A Novel Approach in Ontology Mapping

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

By
Komal Saluja
(800931010)

Under the supervision of:
Dr. Seema Bawa
Professor & Dean
&
Ms. Shalini Batra
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

June 2011

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, **"A Novel Approach In Ontology Mapping"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in Software Engineering submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of **Dr.(Mrs.)Seema Bawa and Ms. Shalini Batra** and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Komal Saluja 07/06/2011
(Komal Saluja)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge

Seema Bawa 07/JUNE/2011
(Dr. (Mrs.) Seema Bawa)
Professor & Dean (Student Affairs)
Department of Computer
Science & Engineering,
Thapar University, Patiala.

(Ms. Shalini Batra)
Assisistant Professor
Department of Computer
Science & Engineering,
Thapar University, Patiala.

Countersigned by

(Dr. Maninder Singh)
Head
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life.

This work would not have been possible without the encouragement and able guidance of my supervisor **Dr. Seema Bawa** and **Ms. Shalini Batra**. I thank my supervisors for their time, patience, discussions and valuable comments. Their enthusiasm and optimism made this experience both rewarding and enjoyable.

I am equally grateful to **Dr. Maninder Singh**, Associate Professor & Head, Computer Science & Engineering Department, a nice person, an excellent teacher and a well – credited researcher, who always encouraged me to keep going with work and always advised me with his invaluable suggestions.

I will be failing in my duty if I don't express my gratitude to **'Dr. Abhijit Mukherjee'**, Director of the University, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable.

Last but not least, I would like to thank my family whom I dearly miss and without whose blessings none of this would have been possible. To my parents, I own thanks for their wonderful love and encouragement. I would also like to thank my brother, since he insisted that I should do so. I would also like to thank my close friends for their constant support.

ABSTRACT

The World Wide Web is now widely used as a universal medium for information exchange but the application of WWW is limited due to information heterogeneity and non semantic nature of HTML. The vision of semantic web provides many new technologies to overcome the limitation of WWW. In advent of the Semantic Web and recent standardization efforts, Ontology has quickly become a popular and core semantic technology. Ontology is seen as a solution provider to knowledge based systems. It facilitates tasks such as knowledge sharing, reuse and intelligent processing by computer agents. A key problem addressed by Ontology is the semantic interoperability problem. Interoperability in general is a common problem in different domain applications and semantic interoperability is the hardest and an ongoing research problem. It is required for systems to exchange knowledge and having the meaning of the knowledge accurately and automatically interpreted by the receiving systems. While Ontology promotes semantic interoperability across systems by unifying their knowledge bases through consensual understanding, common engineering and processing practices, it does not solve the semantic interoperability problem at the global level. Ontology mapping is a mechanism for providing semantic bridges between Ontologies. While Ontology Mapping promotes semantic interoperability across Ontologies, it is seen as the solution provider to the global semantic interoperability problem. Different applications would require different mapping techniques. This thesis analyzes the relations between Ontology, semantic interoperability and Ontology Mapping. An Ontology-based semantic interoperability solution has been developed.

Table of Content

List of Figures

List of Tables

Chapter 1
Introduction

The Internet and the World Wide Web have brought a revolution to information technology and the daily lives of most people. The current World Wide Web has well over billion pages, but the vast majority of them are in human readable format only which are relatively less comprehends by machines. The current shortcomings of Web pages are in large part due to the fact that they are written in Hypertext Markup Language (HTML), a language primarily concerned with marking up the text for formatting, rather than identifying and tagging content. There have been extensions to the HTML standard to allow some additional information to be added to a page, but this is largely superfluous and used to tag the page as a whole, rather than the individual pieces of content on a given page. In the light of these shortcomings in using HTML for creating WWW, the Semantic Web [14] provides many new perspectives and technologies to overcome the limitation of the WWW.

The Semantic Web is a project that intends to create a universal medium for information exchange by giving meaning (semantics) to the content of documents on the Web, in a manner understandable by machines. The idea is to make computers present formatted documents to the user and enable them to deal with the contents.

1.1The Semantic Web

The goal of the Semantic Web, with its vision by Berners-lee (1998), is to develop expressive languages to describe information in forms understandable by machines. The Semantic Web is an extension of the Current Web in which information is given well-defined meaning, better enabling computers and people to work in co-operation [15]. Semantic Web is an effort to enhance Current Web so that computers can process the information presented on WWW, interpret and connect it, to help humans to find required knowledge. The Semantic Web is implemented in the layers of Web technologies and standards. The layers are presented in Figure 1.1.

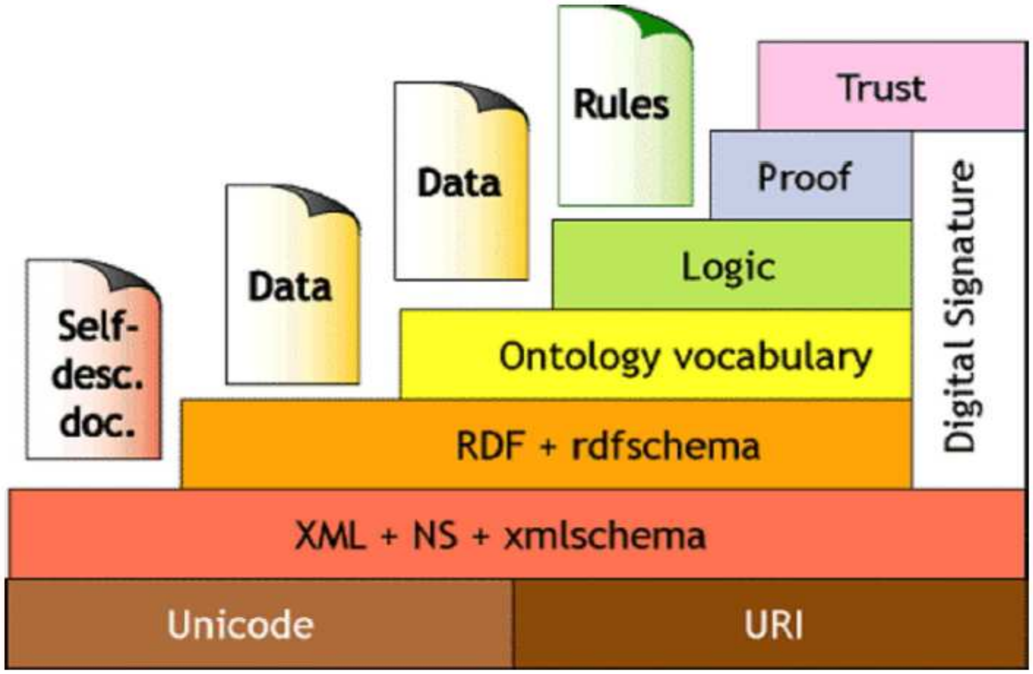


Figure 1.1: Semantic Web “layer cake” [16]

The **Unicode and Uniform Resource Identifier** (URI) layers make sure that international character sets are used and provide means for identifying the objects in the Semantic Web. The most popular URI’s on the World Wide Web are Uniform Resource Locaters (URL).

The **Extensible Markup Language** (XML) layer with namespace and schema definitions make sure the Semantic Web definitions can integrate with the other XML based standards [4]. XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents. XML Schema is a language for restricting the structure of XML documents [4].

The **Resource Description Framework (**RDF) Schema is a simple type modeling language for describing classes of resources and properties between them in the basic RDF model. It provides a simple reasoning framework for inferring types of resources.

The **Ontology layer** (OWL) supports the evolution of vocabularies as it can define relations between the different concepts. In this layer knowledge is expressed as descriptive statements, stating some relationship exists between one thing and another.[6]
A **Digital Signature** is an electronic signature that can be used to authenticate the identity of the sender of a message or the signer of a document. The Digital Signature layer ensures that the original content of the message or document is unaltered.
The top layers **Logic, Proof and Trust** are currently being researched and simple application demonstrations are being constructed. The Logic layer enables the writing of rules while the Proof layer executes the rules and evaluates together with the Trust layer mechanism for applications whether to Trust the given proof or not.

The semantic Web has the aim of resolving Current Web issues by using additional technology to help categorize and organize the human information contained on a page with machine-readable and understandable information that can then be used by applications to help categorize and organize the information.

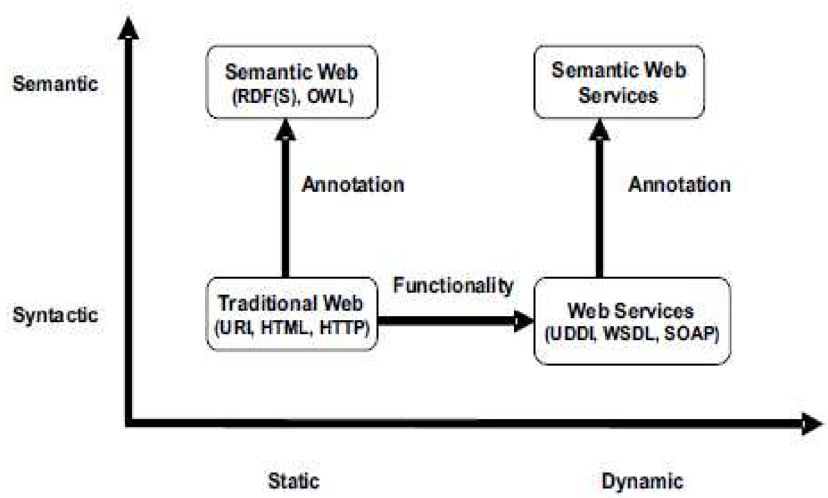


Figure 1.2:The Changing Web[14]

XML, RDF, OWL all of these technologies are used together to help provide semantic information about a given resource and the resources to which it may be linked shown in figure 1.3.Resources are identified by a Uniform Resource Identifier (URI); the RDF and OWL data are then connected to the URI describing not only its content but also the relationship between content items and their relationship with other URIs or content types.

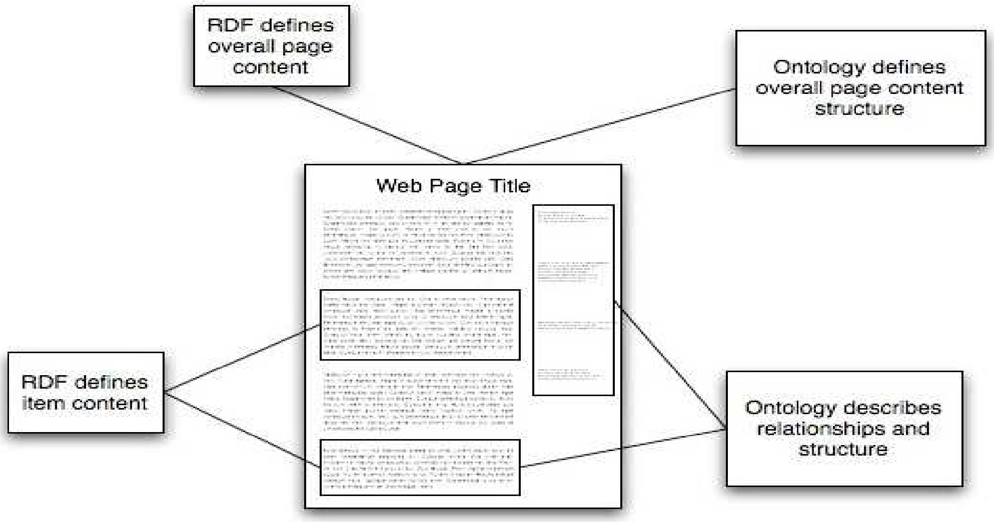


Figure 1.3: Semantic Document Structure [17]

1.2 THE PROBLEM OF INFORMATION HETEROGENEITY

Information Heterogeneity refers to the difference between information. Information heterogeneity occurs at three levels, i.e., syntax, structure and semantics [18]. Syntactic heterogeneity is the simplest heterogeneity problem caused by the usage of different data formats. To solve the syntactic heterogeneity, standardized formats such as XML, RDF and OWL have been widely used to describe data in a uniform way that makes automatic processing of shared information easier. Structural heterogeneity which occurs as a result of the way information is structured even in homogeneous syntactic environments. For example, one source might model trucks but only classify them into a few categories; while the other source might make very fine-grained distinctions between types of trucks based on their physical structure, weight, purpose, etc. Though sophisticated solutions to syntactic and structural heterogeneity have been developed, the problem of semantic heterogeneity is still not solved fully.

Semantic heterogeneity occurs whenever two contexts do not share the same interpretation of information (e.g. homonyms and synonyms). Data sources are being developed by independent organizations so there might be semantic differences between their schemas. In different data sources, a same concept may be represented with different names such as, instructor, teacher or lecturer so these words are synonym to each other. Similarly different concepts in different data sources may be represented by same name which is called polysemy such as the term ‘bank’ refers to a river bank as well as a financial institution. So there is need to cope the semantic heterogeneity by generating correct semantic mapping between global Ontology which works as the global schema and source ontologies elements. For example, as shown in Figure 1.4, Swoogle returns *many* documents when searching for *spring*. The top ranked results show that the same term has many different meanings, e.g. one *spring* means the *season*, and the other *spring* means *ground water*.

Figure1.4: Swoogle snapshot

To solve such semantic heterogeneity among different information systems many approaches such as using synonym sets, term networks, concept lattices, features and constraints have been proposed [28]. However these approaches are not sufficient to solve the problem of semantic heterogeneity in many areas.

1.3 Ontology:

Ontology definition proposed by Gruber "Ontologies are defined as a formal explicit specification of a shared conceptualization". This mean Ontologies provide a way to describe the meaning and relationships of terms so that a shared understanding or a consensus can be acquired among people and machines [9]. The openness of the Web means the volume of published information is growing exponentially resulting in what is commonly termed 'information overload'. To search for relevant data in such a huge amount of collection is becoming very tedious and difficult. Each search engine uses its own secret algorithm when determining, what results to give back and in what order the results should be displayed. It can often be difficult to extract relevant information from the retrieved search results. Sometimes, relevance can only be determined by shifting through the result, one by one. Ontologies help in querying knowledge on the Web by enriching information with descriptions of its meaning. Significantly, these rich descriptions can be interpreted by computer systems, which allow them to interpret the results of Web queries intelligently. Ontologies enables web to share knowledge and provide reasoning support, i.e., verification of formalisms and rules. Checking for constraint violations in knowledge shared through the Ontology can be automated [13].

A key problem addressed by Ontology is the semantic interoperability problem, because the conceptualizations they represent are not commonly shared and agreed by everyone. In order to realize the Semantic Web vision, disparate ontological representations must be connected through mappings.

Ontologies are widely used in the fields of Knowledge Engineering, Artificial Intelligence and Computer Science, in applications such as knowledge management, natural language processing, e-commerce, information integration, bio-informatics, and emerging fields such as the Semantic Web.

1.4 Semantic Interoperability:

Semantic Interoperability is the capability of different information systems to communicate information consistent with the intended meaning of the encoded information. Semantic Interoperability requires information systems that can exchange data and reuse the exchanged data with their intended meanings. [19]

The goal of semantic interoperability across systems is to empower computer systems with the capabilities to intelligently consume new knowledge and use it in such a way that the original creator of the knowledge did not foresee. Semantic interoperability facilitates many types of applications such as database schema integration, catalog integration, query answering, web service composition, P2P information sharing, and agent communication.

Semantic interoperability is a multi-level problem that requires solution at the syntactic level (specification), structural level (scope) and semantic level (intelligence). At the syntactic level, interoperability requires translation between data and knowledge representations e.g. from Resource Description Framework Schema (RDFS) tuple to Web Ontology Language (OWL). The primary objective is to allow machine processing of data from another computer system with a different data format specification.

At the Structural level there is a problem that arises when there is a mismatch between data structures of ontologies. Imagine that there are two travel-based Ontology lattices representing two separate business entities. In this example both entities are involved in the travel domain. Figure 1.5 illustrates that Hotel B has single and double room occupancy and the tree structure starts with “Rooms”. For the same occupancy category, the tree in Hotel A starts from “Price” instead of “Rooms”. Hotel B has two separate specifications for “Price” in its tree. The difference in the lattice structure is mainly due to design styles that differ among Ontology engineers.

