International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-9)

Research Article | **September 2015**

Global Knowledge Information Gathering For Hierarical Ontologies

Ravi Sankar G*, Rajasekhar Reddy V, Arunbabu P
Department of IT, RGMCET
India

Abstract—

The nature of collaborative learning involves intensive interactions among collaborators, such as articulating knowledge into written, verbal or symbolic forms, authoring articles or posting messages to this community's discussion forum, responding or adding comments to messages or articles posted by others, etc. Knowledge collaborators' capabilities to provide knowledge and the motivation to collaborate in the learning process influence the quantity and quality of the knowledge to flow into the virtual learning community. In this paper, we have developed an ontology enabled annotation and knowledge management to provide semantic web services from three perspectives, personalized annotation, real-time discussion, and semantic content retrieval. Personalized annotation is used to equip the collaborators with Web based authoring tools for commenting, knowledge articulation and exertion by extracting metadata from both the annotated content and the annotation itself, and establishing an ontological relation between them. The real time discussion is used as a bridge to link collaborators and knowledge and motivate collaborators for knowledge sharing by building profiles for collaborators and knowledge (in the forms of content and annotation) during every discussion session, and establishing ontological relation between the collaborators and knowledge for the use of semantic content retrieval. The semantic content retrieval then utilizes the ontological relations constructed from the personalized annotation and real-time discussion for finding more relevant collaborators and knowledge.

Keywords— virtual learning , knowledge articulation , exertion, collaborators, extracting

I. INTRODUCTION

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 User profiles represent the concept models possessed by users when gathering web information. A concept model is implicitly possessed by users and is generated from their background knowledge. While this concept model cannot be proven in laboratories, many web ontologists have observed it in user behavior . When users read through a document, they can easily determine whether or not it is of their interest or relevance to them, a judgment that arises from their implicit concept models. If a user's concept model can be simulated, then a superior representation of user profiles can be built. To simulate user concept models, ontology a knowledge description and formalization model—are utilized in personalized web information gathering. Such ontologies are called ontological user profiles or personalized ontologies .

To represent user profiles, many researchers have attempted to discover user background knowledge through global or local analysis Global analysis uses existing global knowledge bases for user background knowledge representation. Commonly used knowledge bases include generic ontologies (e.g., WordNet), thesauruses (e.g., digital libraries), and online knowledge bases (e.g., online categorizations and Wikipedia). The global analysis techniques produce effective performance for user background knowledge extraction. However, global analysis is limited by the quality of the used knowledge base. For example, WordNet was reported as helpful in capturing user interest in some areas but useless for others. Local analysis investigates user local information or observes user behavior in user profiles. For example, Li and Zhong discovered taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups learned personalized ontologies adaptively from user's browsing history. Alternatively, Sekine and Suzuki analyzed query logs to discover user background knowledge. In some works, such as, users were provided with a set of documents and asked for relevance feedback. User background knowledge was then discovered from this feedback for user profiles. However, because local analysis techniques rely on data mining or classification techniques for knowledge discovery, occasionally the discovered results contain noisy and uncertain information. As a result, local analysis suffers from in effectiveness at capturing formal user knowledge.

We can hypothesize that user background knowledge can be better discovered and represented if we can integrate global and local analysis within a hybrid model. The knowledge formalized in a global knowledge base will constrain the background knowledge discovery from the user local information. Such a personalized ontology model should produce a superior representation of user profiles for web information gathering.

© 2015, IJERMT All Rights Reserved

1.1 Purpose

An ontology model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontologies, and attempts to improve web information gathering performance by using ontological user profiles. The world knowledge and a user's local instance repository (LIR) are used in the proposed model. World knowledge is commonsense knowledge acquired by people from experience and education; an LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontologies by adopting user feedback on interesting knowledge. A multidimensional ontology mining method, Specificity and Exhaustivity, is also introduced in the proposed model for analyzing concepts specified in ontologies. The users' LIRs are then used to discover background knowledge and to populate the personalized ontologies. The proposed ontology model is evaluated by comparison against some benchmark models through experiments using a large standard data set. The evaluation results show that the proposed ontology model is successful.

1.2 Scope

The research contributes to knowledge engineering, and has the potential to improve the design of personalized web information gathering systems. The contributions are original and increasingly significant, considering the rapid explosion of web information and the growing accessibility of online documents.

1.3 Motivation

This paper presents the extensive work of, but significantly beyond, an earlier paper published in WI '07. The authors thank the Library of Congress and QUT Library for the use of the LCSH and library catalogs. The authors also thank the anonymous reviewers for their valuable comments. Thanks also go to M. Carey-Smith, P. Delaney, and J. Beale, for their assistance in proofreading and editing the paper.

