International Journal of Enterprise Computing and Business Systems (Online)

http://www.ijecbs.com

Vol. 1 Issue 1 January 2011

A SURVEY ON ONTOLOGY CONSTRUCTION METHODOLOGIES

R.Subhashini

Senior Lecturer,

Department of CSE,

Sona College of Technology,

Salem, India

E-mail : subhageorg@gmail.com

Dr. J. Akilandeswari

Professor and Head,

Department of IT,

Sona College of Technology,

Salem, India

E-mail : akila_rangabashyam@yahoo.co

ABSTRACT

Ontology is defined as partial specification of conceptual vocabulary used for formulating knowledge-level theories about a domain of discourse. Ontology is applied in domains like natural disaster management system, medicine, military intelligence, cooking, enterprise, jobs, agriculture, wikipedia, automobiles and so on. This paper presents a review on various ontology construction methodologies for different domains. This paper also presents the merits and drawbacks in those methodologies.

Keywords: Ontology, ontology encoding, ontology integration.

1. Introduction

Ontology is a formal explicit specification of a shared conceptualization [1]. Gonzales [2] explores the terminologies in the above definition and states that formal is an abstract model of portion of the world; explicit specification means that the constructed ontology must be machine readable and understandable; shared implies consensus of the community towards the ontology that have been built and conceptualization is expressed in terms of the concepts and the properties of the ontology.

Ontology is also expressed as a formal representation of knowledge by a set of concepts within a domain and the relationship between these concepts. Swartout et. al [19] divides ontology into two categories: domain ontology and theory ontology. Domain ontology deals with the formal description of the classes, the relationship between the classes, while theory ontology tends to be abstract and smaller. It mainly deals with time, space, plans etc. Ontologies can be constructed by using three different approaches [20]. They are single ontology approaches, multiple ontology approaches and hybrid ontology approaches. These approaches help in providing the integration task to describe the semantics of the information sources. Ontology construction is an iterative process and involves the following steps [3].

- Design: Specifies the scope and purpose of the ontology. Also reveals the relationship among classes and subclasses.
- Develop: Decides whether construction of ontology has to be done from scratch or to reuse an existing ontology.
- Integrate: Combine the developed ontology with the already existing one.
- Validate and Feedback: The completeness of the constructed ontology is verified with the help of automated tools or by seeking the opinion of the experts.
- Iterate: Repeat the process and incorporate the changes given by the expert.

The paper is organized as follows. Section 2 outlines the advantages and the need for an ontology. Section 3 introduces the components of an ontology; section 4 briefs on steps in ontology construction. Section 5 provides the conclusion.

2. Advantages and Need for Ontology

Ontology provides a common understanding of specific domains that can be communicated between people and application systems. Ontologies can be used to [4]:

- Share a common understanding of the structure of information.

- Enable reuse of already existing domain knowledge instead of creating a new one.
- Make domain assumptions unambiguous
- Examine domain knowledge.

Ontology has gained much importance not only in the field of artificial intelligence but also in the fields like information organization, natural language processing, information retrieval, knowledge representation, knowledge acquisition. Ontologies are popular for the following reasons.

- Reusability- Instead of creating a new ontology the existing ontology can be reused.
- Formal community time- Shared viewpoint over a certain universe of discourse.
- Interoperability- Support communication and co-operation among systems
- Knowledge level validation and verification – checks for the completeness of domain discourse.

3. Components of Ontology

Ontology consists of four main components to represent a domain. They are:

i. Concept represents a set of entities within a domain.
ii. Relation specifies the interaction among concepts
iii. Instance indicates the concrete example of concepts within the domain
iv. Axioms denote a statement that is always true.

Let us take an example of a wine ontology and look at its components. The concepts of the wine ontology are, "Winery, Wine, Wine descriptor, Wine color, etc". The relationships are given as Winery *produces wine*, wine *haswinedescriptor* wine descriptor. The instances of winecolor can be red, rose and white. The axiom in this example is 'a winery must produce atleast one type of wine'.

