ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

Ontology Learning based on Bootstrapping Approach using Improved ET Algorithm for Semantic Web Services with WSDL

Sasanka Bhimavaram[1], Dr. P. Govindarajulu[2]

Research Scholar, Department of Computer Science, Sri Venkateswara University, Tirupati, India[1]

Professor, Department of Computer Science, Sri Venkateswara University, Tirupati, India[2]

Abstract: In this research work, we deliberate how ontologies can be constructed by mining information from databases and Web documents in web service description language. This is a stimulating task, because information mining is typically a noisy effort whereas ontologies generally involve hygienic and crispy data. This means that the mined information has to be cleaned, disambiguated and prepared logically stable to certain step. We discuss approaches that extract ontology in this spirit from the relational databases and also present methodologies that objective to extract ontology from documents or by extension from the entire Web then implemented with the improved enhanced traversal algorithm on the databases such as bank and hospital then obtain better accuracy performance. We will show that information mining and ontology creation can arrive into a successful support loop, where additional mined information pointers to a bigger ontology which helps mining more information.

Keywords: Ontology, Semantic Web Services, Bootstrapping, Improved ET Algorithm, WSDL.

I. INTRODUCTION

A. Information Extraction/Mining

Information Extraction (IE) is the process of extracting structured information from one or multiple given source documents. There are literally hundreds of different IE techniques. Some of these techniques shave become particularly popular. At the same time, the ontology can also help the IE process. This can happen on multiple levels. First, the data that is already present in the ontology can serve as seed data for the IE process.

Thus, the more knowledge has already been extracted, the better the system will be able to extract new information. The better it extracts new information, the more knowledge will be extracted. IE and ontologies could enter a fruitful cycle of knowledge accumulation. We will look at two classes of ontological knowledge extraction systems that go into this direction.

The first class extracts information from relational databases. The second class ventures beyond Web and extracts information from arbitrary documents [3,5].

B. Information Extraction/Mining for Ontologies

Canonicity: The entity names have to be disambiguated and mapped to existing entities of the ontology. The system may also decide to introduce a new entity if the entity does not yet exist in the knowledge base. The same applies to class names and relation names.

Taxonomic organization: The extracted entities have to be placed in a taxonomy of classes.

Consistency: The extracted facts have to be logically consistent with the ontology.

C. Information Extraction from the Web

The relational databases contains already much information, it contains only part of the huge amount of data that is available on the Web. This is why several newer approaches have embarked to extract ontological information from the entire Web. The Web is much more heterogeneous than databases, with different file formats, different languages, different page layouts, and only creeping standardization. Furthermore, the information on the Web exhibits various degrees of credibility. Data may be faulty, incomplete, contradictory or wrong. In addition, the Web is one of the largest computer processable resources at all[2,7].

II. LITERATURE REVIEW

A. Ontology Learning from Databases Semantic Patterns in Data

To construct the accurate semantic resources required by future knowledge-intensive applications, existing databases are undoubtedly among the most reliable sources to be exploited. However, finding ways to significantly ease the task of building highly expressive ontologies from such structured information sources is far from being a straight forward issue. Early methods exclusively based on the transformation of database schemas often result in incomplete ontologies that need to be further refined at the cost of huge manual post-editing efforts. Such manual tasks might be deemed too tedious and costly by many practitioners. To provide an extended automated support to facilitate the production of high quality ontologies from databases, adequate ontology learning methods should be elaborated[1,9].

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

B. Range of Semantic Patterns

We give in this section an overview through selected examples of the structuring patterns that will be further explored in this chapter. We introduce here some interesting structuring patterns without paying to much attention to how they can be automatically identified and used to generate appropriate ontology fragments. Some of the most relevant content-driven transformation techniques will be addressed in next sections and we show how to exploit them in order to complement ontologies derived from database schemas.

