Asian Journal of Information Technology 12 (11-12): 376-386, 2013
ISSN: 1682-3915
© Medwell Journals, 2013

Ontology Extraction and Merging for a Better Semantic Web Integration

[1]V. Rajeswari and [2]Dharmishtan K. Varughese
[1]Faculty of IT, [2]Faculty of ECE, Karpagam College of Engineering, Tamil Nadu, India

Abstract: The internet is a huge ocean of knowledge, embedded in trillions of web pages that are linked and entwined and presents itself as the World Wide Web for anyone to use its resources. The computer used to browse through this maze, usually serve only to deliver and present the content of documents describing the knowledge. People have to communicate with all the sources of relevant information and interpret them by themselves. Semantic web is an effort to enhance current web so that, computers can process the information presented on the internet, interpret and communicate with it to help humans to find required essential knowledge. The aim of this research is to explore the applicability of ontology in simplifying semantic web searches through conventionally adopted different data processing techniques. Researchers present methods for representing data and knowledge in such a way that machines are directed to understand the meaning. One of the basic problems in the development of techniques for the semantic web is the integration of ontologies. The web is constituted by a variety of information sources which are expressed over certain ontology and in order to extract information from such sources their semantic integration and reconciliation in terms of a global ontology is required. In this study, researchers address the fundamental problem of how to specify the mapping between the global and local ontologies.

Key words: Ontology, ontology extraction, RDF, OWL, merging databases, heterogeneous data, linked data, semantic web, data space, data integration, data management

INTRODUCTION

The World Wide Web contains a very large number of repositories of information accumulated over a number of years. They exist in various forms and formats not adhering to any uniform standard or schema, since standardisation mechanisms were non existent at that time. This makes retrieval of the required information for an application a great challenge since computer applications understand only the web page structure and layout and have no access to their intended meaning. The evolving semantic web technologies address precisely this issue and enable users get information from the web by querying these heterogeneous data sources.

The semantic web (Allemang and Hendler, 2005) is used to enhance the existing web with a layer of machine-interpretable metadata. The RDF (Resource Description Framework) data model (Ceballos and Brena, 2005) was proposed for modeling web objects as part of developing the semantic web. It has been used in various applications. For example, Yahoo and DBPedia extract facts from Wikipedia automatically and store them in RDF format to support structural queries over Wikipedia. Basically these principles require the identification of entities with URI references that can be resolved over the HTTP protocol into RDF data (Wang *et al*., 2004) that describes the identified entity. The semantic web's current focus is on ontology (Borgida and Serafini, 2002). Presently, ontology forms the top research topic in various areas such as information integration, knowledge engineering, co-operative information systems and natural language processing.

In system integration, ontology plays an important role, mainly concerned with providing a set of mechanisms for resolving the semantic heterogeneity problems, resolving the queries, hiding the complexity of accessing data from different data sources and describing the contents of all data source as concepts in a global ontology.

ROAD MAP OF THE SEMANTIC WEB

Overview of the semantic web: The World Wide Web consortium initiates the semantic web. The W3C is an international organization which sets standards for the technologies which are used in the World Wide Web. The WWW shares the information infrastructure between people and organizations in the world. The leading research team of Tim Berners Lee seeks to maintain the interoperability and universality of the web. It can be achieved by setting the open standards which are used for efficient information retrieval by the web tools.

Corresponding Author: V. Rajeswari, Faculty of IT, Karpagam College of Engineering, Tamil Nadu, India

The semantic web (Allemang and Hendler, 2005) initiative was started as the web Metadata Working Group in 1998. But subsequently it has become the semantic web activity. It is organized with the view that the semantic web "provides a common framework (Matthews, 2002) that allows data to be shared and reused across application, enterprise and community boundaries". It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework. It integrates a variety of applications using XML. The following groups were framed as a result of the movement in the Semantic work:

- Resource description framework model and syntax specification
- Resource description framework schema specification

The DAML Programme was organized by a DARPA-sponsored initiative of the US. It proposed several influential approaches to the problems which are posed by the semantic web. The contribution of the W3C on behalf of the semantic web has grown indefinitely. Two major working groups of the W3C such as the RDF Core Working Group and the Web Ontology Working Group have produced major sets of recommendations.

The two groups (Correndo *et al*., 2010) such as the Semantic Web Best Practices and Deployment Working Group seek to support and extend the practical application of the semantic web to a number of fields by providing sample tools and general descriptive vocabularies in key areas. The RDF Data Access Working Group is developing languages for querying and processing semantic annotations across the web. Semantic web projects are processed in US, UK and in the rest of the world. The total investment in the semantic web world-wide has been in the tens of millions of pounds. Organizations such as Hewlett-Packard and British Telecom are investing in research programmes of semantic web. Active and recent projects of Semantic Computing Research Group (SeCo) are:

- Finnish National Ontologies on the Semantic Web (FinnONTO)
- Semantic Uniquitous Services (SUBI)
- Linked Data Finland (LDF)
- Linked Open Aalto
- Event-centric Multimedia Content Access Platform

SEMANTIC WEB MODEL

Machine understandable information: The World Wide Web was designed as a space for the wide informative knowledge. The aim of WWW is that it should be used for human to human communication (Allemang and Hendler, 2005). Now the proposed new technique allows the machines to participate in the communication.