International Journal of Enterprise Computing and Business Systems (Online)
http://www.ijecbs.com
Vol. 1 Issue 1 January 2011

3.1. Levels of Formality

Ontology may be specified at different levels of formality depending on their representation.

1. Highly informal- Expressed in an ordinary natural language.

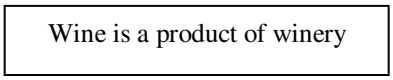


Fig 1: Example for highly informal representation

2. Semi informal-expressed in structured form of natural language

Fig 2: Example for semi informal representation

3. Semi formal- expressed in a formally defined language.

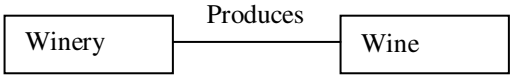


Fig 3: Example for semi formal representation

4. Formal- defines all terms formally, establish theorem and proofs.

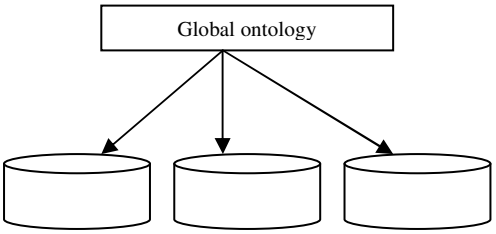
```
<a:owl_objectproperty rdf:about="produces" rdfs:label="produces">
        <rdfs:range rdf:resource="Winery"/>
        <rdfs:domain rdf:resource="Wine"/>
</a:owl_objectproperty>
```

Fig 4: Example for formal representation

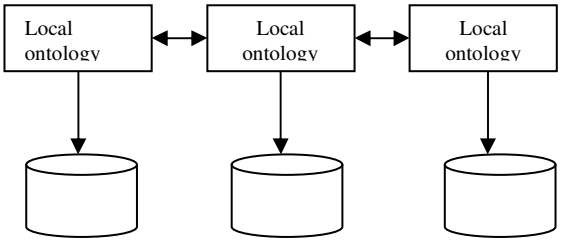
4. Ontology Construction

Ontologies can be constructed using three different approaches; single ontology approach, multiple ontology approach and hybrid ontology approach [20].

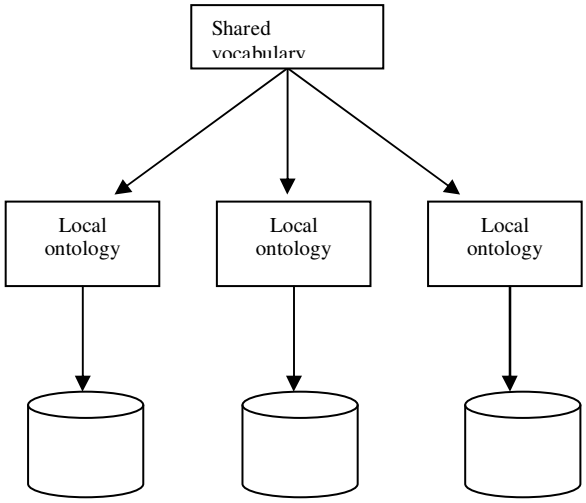
The single ontology approach is the simplest of all and it uses single global ontology for all information sources, which shares the vocabulary and the terminology to specify the semantics. The limitation of this approach is that it does not provide a perfect solution for information integration. This limitation has paved way for multiple ontology approach, where each information source is described by its own ontology thus each source will have its own local ontology. The main drawback of this approach is the construction of individual ontology. The hybrid ontology approach is the combination of single and multiple approaches [20].



(a) Single ontology approach



(b) Multiple ontology approach



(c) Hybrid ontology approach

Fig 5: Types of ontology construction approaches

In general, ontology construction could be done in three ways [3]:

- Manual: Ontology is constructed manually.
- Semi automatic: Human intervention is needed during ontology process.
- Fully automatic: The system takes care of the complete construction.