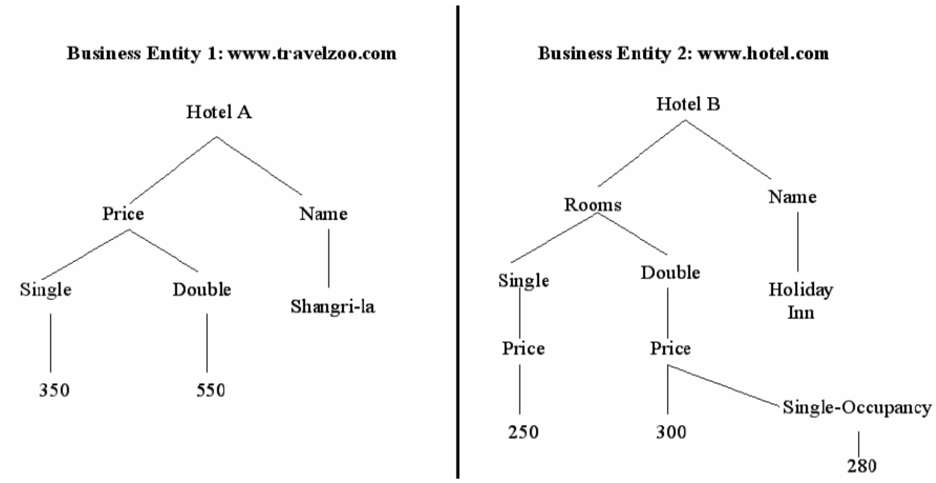


Figure1.5: Structural Interoperability Problems [19]

Semantic level interoperability problem happens when semantically identical information is represented in different data formats or scales. This type of heterogeneity can be further divided into scale and representation conflict. For example the "Price" attribute is represented in "US Dollar" for Hotel A and in Hotel B the scale used is "Euro Dollar". This is defined as a scale conflict. Hotel A denotes "Category" and Hotel B denotes "Class" to represent ratings or rankings of the hotel (i.e. 5 indicates a 5 star hotel). Here Hotel B has used a different rating scale (i.e. A which is in their definition is equivalent to 5 star). This is defined as a representation conflict. Suppose, Hotel B uses the label "Quote" instead of "Price" this is again another case of representation conflict.

Business Entity 1 : www.travelzoo.com

Hotel A – Hotel entity

Name	Location	Category	Price
Shangri-la	City Center	5	100

Note : Scale & representation conflicts

- 5 star hotel is a category 5 (object representation conflict)
- Price in US dollar (scale conflict)

Business Entity 2 : www.hotel.com

Hotel B –Accommodation entity

Name	Location	Class	Price
Holiday Inn	City Center	A	100

Note : Scale & representation conflicts

- Class A is a 5 star hotel (object representation conflict)
- Price in Euro dollar (scale conflict)

Table1.1: Semantic Data Interoperability Problem [19]

1.5 Ontology Mapping:

Ontology mapping is the determination of semantic correspondences between similar elements in different Ontologies. Ontologies are becoming increasingly significant because they provide semantics for annotations in the Semantic Web [20]. Ontology development is very distributed in nature and this has led to a large number of Ontologies that overlap when addressing similar domains. Researchers and knowledge engineers in different domain areas have developed Ontologies, which have semantic interoperability problem. It would be difficult to find correspondence among these ontologies as similar concepts could have been expressed using different naming conventions or structures. For example in the domain of computer security, definitions about attacks can exist in two or more ontologies designed by different Ontology engineers. Some attack definitions could be more detailed in one Ontology but not detailed enough in the other. Thus Ontology Mapping promotes semantic interoperability across ontologies, it is seen as the solution provider to the global semantic interoperability problem.

1.6 Structure of the Thesis

Chapter 2- Provides a review of all the work done in the area of Ontology Mapping.

Chapter 3-Gives the problem statement and methodology used to solve it.

Chapter 4 –Gives a detailed introduction about Ontology, Ontology Mapping, Ontology development Process, Ontology mapping operations and provides a Novel approach in Ontology Mapping.

Chapter 5 –Explains the experiments performed and evaluates the results achieved.

Chapter 6 –Summarizes the work presented in this thesis followed by the features that can be incorporated in future for the enhancement of the Ontology Mapping.
This thesis concludes with references.

Chapter 2
Literature Review

2.1 The Semantic Web: Opportunities and Challenges

The advent of the World Wide Web (WWW) has taken the availability of information to an unprecedented level. The rapid growth of the web poses new problems like anyone can easily publish a new document or add a link to a site with no restrictions on the structure or validity of the new content. The lack of restrictions has been partly the reason of the success of the web. It has kept the expertise necessary to create content for the web low. The overabundance of unstructured and possibly faulty information has made it difficult for the users of the web to find relevant information easily. It also poses scaling difficulties for current web crawlers and search engines.

Current web is like a book, having multiple hyperlinked documents. In book scenario, index of keywords are there but the context in which those keywords are used is missing. There are no formal semantics of keywords in indexes. To check which one is relevant, then read the corresponding pages of that book. It has been determined that inaccessible part of the web is about five hundred times more than what search engines find [1], [2]. There are various web resources that may be very useful in our everyday activities. It is difficult to locate them in current web because they are not annotated properly by the metadata.There are billion pages of information available on the web, and only a few of them can be reached via traditional search engines.

These problems encountered in the current web is somewhat resolved by development of Semantic Web. Semantic web aims to provide data in a format that embeds semantic information, and then seeks to develop sophisticated query tools to interpret and combine this information The Semantic Web is based on treating the Web as a knowledge base defining meanings and relationships. Semantic Web is the web of ontologies having data with formal meanings [3]. This is in contrast to current web which contains virtually boundless information in the form of documents. The Semantic Web, on the other hand, is about having data as well as documents that machines can process, transform, assemble, and even act on it in useful ways.

Following are the applications of Semantic Web:

2.1.1 Semantics-Based Web Search Engines

Search engines are among the most useful resources on the Web and currently there are two types of search engines:

- Large-scale robot-based search engines. These systems rely on robots to retrieve Web pages and store them in a centralized database. The advantage of this mechanism is that it increases recall (the proportion of relevant documents that are actually retrieved) since robots can almost retrieve all web pages on the Web, while the disadvantage is the precision (the proportion of retrieved documents that are actually relevant) of the search result might be low.[5]
- Small-scale reviewer-based search engines. A category hierarchy is created and each category is described by a set of keywords. Reviewers will review each web page and associate it with appropriate categories. The advantage is that precision is increased, but the disadvantage is that recall might be low since it is impossible to review and include every single relevant web page on the Web.[15]

Both types of search engines are based on keywords, and hence are subject to the two well-known linguistic phenomena that strongly degrade a query's precision and recall: polysemy (one word might have several meanings) and synonymy (several terms, i.e. words or phrases, might designate the same concept). A number of stemming algorithms have been developed to address the synonymy issue including suffice removal, strict truncation of character strings, word segmentation, letter bigrams and linguistic morphology. The idea is that different derivations of a word are similar to each other in their forms (e.g. they have the same prefix) and can be traced back to the same root using these stemming methods. However, these methods are subject to the following stemming errors: words with different meaning might be reduced to the same root. For example, words general, generous, generation, and generic might be reduced to the same root. On the other hand, different words with the same meaning cannot be reduced to the same root for example, automobile and car. The situation becomes worse for the large-scale robot-based search engines. Only limited semantics can be derived from the lexical or syntactic content of the web pages. Several systems have been built to overcome these problems based on the idea of annotating Web pages with special HTML tags to represent semantics, including XML (Extensible Markup Language), RDF (Resource Description Framework), and OWL (Web Ontology Language). However the limitation of these systems is that they can only process web pages that are annotated with these HTML tags, and so far there is no agreement upon a universally acceptable set of HTML tags.

XML is a promising technique since it keeps content, structure, and representation apart and is a much more adequate means for knowledge representation. Currently, there are numerous techniques and tools available for XML, e.g., XSL (Extensible Stylesheet Language), XSLT (XSL Transformation) and XML parsers are available in different languages and for different platforms. Using XML, one can describe document types for various domains and purposes. For example, XML documents may represent multi-media presentations, and business transactions [36].

Tatarinov, et al [4], propose that these query languages should be extended with an update capability so that an XML document repository becomes an XML database.

XML will continue to play an important role in the development of the Semantic Web. However, it does not provide a full solution to the requirements of the Semantic Web. XML can represent only some semantic properties through its syntactic structure, i.e., by the nesting or sequentially ordering relationship among elements (XML tags). XML queries need to be aware of this syntactic structure via the document type that is defined by a DTD (Document Type Definition). Although one might derive some sort of semantics from the structure of the documents within the context of the document type, the semantics of each element (XML tag) is not defined and its interpretation totally relies on the implicit knowledge hardcoded in application programs. To develop a Web with semantics, resources on the Web need to be represented in or annotated with structured machine-understandable descriptions of their contents and relationships, using vocabularies and constructs that have been explicitly and formally defined with domain Ontology [36].

RDF is more useful than XML because it provides independent syntax serialization and abbreviation for data modeling, syntax reification and semantic based features like domain independency, vocabulary and privileges in defining terminologies used in schema language RDF (Resource Description Framework) [5] is a URL based syntax data representation which provides a secure and reliable mechanism for metadata exchange between web applications. RDF processes Meta data by making abstract data models based on three object types .i.e., Resource, Property and Statement. Resource is an expression, Property is an attribute describing resource and Statement is a resource having some properties and values but still RDF modeling mechanism is insufficient in expressing various logical statements.

OWL claims to be an extension in RDF in expressing logical statements because it not only describes classes and properties but also provides the concept of namespace, import, cardinality relationship between the classes and enumerated classes. OWL (Web Ontology Language) derived from American DARPA Agent Markup Language (DAML) and based on Ontology, inference and European Ontology Interchange Language (OIL) [6].

2.1.2 Semantics-Based Digital Libraries

Digital multimedia data in various formats has increased tremendously in recent years on the Internet. With the development of digital photography, more and more people are able to store their personal photographs on their PCs. Sharing of picture albums and home video on the Internet becomes more and more popular. Furthermore, many organizations have large image and video collections in digital format available for online access. Film producers want to advertise movies through interactive preview clips. Travel agencies are interested in digital achieves of holiday resorts photographs. Hospitals would like to build medical image databases. These emerging applications for multimedia digital libraries require interdisciplinary research in the areas of image processing, computer vision, information retrieval and database management. Semantics-based retrieval of multimedia digital content is important for efficient use of the multimedia data repositories. Traditional content-based multimedia retrieval techniques describe images/videos based on low level features (such as color, texture, and shape) and support retrieval based on these features. However, human typically does not view images/videos in terms of low-level features. A semantics-based query capability is highly desirable. For example, one might want to formulate a query like "return all the scenes in clip 1 in which a boy is riding a bicycles". Retrieving images/videos based on low-level features cannot provide satisfactory results. Effective and precise multimedia retrieval by semantics remains an open and challenging problem [5].

It is believed that various digital libraries will become another major web resource of the Semantic Web. The challenges here are: (1) The development of efficient and effective classification and indexing mechanism for each type of digital library, and (2) The semantic interoperability between digital libraries of similar types and between a digital library and the Semantic Web.

The various challenges posed by the semantic web can be resolved by application of ontologies which we will be discussing now.

2.2 The Development of Ontologies

Ontologies play an essential role in the development of the Semantic Web. Ontologies are formal models about how perceive a domain of interest and provide a precise, logical account of the intended meaning of terms, data structures and other elements modeling the real world. Ontologies can help in the representation of the content of a web page in a formal manner, so as to be suitable for use by an automated computer agent. Ontologies are core concept of the Semantic Web that allows the representation of data in a machine processable way. Ontologies are often large and complex structures, whose development and maintenance give rise to several sturdy and interesting research problems.

According to Noy & McGuinness [13], ontologies are used to build knowledge bases, knowledge-base is formed by an Ontology and a set of individual instances of its classes). Knowledge bases can be queried by agents in order to enrich, reuse and maintain them.

Tramullas et al. [7] Ontologies are able to operate as repositories to organize information for specific communities. They can be used as a tool for knowledge acquisition, (teamwork can use ontologies as a common support to classify the knowledge of an organization).

Sure et al. [8] presents Semantic integration of heterogeneous information sources such as digital libraries can benefit with the incorporation of ontologies. Some applications use domain Ontology to integrate information resources and others allow each resource to use its own Ontology. The Ontology is one effective solution to semantic heterogeneity problem in web data integration. Explicit semantic information of terms in the Ontology can help to resolve semantic heterogeneities among web data [9]. The subject of Ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called Ontology, is a catalogue of the types of things that are assumed to exist in a domain [9]. Ontology helps to figure out what a specific term means.

For reconciliation of semantic conflicts between web data sources, creation of one global or shared Ontology by integrating or merging local schemas or ontologies is infeasible. In the web context the maintenance and updating of global or shared Ontology is very time consuming and costly because many web data sources are involved and the number of involved web data sources change frequently; web designers and users are free to use their own terms and vocabulary and schemata which are subject to frequent changes.

Despite ontology has provided solution to lots of issue of semantic web but still there are some further challenges in using ontology.

2.2.1 Ontology Challenge

While Ontology promotes semantic interoperability across systems but it does not solve the semantic interoperability problem at the global level. As individuals are increasingly empowered with tools, ontologies will eventually be created more easily and rapidly at a near individual scale. Global semantic interoperability between heterogeneous ontologies created by small groups of individuals will then be required. Semantic interoperability is the ability for disparate system domains to comprehend meanings and terminologies via axiomatic mapping of agreed-upon concepts to create semantically compatible information. Establishing semantic interoperability among heterogeneous and disparate information sources has been a critical research area within the database community for the past two decades [10].

There are two essential issues in semantic interoperability. First, is the identification of semantically-related data with subsequent resolution of their schematic differences. Second, is the access to and usage of large autonomous databases without prior knowledge of their content [9].The first issue involves making semantics explicit and this depends on the context in which a data object is used, and its contextual representation. The second issue requires users to be familiar with the content and structure of the information sources in order to be able to specify a query. This approach provides a uniform method for query translation and heterogeneity resolution in a multidatabase environment [13, 19]. Since the Semantic Web has to deal with more than just structured data sources, this technique was not sufficient.

Capturing explicitly the semantic content of the individual databases is another important challenge. It is important to understand the semantics of each schema component and to capture and reason them using semantics. Semantic data models do not quite capture semantics of a database, and the meta-information (i.e., tacit knowledge) captured during its design phase is not explicitly represented in the resulting database. Hence, such information cannot be completely accessible to applications, queries, or users. As such, semantic data models that capture domain meta-information (i.e., entity classes, relationships, constraints, cardinalities, etc.) alone are not enough to support semantic interoperability among heterogeneous databases [10].

Two widely used approaches for semantic interoperability are the domain Ontology approach and the tacit knowledge capture approach. The domain Ontology approach attempts to construct a global schema and establish mappings between domain schemas and participating local schemas. However, the drawback of this approach is its lack of semantic richness and flexibility [11]. The knowledge capture approach tries to solve the problem of lack of semantic richness by capturing the tacit knowledge within a certain domain in great detail in order to provide a rich conceptualization of data objects and their relationships. Even though it is theoretically valid, in practice it is not feasible due to the inherent complexities of the knowledge domain. Hence, it can only serve restricted application domains. This limits its general applicability. Thus a hybrid approach was initiated [11].

A hybrid approach uses a common Ontology, which specifies a vocabulary to describe and interpret shared information among its users. This approach is similar to the domain Ontology approach as it has a high-level domain model playing the role of shared schema while ensuring the autonomy of the local schemas. However, a domain model is different because domain knowledge captured in the domain model is generally represented in logic using the vocabulary provided by the Ontology editor. An Ontology-based domain model captures much richer semantics and covers a much broader range of knowledge within a target domain [12]. This approach provides a simple formalism to capture only the domain knowledge pertaining to potential semantic conflicts. The advantage of this simplified Ontology is that it is not domain-specific and does not lose any semantic richness. The researchers stress that using both a common Ontology and a semantic data model provides a more complete understanding of the application domain.

Consequently new breakthrough in research was achieved with the advent of Ontology Mapping.