Most of the information on the web is designed for human consumption (Allemang and Hendler, 2005) even if it is extracted from a database with well defined meanings. The semantic web is a web of data. It is also called as global database. The semantic web will bring structure (Matthews, 2002) to the meaningful content of web pages. It creates an environment where Software agents roaming from page to page can readily carry out sophisticate tasks for users. Once the web is provided with a mechanism for defining semantics about resources and links then the intelligent agents automatically process the web rather than the mediation by the people.

The basic assertion model: The general model of the semantic web is the Resource Description Framework. This basic model contains the concept of an "assertion" and "quotation". This is used because most of the RDF applications (Ceballos and Brena, 2005) are for data about data Assertions of that resource are implicit parameter and are known as a property of a resource. Most of the applications which uses metadata can be handled by RDF. Some of the examples of the metadata are card index information (the Dublin Core), Privacy information (P3P) and associations of style sheets with documents.

The schema layer: RDF gives a model of assertions and quotations on which the data can be mapped to any new format. The schema layer is needed to declare the existence of new property. These meta-assertions make it possible to do rudimentary checks on a document. As in SGML, the DTD allows to check whether elements have been used in appropriate positions.

The schema language typically makes simple assertions about permitted combinations. If SGML's DTD is used as a model, the schema can be in a language of very limited power. The constraints expressed in the schema language are easily expanded into more powerful logical layer expressions.

The logical layer: The layer next to Schema layer is the logical layer. The logics are embedded into documents to allow the properties such as:

- Set of rules (Eiter *et al*., 2003) to extract of one type of document from a document of another type
- The checking of a document against a set of rules of self-consistency
- The resolution of a query by conversion from unknown terms into known terms

SEMANTIC WEB ARCHITECTURE

The semantic web architecture is conceived in the form shown in Fig. 1. The different blocks are assigned with respective conceptual functioning.

Unicode and URI: Unicode (Allemang and Hendler, 2005) is a computing industry standard for consistent encoding, representation and handling of text expressed in most of the world's writing systems. A Uniform Resource Identifier (URI) is a string of characters used to identifies a name or a web resource. Such identification enables interaction with representations of the web resource over the World Wide Web using specific protocols. Schemes specify a concrete syntax. The URI (Cruz and Xiao, 2006) is defined by the associated protocols. The semantic web is generally built on syntax which use URIs to represent data. Many triples of URI data are held in databases or interchanged on the World Wide Web.

RDF and RDFSCHEMA: Resource Description Framework (RDF) (Welty, 2003) (Jimenez-Ruiz *et al*., 2008) is the W3C standard for encoding knowledge. The properties of the RDF are as follows:

- A general metadata format used to represent information about internet resources
- Extends the expressive capability of the web
- Augments human-readable web pages with machine-processable information

The RDF Schema permits the RDF vocabulary (Bhatt *et al*., 2004a, b) to describe taxonomies of classes and properties. It also provides definitions for some of the elements of RDF. It sets the domain and range of properties and relates the RDF classes and properties into taxonomies using the RDFs vocabulary.

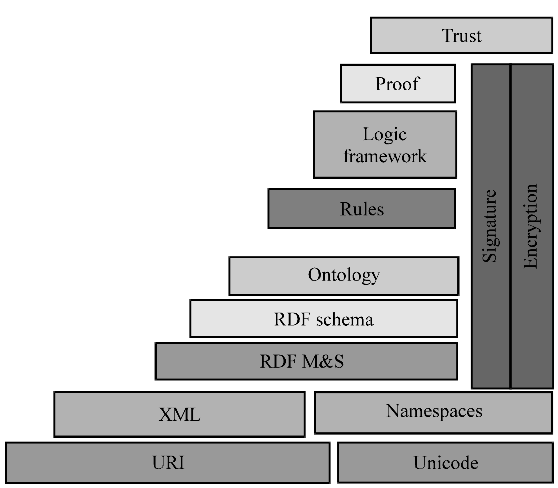

Fig. 1: Semantic web architecture

Ontology: The best known definition of Ontology according to Gruber is ontology is an explicit specification of a conceptualization (Calvanese *et al*., 2004). In this concept, a conceptualization indicates an abstract model of some aspect of the world. An explicit specification means that the model should be specified in some unambiguous language making it amenable for processing by machines as well as by humans. Ontologies are of increasing importance (Reynolds *et al*., 2005) in fields such as knowledge management, information integration, co-operative information systems, information retrieval and electronic commerce.

The application area which has recently seen an explosion of interest is the semantic web where ontology is set to play a key role in establishing a common terminology between agents thus ensuring that different agents have a shared understanding of terms using in semantic markup. The effective use of ontologies requires not only a well-designed and well-defined ontology language but also support from reasoning tools. Formally an ontology O is a symbol system (Cruz and Xiao, 2006) consisting of:

- A set SC of concepts and a set S_R of binary relations specifying pairs (D, R) of domains and ranges (in S_C)
- A set of ontology axioms include introduction of concepts and of relations

Languages for ontology:

- XML is a language for describing documents
- RDF and RDFS are languages (Welty, 2003) for describing the organization of resources on the Web
- SKOS is a the Simple Knowledge Organization System
- SPARQL is an RDF query language (Le *et al*., 2011)
- N3 (Notations3) is designed with human readability in mind
- N-Triples is a format for storing and transmitting data
- OWL is a The Web Ontology Language, a family of knowledge representation languages

MANAGEMENT OF DATA AND META DATA

The components of the data and metadata management are indicated in Fig. 2 and are described subsequently.

