

# ENHANCED SEMANTIC SEARCH MODEL FOR WEB INFORMATION GATHERING BY USING PERSONALIZED ONTOLOGY

Mr. ASMAMAW AREGA SHASHE <sup>#1</sup> , Dr. K.VENKATA RAO <sup>#2</sup>

<sup>#1</sup> M.Sc Scholar, Master of Computer Science,  
College of Engineering, Andhra University, Visakhapatnam, AP, India.

<sup>#2</sup> Professor, Department of Computer Science and Systems Engineering,  
College of Engineering, Andhra University, Visakhapatnam, AP, India.

## ABSTRACT

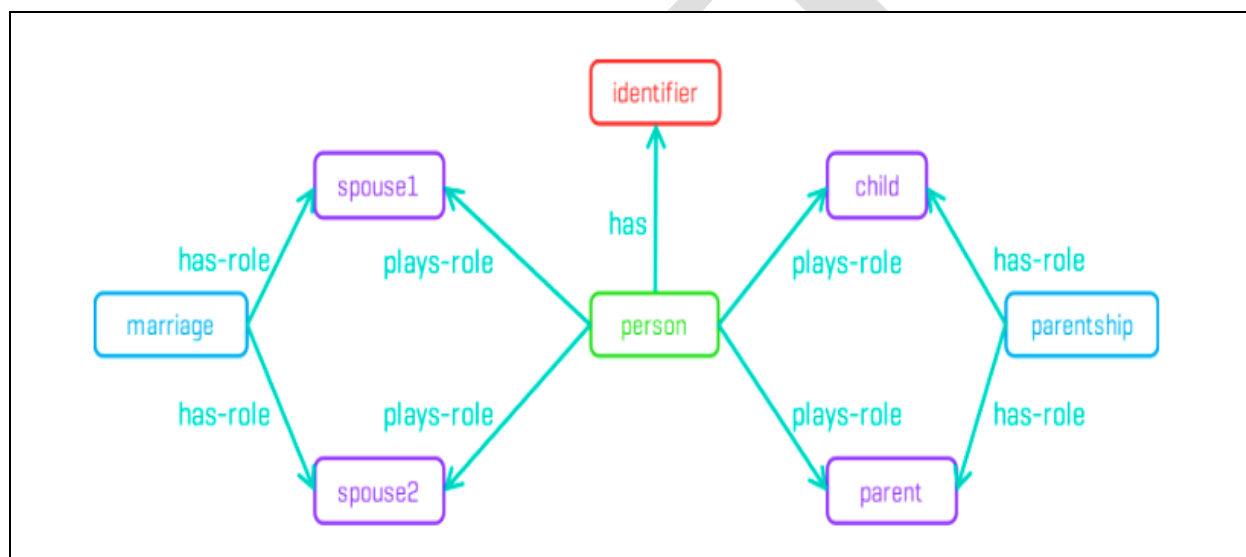
Ontology is the process of extracting the valuable information from a large data source. As we all know that www is an interconnection of several billion of documents formatted by using html. Due to growing and dynamic nature of the web, it has become a challenge to traverse all URLs in the web documents and to handle these URLs. Ontologies are widely used to represent user profiles in personalized web information gathering. However, when representing user profiles, many models have utilized only knowledge from either a global knowledge base or user local information. In this paper, a personalized ontology model is proposed for knowledge representation and reasoning over user profiles. This model learns ontological user profiles from both a world knowledge base and user local instance repositories. By conducting various experiments on our proposed model we finally came to a conclusion that personalized ontology model is best suited for extracting the exact information from web rather than similar data from the web.

**Key Words:** Personalized Ontology, World Wide Web, Data Extraction, Data Repository, Global Knowledge Repository, Local Information Repository.

## I. INTRODUCTION

As we all know that world wide web (www) has rapidly increased its users from the past decades. In the past decades the information available on world wide web has exploded rapidly with a great range of topics and different categories. One thing that remained as a major challenge is how to collect the required information from the web. There are many search engines that are available now in order to extract the useful information for the given search keyword. And almost each and every individual search engines return more than 1,600 results per user query, where only thirty to forty percent of links are relevant or related to the user search query and remaining all are somewhat ir-relevant information that will be matched and displayed.