2.3 Ontology Mapping Approach

With the later research work it was found that Semantic Interoperability problem can be resolved through semantic mapping between ontologies .Therefore, in order to resolve semantic conflicts between web resources, a semantic mapping between their local ontologies terms is needed. This semantic mapping relates similar terms from the two different sources by specifying the correspondence between them. There are approaches for semantic mapping between ontologies that have been proposed by researchers.
These researches exploit the information available in ontologies and map similar terms of two given ontologies to each other by using mapping algorithms. Ontology mapping approach based on two main factors: quality of mapping results and run time complexity of the mapping approach. Some approaches have high quality mapping results but they are not applicable because of the high runtime of their mapping algorithm.

Several integration methods were surveyed to identify the appropriate approach to Ontology mapping. One such effort is the Grid model. It brings together dispersed data and information spaces via a common platform called the Grid infrastructure. Virtual Organization (VO) that participate and interact on the Grid can seamlessly inter-exchange and collaborate on data, thereby resolving much of the disparity issues. Although this method provided some important facts, it did not show how ontologies could be matched [31].

Graph based solutions have also been used in the effort of creating semantic bridges. The Ontology is based on semantic correspondence between concepts that are tied together based on their node locations on the graph. Graph-like structures are used to highlight heterogeneity amongst data in the graph hierarchy. The hierarchy can also be observed as a taxonomy hierarchy. Concepts are matched based on the nearest neighbor approach for similarity matching. A match operator takes two graph-like structures (e.g., database schemas or ontologies) and produces mappings between their data elements in the form of graphs. The graphs that correspond syntactically to each other are then generated. The problem with this method is that it lacks flexibility [30].

The InfoSleuth project uses Inductive Machine Learning (IML), a revolutionary approach that provides good results. It uses agent architecture for creating semantic mappings [37]. The Semantic Web is viewed in the same manner as the Grid where various layers exist and heterogeneous security policies exist among them. Heterogeneous data labels use reference Ontology (RO) [37]. Such access-level integration was only used in commercial tools and the real need was for semantic data-level integration. IML was successfully implemented in their agent architecture for semantic integration of web based information resources. However, this method is based purely on an XML representation of data sources, and does not scale for the Semantic Web environment.

Another method includes Building Inclusion relations between ontologies. But the shortage of this method is that the information and concepts only be reused and can't be revised.

According to M. Ehrig et al. [32] Ontology mapping is a transition process of associated entities according to semantic relation in concept level. Entities in source Ontology can change into entities in object Ontology by mapping. Most effective method of solving Ontology heterogeneity is Ontology mapping. The purpose of Ontology mapping is look for corresponding relations among concepts and design relevant mapping rules. Approaches introduced for Ontology mapping specifically focus on certain shortcomings. There is a need to extend these techniques to resolve Ontology mapping problems.

Lots of researchers has derived various tools and frameworks to give impetus on Ontology Mapping research out of these we will discuss some important ones

2.4 Ontology Mapping Tools:

2.4.1 PROMPT/SMART/PROMPT-DIFF

Noy and Musen have developed several tools for Ontology mapping, alignment and versioning. SMART [11], PROMPT and PROMPTDIFF. All of them are available as plug-ins for the open source Ontology editor, Protégé [27]. These tools perform linguistic similarity matches between concepts and then use Protégé for discovering further matches between them. They distinguish merging and alignment where merging is described as a process to create a single coherent Ontology and alignment as a process that establishes links, which would help align ontologies to reuse information from one another. PROMPT guides the ontologist during the merging or alignment process. PROMPTDIFF is the latest addition to the set of tools. It uses an algorithm, which integrates different heuristic matchers for comparing Ontology versions. Three types of mapping levels are defined such as unchanged (nothing has changed), isomorphic (images of each other), and changed (not images of each other).

2.4.2 CHIMAERA, GLUE and CAIMAN

CHIMAERA is an interactive tool just like PROMPT where the ontologist does merging and is guided by the tool. Ontologies to be merged are analyzed by the tool and if linguistic matches are found, the merge is done automatically; otherwise the user is prompted for further action. The only difference between the two tools is the suggestions they make to the ontologist during the merging process. Doan and colleagues developed GLUE [34]. It uses machine-learning techniques to locate mappings of two or more ontologies. Similar concepts are searched using probabilistic measures. It uses multiple learning strategies and exploits information, either in data instances or in the taxonomic structure. It can also make predictions based on a content learner and a name learner. A meta-learner combines the predictions of the two learners. CAIMAN also uses machine learning techniques and was developed by Lacher and Groh [35]. This tool is quite similar to GLUE.

2.4.3 Cooperative Framework

Fernandez introduces a model for Ontology integration called the cooperative framework [22]. The algorithm proposed is meant to create a global Ontology. The model is intended for two groups of users - normal and expert. The normal user seeks information and provides specific information of concepts and the expert integrates the Ontology in the author's nomenclature. The algorithm is based on taxonomic features and synonym detection of concepts from the source ontologies (SO). Attributes of concepts can also be defined in this framework and the algorithm integrates attributes of the same concept. However, both the concepts that are to be integrated (e.g. PERSON and PEOPLE) must possess the same exact attributes (i.e. age and name). This rigid criterion makes the model less flexible.

2.4.4 OISs Framework

Calvanese and colleagues propose a framework called Ontology Integration Systems (OISs) [29]. The framework is based on Description Logic (DL) knowledge bases and mappings are expressed via queries. There is no explicit mechanism for the notion of queries. Mapping of concepts into Ontology views is first achieved based on the query results. Two approaches for query views are introduced i.e. global-centric and local-centric. Associating each relation in the global schema to one relational query over the source relations specifies the global-centric approach to mapping. The local-centric approach requires reformulation of query in terms of the queries to the local sources. The authors provide examples of using both approaches. The technique is analogous to integration of relational databases. It presents views but doesn't show how unstructured data could be matched.

2.4.5 OntoMapO Framework

Kiryakov and colleagues developed a framework for accessing and integrating upper level ontologies called meta-Ontology (OntoMapO) [32]. It allows a user to import linguistic ontologies onto a Web server, which will then be mapped onto other ontologies. A uniform representation of the ontologies with mappings and a simple meta-Ontology of property types and relation-types should be defined. The framework does not however show how equivalence of concepts is measured.

2.4.6 IFF Framework

Kent introduces the Information Flow Framework (IFF) to support Ontology sharing [33]. It is based on channel theory of Barwise and Seligman. Kent exploits the distinction made in channel theory to formally describe the stability and dynamism of conceptual knowledge organization. There are two basic assumptions and a two-step process in this model. The framework is purely theoretical and there is no method for implementing the two-step process.

Chapter 3
Problem Statement

3.1 PROBLEM DEFINITION

A generic mapping problem occurs when there exists different representations of similar Information. These representations can be physical, like text, pictures, or events that experience. They can also be our own mental representations of these physical objects and events. A mapping must be constructed in order to transform one representation into another. The mapping process is difficult because terminologies are developed by humans and as a result they often encode their biases, cultural differences, and subjective world views.

Current methods do not provide formal semantics to mapping data structures. Discussions in the literature review indicate Ontology mapping methods and techniques. They also rely heavily on Graph-based and structure-based similarity measures, which often times fail to produce appropriate mappings. Given that previous approaches will have difficulty scaling to address Ontologies for the Semantic Web, a consolidated approach is needed to solve the heterogeneity problems among Ontologies. Semantic heterogeneity is the disagreement of meaning or interpretation of similar or related data. The existence of information heterogeneity and the importance of Ontology Mapping in different applications motivate our thesis in the area of Ontology Mapping. Therefore, the ultimate goal of our research is to solve the problem of Ontology Mapping, and thus enable semantic interoperability between different web applications and services in the WWW.

The problem in Ontologies is analogous to translating between spoken languages. When different groups of people use different syntaxes and vocabularies to express insights about the world, translating from one to another does not always preserve the contextual or idiosyncratic components of the original intention of meaning. Meaning with precision is one of the desirables of Ontology use, and identifying what meanings are intended among multiple Ontologies, developed by different groups, is important. Furthermore, this problem, referred to as generating semantic mappings.

3.2 METHODOLOGY

The purpose of the work is to introduce a method for finding semantic correspondence among the ontologies with the intention to support semantic interoperability. The overall purpose is decomposed into the intermediate goals of this work. The goals of this work are to:

1. The survey of Ontology mapping methods step includes an investigation of existing methods of Ontology mapping and an analysis of the process of Ontology mapping, together with the properties characterizing such a process.

2. Propose an architecture for a system to support our approach as well as implement the system in a prototype, and

3. Provide an analysis of the implementation and evaluation of the method

4. Analyze the applicability of the mapping approach in supporting interoperability. In the sequel we will explain how the above objectives have been approached and motivate for the main decisions made during work on the thesis.

4.1 Ontology

An Ontology is a formal explicit specification of a shared conceptualization of a domain of interest. [21]

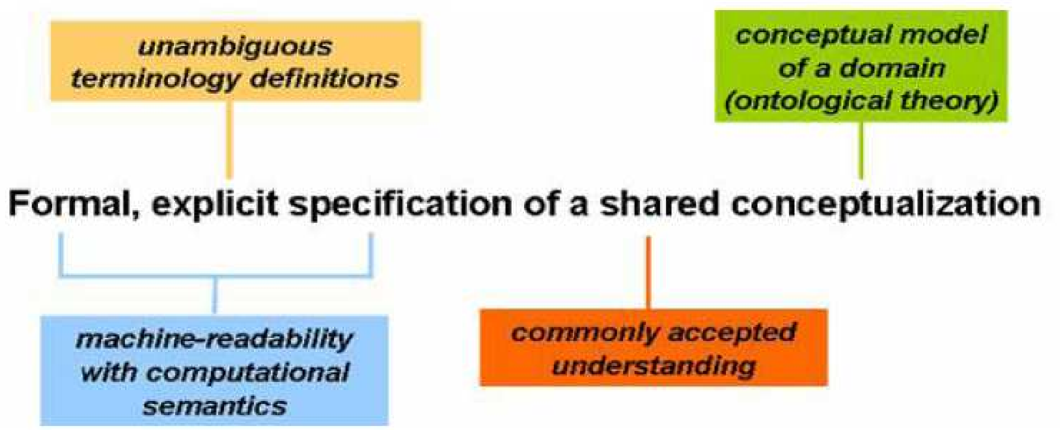


Figure 4.1: Ontology Definition [21].

This definition captures several characteristics of an Ontology as a specification of domain knowledge, namely the aspects of formality, explicitness, being shared, conceptuality and domain-specificity, which require some explanation.

Formality

An Ontology is expressed in a knowledge representation language that provides a formal semantics. This ensures that the specification of domain knowledge in an Ontology is machine processable and is being interpreted in a well-defined way. The techniques of knowledge representation help to realize this aspect.

Explicitness

An Ontology states knowledge explicitly to make it accessible for machines. Notions that are not explicitly included in the Ontology are not part of the machine-interpretable conceptualization it captures, although humans might take them for granted by common sense.

Being shared

An Ontology reflects an agreement on a domain conceptualization among people in a community. The larger the community, the more difficult it is to come to an agreement on sharing the same conceptualization. Thus, an Ontology is always limited to a particular group of people in a community, and its construction is associated with a social process of reaching consensus.

Conceptuality

An Ontology specifies knowledge in a conceptual way in terms of symbols that represent concepts and their relations. The concepts and relations in an Ontology can be intuitively grasped by humans, as they correspond to the elements in our mental model.

Moreover, an Ontology describes a conceptualization in general terms and does not only capture a particular state of affairs. Instead of making statements about a specific situation involving particular individuals, an Ontology tries to cover as many situations as possible, that can potentially occur .

Domain specificity

The specifications in an Ontology are limited to knowledge about a particular domain of interest. The narrower the scope of the domain for the Ontology, the more an Ontology engineer can focus on axiomatising the details in this domain rather than covering a broad range of related topics. In this way, the explicit specification of domain knowledge can be modularized and expressed using several different ontologies with separate domains of interest.

4.2 Ontology Development Process

Activities needed for the development of an Ontology:

• Before a developer starts building an Ontology, the main tasks involved in this project should be made explicit. This includes their proper arrangement, the identification of their estimated duration and the resources needed to complete them in time.

• The purpose and the scope of an Ontology should also be designated beforehand. The reasons why this Ontology is being built and the intended use cases together with their end users are to be found in this activity. The results are recorded in a so-called Ontology requirements specification document.

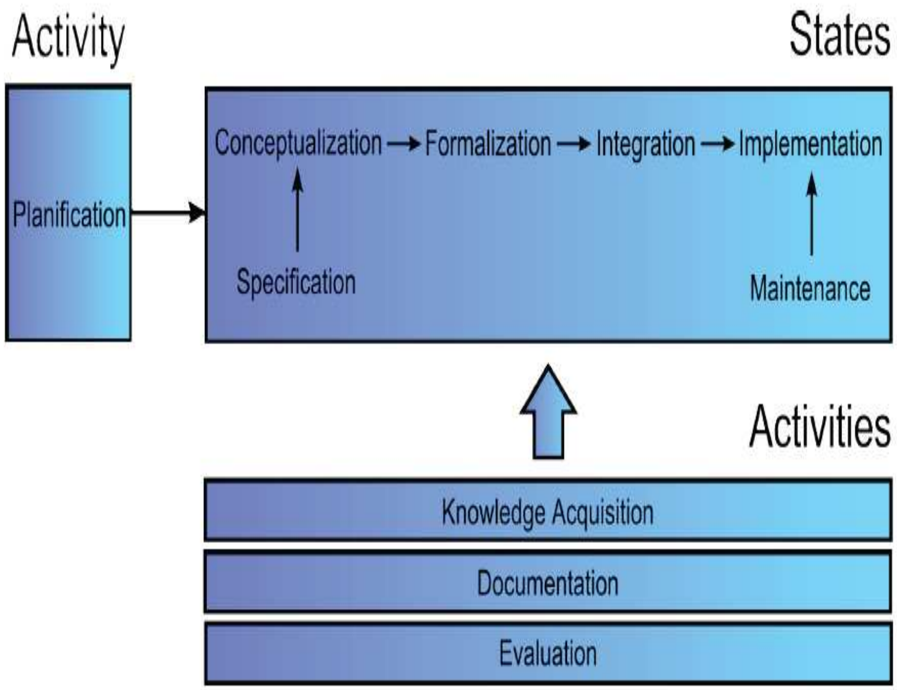


Figure 4.2: Activities and states in the life cycle of an Ontology [22]

• Knowledge extraction techniques used for the development of knowledge base systems should be applied to aggregate knowledge about the domain. The processes and techniques adopted for this purpose should be described accurately afterwards.

• After having acquired enough knowledge about the domain of discourse, a conceptual model is specified which describes the problems and its solutions.

• The conceptual model is then transferred into a formal or semi-compatible model with the use of an appropriate representation system.

• Reusing an existing Ontology should be considered as often as possible and hence suitable candidates should certainly be integrated into the Ontology that is created.

• On the basis of the previously specified formal or semi-compatible model, the vocabulary is implemented in a formal language that is best suited to represent it.

• A crucial precondition for reusing an Ontology is the existence of a detailed documentation. Thus, the activity of documenting an Ontology should be integrated throughout the whole Ontology development process. Without a proper documentation a specific Ontology is not going to be shared and reused on a broad base.

• The evaluation of an Ontology should always be carried out before the Ontology is made available to others. This can be accomplished by using a frame of reference on the basis of which technical judgments are made.

• As end users of the Ontology might ask for the modification or integration of a definition in the Ontology, maintenance is an important activity and the Ontology should be easy to maintain.

4.3 Usage of Ontology:

Information Integration

A promising field of application for ontologies is their use for integrating heterogeneous information sources on the schema level. Often, different databases store the same kind of information but adhere to different data models. An Ontology can be used to mediate between database schemas, allowing to integrate information from differently organized sources and to interpret data from one source under the schema of another [23].