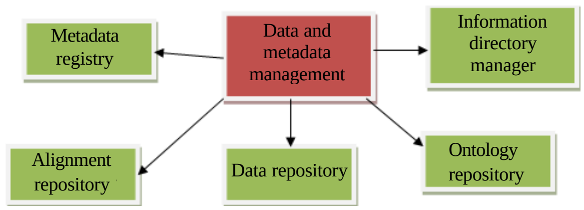

Fig. 2: Data and metadata management

The information directory manager component: This provides functionalities to handle query distribution. It also manages a content provider directory to identify information providers from a query. It also handles the storage and access the data of the distributed ontologies.

The ontology repository components: The component accesses the locally stored ontologies and ontology instances.

The data repository component: This component accesses the locally stored ontologies and annotated ontologies.

The alignment repository component: This component provides the way to locally stored ontologies and access alignments.

The metadata registry component: This access metadata information.

Querying and reasoning: This component generates the query and processes the query.

The query answering component: This component deals with issues related with the logical processing of a query.

The semantic query processor component: This processor takes care of all issues related with the physical processing of a query by providing functionalities to manage query answering over the distributed ontologies.

The semantic query editor component: This is an important component because it takes care of all the issues related with the user interface.

SEMANTIC WEB SERVICES

This semantic web services (Calvanese *et al*., 2004) shown in Fig. 3, include the components that discover, select, mediate, compose, choreograph, ground and profile semantic web services.

The web service discoverer component: It provides the following functionalities:

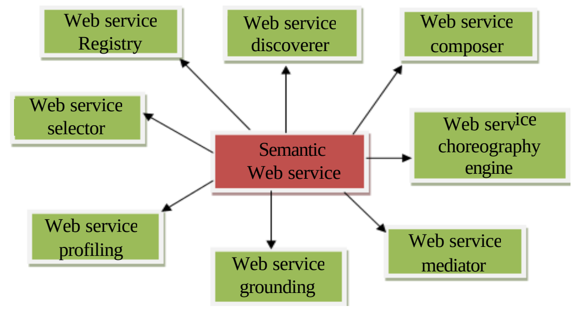

Fig. 3: Semantic web services

- Publish and search service registries
- To control access to registries
- Distribute requests to other registries

The web service selector component: This component checks whether the services can actually fulfill the user's concrete goal.

The web service composer component: This component takes care of the automatic composition of the web services in order to provide new value-added web services.

The web service choreography engine component: This component uses choreography descriptions of service requester and provider to drive the conversation between them.

The web service process mediator component: This component (Eiter *et al*., 2003) provides functionalities to reconcile the public process heterogeneity that can appear during the invocation of web services.

The web service grounding component: This component is responsible for the communication between web services.

The web service profiling component: This component is used to service profiles based on their execution history.

The web service registry component: This component provides functionalities to register semantic web services.

SEMANTIC WEB LANGUAGES

DAML+OIL (Cruz and Xiao, 2006) is an ontology language which is a very important component of the semantic web. This language is used in many applications for storing the semantic pages. DAML+OIL provides modeling primitives commonly found in frame-based languages. DAML+OIL describes the structure of the domain (Ceballos and Brena, 2005) and resource descriptor framework is used to describe the specific instances. Structure is described in terms of classes (concepts) and properties (roles). Ontology in DAML+OIL consists of set of axioms. The classes may be names or expressions. The constructors are used for building class expressions:

- Expressive power of DAML+OIL is determined by
    - Kinds of axiom supported
    - Kinds of class (and property) constructor
    - Supported
- DAML+OIL basic properties
    - DAML+OIL layered on top of RDFS
    - RDFS based syntax
    - Inherits RDFS ontological primitives (subclass, range, domain)
    - Provides much richer set of primitives (equality, cardinality)
- Overview of class expressions
    - DAML+OIL designed to describe structure of domain (schema) Object oriented classes (concepts) and properties (roles)
    - DAML+OIL ontology consists of set of axioms asserting characteristics of classes and properties
        - Ex-Person is kind of animal whose parents are persons
    - RDF (Jimenez-Ruiz *et al*., 2008) used for class/property membership assertions (data)
        - Ex-Jack is an instance of Person <Jack
        - Rosy>is an instance of parent (Table 1)
- DAML+OIL Overview: Axioms (Table 2)

Table 1: DAML+OIL class expressions

Constructors	DL syntax	Example
intersectionOf	$C_1 \sqcap ... \sqcap C_{12}$	Human$\sqcap$Male
uinionOf	$C_1 \sqcup ... \sqcup C_{12}$	Doctor$\sqcup$Lawyer
complementOf	$\neg C$	$\neg$Male
oneOf	$\{\varkappa_1 ... \varkappa_n\}$	{John, mary}
toClass	$\forall P.C$	$\forall$hasChild.Doctor
hasClass	$\exists P.C$	$\exists$hasChild.Lawyer
hasValue	$\exists P. \{\varkappa\}$	$\exists$citizenOf. {USA}
minCardinalityQ	$\geq n$ P.C	≥ 2 hasChild.Lawyer
maxCardinalityQ	$\leq n$ P.C	$= 1$ hasChild.Male
cardinalityQ	$= n$ P.C	$= 1$ hasParent.Female