All the user profiles which are created are in structured manner and they are almost viewed by the un-authorized users by watching over a user's shoulder while he/she is surfing. The profiles are shown to converge and to reflect the actual interests quite well. Web user profiles are widely used by web information systems for user modeling and personalization. User profiles reflect the interests of users [1]. User profiles are used in Web information gathering to capture user information needs in order to get personalized web information for users. When acquiring user profiles, the content, life cycle and applications are taken into consideration since user interests are approximate and unambiguous it is suggested it can be represented by ontologies [2]. On the last decades, the amount of web-based information available has increased dramatically. How to gather useful information from the web has become a challenging issue for users. Current web information gathering systems attempt to satisfy user requirements by capturing their information needs. For this purpose, user profiles are created for user background knowledge description.



**FIGURE 1.REPRESENTS THE BASIC ONTOLOGY MODEL FOR BUILDING AN ONTOLOGY**

From the figure 1, we clearly find out the construction of a basic ontology [3]. From the above figure we can find out how to start designing an ontology. For building an ontology we will be using a genealogy graph for mapping out the family tree of a human race. Here we can find out the complete ontology and the dataset for creating the ontology and the rules and procedures that accompany for creating the ontology. From the above figure the first step for building an ontology is identifying the entities that are available in the graph. So we can clearly tell that identifying the roles of person is the basic idea of our ontology where the person can have distinct roles like marriage, spouse, child, parent and parentship. Once the entities are identified then we will try to identify the roles for each and every entity. He has the following various relations and for each and every attribute it has various roles like has-role and plays-role attributes.

In the literature in order to extract the existing global knowledge bases for user background knowledge representation most commonly we try to use the generic ontologies (e.g.,

WordNet database [4]), thesauruses (e.g., digital libraries), and online knowledge bases (e.g., online categorizations and Wikipedia). In the primitive days in order to extract the useful information we try to use the WordNet for capturing the user interest for some areas and they are treated as useless for other areas. In this thesis, we proposed an ontology model to evaluate this current hypothesis. This model mainly simulates users' concept models by using personalized ontologies and attempts to improve web information gathering performance by using ontological user profiles. The world knowledge base (WKB) and a user's local instance repository (LIR) are used in the proposed model. The WKB is nothing but collecting the information or search query related data from the live data base like google, and for this we need to have internet connection for extracting any data related to user search query keyword. From that WKB we try to construct the personalized ontologies. Here the Local Information Repository (LIR) is mainly used for extracting the user background knowledge i.e. area of interest of that user who got registered for extraction of data. Here the proposed ontology model is evaluated by comparison against various benchmark models. The evaluation results show that the proposed ontology model is successful in design of personalized web information gathering systems.

## II. RELATED WORK

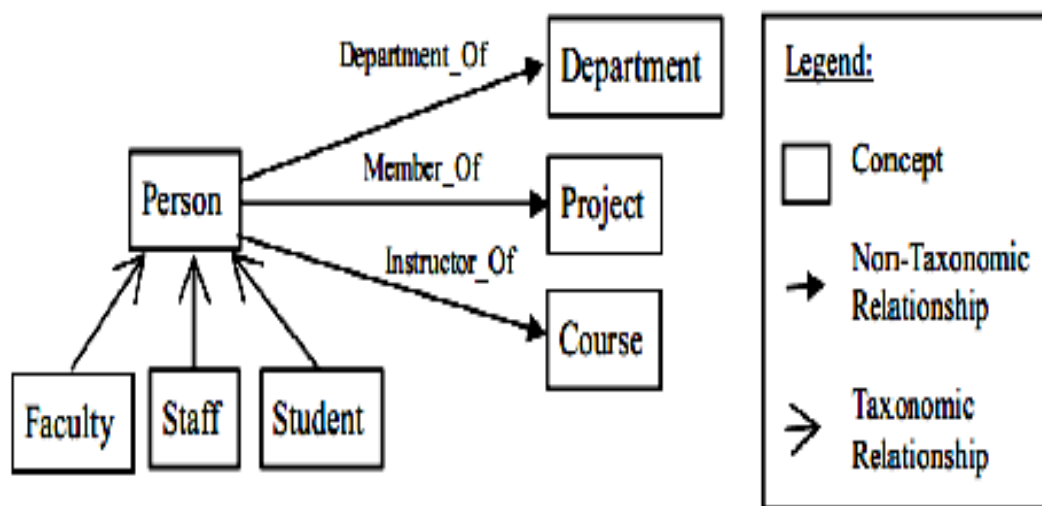
In this section we mainly discuss about the related work that was carried out in identifying the ontology learning environment. Now let us discuss about the ontology learning environment in detail as follows

