ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Automatic Ontology Creation and Management for Semantic Based Data Mining Information System

Dr.R.MALA
Assistant Professor in Computer Science Department
Alagappa University College of Arts and Science,
Paramakudi.

Abstract -- World Wide Web is the largest database which is mostly understandable by human users and not by machines. The interdependency of its components is maintained by lacks of existence of a semantic structure. The web may present the irrelevant information because search on web is based on keyword, matching of URL's/hyperlinks. Hyperlinks make the physical links which are not understood by the machines. So there is a need to create an ontology which captures the meaning of the links. This can be easily understand by machines. The extraction problem in www is still central research goal in web communities like AI, Data Mining data base etc. because of its unstructured data. Information retrieval is synonymous with "determination of relevance". Information retrieval mainly based on user's requirement. This paper discusses general issues to consider and offer one possible process for developing an iterative approach for ontology development. This work mainly focuses on semantic web. This is useful in evaluating and improving web sites and web services.

I. Introduction

Data Mining is the process of finding and extracting new and potentially useful knowledge from data. Data mining is also known as Knowledge discovery in databases (KDD). The terms "Data mining" and "Knowledge discovery in database" are used interchangeably [1]. Data mining is an interdisciplinary field, drawing from different areas including database system, statistics, machine learning, data visualization and information retrieval. The task of data mining involves two primary goals; those goals are prediction and description [2]. Prediction is concerned with using some variables or fields in the database to predict unknown or future values of other variable of interest, while description focuses on finding human-interpretable patterns describing the data.

What Is the Ontology?

The word "ontology" has been recognized in philosophy as the subject of existence. In Artificial Intelligence community, ontology means a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some world phenomena. Ontology concepts and the relationship among those concepts should be explicitly defined. Further, ontology should be machine-readable and the ontology should capture consensual knowledge accepted by the community [3]. Ontology is used for knowledge sharing and reuse. It improves information organization, management and understanding. Ontology has a significant role in the areas dealing with vast amounts of distributed and heterogeneous computer based information, such as World Wide Web, Intranet information systems, and electronic commerce. Ontology will play a key role in the second generation of the web, which Tim Berners-Lee call the "Semantic Web", in which information is given well-defined meaning, and is machine-readable. Search engines will use ontology to find pages with words that are syntactically different but semantically similar [4, 5, and 6].

Semantic web

The current World Wide Web (WWW) has a huge amount of data that is often unstructured and only human understandable. Web is rich with information; gathering and making sense of the data in the web is more difficult because the document of the Web is largely unorganized and unstructured. From the unorganized human readable web data semantic web is how to effectively and efficiently creating a machine-understandable, queriable, information and knowledge layer. If computer can understand the meaning behind the information, it can learn what we are interested in and it help us better find what we want. Since the semantic Web mainly focuses on the data and information. Data in the Semantic Web is well defined and linked in a way that can be used for more effective

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

discovery, automation. The nature of most data on the Web is unstructured that only understand by humans, the amount of data is very huge on the web that processed efficiently by machines.

The goal of the Semantic Web is to develop allowing standards and technologies designed for both user and machines understandable. Semantic web information can support data integration, data discovery, navigation, and automation of tasks.

II. Related Work

Usually the ontology building is performed manually, but researchers try to build ontology automatically or semi automatically to save the time and the efforts of building the ontology. We survey in this section the most important approaches that generate ontologies from data.

Ontobase Ontology Repository [7] is an implementation of a design that allows users and agents to retrieve ontologies and metadata through open Web standards and ontology services. The Ontobase provides a knowledge management mechanism by maintaining structural and semantic information about each data source, recording the relationship between attributes of the data sources with terms from a business domain, and computing contextual information gleaned from these linkages and other resource related information. Another method [8] based on WordNet [9] has been presented to merge the heterogeneous domain ontologies.

WordNet uses a dictionary to detail the relationships between the concepts like the synonym, antonym, hypernym and hyponym. The main idea was to merge the taxonomies, because they are central components of ontologies. After evaluation, it was determined that the new methodology is very efficient in merging between heterogeneous ontologies.

Clerkin et al. used concept clustering algorithm (COBWEB) to discover automatically and generate ontology. They argued that such an approach is highly appropriate to domains where no expert knowledge exists, and they propose how they might employ software agents to collaborate, in the place of human beings, on the construction of shared ontologies [10].

Blaschke et al. presented a methodology that creates structured knowledge for gene-product function directly from the literature. They apply an iterative statistical information extraction method combined with the nearest neighbor clustering to create ontology structure [11].

Formal Concept Analysis (FCA) is an effective technique that can formally abstract data as conceptual structures [12]. Quan et al. proposed to incorporate fuzzy logic into FCA to enable FCA to deal with uncertainty in data and interpret the concept hierarchy reasonably, the proposed framework is known as Fuzzy Formal Concept Analysis (FFCA).They use FFCA for automatic generation of ontology for scholarly semantic web [13].