Table 2: DAML+OIL axioms

Axioms	DL syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human$\sqsubseteq$Animal$\sqcap$Biped
sameClassAs	$C_1 \equiv C_2$	Man$\equiv$Human$\sqcap$Male
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter$\sqsubseteq$hasChild
samePropertyAs	$P_1 \equiv P_2$	Cost$\equiv$price
sameIndividualAs	$\{\varkappa_1\} \equiv \{\varkappa_2\}$	{President_Bush}$\equiv${W_Bush}
disJoinWith	$C_1 \sqsubseteq \neg C_2$	Male$\sqsubseteq\neg$Female
differentIndividualFrom	$\{\varkappa_1\} \sqsubseteq \neg \{\varkappa_2\}$	{John}$\sqsubseteq\neg${peter}
inverseOf	$P_1 \equiv P^{-}{}_2$	hasChild$\equiv$hasParent^{-}
transitiveProperty	$P^{+} \sqsubseteq P$	Ancestor$^{+}\sqsubseteq$Ancestor
uniqueProperty	$T \sqsubseteq \leq 1P$	$T \sqsubseteq \leq 1$hasMother
unambigousProperty	$T \sqsubseteq \leq 1P^{-}$	$T \sqsubseteq \leq 1$isMotherOf^{-}

ONTOLOGY EXTRACTION AND MERGING

For the past few years, information on the World Wide Web was mainly intended for direct human consumption. However, to facilitate new intelligent applications such as meaning-based search and information brokering, the semantics of the data on the internet should be accessible for machines. Therefore, methods and tools to create such a semantic web have generated wide interest.

Ontology which has been a field of philosophy since Aristotle has become a buzz-word in information and knowledge-based systems research (Guarino and Welty, 2000). Various publications in knowledge engineering, natural language processing, cooperative information systems, intelligent information integration and knowledge management report about the application of ontologies in developing and using systems. In general, ontologies provide a shared and common understanding of a domain that can be communicated between people, heterogeneous and distributed application systems. They have been developed in artificial intelligence to facilitate knowledge sharing and reuse.

Ontology creators attempt to model certain domains accurately and completely however this often leads to large-scale ontologies. For example, the Unified Medical Language System ontology has >800,000 concepts and 9,000,000 relationships. Large-scale ontologies are hard to maintain and use. An application may only need a small part of a large-scale ontology (Borgida and Serafini, 2002). Using the whole ontology will greatly increase complexity and redundancy and reduce efficiency. Extracted from the original ontology, a smaller and simpler sub-ontology can make the application more efficient (Wouters *et al*., 2005). The sub-ontology only contains the particular parts of the whole ontology required by the application; the application do not benefit from other outlying information. Early researches study generating application-focused databases from large ontologies. Sub-ontology extraction is a new research area. Some researchers point out sequential extraction process called Materialized Ontology View Extraction (Wouters *et al*., 2002, 2005; Reynolds *et al*., 2005). Under application requirements, MOVE support optimization schemes to guarantee sub-ontology with high quality. However, it proves to be computationally expensive. Subsequent researchers proposed a distributed approach (Bhatt *et al*., 2004b) to the subontology extraction process to decrease cost of sub-ontology extraction from large complex ontology. They also analyze the semantic completeness issue (Bhatt *et al*., 2004a, b) in their method. GoPubmed System (Delfs *et al*., 2004) presents the relevant sub-ontology for browsing GeneOntology it shows the extraction of sub-ontology is profitable. Current methods only focus on extracting sub-ontology from single ontology. However, in many practical cases, especially in the web environments, researchers often face multiple ontologies in one application. Some of the related ontologies or the integration of them may be too large. It faces the same problem of complexity and efficiency. Using sub-ontology can solve the problem. Nevertheless there are no existing methods to extracting sub-ontology from multiple ontologies. The methods for single ontology are not able to solve many new problems with multiple ontologies.

Extracting from multiple ontology: Extraction of ontology for a given domain from different sources of similar ontologies is a challenge that researchers face frequently. Though methods exist for standardising ontlogies as of now, temporal and spatial factors do influence the semantics used for buildine the ontologies. It may be possible in the near future to overcome this difficulty to a maximum extent with extensive research.

Need for Unified Ontology Model: There are many different ontology languages on the web such as OWL (Correndo *et al.*, 2010), DAML-OIL and Ontolingua. They are different in syntax and structure and based on different logic foundations. Therefore, the analyzing, extracting and integrating methods of them are different. Translating ontologies into a unified internal representation i.e., a unified ontology model is necessary. In the idea, researchers first translate different ontology language code into a unified model and then do the extraction and integration based on this model. Therefore, both the extraction and integration process can ignore the different language issues. The result ontology should be able to translate to the language as user demands. The unified model must involve most of the features in typical Ontology languages in order to represent ontologies as exactly as possible. It also has to be simple because researchers need to do many extraction and integration research on it. Especially, to make the extraction easier, the elements in ontology need to be organized in a modularized model. The model should also be visualized for browsing ontologies. Users can browse through ontology in order to understand its scope, structure and content and rapidly search it for terms of interest and related terms.

