



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

Ontology Based Searching Meta Data Extraction

S.Mohana Prakash¹, S.Sathish Kumar², R.Vaishnavi³, E.Anjaline Suganya⁴, M.Ramakrishnan⁵

UG Scholar, Department of Computer science & Engineering, Sree Shakthi Engineering College,
Coimbatore, Tamilnadu, India^{1,2,3,4}

Assistant Professor, Department of Computer science & Engineering, Sree Shakthi Engineering College,
Coimbatore, Tamilnadu, India⁵

ABSTRACT: In present technology, the search queries returns a large set of results produced in the web page, and also searching a relevant result tends to be a tedious process for the user. The user is looking for something but the result produced will contain information from various domains. So the user needs to search for the exact result for a long time by surfing [1]. But the developer cant able to design the webpage as per the user point of view because each individual user can have different aspects so can't able to design as per the user view. This will result in the increased time complexity. To overcome this problem, we use the concept annotation (Web content extraction) by means of using ontology. That is, we depict the system that create and handle the collection of objects that is the user searching in web page can get annotate and find the searching methodology[7]. There are many approaches that process ontology according to semantic Web based upon the characteristic of ontology data on Web. We create the web ontology by using the language RDF (resource description framework) and OWL (Web ontology language). By creating the ontology model so as to use hierarchical vocabularies, we eliminate redundancy of an expression and reusability can be improved. Furthermore, by means of this model will help to reduce the users search time and exact information will be retrieved based on the extraction of the information provided by the user in the webpage.

KEYWORDS: Annotation, Web Ontology Language, Resource Description Framework, Semantic, Web Ontology.

I. INTRODUCTION

A number of databases can be accessible through HTML form-based search interfaces. The data units are extracted from the database by both statically and dynamically for human browsing. There are two techniques in the concept of annotation they are mainly Annotation-Meta data, which are extracted and assign meaning full labels and also annotation wrapper which is another extracting technique [7]. It is possible to be processed in very high data functionalities in the web pages. In the base content it has the functionality, in which the tools are used to annotate. In our process, we done based on the functionality in which the tags are getting annotated by means of metadata functionality. The various annotations are performed based on the user search.

Although the use of ontology is not proposed as a substitute for database technology, a database is still more powerful than ontology for storing large-scale data sets. However, ontology can be used with a database to provide a conceptual vision of heterogeneous data sources distributed in a number of databases with an interface built on an ontological model. Thus, we need a system that utilizes both database and ontology techniques. However, while databases are widely available, the corresponding ontology's are not. Furthermore, constructing ontology from scratch is tedious, Time-consuming, error-prone and labour-intensive, while building one by hand presents the same difficulties [7]. The proposed solution therefore starts by transforming a given database to ontology with some rules as guidelines, which can be used for manual transformation or as the basis for an automatic transformation process.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

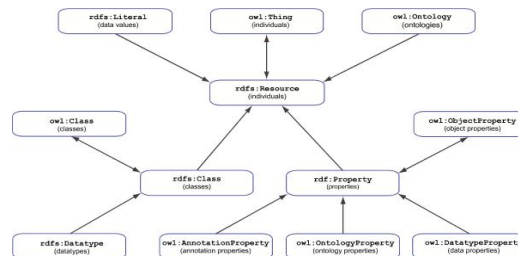


Fig 1 OWL File Representation

Based on the representation of the OWL file the category are separated up and forms the semantic annotation and its meanings.

II. LITERATURE SURVEY

A. Anno Search: Image Auto-Annotation by Search

In this survey paper [4], we use Anno Search to extract image using data mining technologies. To search the images we have two main steps:

- 1) Searching the similar image or visually similar image on the web page,
- 2) Mining metadata from web.

To search the image we need at least one keyword to find the similar image, after retrieve the same image using content-based searching. And also to search the image by using the descriptions likes the URLs, titles and text. In this it increase the efficiency and dimensional of the image are mapped using hash code to speed up the content-based searching. The result shows the real web images are effectiveness and efficiency of this already proposed algorithm.