Information Retrieval

Motivated by the success and key role of Google in the World Wide Web, information retrieval on web documents is a major field of application for ontologies. The idea behind Ontology-based information retrieval is to increase the precision of retrieval results by taking into account the semantic information contained in queries and documents, lifting keywords to ontological concepts and relations. When interpreted according to example geographic Ontology, a query like “capital of Germany” would yield documents that are about Berlin, the capital of Germany. Some of the false positive matches that keyword-based retrieval systems typically produce, such as documents about the German venture capital market, can be filtered out this way.

Semantically Enhanced Content Management

In many areas of computation the data that is actually computed is annotated with meta data for various purposes. Ontologies provide the domain-specific vocabulary for annotating data with meta data. The formality of Ontology languages allows for an automated processing of this meta data and their grounding in knowledge representation facilitates machine-interpretability.

Knowledge Management and Community Portals

In companies or other organised associations, or in communities of practice, individual knowledge can be viewed as a strategic resource that is desirable to be shared and systematically maintained, which is referred to as knowledge management. Ontologies provide a means to unify knowledge management efforts under a shared conceptual domain model, connecting technical systems for navigating, storing, searching and exchanging community knowledge.

Expert Systems

In various domains, such as medical diagnosis or legal advice in case-law, it is desirable to simulate a domain expert who can be asked sophisticated questions. In an expert system, this is achieved by incorporating a thoroughly developed domain Ontology that formalizes expert knowledge. Domain-specific questions can then be answered by reasoning over such highly specialized knowledge. An expert system for the geographical domain could answer questions like "Which is the German city closest to the French border?"Or "Through which cities does the river Rhein flow?"

4.4 Modularization of Ontologies

Ontologies is also reusing of knowledge. Once Ontology is created for a domain, it should be reusable for other applications in the same domain. To simplify both Ontology development and reuse, modular design is beneficial. The modular design uses inheritance of ontologies - upper ontologies describe general knowledge, and application ontologies describe knowledge for a particular application.

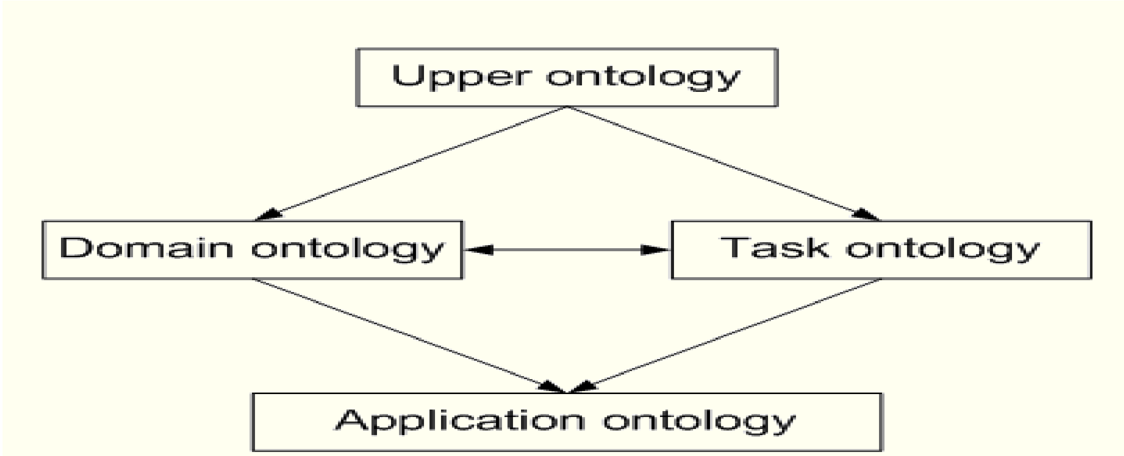


Figure 4.3: Modularization of Ontology [23]

Depending on the scope of the Ontology, Ontology may be classified as follows:

- upper, generic, top-level Ontology - describing general knowledge, such as what is time and what is space
- domain Ontology - describing a domain, such as medical domain or electrical engineering domain, or narrower domains, such as personal computers domain
- task - Ontology suitable for a specific task, such as assembling parts together
- application - Ontology developed for a specific application, such as assembling personal computers

Top-level Ontologies

Top-level ontologies – also called upper ontologies or foundational ontologies – attempt to describe very abstract and general concepts that can be shared across many domains and applications. They borrow from philosophical notions, describing top-level concepts for all things that exist, such as “physical object” or “abstract object”, as well as generic notions of common-sense knowledge about phenomena as time, space, processes, etc. They are usually well thought out and extensively axiomatized.

Domain Ontologies and Task Ontologies

These types of ontologies capture the knowledge within a specific domain of discourse, such as medicine or geography, or the knowledge about a particular task, such as diagnosing or configuring. In this sense, they have a much narrower and more specific scope than top-level ontologies. In the ideal case, the conceptualization in a domain Ontology is kept strictly task-independent, while the notions in a task Ontology are described neutrally with respect to a domain. Much work has been done in the development of domain ontologies in medicine, genetics, geographic and environment information, tourism, as well as cultural heritage and museum exhibits. Task ontologies have been devised, e.g., for scheduling and planning tasks, monitoring in a scientific domain, intelligent computer-based tutoring, missile tracking, execution of clinical guidelines, etc.

Application Ontologies

Further narrowing the scope, application ontologies provide the specific vocabulary required to describe a certain task enactment in a particular application context. They typically make use of both domain and task ontologies, and describe, e.g., the role that some domain entity plays in a specific task. For example, a particular physical entity in some engineering domain may play the role of a replaceable unit in a machine diagnosis and maintenance task, and at the same time play the role of a spare resource in a configuration or production process. Altogether, say that the lattice indicated in Figure 4.3 represents an inclusion hierarchy: the lower ontologies inherit and specialize concepts and relations from the upper ones. The lower ontologies are more specific and have thus a narrower application scope, whereas the upper ones have a broader potential for reuse.

4.5 Advantages of Developing Ontology

Major benefits of developing Ontology are: -

1. **Sharing common understanding of the structure of information among people or software agents**: It is one of the most common goals in developing ontologies. For example several different web sites contain medical information or provide e-commerce services. These web sites share and publish the same underlying Ontology of the terms they use, and then computer agents can extract and aggregate information from different sites. The agents can use this aggregated information to answer user queries or as input data to other applications.

2. **Enabling reuse of domain knowledge:** It is one of the driving forces behind recent surge in Ontology research. For example, models for many different areas need to represent the opinion of time. This representation includes the planning of time intervals, points in time, relative measures of time, etc. If one group of researchers develops such Ontology in detail, others can simply reuse it for their area. To build a large Ontology need to integrate several existing ontologies describing portions of the large domain. General Ontology can be reused and extend it to describe desired Ontology.

3. **Making explicit domain assumptions:** Explicit specifications of domain knowledge are useful for new users who want to learn the specific area.

4. **Separating the domain knowledge from the operational knowledge**: It is another common use of ontologies. It describes a task of configuring a product from its components according to a required specification and implements.

5. Analyzing domain knowledge: Analyzing is possible once a declarative specification of the terms is available. Formal analysis of terms is valuable when attempting to reuse existing ontologies and extending them for solving problem purposes [23].

4.6 Ontology Mapping:

Ontology mapping is also known as Ontology matching, or Ontology alignment. Ontology mapping is different from Ontology merging. Ontology mapping tries to make the source ontologies consistent and coherent with one another while keeping them separate. In contrast Ontology merging aims to create a single coherent Ontology that includes the information from all the sources. There are many different definitions of Ontology mapping, depending upon its application and its intended outcome. Sample definitions of Ontology mapping include:

1. Ontology mapping is “a set of formulae that provide the semantic relationships between the concepts in the models” [28].

2. Ontology mapping is used to “establish correspondences among the source ontologies, and to determine the set of overlapping concepts, concepts that are similar in meaning but

have different names or structure, and concepts that are unique to each of the sources" [29].

3. Ontology mapping aims to "map concepts in the various ontologies to each other, so that a concept in one Ontology corresponds to a query (i.e. view) over the other ontologies" [30]

4. "Given two ontologies O1 and O2, mapping one Ontology onto another means that for each entity (concept C, relation R, or instance I) in Ontology O1, we try to find a corresponding entity, which has the same intended meaning, in Ontology O2".

Take a simple example imagine the concepts of *apple* as a fruit and *apple inc.* as a company are modeled into one Ontology, while *apple* as a company is modeled into a second Ontology. In a communication dialogue where one person equipped with the second Ontology attempts to convey the concept of *apple* as a company to another person equipped with the first Ontology, the process of finding semantic correspondence between *apple* as a company (in the second Ontology) and *apple inc.* as a company (in the first Ontology) over the concept pair of *apple* as a company (in the second Ontology) and *apple* as a fruit (in the first Ontology) is referred as Ontology mapping.

Ontology mapping is the process of identifying correspondences between two ontologies. It is seen as the solution provider to the problem of semantic interoperability.

"Mapping could provide a common layer from which several ontologies could be accessed and hence could exchange information in semantically sound manners".

The process of Ontology mapping embraces a set of components that includes alignment, alignment representation, alignment discovery, articulation, translation and merging or integration of ontologies.

An Ontology alignment normally refers to a collection of mapping relations between concepts of two ontologies. The task of Ontology mapping is to discover an alignment between two ontologies. Different representation schemes (e.g. numeric) are used for encoding Ontology alignments depending on the nature of the member mapping relations. Articulation, on the other hand, refers to an intermediate Ontology bridging two ontologies where mapping between the two ontologies is expressed as a pair of mapping processes: between the first Ontology and the articulation, and between the second Ontology and the articulation. Ontology translation is the process of applying the Ontology mapping between two ontologies in translating instances from each other. Merging or integration of ontologies refers to the fusion or composition of ontologies to build new ones. Articulation is often used to define a way in which the fusion is to be executed such that the resultant merged Ontology semantically respects the articulation.

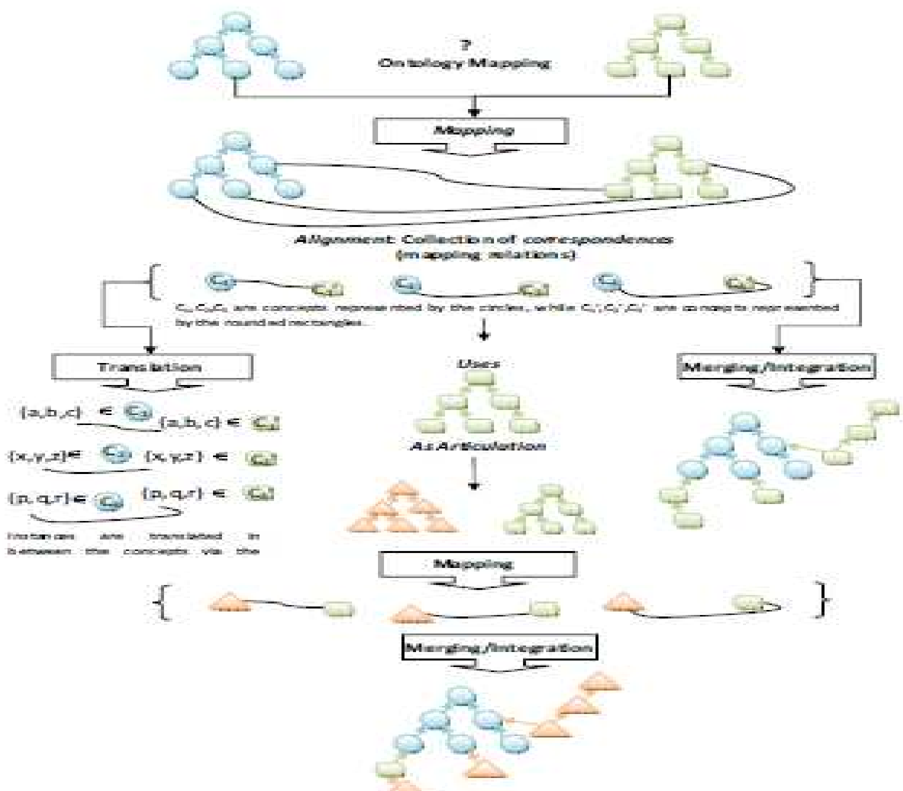


Figure 4.4: Overview of the components related to Ontology Mapping [24]

4.7 Mapping Operations:

When talking about Ontology interoperability the following, are considered relevant Operations:

Ontology mapping/matching

Ontology alignment

Ontology translation

Ontology transformation

Ontology merging/integrating

Ontology checking

4.7.1 Ontology mapping

Mapping from one Ontology to another one is expressing of the way how to translate statements from Ontology to the other one. Often it means translation between concepts and relations. In the simplest case it is mapping from one concept of the first Ontology to one concept of the second Ontology.

Establishing mappings between two ontologies means that for each entity (concept, relation, attribute,etc)in one Ontology we try to find a corresponding entity in the second Ontology, with the same or the closest intended meaning. Mappings can be established after an analysis of the similarity. The mapping process does not modify the involved ontologies and produce, as output, only a set of corespondences.

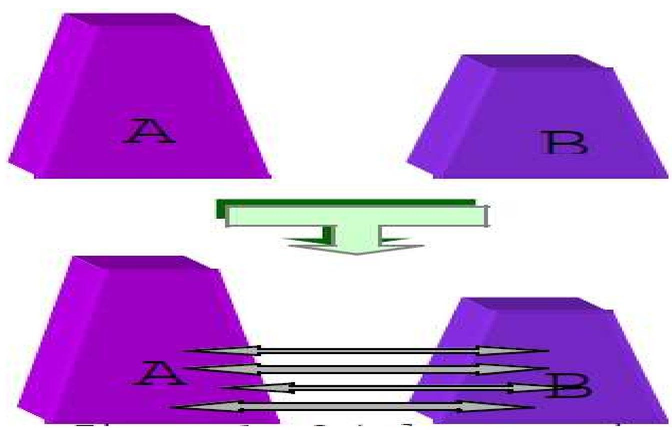


Figure 4.5: Ontology Mapping [24]

4.7.2 Ontology alignment

Ontology alignment as the process of bringing two or more ontologies into mutual agreement, making them consistent and coherent with one and another; this process may require a transformation of the involved ontologies eliminating the “non-needed” information while the missing information must be integrated. In contrast with mapping, this operation might result in changes of one or more of the involved ontologies

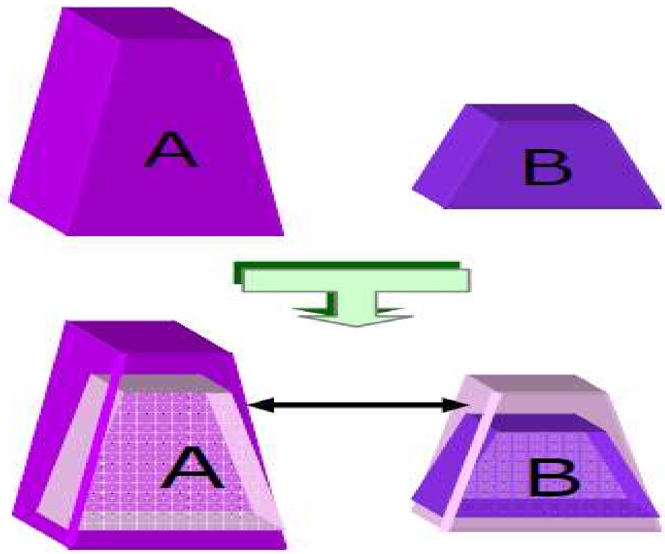


Figure 4.6: Ontology Alignment [25]

4.7.3 Ontology translation

Sometimes there's the need to change the formalism in which a particular Ontology is expressed, for example if decide to reuse an Ontology (or part of an Ontology) using a tool or language that is different from those ones in which the Ontology is available; a good translation will leave the semantics of the translated Ontology unaltered, or as closest as possible, to the original.

4.7.4 Ontology transformation

This process consist in changing the structure of an Ontology leaving unaltered its semantics (lossless transformations) or modifying it slightly (if we have a loss of information we talk about "lossy" transformations) to make it suitable for different purposes other than the original one.