Extraction and integration: There are mainly two approaches to extract sub-ontology from multiple ontologies:

- Integrate all the original ontologies and then extract the demanded sub-ontology from the integrated one
- Extract sub-ontology from each ontology separately and then integrate them into the demanded one

The former can use current methods in single ontology. However, there are two disadvantages: the difficulty in ontology integration and the complexity in ontology extraction. The ontology integration problem (Keet, 2004) is one of the most difficult problems to be solved on the semantic web, especially integrating large ontologies. Current methods of extracting sub-ontology prove to be computationally expensive (Allemang and Hendler, 2005). Even dealing with a single large ontology, many optimization schemes have to be applied.

The latter wipes off the outlying information first so, the sub-ontologies to be integrated are much smaller than original ones. It is better than the former approach at least in aspect of efficiency. The research uses this approach. However, there are also two problems in this approach: how to divide requirements in order to guide extraction in ontology and how to get the “outlying” information required in integration process. Researchers discuss about them in the ensuing sections. When extracting sub-ontology from single ontology, user requirements can directly conduct the extraction. Nevertheless, it is hard to decide requirements to extracting sub-ontology from a certain ontology in multiple ontologies. The key problem is how to divide the requirements or even whether to divide them.

EXTRACTION OF SUBONTOLOGY

The subontologies should be as complete as possible. The redundant information in sub-ontologies may help the integration process. An opposite idea is to make sub-ontologies as small as possible. Requirements are divided into many small parts; each part is to guide extracting sub-ontology from certain ontology. Notice that it may extract several sub-ontolgies from an ontology. The various stages in this process are shown in Fig. 4-8. The sub-ontologies are independent of each other but there was information that helped to integrate them. However, researchers chose the latter method that of choosing smaller ontologies in the research. Comparing to the benefit of redundant information, the disadvantages are more remarkable. It does many redundant researchs in extraction and needs to remove all the redundant information in integration. As both extraction and integration are difficult and computationally expensive, the costs are often unacceptable. Moreover, the redundant information rarely contains all the information