Ontology construction involves six basic steps.

1. Ontology Scope
2. Ontology capture
3. Ontology encoding
4. Ontology integration
5. Ontology evaluation
6. Ontology documentation.

4.1 Ontology Scope

The main objective of this phase is to identify the need and purpose for ontology construction. This phase takes care of identifying whether the ontology has to be built from the scratch or an existing ontology can be reused. It also identifies the range of the users and the type of questions which the ontology should answer.

In [9] construction of TextOntoEx is explained. The authors had considered natural text as input.

In [18] the scope was identified as building an ontological structure for the coronary heart disease. The medical ontology was constructed by considering western medicine and traditional medicine.

The scope of [6] was to construct the ontology in military intelligence automatically by using existing resources, thesaurus and databases.

Chen Huei Chou [15] developed an ontology structure of elements for natural disaster management system. Further, a website was developed for disaster management. The input was taken from an inventory of 6,032 web pages identified from 100 disaster management websites.
David Sanchez [8] addressed the scope as constructing medical ontology from the web with respect to cancer.. This is accomplished with the help of automatic and unsupervised ontology learning methods.

Fernando Batista [7] identified the domain as cooking, since they found many interesting factors in this domain. The inputs were gathered from cooking books like “Pantagruel” and “O grande livro ilustrado da Culin´aria” and websites like Wikipedia, Publicidade etc.

4.2 Ontology Capture

In this phase the key concepts and their relationships in the domain of interest are identified. It should also produce precise definition for the identified concepts and the relationship between the classes and the sub classes.

During ontology construction, concepts are modeled as classes or sub classes, the relations are represented as is-a, part-of and so on that exists between classes and sub classes. Instances are the values of the attributes of the class that describes the necessary properties.
TextOntoEx is a tool developed for automatic ontology construction from natural English text [9]. Most of the existing ontology construction tools support taxonomic, equivalence relationship (E.g. is-a, part-of) but does not support non taxonomic relations, conceptual relationships (E.g. causes, caused by, treat, treated by, material of etc). The author mainly focuses on the construction of ontology with non taxonomic relationship. It is done by using semantic pattern based approach. Semantic patterns are constructed by making use of symbols. The symbol < > is used to represent abstract ontological classes, "< >" is used to represent verb group, "<POS>" to express prepositions and conjunctions. Consider for an example, "Yellow Stripping on second, third or older leaves" is the normal English sentence. The semantic pattern for this sentence is constructed as <color> <shape> on [<Ordinal numeral.POS>] <plant part>.

The limitation of this work is that a small text corpus of about 65 sentences is randomly selected. The size of the text corpus can be maximized thereby more semantic patterns can be matched.

The source of information in [6] is a thesaurus and Oracle database. In this paper, concepts are identified either by pattern extraction or association rule or by concept clustering. The extracted words are built as machine readable dictionary and are converted into a knowledge base.

The work of Fernando Batista [7] comprises of four main modules and two auxiliary modules. The main module includes the identification of key concepts of the cooking domain such as actions, foods, recipes and utensils. The auxiliary modules are measures and equivalences. The main relationship between classes is is- a relation.

David Sanchez [8] builds a lexicon by extracting the domain concepts. Both the taxonomic and the non taxonomic relationship among the classes are possible. Initially taxonomy of concepts was constructed using is- a relations. (E.g breast cancer is-a cancer)

In paper [15], the identification of web elements was done based on grounded theory. The grounded theory is the discovery of theory from data. The main concept of this theory is to generate theory and conceptual abstraction from collected data through a constant comparison process. As qualitative data elements are compared iteratively, new abstract concepts and categories emerge.

4.3 Ontology encoding

Ontology encoding mainly emphasizes on choosing a representation language that will support the ontology. There are many representation languages. Some of them are SHOE (Simple HTML Ontology Extensions), OWL (Web Ontology Language), RDF (Resource Description Framework), XOL etc. The OWL and RDF is defined over XML. In other words coding the ontology is called as ontology encoding.