Dahab et al. presented a framework for constructing ontology from natural English text namely TextOntEx. TextOntEx constructs ontology from natural domain text using semantic pattern-based approach, and analyzes natural domain text to extract candidate relations, then maps them into meaning representation to facilitate ontology representation [14].

Wrobel et_al. used different ways to build ontologies automatically, based on data mining outputs represented by rule sets or decision trees. They used the semantic web languages, RDF, RDF-S and DAML+OIL for defining ontologies [15].

Semantic Web Approaches

Diana Cerbu et al., [16] proposed two developing area Semantic Web and Web Mining. The author's proposed how these two areas can be combined with three different approaches to semantic based web mining an approach to pattern mining; a text classification algorithm is called AdaBoost and a framework for creating well customized content on the web by using web mining and semantic ontologies.

Thomas Fischer et al., [17]. Motivated the application of relational data mining algorithms in semantic web. They have outlined important differences to the knowledge discovery process. The modelling, selection and transformation are different to the standard approach. The knowledge discovery process recommended choosing parts of semantic data to fully use information and background information derived from a web sources.

Nizar R. Mabroukeh et al., [18] proposed a generic framework called SemAware that integrates semantic information into web usage mining. Semantic information can be combined into the pattern discovery.

A semantic distance matrix is used in the agreed sequential pattern mining algorithm to trim the search space and partially relieves the algorithm from support counting. A 1st-order Markov model is used to build the mining process and enriched with semantic information.

Yao et al., [19] proposed a framework designed for an intelligent agent that dynamically gives the recommended to the web site's users by learning from web usage data and users' behaviour is called PagePrompter,. Like a guide, an agent supports a user in navigating the web site. PagePrompter can also be used as a tool for understanding user behaviour, the design of web sites, system performance analysing, web site designer for improving web sites and generating an adaptive web site.

III. Problem Scope And Statement

The traditional task of the knowledge engineer is to translate the knowledge of the expert into the knowledge base of the expert system. Ontology is used to represent knowledge of domain expert by Knowledge engineer. Due to of the difficulty to find a domain expert and the needing for updating the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

knowledge represented in the ontology frequently, it is proposed a system for building ontology automatically from the database.

Ontologies are back bone for Semantic Web based systems. Developing Ontology is not an easy task and there is no correct ontology for any domain. The issues related to ontology domain engineering are ontology consistency check, integration, ontology mapping, ontology translation, and ontology reuse.

Ontology is thee model for the real world and the concepts of ontology reflect this reality [20]. After the definition of an initial version of the ontology, it must be evaluated with the applications and discussed with experts of that domain. This process of iteration will be continued throughout the lifecycle of ontology. By providing standards for ontology, ontologies for same subject will be easy to create and reuse. Some ontology fundamental rules should be followed in ontology creation process. These rules are not strong [21], but they can help in making design decision in many cases. The main concept behind the development of any ontology is

For designing a domain there is no one correct way, an alternative must be needed. This solution always depends on the application used by the ontology.

Ontology development process is iterative.

Concepts in the ontology should be close to objects and relationships among domain selected for development. In a domain, mostly the sentence having nouns are objects and verbs relationships.

Proposed Methodology

A general ontology merging process involves six steps: feature engineering, selection of next search steps, similarity computation, similarity aggregation, interpretation and iteration of this whole process.

Merging the heterogeneous ontologies based on WorldNet has four distinct phases. In the first phase, WorldNet [9] is used as a similarity measure in different ontologies. Then in a second phase, selection of the mass similar concept is performed. Similarity is computed in third phase. Finally the reconstruction of the new ontological hierarchy is performed.

For developing ontologies [22] there is no correct way and methodology. The proposed methodology discussed about the general issues and steps to develop ontology. The proposed method also discusses about the decision modeling for a designer and implications of different solutions [23, 24, and 25]. The proposed development method describes an iterative approach to ontology development.

Step 1. Establish the domain and scope of the ontology

The first step starts with definition of domain and its scope. In this step several basic questions are asked. What domain the proposed ontology will cover? Where the proposed ontology is going to used? What type of questions to be answered by the proposed ontology? Who is the user? And who will maintain the ontology? The answers to these questions may change during the ontology-design process, but they will help to limit the scope of the model.

Step 2. Reusage and Extension of ontologies

Analyze the existing ontology whether it is reused for current application. If the proposed system needs to interact with the other applications, the existing ontology can be reused.If it is necessary the existing ontology can be refined to our existing domain and task [26].

Step 3. List important terms in the ontology

List all the terms to make statements and give explanation to the users about the terms. List the terms to talk about. List the properties of terms.

Step 4. Class definition and organization of the class

First define the class. From the list created from step3, the items are selected. It describes the terms having independent existence called classes. Later it will become anchors in class hierarchy. Classes are also organized into a hierarchy.

Step 5. Classes-slots: Definition

The classes alone will not provide enough information to answer the questions in Step 1. Once the classes are defined the internal structure of concepts must be described.