1.3.1 Definitions

The proposed ontology model was evaluated by objective experiments. Because it is difficult to compare two sets of knowledge in different representations, the principal design of the evaluation was to compare the effectiveness of an information gathering system (IGS) that used different sets of user background knowledge for information gathering. The knowledge discovered by the ontology model was first used for a run of information gathering, and then the knowledge manually specified by users was used for another run. The latter run set up a benchmark for the evaluation because the knowledge was manually specified by users. Under the same experimental conditions, if the IGS could achieve the same (or similar) performance in two different runs, we could prove that the discovered knowledge has the same quality as the user specified knowledge. The proposed ontology model could then be proven promising to the domain of web information gathering.

1.4 Overview

An ontology model is proposed for representing user background knowledge for personalized web information gathering. The model constructs user personalized ontologies by extracting world knowledge from the LCSH system and discovering user background knowledge from user local instance repositories. A multidimensional ontology mining method, exhaustivity and specificity, is also introduced for user background knowledge discovery. In evaluation, the standard topics and a large test bed were used for experiments. The model was compared against benchmark models by applying it to a common system for information gathering. The experiment results demonstrate that our proposed model is promising. A sensitivity analysis was also conducted for the ontology model. In this investigation, we found that the combination of global and local knowledge works better than using any one of them. In addition, the ontology model using knowledge with both is-a and part-of semantic relations works better than using only one of them. When using only global knowledge, these two kinds of relations have the same contributions to the performance of the ontology model. While using both global and local knowledge, the knowledge with part-of relations is more important than that with is-a.

The proposed ontology model in this paper provides a solution to emphasizing global and local knowledge in a single computational model. The findings in this paper can be applied to the design of web information gathering systems. The model also has extensive contributions to the

fields of Information Retrieval, web Intelligence, Recommendation Systems, and Information Systems. we will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

II. FRAMEWORK

Based on existing ontology mining approaches and our previous work for mining ERP patterns, we summarize and propose the following four general procedures for mining the concepts and their relationships in domain ontologies:

© 2015, IJERMT All Rights Reserved

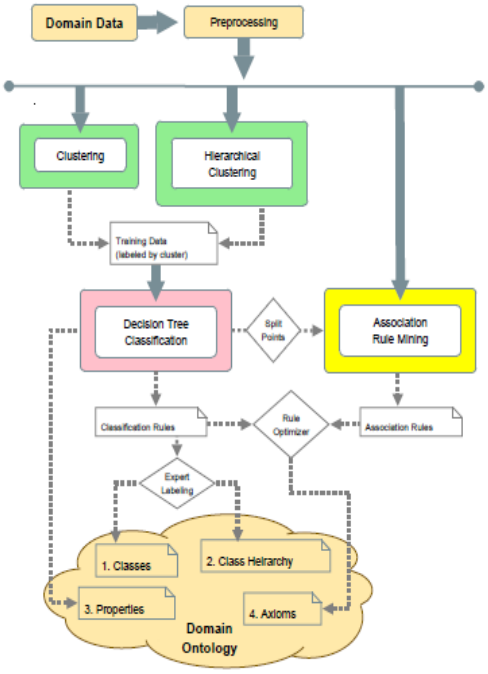


Figure 2: A semi-automatic framework for mining domain ontologies.