Ontology encoding is done by constructing semantic pattern with the help of pattern editor [9]. Semantic pattern is nothing but a general format to express a specific meaning. The author had suggested three steps for automatic ontology construction. They are

- Constructing semantic patterns using pattern editor
- Selecting domain natural text
- Extracting domain ontology from natural text.

Xu Binfeng,Luo [18] uses Protégé 2000 with RDFS as the representation language to construct an ontology for a medical knowledge base. The structure of the medical ontology will be a combination of western medicine and traditional medicine. Protégé is a knowledge based development framework that offers classes, slots, facets and instances as the building blocks for representing knowledge.

Chen Huei Chou [15] identifies the instances and relationships of the web elements and coded using XML. The XML schema is visualized as a DOM tree.

Protégé IDE was used to construct the ontology for the cooking domain [7]. Protégé is capable of generating OWL or RDF.

In paper [6], the author had taken thesaurus and Oracle database as the resource. The thesaurus contains 61895 terms and three kind of relationship between concepts. Protégé IDE is used and the tool and OWL is used as the ontology language.

4.4 Ontology integration

The use of ontology integration is to combine the constructed ontology with the existing one. This plays a vital role because the generated ontology has to be combined with the existing one to fulfill the work. This task will be made easier if the available ontology make explicit assumptions.

In paper [7], new ontology was built because the existing ontology did not match with the scope of this project. Some of the related ontologies are:

- USDA National Nutrient Database for Standard Reference- This has 7146 food items and 13 food components.
- AGROVOC- is a multi lingual thesaurus and has 17000 concepts and three types of relation.
- Grace et al- presented a specialized wine ontology that covers maceration, fermentation, processes, grape maturity state, wine characteristics and classification.
- Noy and McGuinness- presents the development of wine, food and the combination of wine with meals ontology.
- Villarias – describes an ontology of culinary recipes.

In most of the papers, new ontologies are constructed instead of reusing the existing ontologies. The reason for this is the need and the scope of the new ontology demands for the new classes, relations, instances and axioms and does not match with the related ontologies in the respective areas.

4.5 Ontology evaluation

The constructed ontology should be evaluated by taking some evaluation criteria. These criteria can be categorized into two types namely generic criteria and specific criteria. The generic criteria deals with factors like clarity, consistency, reusability. The specific criteria check the generated ontology against the purpose and user requirements. In paper [7], evaluation was done in two ways. Internal evaluation was done by the team of ontology builders and the external evaluation was done by the client. Competency questions are asked to evaluate the correctness of the ontology.

In paper [9], evaluation was done at two levels. The first level was concerned about the relevance of the extracted semantic pattern and the second level was concerned about the method used for extraction.

In [8], evaluation was done taking two parameters into consideration. The factors are: precision and recall. The precision for taxonomic relationship is 74% and for the non taxonomic relationship it is 71.1%.

The evaluation of the constructed ontology in paper [15] was done by comparing with eight government disaster management websites. Two websites from four regions are picked and the ontology coverage was evaluated.

4.6 Ontology documentation

The last phase is ontology documentation. Effective knowledge sharing requires adequate documentation. This phase is very important because ontologies can be reused only if it is properly documented. Documentation should be done with atmost care and must record all the assumptions that are made explicitly.

An example of documenting wine ontology is given here.

Concept	Description
Winery	Place where wine is produced.
Wine	Portable liquid made from grapes.
Desert wine	A kind of wine

Relations	Description
produces	Links wine to its producer
hasdrink	Associates meal course to a wine.

Instances	Description
Bancrof	A winery
Red	A wine color.
White	A wine color.