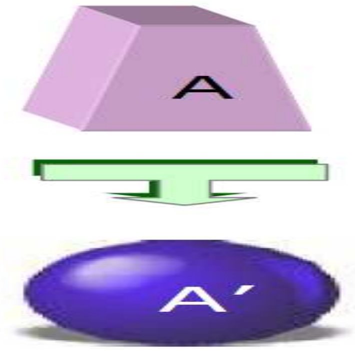


Figure 4.7: Ontology translation/transformation [24]

4.7.5 Ontology merging/integrating

When build a new Ontology starting from two or more existing ontologies then talk about Ontology merging while when reusing existing ontologies, assembling extending and specializing them, this is usually referred as Ontology integration

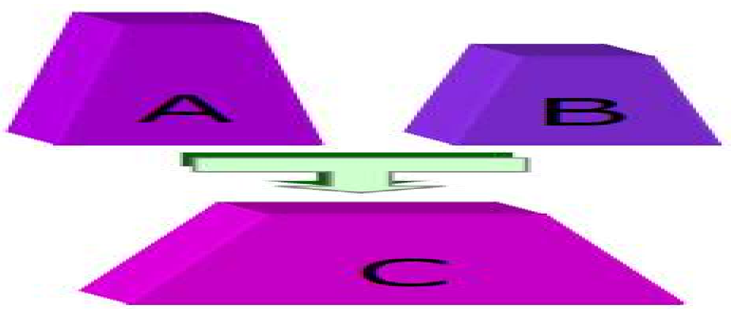


Figure 4.8: Ontology Merging [24]

4.7.6 Ontology checking

Once an Ontology has been created (as the result of an integration/merging) or transformed (as the result of a translation/transformation/alignment), the result must be checked in order to identify possibly inconsistencies or information losses. Validation of an ontologies is commonly performed by reasoners.

4.8 Ontology Mismatches:

Heterogeneity between ontologies are called Ontology mismatches and are categorized into: language level and Ontology level mismatches. Mismatches at the language level occur when ontologies written in different Ontology languages are combined. Language level mismatches are about the difference in language syntax, semantics of language constructs, and formalism mechanisms Mismatches at the Ontology level happen when two or more ontologies that describe partly overlapping domains are combined. The Ontology level mismatches occur because of differences in conceptualization and explication of ontologies. A conceptualization or explication mismatch is a difference in the way a domain is interpreted or specified [26]. The major Ontology level mismatches are as follows:

Domain of description: This is a mismatch in the level of detail to which that domain is modeled. For example in the university domain one Ontology might represent staff but only classify them into a few categories, while another Ontology might make further distinctions between categories of staff.

Terminology: two ontologies may use different names to represent the same term (e.g. "price" and "cost"), or the same name to represent different terms.

Paradigm: Different paradigms can be used to represent terms (e.g. "date of birth" and "age").

In order to realize the interoperation of heterogeneous ontologies, there are methods [31]:

Building inclusion relations between ontologies.

Object Ontology simply includes source Ontology. All data and concepts exist in both source Ontology and object Ontology.

The shortage of this method is that the information and concepts only be reused and can't be revised.

Multiple Ontology Approach

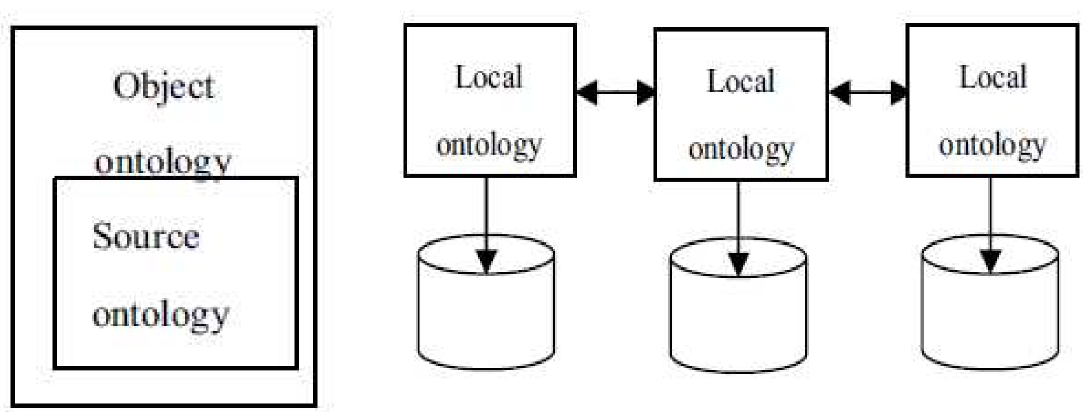
Source data would have their own ontologies and no shared vocabulary would exist among them. Inter-Ontology mapping for definitions has to be determined before data can be shared across local ontologies. In other words static mappings are required. Since a global Ontology is not maintained here, the semantic richness of this method is inferior to the hybrid approach

Domain Ontology Approach

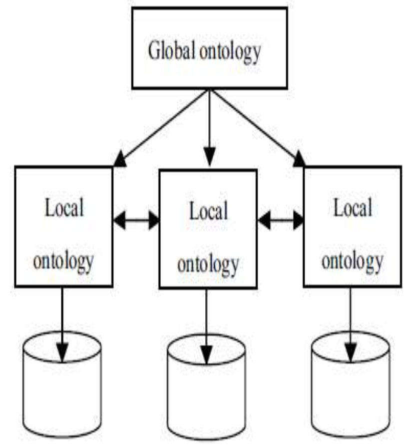
Domain Ontology includes all ontologies and data sources. That is to say, Looking for a global Ontology. It can be inquired in any circumstance. Global Ontology affords a shared vocabulary for specific semantic explanation. All information sources are related to global Ontology. Therefore they have same semantic. The shortage of this method is that it is not semantically rich and does not support inter-Ontology mappings of local schemas so it is difficult to realize.

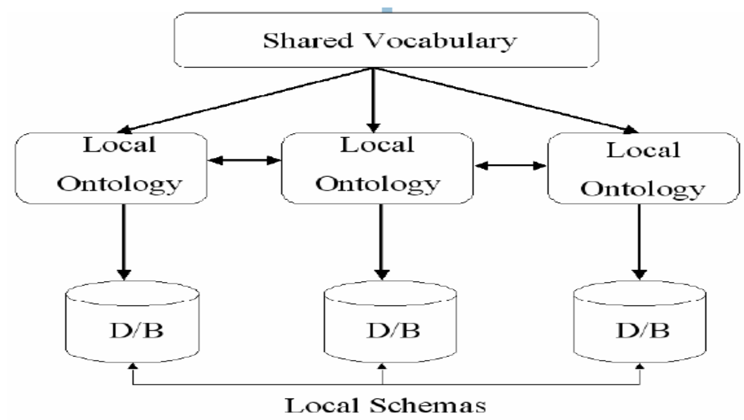
Hybrid Approach

Autonomy of the local schemas is maintained as the source data has its own Ontology. Shared vocabulary is achieved via inter-Ontology mapping. Inter-Ontology mappings must be defined and this helps capture richer semantics. This method is promising and is quite different from conventional approaches. A simple formalism is provided here to capture only the domain knowledge pertaining to potential semantic conflicts. The advantage of this simplified Ontology is that it is not domain-specific and does not lose semantic richness. Local ontologies can maintain domain specific definitions and at the same time have the flexibility to share definitions from an upper-Ontology.



a) Ontology Inclusion b) multiple Ontology approach





c) Domain Ontology approach d) Hybrid approach

Figure 4.9: Shows the Architect of Methods.

4.9 Ontology Development Tool:

A number of Ontology development tools like Protégé-2000, Oiled, Apollo, Ontosaurus, OntoStudio etc.[27] are available for knowledge elicitation in the fulfillment of dream of new web i.e. Semantic Web. Each tool has its own advantage and disadvantage. This thesis mainly analyze on the work of Protégé tool.

Protégé tool

An Ontology model of Java program is created according to Java program specification using Protégé .Protégé is the latest tool in an established line of tools developed at Stanford University for knowledge acquisition. It allows users to construct domain Ontology, to customize knowledge–acquisition forms and to enter domain knowledge. It is able to operate as a platform to extend access to other knowledge based systems, embedded applications, or as a library which can be used by other applications to access and visualize knowledge bases. It has a powerful tailorability and it enables knowledge acquisition. Protégé offers a graphical user interface which allows Ontology developers to focus on conceptual modeling without requiring to know syntax of an output language, such as RDFS. It helps knowledge engineers and domain experts to perform knowledge management tasks.

Ontology developers can access relevant information quickly whenever they need it, and can use direct manipulation to navigate and manage an Ontology. Tree controls allow quick and simple navigation through a class hierarchy. Protégé uses forms as the interface for filling in slot values. The knowledge model of Protégé is JDBC-compatible. It includes support for classes and the class hierarchy with multiple inheritance; template and own slots; specification of pre-defined and arbitrary facets for slots, which include allowed values, cardinality restrictions, default values, and inverse slots; meta classes and meta class hierarchy.

In addition to highly usable interface, two other important features distinguish Protégé from most Ontology-editing environments: its scalability and extensibility.
The first and most important problem was the lack of an appropriate reasone that could be plugged into a rule-enriched Ontology created with Protégé. and entering data proved very tedious when making instances of a superclass belong to an orthogonal subclass partition.

4.10 Proposed Model:

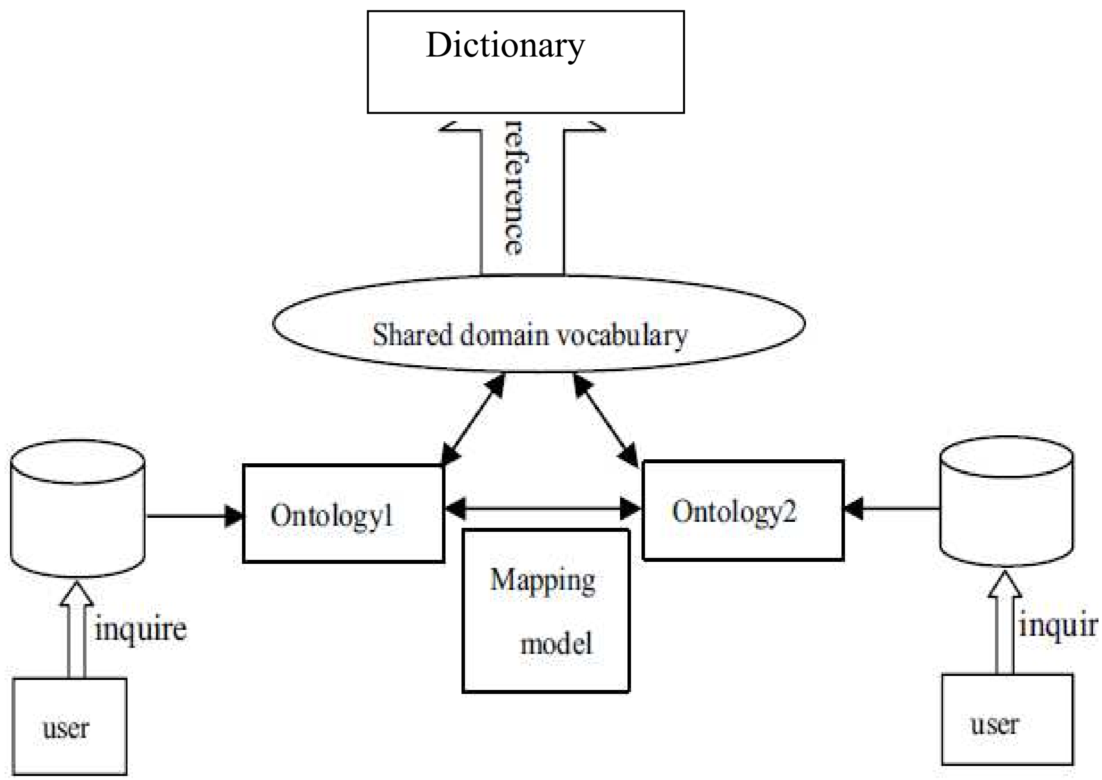


Figure 4.10: Proposed Model

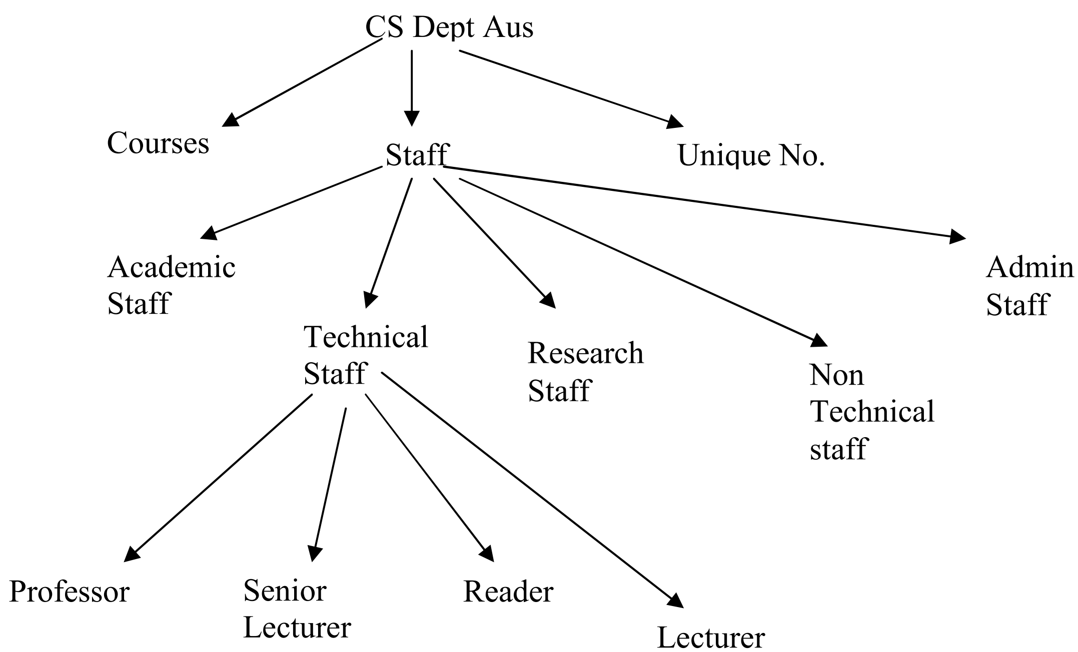
In this proposed approach use the synonyms. In this method, external help is mostly employed in the form of a dictionary, thesaurus for a particular domain. In this case dictionary, acts as the intermediary between the ontologies to be mapped as it will store words with similar meanings.

Components of Proposed Model are:

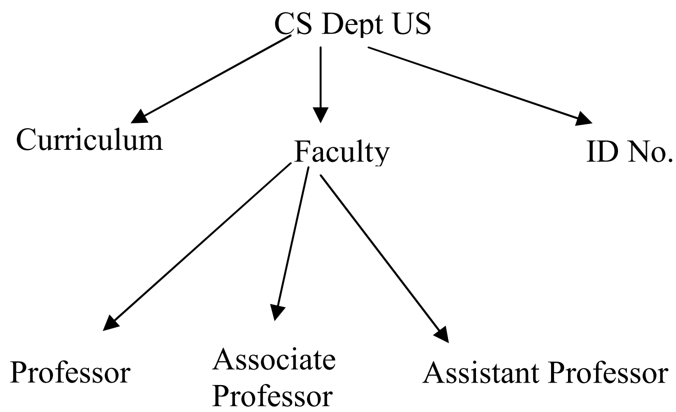
(1) Shared domain vocabulary: It is a whole term vocabulary of shared domain system. Every row includes synonym. When computing concept similarity, refer shared domain vocabulary and only compute similarities of synonym.

(2) Local Ontology: Different users build their local Ontology according to own need. The construction of local Ontology refers the shared domain vocabulary. Therefore, there are semantic relations among concepts.

(3) Mapping model: the task of this model is computing conceptual similarities. When computing, semantic similarity and describe similarity are computed with improved computing method. Semantic similarity describes the degrees of similarity between two concepts. Descriptive similarity explains the similarity of concepts from the view of attributes and relations.