C. Preliminary Definitions

A relational database schema D is defined as a finite set of relation schemas D= $\{R_1,\ldots,R_n\}$ where each relation schema R_i is characterized by its finite set of attributes $\{A_{i1},A_{i2},\ldots,A_{im}\}$. A function pkey links to every relation its primary key which is a set of attributes $K \subseteq R$.A relation r on a relation schema R is a set of tuples which are classifications of |R| values. Correspondingly, a database d on D is defined as a set of relations d = $\{r_1,r_2,\ldots,r_n\}$. By resolution, if a relation schema is represented by a capital letter, the corresponding lower case letter denotes an instance of the relation schema. A projection of a tuple t on a set of attributes $X \subseteq R$, denoted t[X], is a restriction on t, resulting in the subsequence with values corresponding to attributes of X. the projection of a relation r on X, denoted $\pi x(r)$, is defined by $\pi x(r)$ = {t{X]|t€ r}.

The concept of inclusion dependency is used to account of correlations between relations. An inclusion dependency is an expression $R[X] \subseteq S[Y]$ where X and Y are respectively attribute arrangements of R and S relation schemas with the constraint |X| = |Y|. The dependency holds between two instances r and s of the relation schemas for each tuple u in r there is a tuple V in S such that u[X] = v[Y]. Informally, an inclusion dependency is a suitable approach to state that data items are derivative from a new relation. Foreign key relationship can be defined as additional dependencies filling the additional property: Y= pkey(S). The notation $R[X] \subseteq S[pkey(S)]$ will be used for these definite dependencies. Proper metaphors of ontology remains are articulated in OWL theoretical syntax [2,8].

III. IMPLEMENTATION OF MODULES WITH WSDL

1) Ontology Bootstrap equation

The best known application of the bootstrap is to estimating the mean, μ say, of a population with distribution function F, from data drawn by sampling randomly from that population. Now,

$$\mu = \int x \, dF(x)$$

The sample mean is the same functional of the empirical distribution function, i.e. of

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$$

Where$X_1, \ldots, X_n$ denote the data. Therefore the bootstrap estimator of the population mean, μ is the sample mean.

$$\mu = \int x \, dF(x) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Likewise, the bootstrap estimator of a population variance is the corresponding sample variance; the bootstrapping estimator of a population correlation coefficient is the corresponding empirical correlation coefficient; and so on.

More generally, if $\theta_0 = \theta(F)$ denotes the true value of a parameter ,where θ is a functional,

Then$\theta_0 = \theta(F)$ is the bootstrap estimator of θ_0.

2) Performance Similarity

We use for effectiveness of proposed system from parameters of the accuracy which is the probability of the services retrieved that are significant to the user's information requirement and recall is the probability of the services that are significant to the query that are effectively retrieved.

A={relevant_Webservices}

B={retrived_Webservices}

$$\text{Precision}=\frac{|A \cup B|}{|B|}|$$

$$\text{Recall}=\frac{|A \cap B|}{|A|}$$

Algorithm: Improved Top search phase of the "Improved Enhanced Traversal" technique

Step 1: enhanced_top_subs?(y,c)= Σ_π,w
Step 2: if y marked as 'positive' then
Step 3: result of w1≡ w2 → true
Step 4: else if y marked as 'negative' then
Step 5: result → false
Step 6: else if for all
z € predecessors(y) always
enhanced_top_subs(z,c)
Step 7: subs?(y,c) then w1 = w2
Step 8: mark (y,'positive')
Step 9: result $R \subseteq \Sigma_\pi$ → true
Step 10: else if mark(y, 'negative')
Step 11: else σ: w→ Σ_π
Step 12: result → false

3) Ontology Evolution

Ontology evolution is the useful module where, the descriptor is more authenticated using the textual service descriptor. The analysis is based on the improvement that a Web service can be divided into two descriptions: the WSDL description and a textual description of the Web service in free text. The WSDL descriptor is analysed to extract the context descriptors and possible concepts as described [4,11].