Step 6. Define the facets of the slots

Slots can have different facets describing the value type, range of values, total values (cardinality), and other features of the values the slot can take.

The above six steps are repeated until the ontology is fit for the particular application.

IV. Conclusion and Future Implementation

In this paper, proposed methodology for building ontology is used for specific application and refined according to or requirement. For Future enhancement the domain knowledge can be acquired from domain experts. The only correct requirements can be satisfied for a user. In future it is also taken into consideration in building of ontology for unstructured data from WebPages and documents.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

References

[1] Frawley, W., Piatetsky-Shapiro, G., and Matheus, C., Knowledge Discovery in Databases: An Overview. Ai Magazine, Vol. 13 (1992), pp.57-70.
[2] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery: An overview. In Advances in Knowledge Discovery and Data Mining, pp. 1 --34. AAAI Press, Menlo Park, CA, 1996.
[3] Gruber, T.R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5, 199-220.
[4] Berners-Lee, T., Weaving the Web, Harper, San Francisco, 1999
[5] Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M. and Horrocks, I. (2000) 'The semantic web: the roles of XML and RDF', IEEE Internet Computing, Vol. 4, No. 5, pp.63–74.
[6] Ding, Y., and Foo, S., (2002). Ontology Research and Development: Part 1 – A Review of Ontology Generation. Journal of Information Science 28 (2).
[7] Ding Pan, Yan Pan, "Using Ontology Repository to Support Data Mining", The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006, Volume: 2, On page(s): 5947-5951.
[8] Hyunjang Kong, Myunggwon Hwang, Pankoo Kim, "A New Methodology for Merging the Heterogeneous Domain Ontologies based on the WordNet", International Conference on Next Generation Web Services Practices, 2005. NWeSP 2005. page(s): 6
[9] Sahoo, K. Vidyasagar, V.E., (2003) "Kannada WordNet - a lexical database", Conference on Convergent Technologies for Asia-Pacific Region, Volume 4, Page(s):1352 - 1356 Vol.4
[10] Clerkin, P., Cunningham, P., and Hayes, C., Ontology Discovery for the Semantic Web Using Hierarchical Clustering, , Trinity College Dublin, Ireland, TCD-CS-2002-25
[11] Blaschke, C., & Valencia, A., Automatic Ontology Construction from the Literature, Genome Informatics, Vol. 13, pp 201–213, 2002.
[12] Ganter, B.; Stumme, G.; Wille, R. (Eds.) (2005). Formal Concept Analysis: Foundations and Applications. Lecture Notes in Artificial Intelligence, no. 3626, Springer-Verlag. ISBN 3-540-27891-5.
[13] Quan, T. T., Hui, S. C., Fong, A. C. M., and Cao, T. H. (2004). Automatic generation of ontology for scholarly semantic Web. In: Lecture Notes in Computer Science. Vol. 3298. (pp. 726–740).
[14] Dahab, M. Y. Hassan, H., and Rafea, A.., TextOntoEx: Automatic ontology construction from natural English text, Expert Systems with Applications (2007), doi:10.1016/j.eswa.2007.01.043.
[15] Wuermli, O., Wrobel, A., Hui S. C. and Joller, J. M. "Data Mining For Ontology_Building: Semantic Web Overview", Diploma Thesis–Dep. of Computer Science_WS2002/2003, Nanyang Technological University.
[16] Diana Cerbu, Romania Konstanz 2008. Semantic Web Mining journal of web service and Semantic web.
[17] Thomas Fischer,Johannes Ruhland 2010. Towards Knowledge Discovery in the Semantic Web, MKWI – Business Intelligence, vol 2 pp 151 -166
[18] Nizar R. Mabroukeh and Christie I 2009 .Using Domain Ontology for Semantic Web Usage Mining and Next Page Prediction,. Ezeife CIKM'09, 2–6.
[19] Y.Y. Yao, H.J. Hamilton, and Xuewei Wang PagePrompter 2008. An Intelligent Agent for Web Navigation Created Using Data Mining, Volume 27 Issue 3, Pages 59 - 74.
[20] McGuinness, D.L. and Wright, J. (1998). Conceptual Modeling for Configuration: A Description Logic-based Approach. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing - special issue on Configuration.
[21] Chimaera, Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B. and Benjamins, V.R. (2000). Ontology Environment. www.ksl.stanford.edu/software/chimaera
[22] WonderTools? A comparative study of ontological engineering tools. International Journal of Human-Computer Studies 52(6): .
[23] Farquhar, A. (1997). Ontolingua tutorial. http://ksl-web.stanford.edu/people/axf/tutorial.pdf
[24] Gómez-Pérez, A. (1998). Knowledge sharing and reuse. Handbook of Applied Expert Systems. Liebowitz, editor, CRC Press.
[25] Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5: 199-220.
[26] Hendler, J. and McGuinness, D.L. (2000). The DARPA Agent Markup Language. IEEE Intelligent Systems 16(6): 67-73.