1. Classes (Clustering-based Classification: If there exists n clusters in existing domain dataset D based on some clustering algorithm, we first define n candidate classes for the domain and assign them arbitrary name, such as "C1," "C2" for each class. After the data instances of each cluster are labeled by the as- signed class name, each class (i.e., each cluster) will be formally defined by classification rules. The arbitrary class names may be updated to more meaningful ones by domain experts based on their understanding of classification rules. 2. Class Taxonomy (Hierarchical Clustering: More granular classes and their taxonomy (hierarchy) will be de- termined by a hierarchial clustering algorithm. Again, we can assign arbitrary names for each new class for the clustering-based classification process. Domain experts may give more meaningful names for new classes based on classification rules. 3. Properties (Classification: The classification process for defining rules for classes will also be used to deter- mine candidate properties between different classes or between classes and data types. 4. Axioms (Association Mining and Classification: The association rules between different properties will be used for defining the axioms (rules) between proper- ties. And the classification rules will also be defined as the axioms (rules) among different classes and properties. The interaction of classification and association rule mining will be used for rule optimization . All of the above four procedures and their interactions are shown in Figure 2 and the outputs (i.e., classes, class hierarchy, properties, axioms) are put together into a domain ontology. It is a semi-automatic framework because we need "expert labeling" to give meaningful names for classes. The input data are put into some semi-structured formats, such as the spreadsheet, after data preprocessing. Otherwise, some statistical or text processing step needs to be done as a part of data preprocessing. To further explain why our ontology mining framework based on the four general procedures makes sense, we first suppose there exists a domain ontology (i.e., semantics of data) for a set of data instances in some specific domain(e.g., ERP). Our goal is to find what classes, properties and axioms can be mined to compose that domain ontology. From a machine learning point of view, the domain ontology is the target function to be learned and it includes several components such as classes, class hierarchy, properties and axioms. A reasonable assumption is that the data instances which belong to the same class must be similar by sharing some properties, the data instances which belong to different classes must be dissimilar. Therefore, determining what and how many classes should be included in an ontology is typically a clustering problem. It is a natural extension that finding the hierarchy of classes (clusters) is a hierarchial clustering problem. On the other hand, what properties and values the data instances in the same class should share is a typical classification problem. The selection of attributes for classification (e.g., information gain selection) can be used for property selection in ontology mining. The classification rules can also be treated as the relationships (axioms) of properties and classes. The association rules between different properties can be treated as relationships (axioms) of two or multiple properties themselves, which will be a good complementary for the ontology.

2.1 Related Work:

The following section presents related work on ontologies and personalization. Since we create our user profiles automatically using text classification techniques, we will also review research in this area

Classification

Classification is one approach to handling large volumes of data. It attempts to organize information by classifying documents into the best matching concept(s) from a predefined set of concepts. Several methods for text classification

© 2015, IJERMT All Rights Reserved

have been developed, each with a different approach for comparing the new documents to the reference set. These include: comparisons between a variety of frequently-used vector representations of the documents Machines use of the joint probabilities of 4 the words being in the same document (Naive Bayesian); decision trees; and neural networks. A thorough survey and comparison of such methods is presented. Classification has been applied to newsgroup articles, Web pages, and other online documents. The system described in classifies NETNEWS articles into the best matching news groups. The implementation uses the vector space model to compare new articles to those articles manually associated with each news group. The system presented in is based on a probabilistic description-oriented representation of Web pages, and a probabilistic interpretation of the *k*-nearest neighbor classifier. It takes into account: 1) Features specific to Web pages (e.g., a term appears in a title, a term is highlighted), 2) Features standard to text documents, such as the term frequency. The *k*-nearest neighbor approach has also been used by in a system that uses classification techniques to automatically grade essays.

Ontologies

One increasingly popular way to structure information is through the use of ontologies, or graphs of concepts. One such system is *Onto Seek* , which is designed for content-based information retrieval from online yellow pages and product catalogs. Onto Seek uses simple conceptual graphs to represent queries and resource descriptions. The system uses the *Sensus* ontology, which comprises a simple taxonomic structure of approximately 70,000 nodes. The system presented in uses *Yahoo!* as an ontology. The system semantically annotates Web pages via the use of Yahoo! categories as descriptors of their content. The system uses *Telltale* as Its classifier. Telltale computes the similarity between documents using *n*-grams as index terms. The ontologies used in the above examples use simple structured links between concepts. A richer and more powerful representation is provided by *SHOE* is a set of Simple HTML Ontology Extensions that allow WWW authors to annotate their pages with semantic content expressed in terms of ontology. SHOE provides the ability to define ontologies, create new ontologies which extend existing ontologies, and classify entities under an "is a" classification scheme.

2.2 Technology Used:

About Java Platform

The Java platform consists of the Java application programming interfaces (APIs) and the Java virtual machine (JVM).

The following Java technology lets developers, designers, and business partners develop and deliver a consistent user experience, with one environment for applications on mobile and embedded devices. Java meshes the power of a rich stack with the ability to deliver customized experiences across such devices.

Java APIs are libraries of compiled code that you can use in your programs. They let you add ready-made and customizable functionality to save you programming time.

Java programs are run (or interpreted) by another program called the Java Virtual Machine. Rather than running directly on the native operating system, the program is interpreted by the Java VM for the native operating system. This means that any computer system with the Java VM installed can run Java programs regardless of the computer system on which the applications were originally developed.

In the Java programming language, all source code is first written in plain text files ending with the .java extension. Those source files are then compiled into .class files by the javac compiler. A .class file does not contain code that is native to your processor; it instead contains bytecodes — the machine language of the Java Virtual Machine (Java VM). The java launcher tool then runs your application with an instance of the Java Virtual Machine.