4) Domain Extraction:

In this module we extend the data extraction procedure using Web service that allows domain particulars to be recognized based on the domain name, that maintains a web services associated with operations and services.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

Finally it extracts the URL's list for user specific domains and provides those URL's to access.

5) Creating Required Tables

The CREATE TABLE command can also be entered at the mysql> prompt or can be written into a file and sent into MySQL. The latter is preferable because you keep hold of records of how created the tables. The tables may be created as follows:

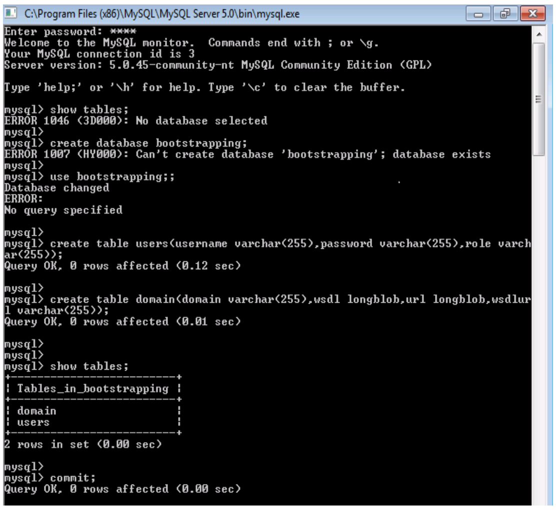

Fig.1: Creation of tables for bootstrapping

6) Adding data to MySQL

Once you have created tables, you can start filling it with data. One easy way of adding a lot of data is by using the mysql import system command. This will read in a text file where data for each table row are divided by newlines and data for every column are separated by tabs and in the identical order as the columns were defined. The file should be named according to the principle tablename.txt.table, replacing "tablename" appropriately.

7) Activity Diagram

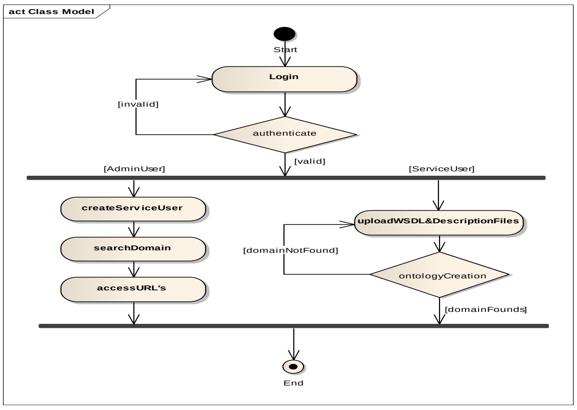

Fig.2: Activity diagram for ontology creation

IV. COMPARISON OF TRAVERSAL AND IMPROVED ET ALGORITHM FOR SEMANTIC WEB SERVICES

A. Token Extraction

In this module we develop the token extraction or mining process using WSDL (Web Service Description Language) using Bank Application Service and Hospital data. WSDL document with the token record is bolded. Here the extracted token record or list serves as a baseline. These tokens are extracted from the WSDL document of a Web service.

TABLE I

Traversal Algorithm	Improved Enhanced Traversal Algorithm
52%	70%
79%	85%
45%	90%
38%	82%
61%	93%

Table 1: Performance Accuracy Comparison of Traversal and Improved Enhanced Traversal Algorithm using Bank Application Service

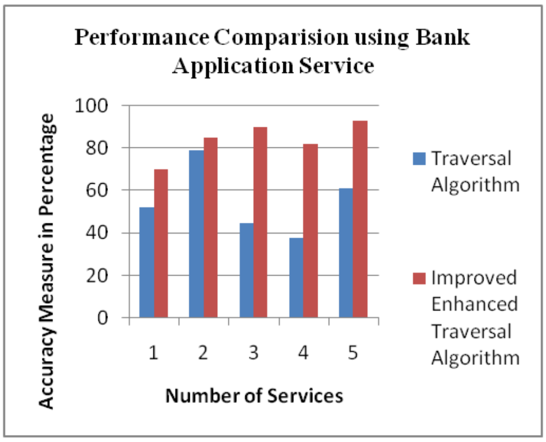

Fig.3: Performance comparison using Bank Application Service

TABLE II

Traversal Algorithm	Improved Enhanced Traversal Algorithm
60%	67%
56%	72%
74%	96%
59%	85%
81%	94%