Ontology 1



Ontology 2

Figure 4.11: Architecture of Ontology

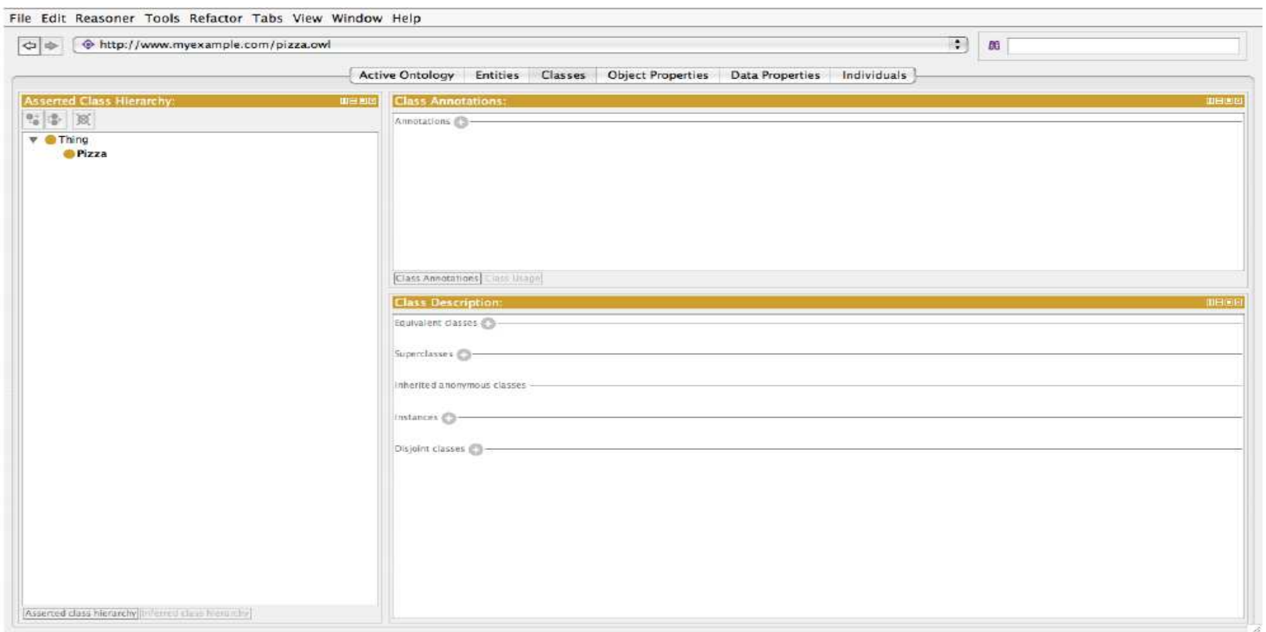
If a user of O1 wants some information about a concept in O2, there is a mapping between O1 and O2. Taking Dictionary as reference and obtains a shared vocabulary of concepts, as table: At first, using shared thesaurus vocabulary and obtaining the original similarity of thesaurus. Then further compute similarities of concepts.

People	Staff , Academic Staff, Admin Staff, non technical staff, research staff, technical staff
Job	Professor, Senior Lecturer, Reader, Lecturer, Chairman, Vice chancellor, Assistant, Dean, Lab Incharge, Peon, Scientist grade, System Analyst, Team Leader
Course	Under Graduate Course, Post Graduate Course
Name	First Name, Last Name
Unique Number for student	Id, Roll No. Identification No,Registration No.

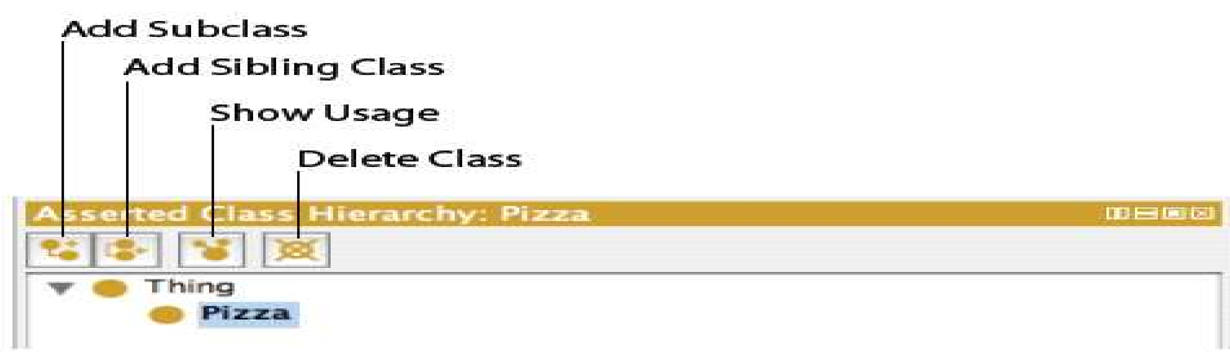
Table 4 1: Shared Vocabulary

Chapter 5
Implementation and Experiment Results

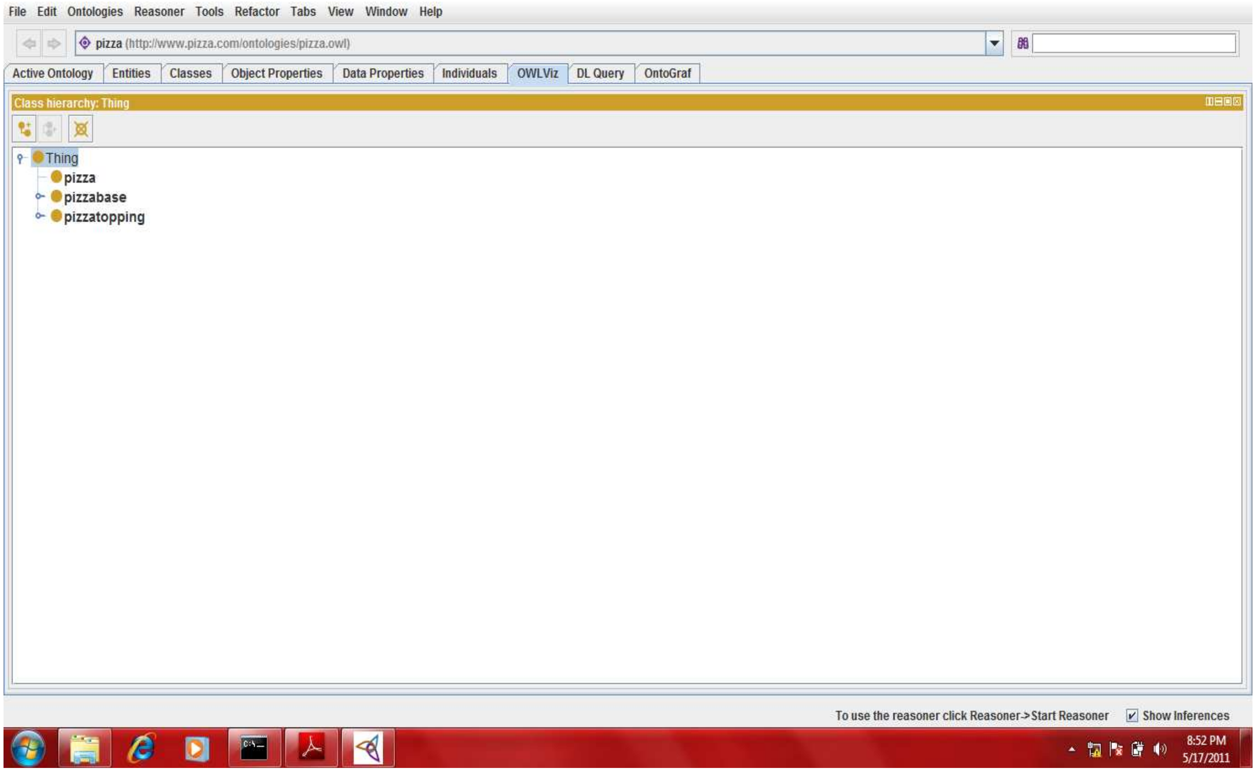
5.1 Protégé tool



Snapshot 5.1 The Class Tab

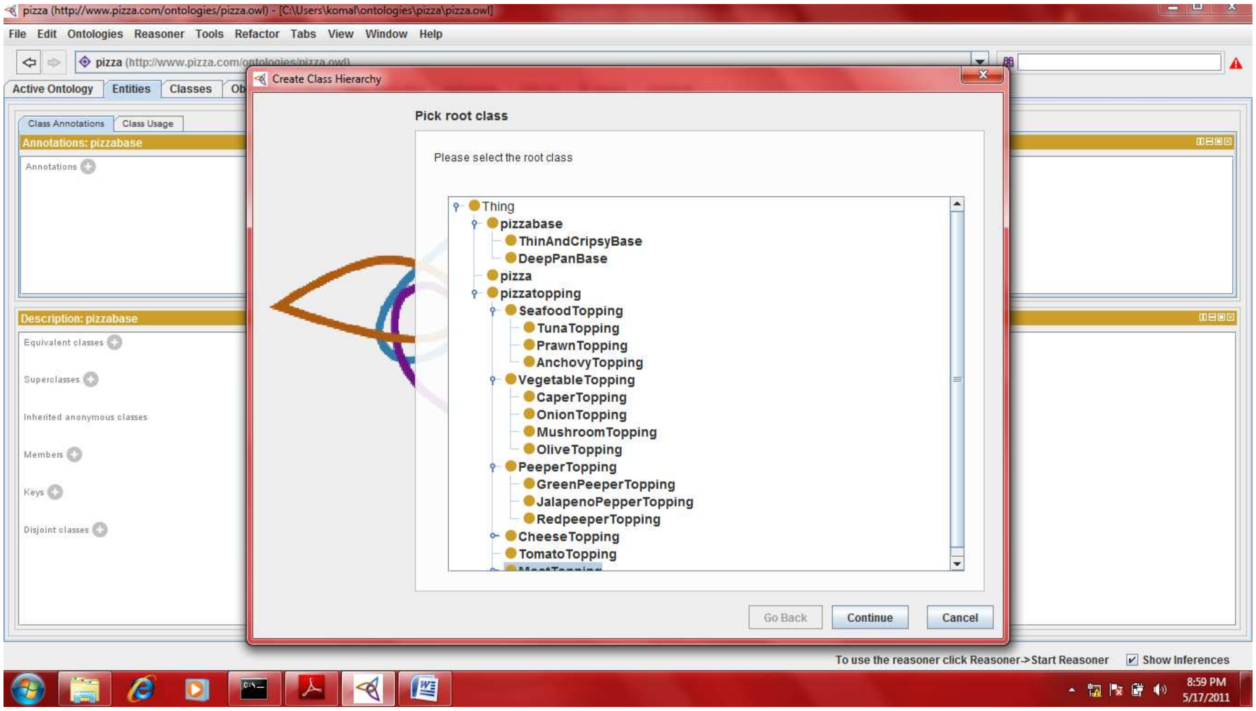


Snapshot 5.2: The Class Hierarchy Pane



Snapshot 5.3 Create named class

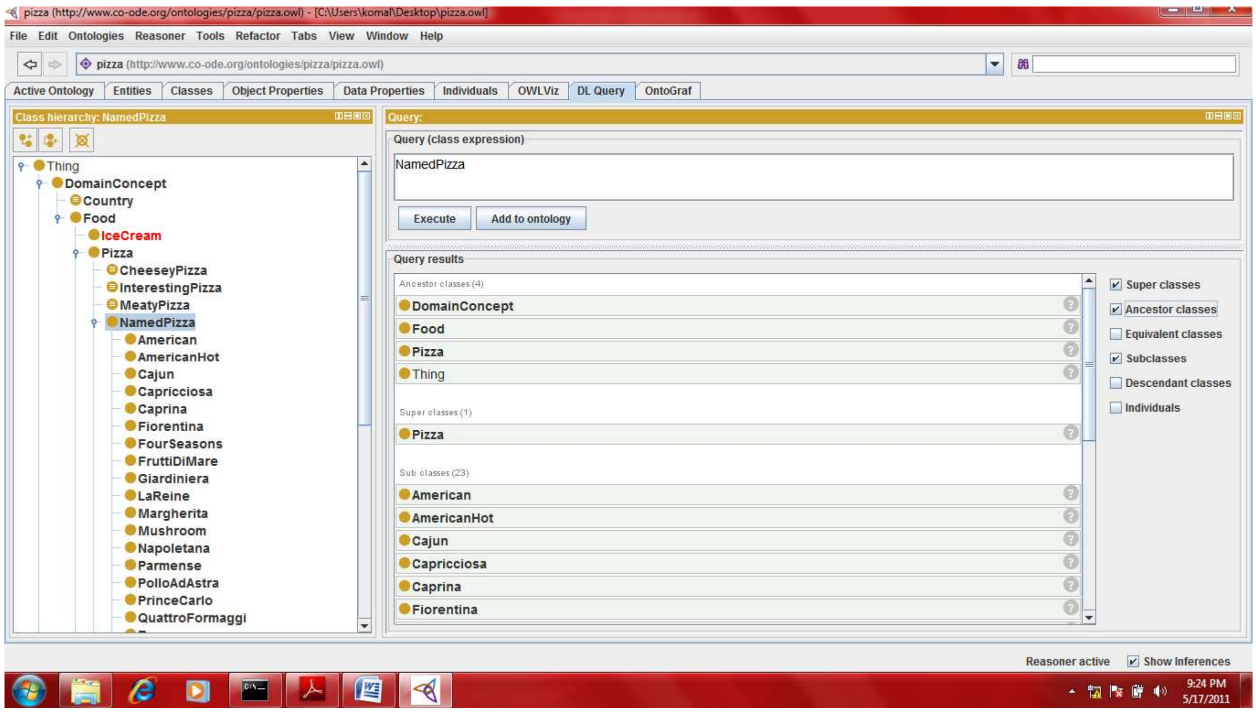
Select the class tab use the 'Create subclass' and 'Create sibling class' buttons to create Pizza, PizzaBase and PizzaTopping.



Snapshot 5.4 Create Subclasses

Create some subclasses of PizzaTopping to represent high level categorizations of pizza toppings.

Add additional classes to represent different kinds of pizza toppings.
Seafood Topping, Vegetable topping, Peeper Topping, Cheese Topping.



Snapshot 5.5: Query Pane

In DL Query first start Reasoner and then select query.

5.2 Pseudocode of our Proposed Model:

Step 1: Input from the user.