5. Conclusion

This paper explains the importance and the purpose of constructing an ontology. The component modules involved in constructing an ontology is also briefed. Six steps associated with the construction

of ontology namely ontology scope, ontology capture, encoding, integration, evaluation and ontology documentation with respect to domains like natural disaster management system, medicine, military intelligence and cooking are elucidated.

References

1. Thomas R. Gruber, 1995. Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies vol.43: 907-928.
2. Roberto Garzia Gonzales. 2005. A Semantic Web Approach to Digital Rights Management, PHD Thesis, Universitat Pompeu Fabra: http://rhizomik.net/~roberto/thesis/Thesis.pdf.
3. Maizatul Akmar Ismail, Mashkuri Yaacob & Sameem Abdul Kareem, 2006. Ontology Construction: An Overview, Malaysia.
4. N. Noy and D. L. McGuinness, 2001. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Medical Informatics, Stanford
5. Alberto H.F.Laender, Berthier A.Roberio Neto, Altigran S.da Silva, Juliana S.Teixeira, 2002. A Brief Survey of Web Data Extraction Tools. ACM SIGMOD Record , Volume 31, Issue 2: 84 – 93
6. Bing ru Yang , De quan Zheng Jing Yang, Li Liu , Mei ying Jia, Wei cong Sun, 2009. Automatic Ontology Construction Approaches and its applications on Military Intelligence, Asia Pacific Conference on Information Processing: 348-351.
7. Fernando Batista, Joana Paulo, Nuno Mamede, Paula Vaz, Ricardo Ribeiro, 2006. Ontology construction: cooking domain.
8. Antonio Moreno,David Sánchez, 2008. Learning medical ontologies from the Web. LNCS Knowledge management for Health care procedures:32-45.
9. Ahmed Rafea, Hesham A. Hassan, Mohamed Yehia Dahab, 2006. TextOntoEx: Automatic Ontology Construction from Natural English Text. International conference of Artificial Intelligence and Machine Learning.
10. Ying Ding and Schubert Foo, 2001. Ontology research and development. Part 1 - A Review of ontology generation. Journal of Information Science : 123-136.
11. Ahmad Abdollahzadeh Barforoush, Sima Soltani, 2009. Web Pages Classification Using Domain Ontology and Clustering. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI).
12. Antonio Moreno and David Sánchez,2008. Learning non-taxonomic relationships from web documents for domain ontology construction. Data and Knowledge Engineering.

13. Christine W. Chan, 2004. From Knowledge Modeling to Ontology Construction. International Journal of Software Engineering and Knowledge Engineering (IJSEKE).
14. Gaoying Cui, Qin Lu, Wenjie Li, Yirong Chen. Corpus Exploitation from Wikipedia for Ontology Construction: 2125-2132
15. Chen-Huei Chou, Fatemeh Zahedi, Huimin Zhoa, 2008. Ontology for Developing websites for natural disaster management: methodology and implementation.
16. Harith Alani, 2006. Position paper: Ontology construction from online ontologies. Proceedings of the 15th international conference on World Wide Web : 491 - 495.
17. Antonio M. Rinaldi, 2009. An Ontology-Driven Approach for Semantic Information Retrieval on the Web. In ACM Transactions on Internet Technologies, Volume 9, Article 10.
18. Xu Binfeng,Luo, Xiaogang Peng Cenglin, Huang Qian, 2007. Based on ontology: construction and application of medical knowledge base. IEEE International conference on complex medical engineering: 586- 589.
19. Bill Swartout, Ramesh Patil, Kevin Knight, Tom Russ,1996. Toward Distributed Use of Large-Scale Ontologies. AAAI Technical Report SS-97-06.
20. S. Hubner, H.Neumann, H.Stuckenschmidt, G.Schuster, T.Vogele, U.Visser, H.Wache, 2001. Ontology-Based Integration of Information-A survey of Existing Approaches.

International Journal of Enterprise Computing and Business Systems (Online)

http://www.ijecbs.com

Vol. 1 Issue 1 January 2011