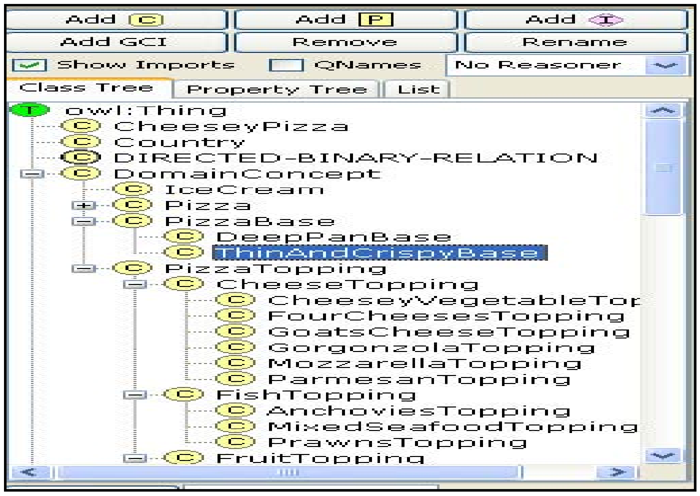

Fig. 4: Extracting the Class ThinAndCrispyBase for SubOntology

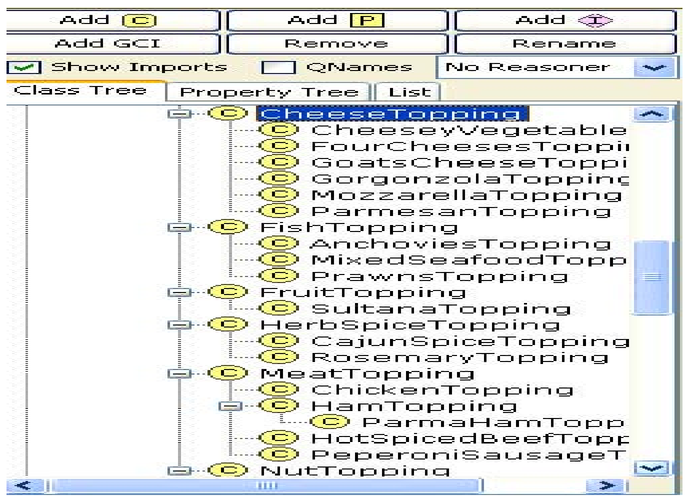

Fig. 5: OWL class for ThinAndCrispyBase

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#">
  <!ENTITY pizza "http://www.co-ode.org/ontologies/pizza/pizza.owl">
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
]>
<rdf:RDF xml:base="&pizza;"
         xmlns:owl="&owl;"
         xmlns:rdf="&rdf;"
         xmlns:rdfs="&rdfs;">
  <owl:Class rdf:about="#CheeseTopping">
    <rdfs:label xml:lang="pt">CoberturaDeQueijo</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PizzaTopping"/>
    </rdfs:subClassOf>
    <owl:disjointWith>
      <owl:Class rdf:about="#FishTopping"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:about="#FruitTopping"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:about="#HerbSpiceTopping"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:about="#MeatTopping"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:about="#NutTopping"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:about="#SauceTopping"/>
```

Fig. 6: Extracting the class CheeseToppingFor SubOntology

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#">
  <!ENTITY pizza "http://www.co-ode.org/ontologies/pizza/2005/05/16/pizza.owl">
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
]>
<rdf:RDF xml:base="&pizza;"
         xmlns:owl="&owl;"
         xmlns:rdf="&rdf;"
         xmlns:rdfs="&rdfs;">
  <owl:Class rdf:about="#ThinAndCrispyBase">
    <rdfs:label xml:lang="pt">BaseFinaEQuebradica</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PizzaBase"/>
    </rdfs:subClassOf>
    <owl:disjointWith>
      <owl:Class rdf:about="#DeepPanBase"/>
    </owl:disjointWith>
  </owl:Class>

  <owl:AnnotationProperty rdf:about="&rdfs;label"/>
</rdf:RDF>
```

Fig. 7: OWL class for CheeseTopping

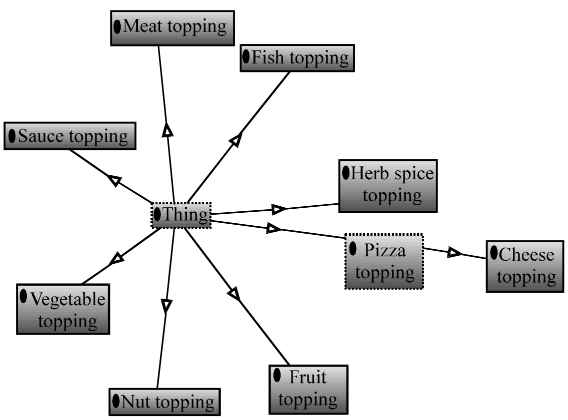

Fig. 8: Graphical representation of extracted ontology

needed in integration. A better way is to separate it from sub-ontologies. Researchers use bridge ontology to maintain the information used in integration more conveniently and completely.

MERGING OF ONTOLOGIES

Bridge ontology: The ontology integration is a laborious work and may have not effective solutions as ontologies created by integrating are weak during ontology evolution. Distributed description logic (Borgida and Serafini, 2002) is one of method for tackling this problem and has a simple bridge rule to express the concept subsumption relations between ontologies. Bridge ontology (Delfs *et al*., 2004) can describe more refined relations. The bridge ontology is a peculiar ontology and has the ability of expressing the complex relations between multiple ontologies. It can be created and maintained conveniently and is effective in the applications based on multiple ontologies. It has the advantages of low-cost, scalable, robust in the web circumstance, avoiding the unnecessary ontology extending and integration and promoting ontology reuse. The bridge information can provide information needed in the integration process.

Adopted framework: The framework contains four processes as indicated in Fig. 9 and 10, to extract sub-ontology form multiple ontologies.

- Convert all the ontologies into a unified Ontology model
- Divide the requirements into sub-requirements
- Extract sub-ontologies based on the sub-requirements respectively
- Integrate the sub-ontologies

Extracting sub-ontology from original ontology is the key process of the framework. The input of this process is a ontology and a sub-requirement, the output is the subontology extracted from the original ontology according to the sub-requirement.

INTEGRATING SUBONTOLOGIES INTO NEW ONTOLOGY

The integration of sub-ontologies to form a new ontology in the domain of interest is achieved as indicated in Fig. 11-13 in a series of processes. The ontology obtained through the integration of the sub-ontologies is a combined model with the attributes and instances of the constituent ontologies.