### ONTOLOGY LEARNING

As we all know that GIR is mainly used for extracting the knowledge from web. They are many models that are available in literature for web information gathering. If we take an author like Gauch [7] and another author like Sieg [6], they both learned personalized ontologies from the Open Directory Project (ODP) to specify the user's preference in web search. These works effectively discovered user background knowledge; however, their performance was limited by the quality of the global knowledge bases. Also they were many learning models that were used for mining user background knowledge from user local information. In some rare cases, ontologies are defined to also include instances of concepts and relationships [7]. For this purpose, it is defined as ontology to be a set of concepts  $C$  and relationships  $R$ . The relationships in  $R$  can be either taxonomic or non-taxonomic. For example, Fig.2 depicts a simple University ontology consisting of a set of concepts

From the below figure 2, we can find out the example of a person in a university. The person is mainly classified into three types in that university like Faculty, Staff and Student.

$C_{univ} = \{\text{Person, Faculty, Staff, Student, Department, Project, Course}\}$ , and



**FIGURE 2.REPRESENTS THE SAMPLE ARCHITECTURE OF AN ONTOLOGY MODEL**

A set of relationships

```
Runiv={Department_Of(Person,Department),
Member_Of(Person,Project),
Instructor_Of(Course,Person),
Superclass_Of(Faculty,Person),
Superclass_Of(Staff,Person),
Superclass_Of(Student, Person)}.
```

Superclass\_Of represents the taxonomic relationship while the rest are not. With this definition, the instances of ontology refer to the instances of its concepts and relationships. If each concept instance exists in the form of a Web page, a relationship instance will then exist in the form of a Web page pair. This view has been adopted in most the Web classification research. In practical terms, developing ontology includes:

- Defining classes in the ontology,
- Arranging the classes in a taxonomic (subclass–super class) hierarchy,
- Defining slots and describing allowed values for these slots,
- Filling in the values for slots for instances. Then create a knowledge base by defining individual instances of these classes filling in specific slot value information and additional slot restrictions.

Then create a knowledge base by defining individual in-stances of these classes filling in specific slot value information and additional slot restrictions.

### **III. PROCEDURE FOR NOVEL PERSONALIZED ONTOLOGY CREATION**

In this section we will mainly discuss about the proposed personalized ontology creation for extracting the useful or related data for the given area of interest. Now let us discuss about the personalized ontology creation in detail.

#### **MAIN MOTIVATION**

Personalized ontology's are a conceptualization model that formally describes and specifies user background knowledge. From observations in daily life, we found that web users might have different expectations for the same search query. For example, for the domain like "Health" a person may demand different information of various medical aids. Sometimes even the same user may have different expectations for the same search query if applied in a different situation. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. Therefore, domain wise searching of urls is suggested.

#### **WORLD KNOWLEDGE REPRESENTATION**

World Knowledge Representation is the process of extracting the valuable information from a large data source. World knowledge is mainly composed with knowledge possessed by people and acquired through their experience and education. Also, the term "world knowledge" is nothing but used for lexical and referential analysis. As we know that in our proposed application the WKR is nothing but user query keyword is send to the google server and in turn it will try to extract the related information from that WKB[8]-[10].

#### **ONTOLOGY CREATION**

General the user interactions are mainly extracted from web via user interests. For extracting the user interaction, we need to develop an ontology model to assist the users. Generally all the interactions are extracted in the form of URLs ,where the URLs are of two types: Positive URL and other is negative URL. Thus for a given topic we have both positive URLs and negative URLs of equal size. So when a request request a search topic, he will get both set of positive and negative URLs. We mainly concentrate on crawling or finding only the positive and related URLs rather than all ir-related URLs. Hence such a type of extraction is most related for ontology creation.