Also, the image is searching for similar and visually similar images using web mining from the description. It contains three main steps:

- 1) Text-based searching is performed to retrieve similar images.
- 2) Then content-based searching is performed to retrieve visually similar images.
- 3) Finally, the images are clustered according to the description. A hash code algorithm is used to speed up the content-based searching. On 2.4 million experiments are taken to prove the proposed approach is effectiveness and efficient. In the future, we have to work for a large number of databases and to handle the problem to annotate the image without using the keyword.

B. A Translation Approach to Portable Ontology Specifications

A definition of classes, relation, functions, and other objects are represented for a specified domain is called as ontology [5]. This translation approach to ontology specification describes for defining ontology's for a portable system. For predicate calculus are translated in a system called ontolingua, to retaining the computational implementations to share and reuse ontology's. Several technical problems are discussed to translate this approach for portability addresses. First technique is that to describe how to preserve the declarative content. Another is that to translate very expensive language to a restricted language. By using this technique it preserving the computational efficiency of that implemented system. In this paper it describes how the problems are addressed by ontolingua.

The replacement is not for representing system like Algernon or Loom using the ontolingua. Several implementations of representation systems are motivated to trying the ontology's are portable, rather than trying with one systems, those systems provides different computational and different costs. Ontolingua does not support



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

query processing. It is incomplete and inherently with KIF language. It is portable in only a single system, writing definitions are supported in all the systems. The set of most common idioms are supported in a target system may defined in frame ontology. KIF expressions are only translated by using the ontolingua. It does not support some user defined second-order relations it only defined in the frame ontology. Ontolingua does not translate target information, it gives some information issues. Ontolingua is a translation tool and it is a domain-independent. Frame ontology and set theory in KIF called as representation ontology's.

C. A Brief Survey of Web Data Extraction Tools

Data extraction from web pages are addressed the problem from several years. In a traditional database the data can be handled in many ways [10]. Natural language processing, languages and grammars, machine learning, information retrieval, database and ontology's use these technologies. This web data extraction tools are used to extract the data using the web tools. By using the keyword searching is more efficient than browsing. But browsing is not a suit for searching the data.

In this paper it extracts the data using wrappers from web database. The main goal is to perform wrapper development process by using these extraction tools which is used to write the code using the languages like Perl and java. It supports most of the features for generation of wrappers and the data extraction process. Quantitative analyses are performed to show performance of extraction tools. And moreover some of the tools are not used for the extraction of data, some tools are also provides the obtained results.

D. Ontology Based Context Modelling and Reasoning using OWL

For supporting logic based context reasoning and computing environment we use this CONON (context ontology) modelling method. It shows a hierarchical manner for an upper ontology about basic context and for adding the domain-specific ontology [11]. We have the context information about to reason over low-level, derive explicit content to implicit context. We have rapid advance technologies in computing for recent years. Pervasive computing is context-awareness is an important step for and widely environment. Context information completely gives the information like in-surmountable task. The main of this context based modelling includes set of upper level entities, and add the some concepts for different applications. We separate our context model into two main types upper ontology and specific ontology. The upper ontology is high level ontology which contains some general features of contextual entities. Specific ontology describes the set of ontology and their features in sub-domain. In this survey paper ontology based context modelling is supporting for context modelling and reasoning computing environments. CONON and context reasoning schemes are implemented. It is feasible while taken for a performance study in computing environments.

E. Towards the Self-Annotating Web

This paper describes the PANKOW (pattern-based annotation through knowledge on the web) technologies [12]. This annotation was done in manual of two human subjects. Onto Mat approach is implemented for semantic web to show the best results. This technology does not require the lab manual definition. These results provide high efficiency to annotate the web. PANKOW contains many steps to annotate web, it gives high efficiency, effectiveness and range can be proved.

Google helps the web server API. Self-annotating web is not possible. It shows large number of queries against the Google API. Machine learning technologies are used to increase the weights of direction of search. In addition, it is used to reduce the amount of queries in the Google server API. It was more intelligence into account the ontological hierarchy.

F. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment

Content for ontology's were developed in many domain areas. Ontology plays an vital role in world wide web and they provide semantics for annotation in web pages. The large numbers of ontology were led to

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

develop the distributed nature of ontology [13]. Ontology to be reused, first it need to merged or align. This merged and align can be done in manually. It uses the algorithm, PROMPT to provide a semi-automatic to merging and alignment. PROMPT performs some tasks automatically and guide the user to perform some tasks. It is a highly general knowledge model in all the platforms. But in this paper strategies and algorithms described as OKBC-compliant knowledge model. so result is applicable for knowledge representation for ontology.

III. EXISTING SYSTEM

In the current scenario they have used annotator tool of six different types. Those tools will automatically assigns the labels to the data units within the SRRs returned from WDBs. It does automatic annotation on the data set that **first aligns** the data units and its attributes. It also uses annotation wrapper to extract the content. It formats the main process as like a clustering-based shifting technique to align data units into different groups and also employ those six basic annotators by which each annotator can independently assign labels to data units, based on certain features of the data units. But this model is highly flexible.

IV. PROPOSED SYSTEM

The proposed method used ontology instead of annotation tool which sets a rule accordingly, for to perform the concept of annotation. The extraction of tags by means metadata. The annotation is done by means of application. Here we can use many web pages or a web database for annotation. Multiple pages can be annotated. While extracting the tag we can easily identify the content of the web. Ontology can set an efficient rule. Compare to the tool annotation this will be efficient. The overall functional diagram shows below.

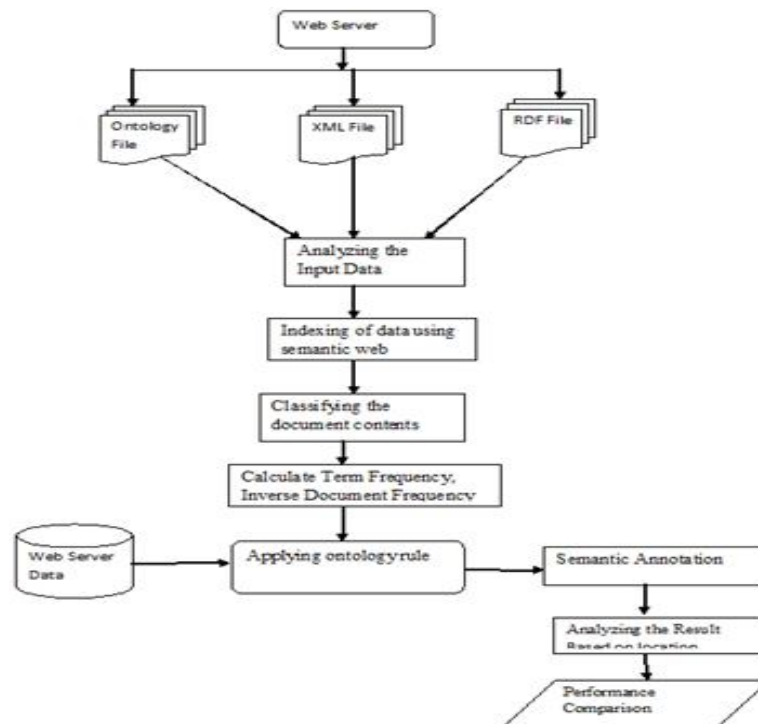


Fig: Architecture Diagram



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

The XML file or an ontology file and RDF file can be included for analyse the annotation process and to find the semantic meaning by the formation on ontology rule.

V. EXPERIMENTAL RESULT

There are four modules; the overall process includes that, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first. The user can search the content that will show the results in a web page. User can search any type of content that he wants just like Google search. The Searched content just displayed with the related web links. Just click on the link it goes to that related website. The searched contents are not aligned or processed in ordinary search engines. They just fetch the links related to our search but in this module we can customize our search by manipulating data units and text nodes. Depending upon our selection it will process and fetch the content for our wishes. In this module, admin are having authentication and security to access the detail which is presented in the ontology system. Once admin enter with proper validation, he can upload the web contents and also web links for the different categories and also he can update it.