Because the Java VM is available on many different operating systems, the same .class files are capable of running on Microsoft Windows, the Solaris TM Operating System (Solaris OS), Linux, or Mac OS.

III. SYSTEM ANALYSIS

The **Systems Development Life Cycle (SDLC)**, or *Software Development Life Cycle* in systems engineering, information systems and software engineering, is the process of creating or altering systems, and the models and methodologies that people use to develop these systems.

In software engineering the SDLC concept underpins many kinds of software development methodologies. These methodologies form the framework for planning and controlling the creation of an information system the software development process.

© 2015, IJERMT All Rights Reserved

3.1 Software Model Or Architecture Analysis:

Structured project management techniques (such as an SDLC) enhance management's control over projects by dividing complex tasks into manageable sections. A software life cycle model is either a descriptive or prescriptive characterization of how software is or should be developed. But none of the SDLC models discuss the key issues like Change management, Incident management and Release management processes within the SDLC process, but, it is addressed in the overall project management. In the proposed hypothetical model, the concept of user-developer interaction in the conventional SDLC model has been converted into a three dimensional model which comprises of the user, owner and the developer. In the proposed hypothetical model, the concept of user-developer interaction in the conventional SDLC model has been converted into a three dimensional model which comprises of the user, owner and the developer. The ―one size fits all‖ approach to applying SDLC methodologies is no longer appropriate. We have made an attempt to address the above mentioned defects by using a new hypothetical model for SDLC described elsewhere. The drawback of addressing these management processes under the overall project management is missing of key technical issues pertaining to software development process that is, these issues are talked in the project management at the surface level but not at the ground level.

3.2 What is SDLC?

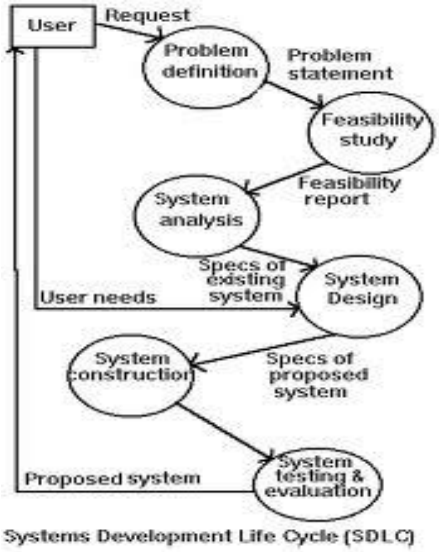
A software cycle deals with various parts and phases from planning to testing and deploying software. All these activities are carried out in different ways, as per the needs. Each way is known as a Software Development Lifecycle Model(SDLC)[2].A software life cycle model is either a descriptive or prescriptive characterization of how software is or should be developed. A descriptive model describes the history of how a particular software system was developed. Descriptive models may be used as the basis for understanding and improving software development processes or for building empirically grounded prescriptive models.

SDLC models * **The Linear model (Waterfall)** - Separate and distinct phases of specification and development. - All activities in linear fashion. - Next phase starts only when first one is complete. * **Evolutionary development** - Specification and development are interleaved (Spiral, incremental, prototype based, Rapid Application development). - Incremental Model (Waterfall in iteration), - RAD(Rapid Application Development) - Focus is on developing quality product in less time, - **Spiral Model** - We start from smaller module and keeps on building it like a spiral. It is also called Component based development. * **Formal systems development** - A mathematical system model is formally transformed to an implementation. * **Agile Methods.** - Inducing flexibility into development. * **Reuse-based development** - The system is assembled from existing components.

The General Model

Software life cycle models describe phases of the software cycle and the order in which those phases are executed. There are tons of models, and many companies adopt their own, but all have very similar patterns. The general, basic model is shown below:

General Life Cycle Model



Systems Development Life Cycle (SDLC)

Each phase produces deliverables required by the next phase in the life cycle. Requirements are translated into design. Code is produced during implementation that is driven by the design. Testing verifies the deliverable of the implementation phase against requirements.

3.3 Existing System

Local analysis investigates user local information or observes user behavior in user profiles. For example, Li and Zhong discovered taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups learned personalized ontologies adaptively from user's **browsing history**. Alternatively, Sekine and Suzuki

© 2015, IJERMT All Rights Reserved

analyzed query logs to discover user background knowledge. In some works, such as , users were provided with a set of documents and asked for **relevance feedback**. User background knowledge was then discovered from this feedback for user profiles. However, because local analysis techniques rely on data mining or classification techniques for knowledge discovery, occasionally the discovered results contain **noisy and uncertain information**. As a result, local analysis suffers from **in-effectiveness** at capturing formal user knowledge.