Table 2: Comparison of Traversal and Improved Enhanced Traversal Algorithm using Hospital data service

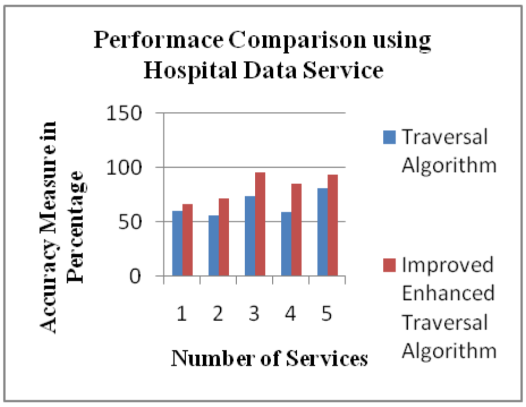

Fig. 4: Performance comparison using Hospital Data Service

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

The service is used as an initial step in our example in building the ontology of Bank database and Hospital data distinctly. Additional services will be used later to illustrate the process of expanding the ontology. Here we have implemented Traversal Algorithm and Improved Enhanced Traversal Algorithm with Bank Application Service and Hospital data, then compared the accuracy performance [3,10]. It will evaluate the list of services and operations from the WSDL File.

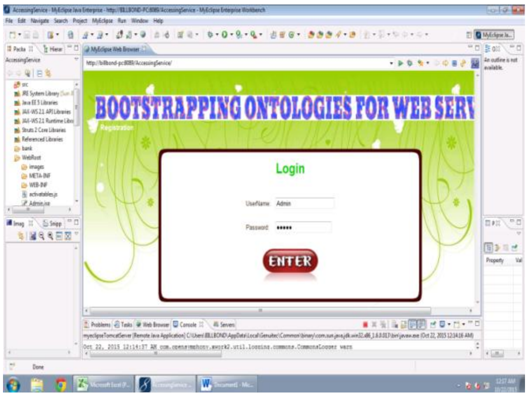

Fig.5: Admin Login Page for Bootstrapping Ontology

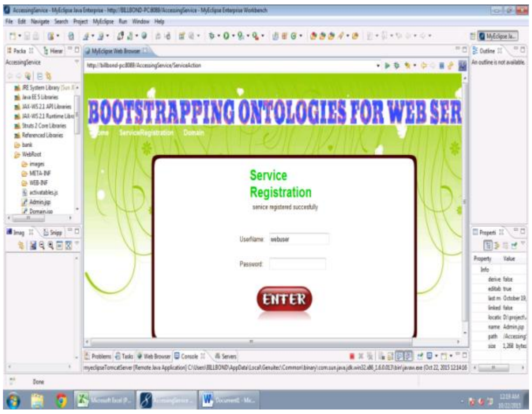

Fig.6: Service Registration Process

V. CONCLUSION

We deliberated about ontology learning for relational databases with a focus on methods for the identification of semantic patterns in the stored data. Ontology learning in this context should not be conceived as an endogenous process that only aims at discovering domain semantics from the data and metadata of some source legacy databases. In this research, we applied improved enhanced traversal algorithm with bootstrapping mechanism to relational databases such as bank application service and hospital data for constructing the web services to permit web service clients to discover related services easily for finding the best set of semantic web services. It was shown that the proposed system gives better accuracy performance for bootstrapping ontology creation and semantic web services. Moreover, Interoperability of the resulting ontologies can significantly be increased by mapping the application specific concepts extracted from these databases to equivalent or closely related concepts from widely shared reference ontologies. In future work, we will incorporate information extraction techniques with improved enhanced traversal algorithm to discover complex axioms from these heterogeneous sources is a major challenge for bootstrapping ontology learning system with different set of semantic web services.