VI. CONCLUSION

Thus when searching any content in a search engine by means of the web page, it will group the content into different category which are related to what we are searching about. It also provides data unit level annotation which means order or group the content which belongs to our wish by means of indexing in which the annotation will done for both the levels of the page. This can be done by various levels in the web page in which the metadata of the page is getting extracted and thus the result produce will be effective for the best web page content development. This concept will be used to make good decision as per the user search beyond the other web pages. In future we are going to develop this with the sentiment analysis.

REFERENCES

1. Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng, Member, IEEE, and Clement Yu, Senior Member, IEEE "Annotating Search Results from Web Databases", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013.
2. T.Gruber, A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, June 1993, pp.199-220.
3. B.Swartout, R.Patil, K.Knight and T. Russ, "Toward Distributed Use of Large-Scale Ontologies", Ontological Engineering AAAI-97 Spring Symposium Series, 1997, pp. 138-148.
4. Xin-Jing Wang, Lei Zhang, Feng Jing, Wei-Ying Ma "AnnoSearch: Image Auto-Annotation" Microsoft Research Asia, 49 Zhichun Road, Beijing (100080), China.
5. Thomas R. Gruber, "A Translation Approach to Portable Ontology Specifications", KNOWLEDGE SYSTEMS LABORATORY
6. Computer Science Department Stanford University Stanford, California, 1993.
7. N.Noy and D.McGuinness, "Ontology Development 101: A guide to creating your first ontology", Report, Stanford University, Stanford, CA, 94305.
8. M.Li, X.Du and S.Wang, "Learning ontology from Relational Database", in Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, August 2005, 18-21.
9. S.Benslimane, D.Benslimane and M.Malki, "Acquiring OWL Ontologies from Data-Intensive Web Sites", USA.ACM, Palo Alto, California, July 11-14, 2006.
10. R.Shanmugapriya and L.Jaganraj, "A Study on Ontology Query Result Based On Semantic Web", Vol. 2, Issue 10, October 2013, IJARCCCE, ISSN (Online): 2278-1021, Page No: 3841-3844.
11. Alberto H. F. Laender Berthier A. RibeiroNeto Altigran S. da Silva_ Juliana S. Teixeira, "A Brief Survey of Web Data Extraction Tools"
12. Department of Computer Science Federal University of Minas Gerais 31270901.
13. Xiao Hang Wang, Tao Gu1, Da Qing Zhang, Hung Keng Pung, "Ontology Based Context Modeling and Reasoning using OWL" Institute for Infocomm Research, School of Computing, National University of Singapore, Singapore 119260.
14. Philipp Cimiano, Siegfried Handschuh, Steffen Staab, "Towards the SelfAnnotating Web" Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany.



ISSN(Online): 2320-9801

ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, December 2014

15. Natalya Fridman Noy and Mark A. Musen ,”PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment “ Stanford Medical Informatics, Stanford University, Stanford, CA 94305-5479.

BIOGRAPHY

S.Mohana Prakash is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College, Coimbatore, under Anna University. Area of Interest is Web Content Mining and Networking.

S.Sathish Kumar is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College, Coimbatore, under Anna University. Area of Interest is Web mining,

R.Vaishnavi is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College of, Coimbatore, under Anna University. Area of Interest is Web Mining.

E.Anjaline Suganya is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College, Coimbatore, under Anna University. Area of Interest is Data mining.

M.Ramakrishnan completed his PG in the Dept of Computer Science and Engineering, and currently working as the assistant professor in Sree Sakthi Engineering College, Coimbatore, under Anna University. He has guided many UG and PG student in his experience. His Area of Interest is Networking, Data Mining, Cloud Computing, Image Processing, Network Security, Compiler Design, Operating System, and Web Content Mining.