Step 2: "Select * from vewDictonary”

Step 3: If (MappingString.Length = 0) Then

    MsgBox("Dictonary does not have this word")

    Exit Sub

Step4: Else

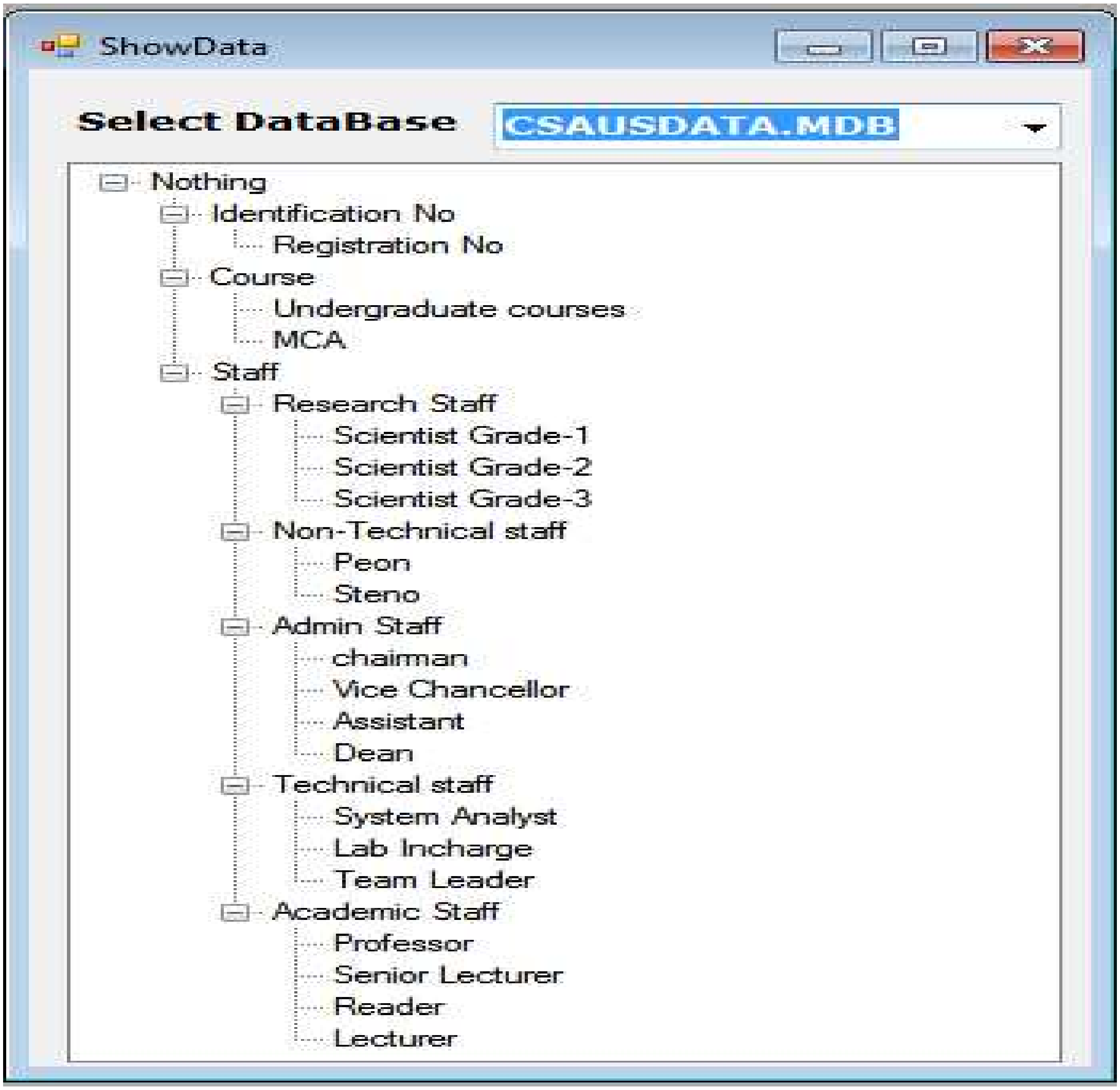
(chkNearWord.Checked = True) then

("Select distinct WordName,childword as 'synonym in US')

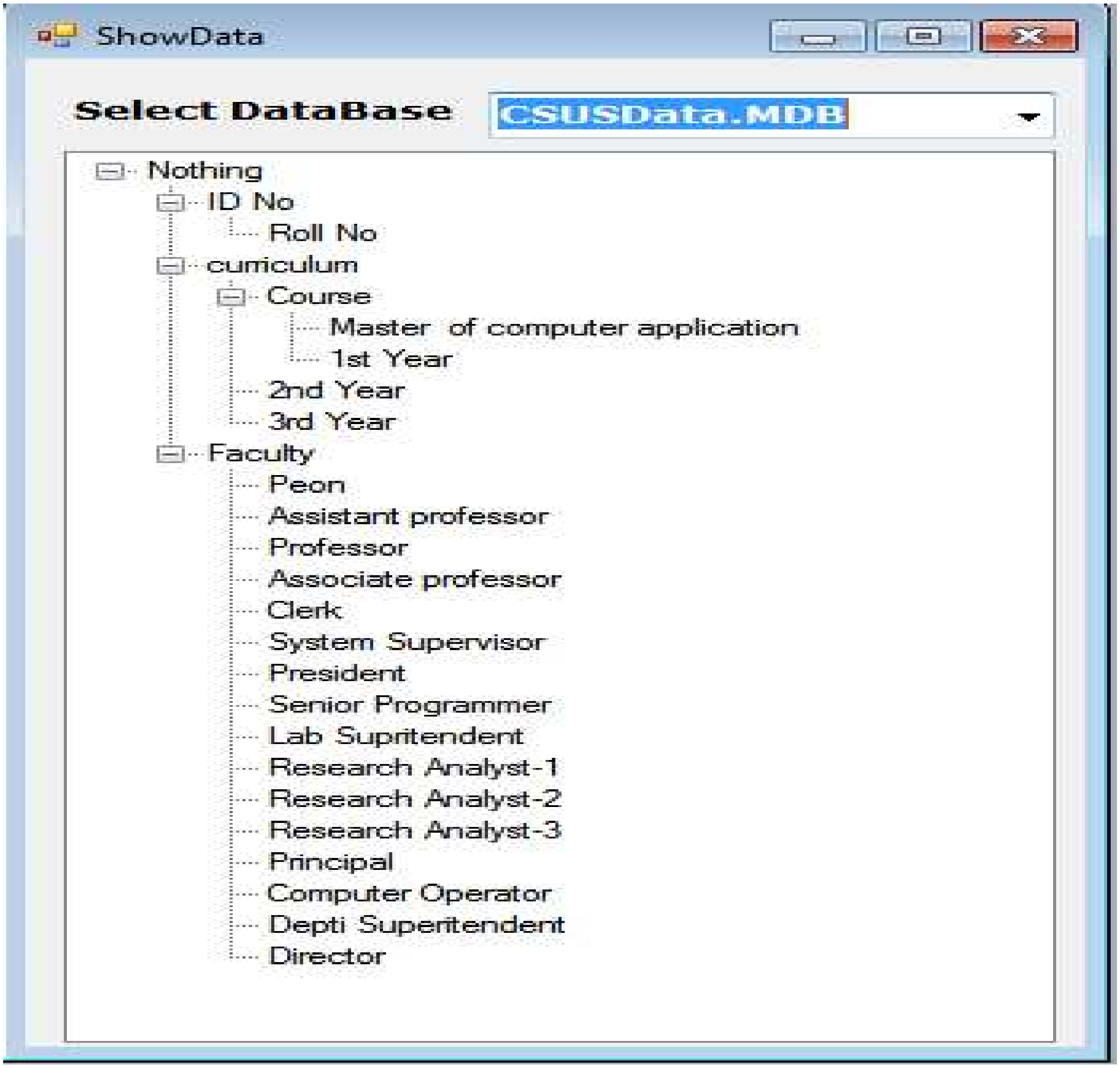
  Else if

    ("Select distinct WordName,childword as 'synonym in AUS').

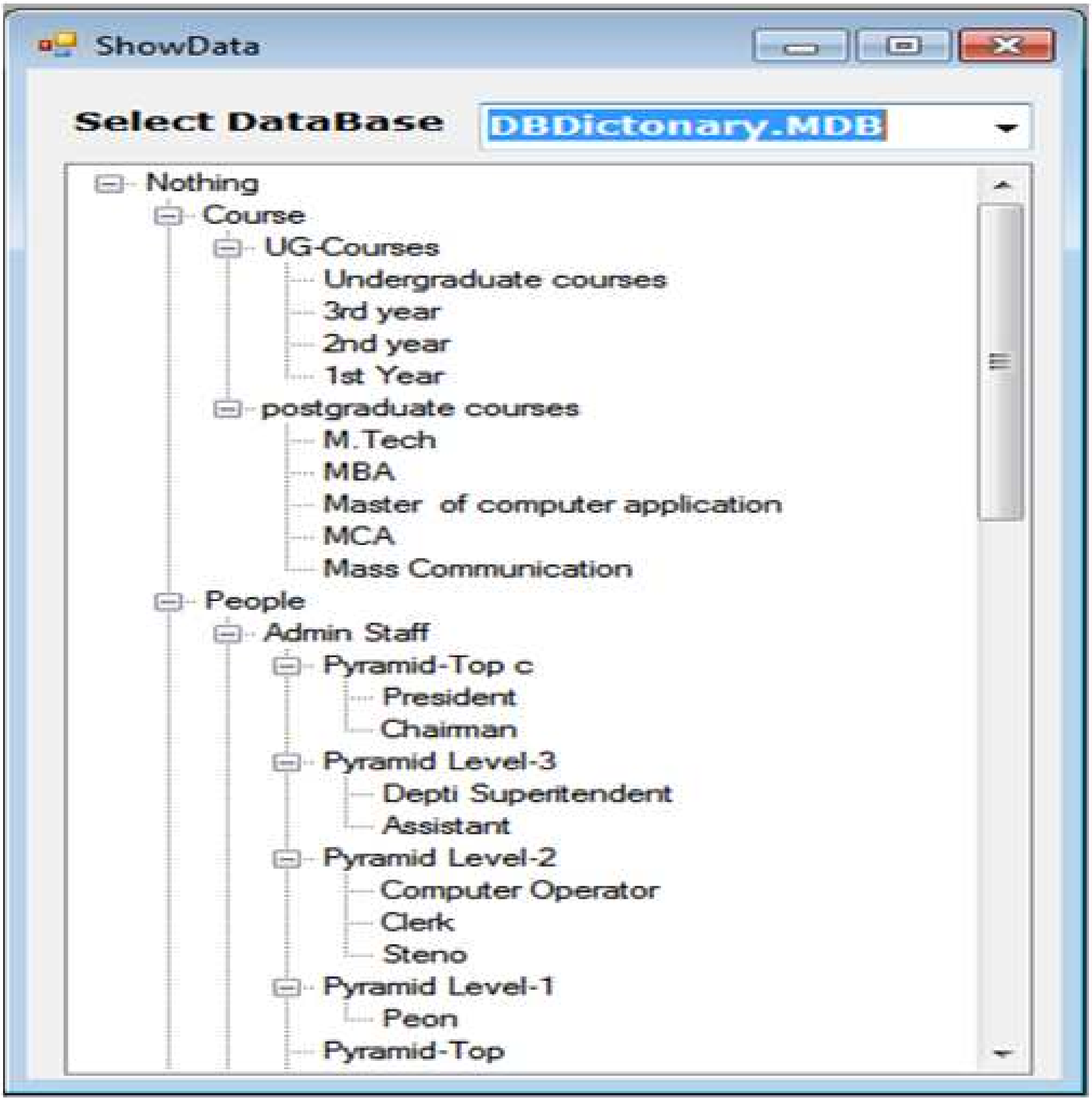
Step5:Print the word.



