Porter. Knowledge Patterns. In KR'2000 (Proc 7th Int. Conf), pp. 591-600, Eds: A. Cohn, F. Giunchiglia, B. Selman, Canada: Kaufmann, 2000.
4. M.Uschold, M.Healy, K.Willianson, P.Clark, S.woods. Ontology Reuse and Application. In Proceedings of the First International Conference, pp.179-192, Ed: N.Guarino, Formal Ontology in Information System, Italy: Padova, 1998.
5. N.Guarino. Formal Ontology and Information System. In Proceedings of the First International Conference, pp.3-15, Ed: N.Guarino, Formal Ontology in Information System, Italy: Padova, 1998.
6. Cao Cungen, Shi Qiuyan. Acquiring Chinese Historical Knowledge From Encyclopedic Text. In ICYCS'2001 (Proc 6th Int. Conf), Vol(3), pp.1194-1198, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001.
7. Cao Cungen. Medical Knowledge Acquisition From the Electronic Encyclopedia of China. Lecture Notes in Artificial Intelligence. China: Beijing, 2001.
8. Gu Fang, Cao Cungen. Biological Knowledge Acquisition From the Electronic Encyclopedia of China. In ICYCS'2001 (Proc 6th Int. Conf), Vol(3), pp.1199-1203, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001.
9. Zhang Dehai, Cao Cungen. Acquiring Knowledge of Chinese Cities From the Encyclopedia of China. In ICYCS'2001 (Proc 6th Int. Conf), Vol(3), pp.1199-1203, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001.
10. Tian Wen, Cao Cungen. A framework for Extracting Knowledge of Human Blood System From Medical Text. In ICYCS'2001 (Proc 6th Int. Conf), Vol(1), pp.501-505, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001

11. Lei Yuxia, Cao Cungen, Sui Yuefei. Acquiring Military Knowledge From Texts in the Electronic Encyclopedia of China. In ICYCS'2001 (Proc 6th Int. Conf), Vol(1), pp.366-370, Eds: Wu Zhaohui, Zhang Yueting, etc. China: Beijing, 2001
12. D.B.Lenat. "Cyc: A Large-Scale Investment in Knowledge Infrastructure". Communications of the ACM 38(11), pp.33-38, 1995.
13. V.Chaudhri, A.Farquhar, R.Fikes, et.al: Open Knowledge Base Connectivity (OKBC) 2.0.3. Stanford University Tech Report, 1998.
14. P.Vossen, 1996: Right or Wrong. Combining lexical resources in the EuroWordNet project. In: M. Gellerstam, J. Jarborg, S. Malmgren, K. Noren, L. Rogstrom, C.R. Papmehl, Proceedings of Euralex-96, Goetheborg, pp.715-728, 1996.
15. R.Fikes, A.Farquhar, J.Rice. The Ontolingua Server: A Tool for Collaborative Ontology Construction. Knowledge Systems Laboratory, September, 1996.
16. S.D.Richardson, W.B.Dolan, L.Vanderwende. MindNet: Acquiring and Structuring Semantic Information from Text. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, ACL, 1998, CONF 17, Vol.(2), pp1098-1102.
17. L.Vanderwende. The Analysis of Noun Sequences Using Semantic Information Extracted from On-line Dictionaries. Ph.D. Dissertation, Georgetown University, Washington, DC, 1996.
18. P.Vossen. Grammatical and Conceptual Individuation in the Lexicon. Ph.D. Dissertation, University of Amsterdam, 1995.
19. Ibnu Samsi, "the difference between information and knowledge", October 09, 2000,http://www2.darwinmag.com/
20. Feigenbaum, E., Knowledge Engineering: the Applied Side of Artificial Intelligence. STAN-CS-80-812 Department of Computer Science, Stanford University, 1980.
21. Mark Wieman, "Technology, Entertainment,Design Richard Saul Wurman", http://frontwheeldrive.com/richard_saul_wurman.html
22. Drue Miller, "Interview with Nathan Shedroff", http://www.nathan.com/thoughts/interview.html
23. RAJ REDDY, "Three Open Problems in AI" Journal of the ACM, Vol. 50, No. 1, January 2003, pp. 83-86.
24. Bergamaschi S, Lodi S., Sartori C., Description Logics as core of a Tutoring System, P.F. Patel-Schneider ed., Description Logic workshop, Bonn, 1994.
25. Guarino, N. 1992. Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases. Data & Knowledge Engineering, 8: 249-261.
26. Guarino, N. 1994. The Ontological Level. In R. Casati, B. Smith and G.White (eds.), Philosophy and the Cognitive Science. Hölder-Pichler-Tempsky, Vienna: 443-456.
27. Guarino, N. 1995. Formal Ontology, Conceptual Analysis and KnowledgeRepresentation. International Journal of Human and Computer Studies, 43(5/6): 625-640.
28. Guarino, N. and Boldrin, L. 1993. Ontological Requirements for Knowledge Sharing. In Proceedings of IJCAI workshop on Knowledge Sharing and Information Interchange. Chambery, France: 1-5.
29. Guarino, N., Carrara, M., and Giaretta, P. 1994. Formalizing Ontological Commitment. In Proceedings of National Conference on Artificial Intelligence (AAAI-94). Seattle, Morgan Kaufmann.
30. Guarino, N., Carrara, M., and Giaretta, P. 1994. An Ontology of Meta-Level Categories. In D. J., E. Sandewall and P. Torasso (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94). Morgan Kaufmann, San Mateo, CA: 270-280.

31. Guarino, N. and Giaretta, P. 1995. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) Towards VeryLarge Knowledge Bases: Knowledge Building and Knowledge Sharing 1995. IOS Press, Amsterdam: 25-32.
32. Knight, K. and Luk, S. 1994. Building a Large Knowledge Base for Machine Translation. In Proceedings of American Association of Artificial Intelligence Conference (AAAI-94). Seattle, WA.