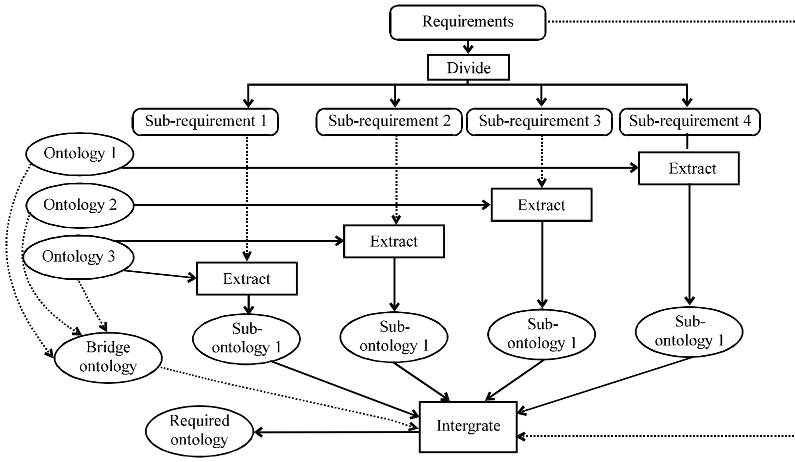

Fig. 9: Extraction of ontology from multiple ontologies

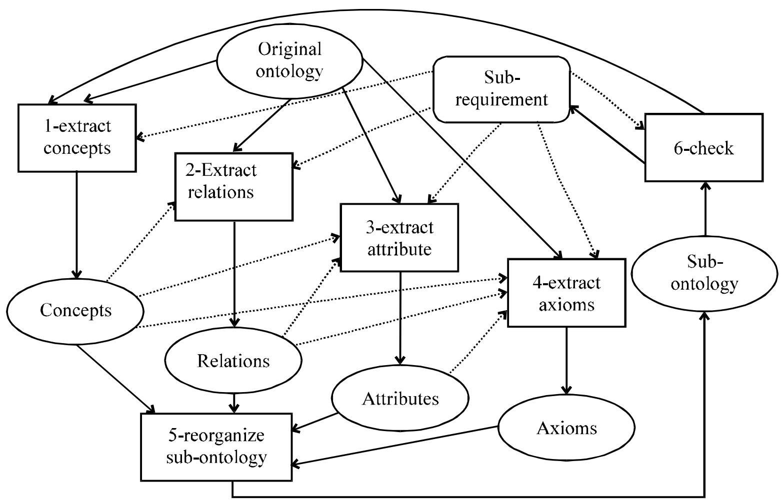

Fig 10: Integration of sub-ontologies

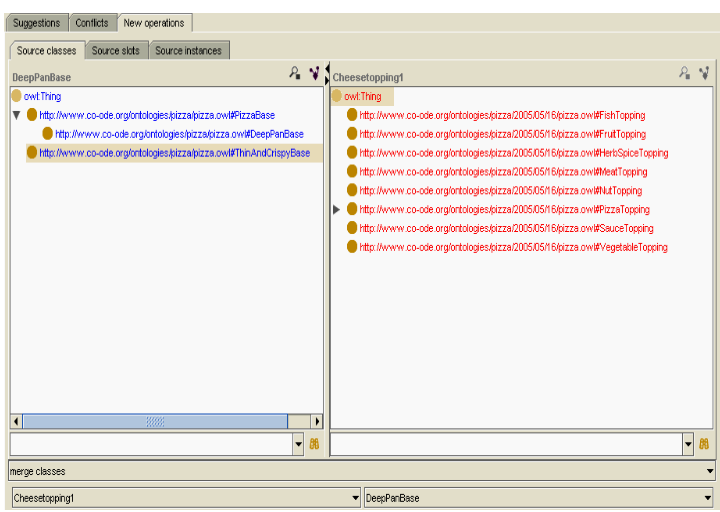

Fig. 11: Processing of Two subOntologies

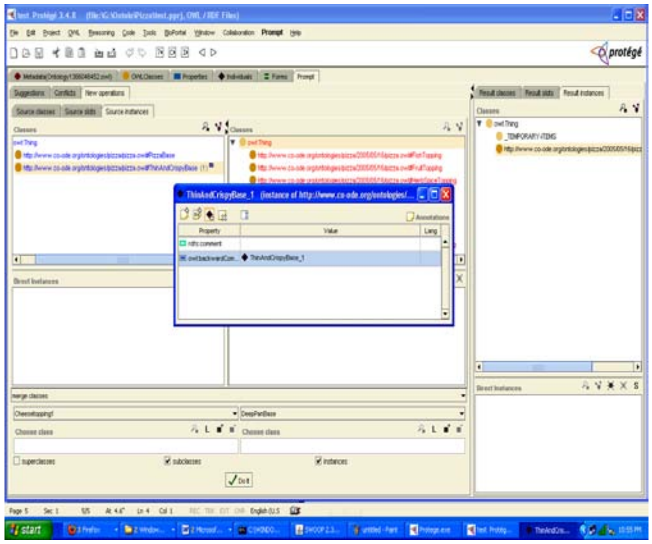

Fig. 12: Adding instances to the ontology

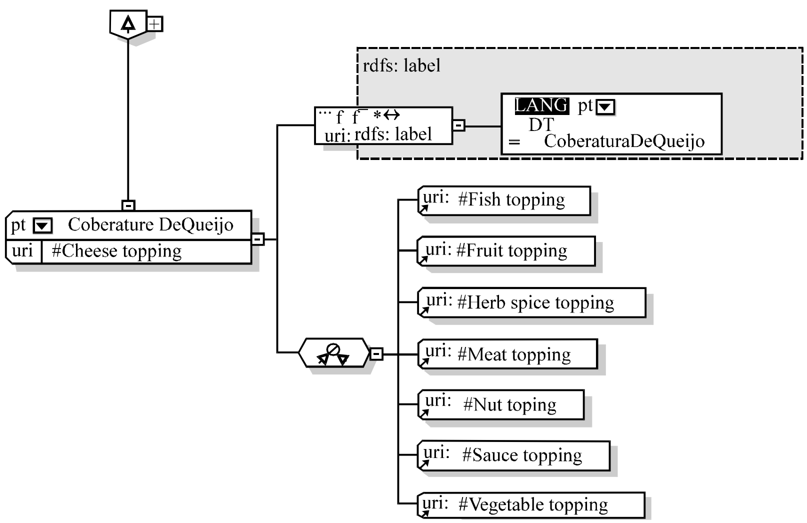

Fig. 13: RDF graph of ontology

CONCLUSION

This research is an attempt to illustrate the applicability of the conventional data merging and matching techniques as applied to ontological groupings of data that is proving to be the back bone of the semantic web. It is shown that object oriented approach that is provided by OWL and RDF can be effectively manipulated to suit the requirements with the ontology tools such as Protege. In the process, results that show possibilities of optimization through merging of ontologies are obtained. It is an initiative in the process of developing algorithms and methodologies for finding solutions through ontology arithmetics. This research is expected to progress further to utilise the full potential of the tools used to build effective and optimized ontologies on the required domain.