A pattern is a character string. All keywords can be written in both the upper and lower cases. It is used to extract hidden information from not-structured or semi-structured data. This aspect is fundamental because much of the web information is semi-structured due to the nested structure of HTML code, much of the web information is linked, and much of the web information is redundant. It should not include images, tags, and buttons. The extracted content should be stored in some file[11]. But it should not include any HTML tags. The constructed

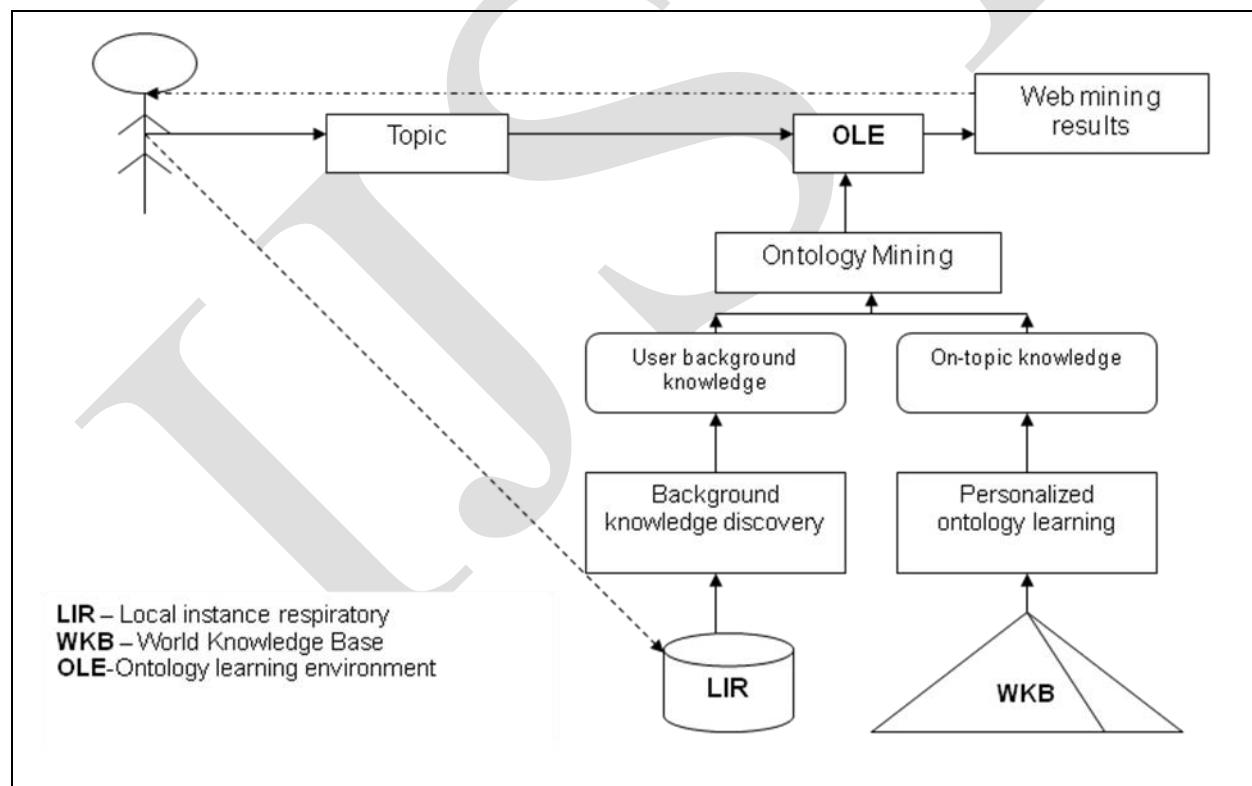
ontology is personalized because the user selects positive subjects for personal preferences and interests as by selecting do-main names. This model is developed for four domains as–

1. General
2. Health
3. Education
4. Entertainment

It also allows entering 4 URL addresses at a time which pro-vides parallel processing for finding relative URL's.It also avoids time delay since providing parallel processing of input. It also counts every time how many URLs are searched at once, type of protocol, Hash code, web page content ,time of download. It also maintains local database which is used when user is offline and world knowledge base is searched when user is online.