3.3.1 Drawbacks

We will investigate the methods that generate user local instance repositories to match the
representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, in Section 4.2, strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

3.4 Proposed System

We can hypothesize that user background knowledge can be better discovered and represented if we can integrate global and local analysis within a hybrid model. The knowledge formalized in a global knowledge base will constrain the background knowledge discovery from the user local information. Such a personalized ontology model should produce a superior representation of user profiles for web information gathering.

In this paper, an ontology model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontologies and attempts to improve web information gathering performance by using ontological user profiles. The world knowledge and a **user's local instance repository** (LIR) are used in the proposed model. World knowledge is commonsense knowledge acquired by people from experience and education; an LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontologies by adopting user **feedback on interesting knowledge**. A multidimensional ontology mining method, Specificity and Exhaustivity, is also introduced in the proposed model for analyzing concepts specified in ontologies. The users' LIRs are then used to **discover background knowledge** and to populate the personalized ontologies. The proposed ontology model is evaluated by comparison against some **benchmark models** through experiments using a **large standard data set**. The evaluation results show that the proposed **ontology model is successful**.

3.2.1 Advantages

Which one is more important: the WKB or LIRs? The Loc model using only user LIRs had substantially low performance, compared with the GP, GI, and GIP models using only the WKB. Thus, The WKB is more important than user LIRs. In addition, the GP, GI, and GIP models using the WKB also have the knowledge with is-a and/or part-of semantic relations. The Loc model, however, has no such relations specified. Hence, it is reasonable to conclude that a part of
the improvement achieved by the GP, GI, and GIP models is due to the is-a and/or part-of knowledge. We then have an extensive finding: the knowledge with is-a and/or part of relations is an important component of the ontology model.

3.3 Algorithm: Analyzing semantic relations for specificity:

Terms Expansion:
Tax: Taxonomic structure can be provide the edge communication
Rel: can be provides the Boolean operations
Aim:
1. It can be identifies the specificity of information like leaves identification process
2. Using leaves maintain the two types of relationship operation like is-a and part-of.
3. In between of two types of relationship to maintain the union operation.
4. All the subjects to related objects to arrange in the form of tree format of structure.

IV. CONCLUSION AND FUTURE ENHANCEMENTS

4.1Conclusion

An ontology model is proposed for representing user background knowledge for personalized web information gathering. The model constructs user personalized ontologies by extracting world knowledge from the LCSH system and discovering user background knowledge from user local instance repositories. A multidimensional ontology mining method, exhaustivity and specificity, is also introduced for user background knowledge discovery. In evaluation, the standard topics and a large test bed were used for experiments. The model was compared against benchmark models by applying it to a common system for information gathering. The experiment results demonstrate that our proposed model is promising. A sensitivity analysis was also conducted for the ontology model. In this investigation, we found that the combination of global and local knowledge works better than using any one of them. In addition, the ontology model using knowledge with both is-a and part-of semantic relations works better than using only one of them. When using only global knowledge, these two kinds of relations have the same contributions to the performance of the ontology model. While using both global and local knowledge, the knowledge with part-of relations is more important than that with is-a.

© 2015, IJERMT All Rights Reserved

4.2 Future Enhancements

The proposed ontology model in this paper provides a solution to emphasizing global and local knowledge in a single computational model. The findings in this paper can be applied to the design of web information gathering systems. The model also has extensive contributions to the
fields of Information Retrieval, web Intelligence, Recommendation Systems, and Information Systems. we will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley, 1999.
[2] .E.P. Box, J.S. Hunter, and W.G. Hunter, Statistics For Experimenters. John Wiley & Sons, 2005
[3] C. Buckley and E.M. Voorhees, "Evaluating Evaluation Measure Stability," Proc. ACM SIGIR '00, pp. 33-40, 2000
[4] Z. Cai, D.S. McNamara, M. Louwerse, X. Hu, M. Rowe, and A.C. Graesser, "NLS: A Non-Latent Similarity Algorithm," Proc. 26th Ann. Meeting of the Cognitive Science Soc. (CogSci '04), pp. 180-185, 2004.
[5] L.M. Chan, Library of Congress Subject Headings: Principle and Application. Libraries Unlimited, 2005.
[6] .A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," Proc. ACM SIGIR ('07), pp. 7-14, 2007.
[7] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web," Proc. 11th Int'l Conf. World Wide Web (WWW '02), pp. 662-673, 2002.

© 2015, IJERMT All Rights Reserved