REFERENCES

[1] Amjad Umar and Adalberto Zordan. Enterprise Ontologies for Planning and Integration of Business: A Pragmatic Approach. IEEE Transactions on Engineering Management, vol. 56, no. 2, pages 352–371, 2009.

[2] C. Golbreich, S. Zhang, and O. Bodenreider. The foundational model of anatomy in OWL: Experience and perspectives. *Journal of Web Semantics*, 4(3):181–195, 2006.

[3] D. Gessler , G. Schiltz, G. May, S. Avraham, C. Avraham, D. Grant and R. Grant. SSWAP: A Simple Semantic Web Architecture and Protocol for semantic web services. *BMC Bioinformatics*, 10(1):309, 2009.

[4] E. Al-Masri and Q.H. Mahmoud, "Investigating Web Services on the World Wide Web," Proc. Int'l World Wide Web Conf. (WWW '08), 2008.

[5] Gomadam K., Ranabahu A. and Sheth A.. SA-REST: Semantic Annotation of Web Resources. Member submission, W3C, April 2010.

[6] G. Zhang, A. Troy, and K. Bourgoin, "Bootstrapping Ontology Learning for Information Retrieval Using Formal Concept Analysis and Information Anchors," Proc. 14th Int'l Conf. Conceptual Structures (ICCS '06), 2006.

[7] G. Rizzo, R. Troncy, S. Hellmann, M. Bruemmer: NERD meets NIF: Lifting NLP Extraction Results to the Linked Data Cloud, Proc. of 5th Workshop on Linked Data on the Web (LDOW), 2012.

[8] H. Davulcu, S. Vadrevu, S. Nagarajan, and I. Ramakrishnan, "OntoMiner: Bootstrapping and Populating Ontologies from Domain Specific Web Sites," IEEE Intelligent Systems, vol. 18, no. 5, pp. 24-33, 2003.

[9] N.F. Noy and M.A. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment", In Proceedings of AAAI/IAAI, pp.450-455, 2000.

[10] P. Kungas; M. Matskin. Web Services roadmap: the Semantic Web perspective, Proceedings of the International Conference on Internet and Web Applications and Services, 2006.

[11] Q.A. Liang and H. Lam, "Web Service Matching by Ontology Instance Categorization," Proc. IEEE Int'l Conf. on Services Computing (SCC '08), pp. 202-209, 2008.

[12] Shanfeng Qi, Xinhuai Tang, and Delai Chen, "An Automated Web Services Composition System Based on Service Classification and AI Planning," in 2012 Second International Conference on Cloud and Green Computing, Xiangtan, pp. 537 – 540, 2012.

[13] T.R. Gruber, "A Translation Approach to Portable Ontologies," nowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993.

[14] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web A new form of Web content that is meaningful to computers will unleasha revolution of new possibilities. 2001.

[15] Y. Zhao, J. Dong, and T. Peng, "Ontology Classification for Semantic-Web-Based Software Engineering," IEEE Trans. Services Computing, vol. 2, no. 4, pp. 303-317, Oct.-Dec. 2009.

BIOGRAPHIES

Sasanka Bhimavaram, has completed M.Sc from Sri Venkateswara University, Tirupati, Andhra Pradesh, India. Now Pursuing Ph.D in the department of Computer Science, Sri Venkateswara University, Tirupati, India. He Published research papers in International Journals and the areas of research interests include Data Mining, Web Technology, Semantic Web Mining, Cloud Computing, etc.

Dr.P.Govindarajulu, worked as a professor in the department of Computer Science, Sri Venkateswara University, Tirupati. He has published several research papers in various national, International journals, conferences and workshops. His research areas of interests include Image Processing, Data Mining, Cloud Computing, Web Mining, etc.