#### IV. SYSTEM ARCHITECTURE PHASE

In this section we will mainly discuss about the proposed personalized ontology model and its architecture. Here we will discuss about the proposed architecture in detail as follows:



**FIGURE 3. REPRESENTS THE PROPOSED PERSONALIZED ONTOLOGY MODEL**

From the figure 3, we can clearly get an idea that our proposed personalized ontology model has two sources for extracting the information. One is LIR (Local Information Repository) and other is WKB (World Knowledge Base).From the above figure we can clearly find out that a

user gives a topic or keyword as a search query and then the query enters into OLE (Ontology Learning Environment) for extracting the related data. Now the OLE will undergo the process of ontology mining by the two sources like LIR and WKB. From the LIR the user's AOI or area of interest is been extracted and from the WKB the search input query is extracted from the google. Once the query data which is extracted from WKB directly come to the buffer reader where the main logic is performed at the buffer area. In the buffer area all the query related links will be extracted and they will be filtered with the matched user interest field which was specified by the users at the time of registration. Now both the fields will be undergo the process of filtering and only the links which match will both LIR and WKB are extracted and all other links will be discarded at the buffer level. This is the way how each and every user individual area of interest is taken into consideration for extracting the valuable information from the web.

## **V.CONCLUSION**

We finally implemented a personalized ontology model for representing user background knowledge (I.e. individual area of interest of user's) for personalized web information gathering. The model constructs user personalized ontologies by extracting world knowledge and discovering user background knowledge from user local instance repositories. A multidimensional ontology mining method, exhaustively and specificity, is also introduced for user background knowledge discovery. As an extension we also included the concept of rating chart for the proposed application which will clearly show the number of users who were interested in extracting the appropriate information based on their individual area of interest. This rating chart will tell the performance of our proposed application with respect to their individual area of interest.

## **VI. REFERENCES**

- [1] Trajkova and S. Gauch, "Improving Ontology-Based User Profiles," Proc. Conf. Recherche d'Information Assistee par Ordinateur (RIAO '04), pp. 380-389, and 2004.
- [2] N. Zhong, "Representation and Construction of Ontologies for Web Intelligence," Int'l J. Foundation of Computer Science, vol. 13, no. 4, pp. 555-570.
- [3] <https://grakn.ai/pages/documentation/building-an-ontology/basic-ontology.html#identifying-entity-types>
- [4] G. M. Voorhees and Y. Hou, "Vector Expansion in a Large Collection," Proc. First Text Retrieval Conf., pp. 343-351.
- [5] A. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07), pp. 525-534, 2007.
- [6] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 554-568, Apr. 2006.

- [7] T. Tran, P. Cimiano, S. Rudolph, and R. Studer, „Ontology-Based Interpretation of Keywords for Semantic Search,“ Proc. Sixth Int’l Semantic Web and Second Asian Semantic Web Conf. (ISWC ’07/ASWC ’07), pp. 523-536, 2007.
- [8] S. Gauch, J. Chaffee, and A. Pretschner,“ Ontology-Based Personalized Search and Browsing, Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.
- [9] L.A. Zadeh, “Web Intelligence and World Knowledge—The Concept of Web IQ (WIQ),” Proc. IEEE Ann. Meeting of the North American Fuzzy Information Soc. (NAFIPS ’04), vol. 1, pp. 1-3, 2004.
- [10] R R.Navigli, P.Velardi, and A.Gangemi, “Ontology Learning and Its Application to Automated Terminology Translation, IEEE Intelligent Systems, vol. 18, no. 1, pp. 22-31, Jan./Feb. 2003.
- [11] J.D. King, Y. Li, X. Tao, and R. Nayak, „Mining World Knowledge for Analysis of Search Engine Content,“ Web Intelligence and Agent Sys-tems, vol.5, no.3, pp.233-253, 2007.

## VII .ABOUT THE AUTHORS



**Mr.ASMAMAW AREGA SHASHE** is currently pursuing his 2 Years M.SC in Computer Science at College of Engineering,Andhra University,Visakhapatnam,AP,India. He has more than 10 years of teaching experience in MADDA WALABU UNIVERSITY.His area of interests includes Networking and Web Mining.



**Dr. K.VENKATA RAO** is a Honorary director of A.U Computer Centre and a Professor, in Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam, AP, India. He has more than 23 years of experience in teaching field. His research interest includes Image Processing, Security and Big data analyst.