REFERENCES

Allemang, D. and J. Hendler, 2005. Semantic web for the Working Ontologist: Effective Modeling in RDFS and OWL. 2nd Edn., Morgan Kaufmann Publication, USA.

Bhatt, M., C. Wouters, A. Flahive, W. Rahayu and D. Taniar, 2004a. Semantic completeness in sub-ontology extraction using distributed methods. Proceedings of the International Conference on Computational Science and Its Applications, Volume 3045, May 14-17, 2004, Assisi, Italy, pp: 508-517.

Bhatt, M., A. Flahive, C. Wouters, W. Rahayu D. Taniar and T. Dillon, 2004b. A distributed approach to sub-ontology extraction. Proceedings of the 18th International Conference on Advanced Information Networking and Application, Volume 1, March 29-31, 2004, USA., pp: 636-641.

Borgida, A. and L. Serafini, 2002. Distributed Description Logics: Directed Domain Correspondences in Federated Information Sources. In: On the Move to Meaningful Internet Systems 2002: CoopIS, DOA and ODBASE. Meersman, R. and Z. Tari (Eds). LNCS., 2519, Springer, Berlin/Heidelberg, ISBN: 978-3-540-00106-5, pp: 36-53.

Calvanese, D., G. De Giacomo and M. Lenzerini, 2004. A framework for ontology integration. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.2599&rep=rep1&type=pdf.

Ceballos, H. and R. Brena, 2005. Finding compromises between local and global ontology querying in multiagent systems. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.1015&rep=rep1&type=pdf.

Correndo, G., M. Salvadores, I. Millard, H. Glaser and N. Shadbolt, 2010. SPARQL query rewriting for implementing data integration over linked data. Proceedings of the 2010 EDBT/ICDT Workshops, ACM, March 22-26, 2010, Lausanne, Switzerland.

Cruz, I.F. and H. Xiao, 2006. The role of ontologies in data integration. http://citeseerx.ist.psu.edu/ viewdoc/ download?doi=10.1.1.60.4933&rep=rep1&type=pdf.

Delfs, R., A. Doms, A. Kozlenkov and M. Schroeder, 2004. GoPubMed: Ontology-based literature search applied to gene ontology and PubMed. Proceedings of the German Conference on Bioinformatics Bielefeld, Germany, October 4-6, 2004, Springer, pp: 169-178.

Eiter, T., G. Ianni2 T. Krennwallner and A. Polleres, 2003. Rules and ontologies for the semantic web. http://www.researchgate.net/publication/221510784_Rules_and_Ontologies_for_the_Semantic_Web/file/d912f50b5f87ae2493.pdf.

Guarino, N. and C. Welty, 2000. A formal ontology of properties. Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management Methods, Models and Tools, Volume 1937, October 2-6, 2000, Juan-les-Pins, France, pp: 97-112.

Jimenez-Ruiz, E., B.C. Grau, I. Horrocks and R. Berlanga, 2008. Ontology integration using mappings: Towards getting the right logical consequences. Technical Report, December 12, 2008. http://krono.act.uji.es/people/Ernesto/contentmap/ContentMapTechReport.pdf.

Keet, C.M., 2004. Aspects of ontology integration. http://www.meteck.org/AspectsOntologyIntegration.pdf.

Le, W., S. Duan, A. Kementsietsidis, F. Li and M. Wang, 2011. Rewriting queries on SPARQL views. Proceedings of the 20th International Conference on World Wide Web, ACM, March 28- April 1, 2011, Hyderabad, India, Pages: 655-664.

Matthews, B., 2002. Semantic web technologies. JISC Technology and Standards Watch. http://www. jisc.ac.uk/uploaded_documents/jisctsw_05_02bpdf.pdf.

Reynolds, D., C. Thompson, J. Mukerji and D. Coleman, 2005. An assessment of RDF/OWL modeling. Digital Media Systems Laboratory, HP Laboratories Bristol, HPL-2005-189, October 28, 2005. http://www. hpl.hp.com/techreports/2005/HPL-2005-189.pdf.

Wang, P., B.W. Xu, J.J. Lu, D.Z. Kang and Y.H. Li, 2004. A novel approach to semantic annotation based on multi-ontologies. Proceedings of the 2004 International Conference on Machine Learning and Cybernetics, Volume 3, August 26-29, 2004, Shanghai, pp: 1452-1457.

Welty, C., 2003. IBM research (also W3C semantic web best practices WG). Semantic Web Ontologies. http://www.daml.org/meetings/2005/04/pi/Ontologies.pdf.

Wouters, C., T. Dillon, W. Rahayu and E. Chang, 2002. A practical walkthrough of the ontology derivation rules. Proceedings of the 13th International Conference on Database and Expert Systems Applications, Volume 2453, September 2-6, 2002, France, pp: 259-268.

Wouters, C., T. Dillon, W. Rahayu, E. Chang and R. Meersman, 2005. A practical approach to the derivation of materialized ontology view. http://www.igi-global.com/chapter/practical-approach-derivation-materialized-ontology/31126.