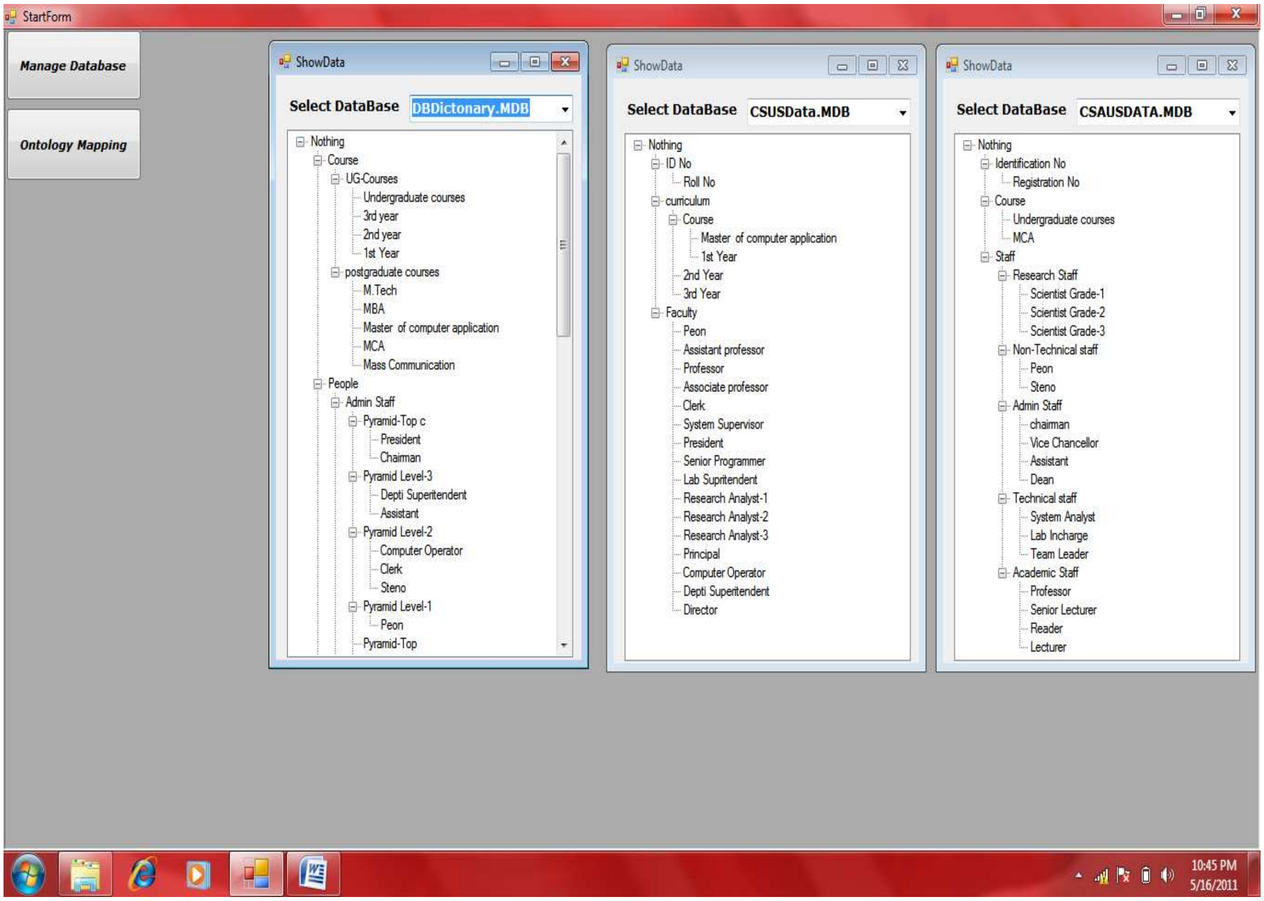
Snapshot 5.6: AUS Data



Snapshot 5.7: US Data

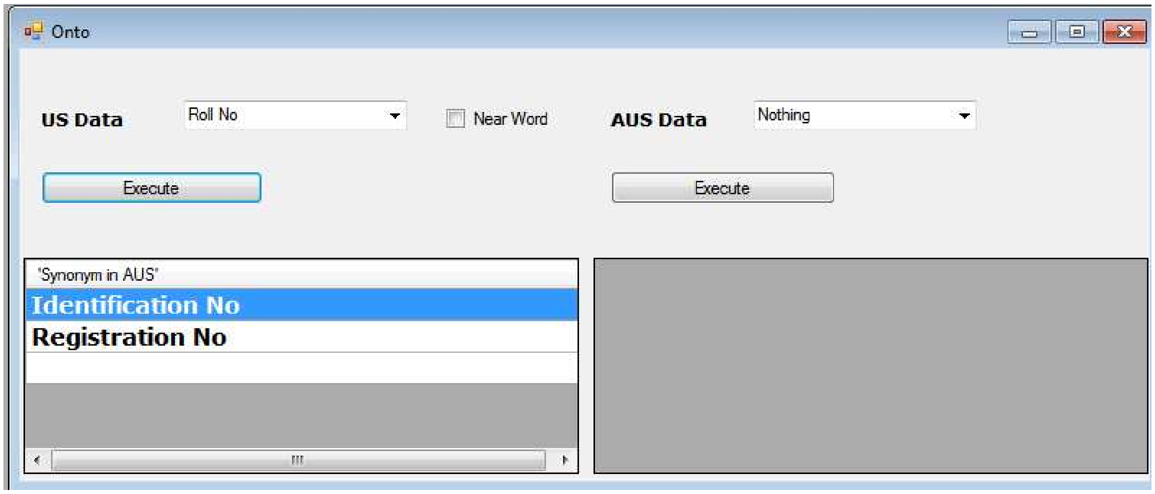


Snapshot 5.8: Dictionary

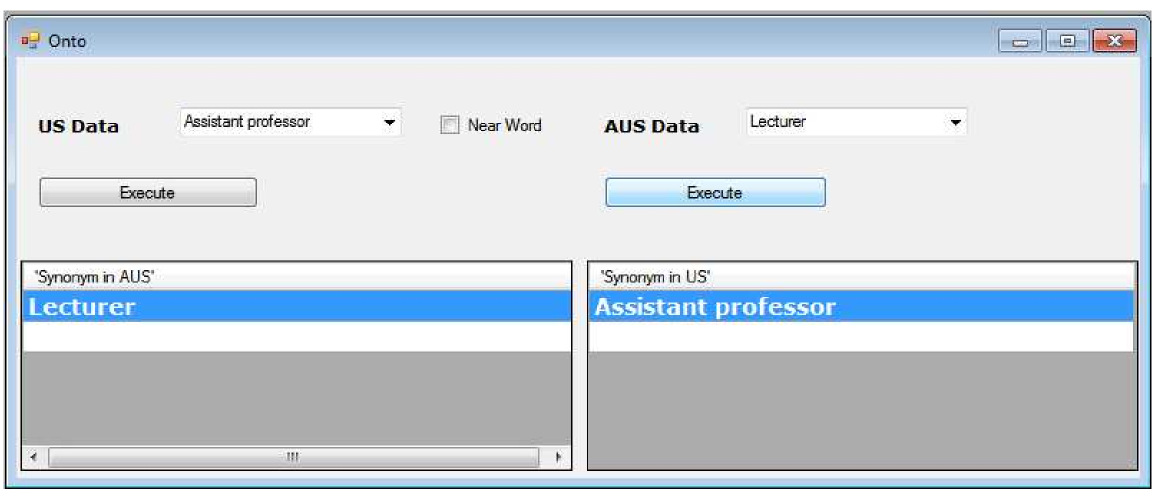


StartForm
Manage Database
Ontology Mapping
ShowData
Select DataBase
DBDictonary.MDB
Nothing
Course
UG-Courses
Undergraduate courses
3rd year
2nd year
1st Year
postgraduate courses
M.Tech
MBA
Master of computer application
MCA
Mass Communication
People
Admin Staff
Pyramid-Top c
President
Chairman
Pyramid Level-3
Depti Superitendent
Assistant
Pyramid Level-2
Computer Operator
Clerk
Steno
Pyramid Level-1
Peon
Pyramid-Top
ShowData
Select DataBase
CSUSData.MDB
Nothing
ID No
Roll No
curriculum
Course
Master of computer application
1st Year
2nd Year
3rd Year
Faculty
Peon
Assistant professor
Professor
Associate professor
Clerk
System Supervisor
President
Senior Programmer
Lab Supritendent
Research Analyst-1
Research Analyst-2
Research Analyst-3
Principal
Computer Operator
Depti Superitendent
Director
ShowData
Select DataBase
CSAUSDATA.MDB
Nothing
Identification No
Registration No
Course
Undergraduate courses
MCA
Staff
Research Staff
Scientist Grade-1
Scientist Grade-2
Scientist Grade-3
Non-Technical staff
Peon
Steno
Admin Staff
chairman
Vice Chancellor
Assistant
Dean
Technical staff
System Analyst
Lab Incharge
Team Leader
Academic Staff
Professor
Senior Lecturer
Reader
Lecturer
10:45 PM
5/16/2011

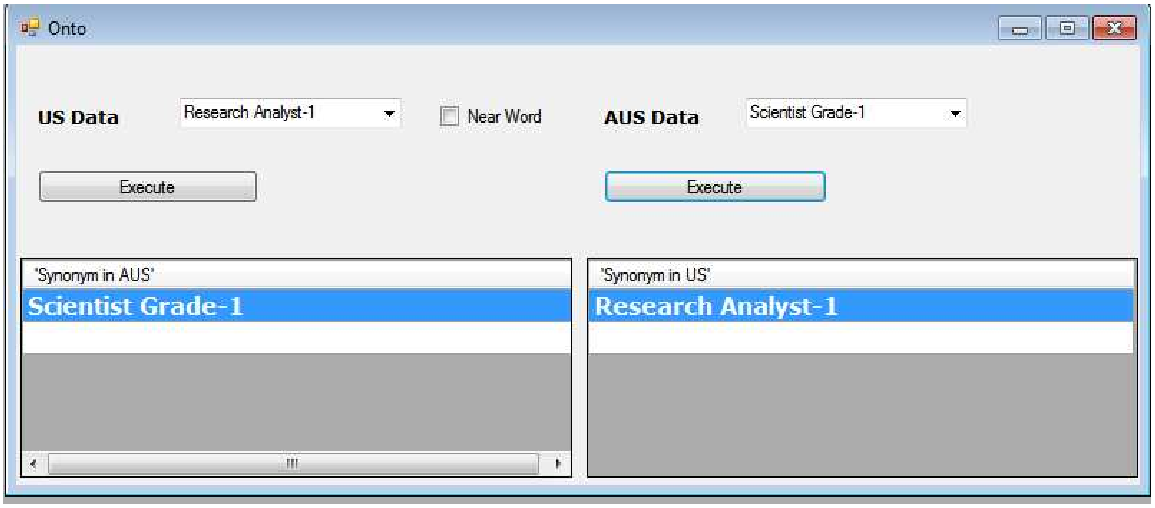
Snapshot 5.9



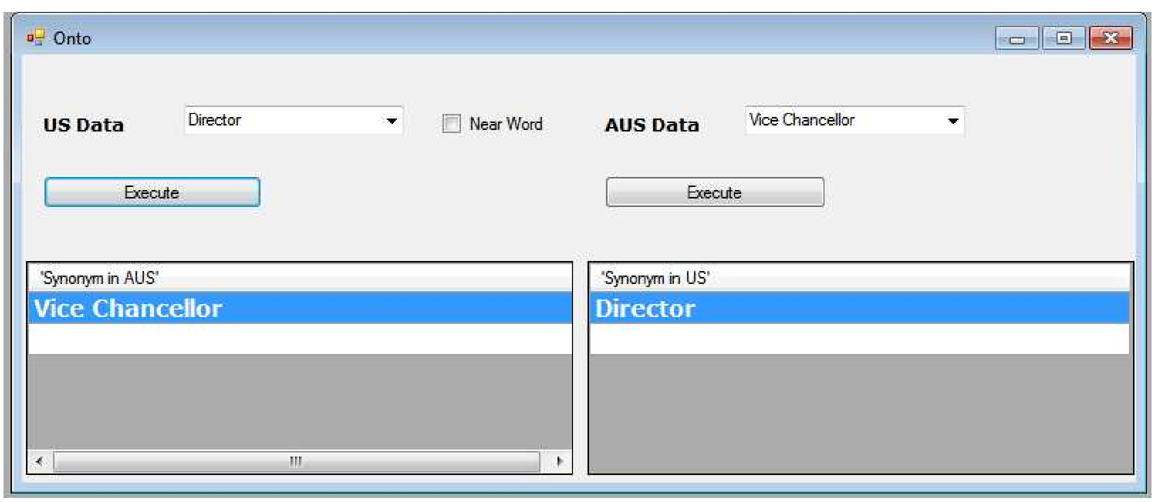
Snapshot 5.10:Roll No in US smilar to Identification No and Registration No in Australia



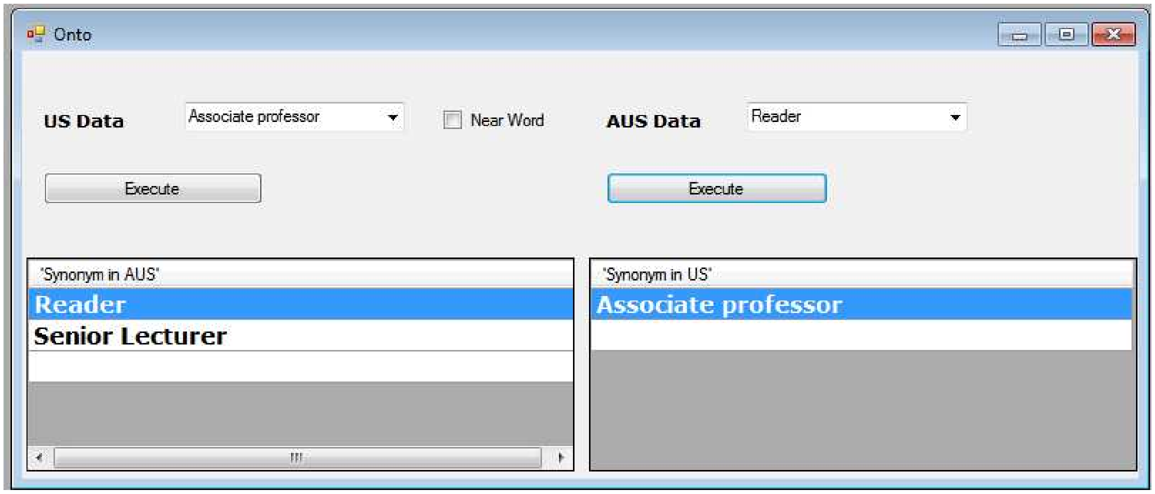
Snapshot 5.11: Assistant Professor in US smilar to lecturer in Australia



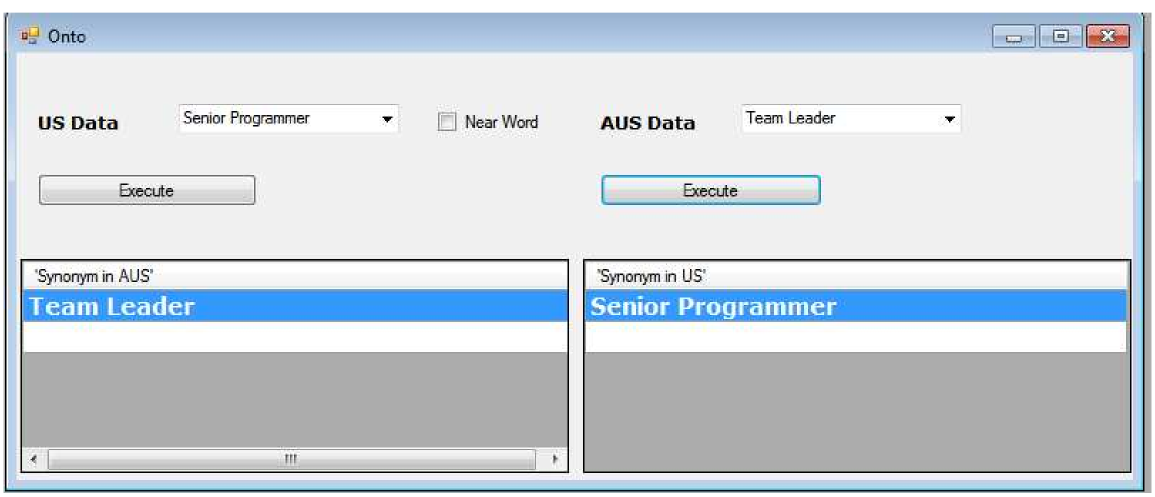
Snapshot 5.12: Research Analyst-1 in US smilar to Scientist Grade-1 in Australia



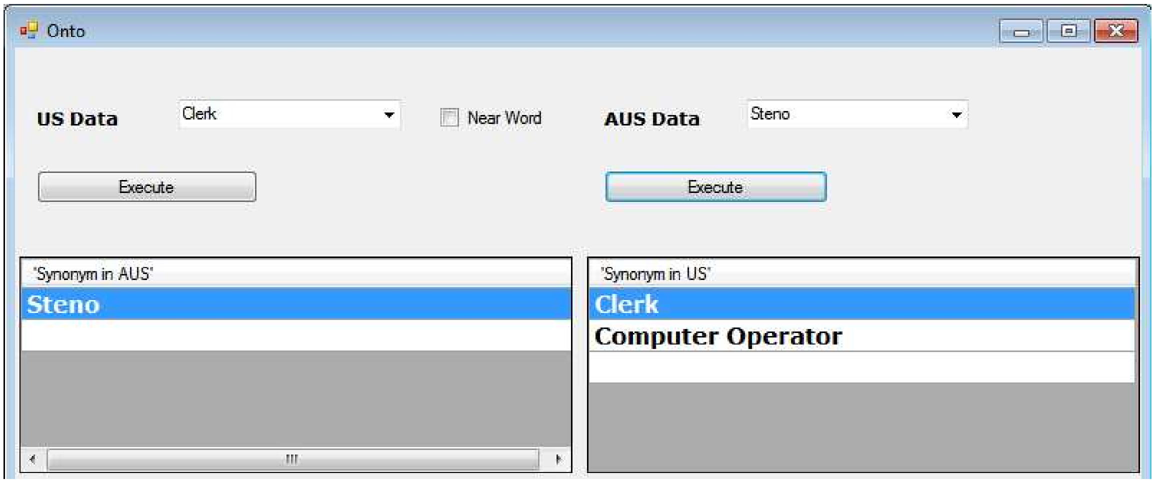
Snapshot 5.13: Director in US smilar to Vice Chancellor in Australia



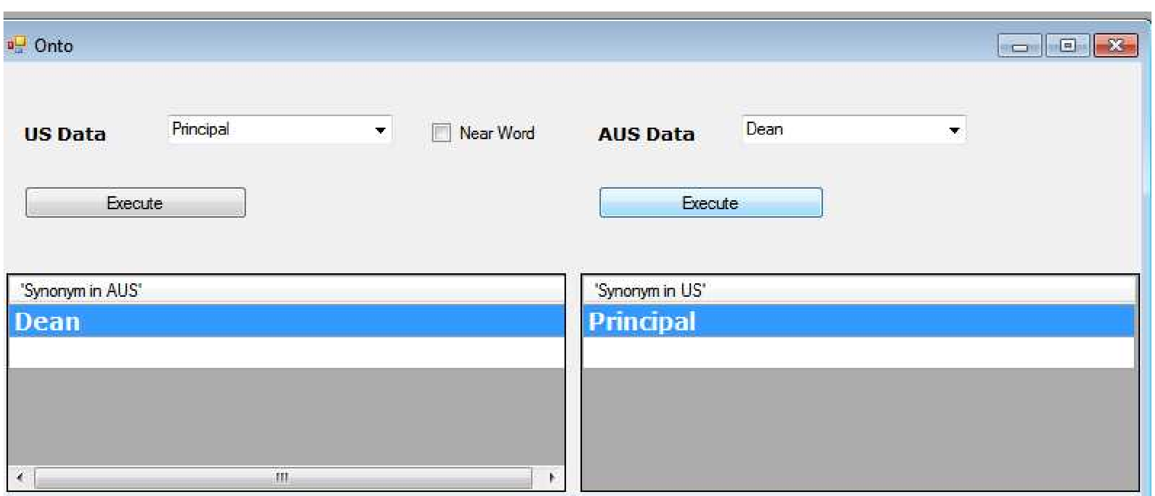
Snapshot 5.14: Associate Professor in US smilar to Reader and Senior Lecturer in Australia



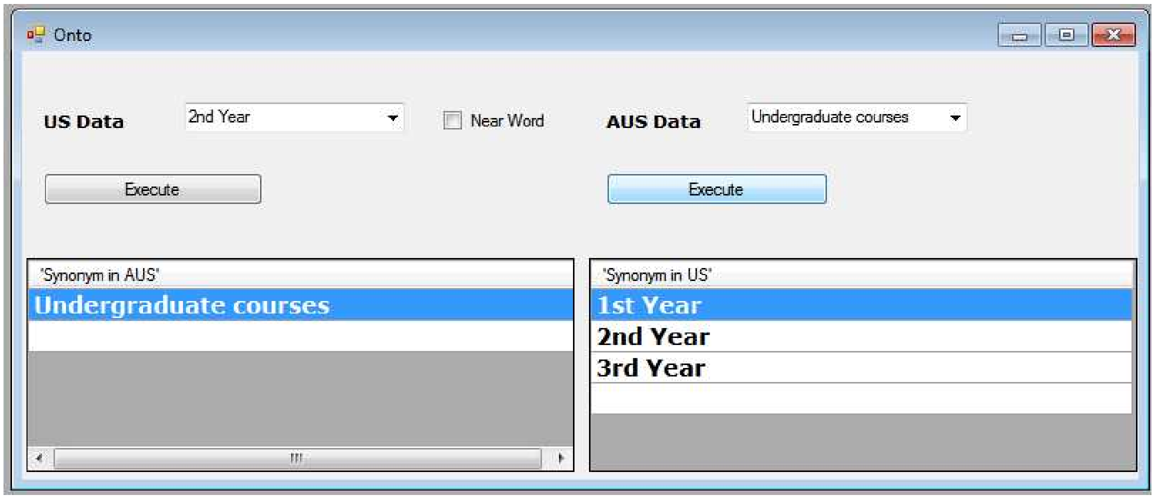
Snapshot 5.15: Senior Programmer in US smilar to Team Leader in Australia



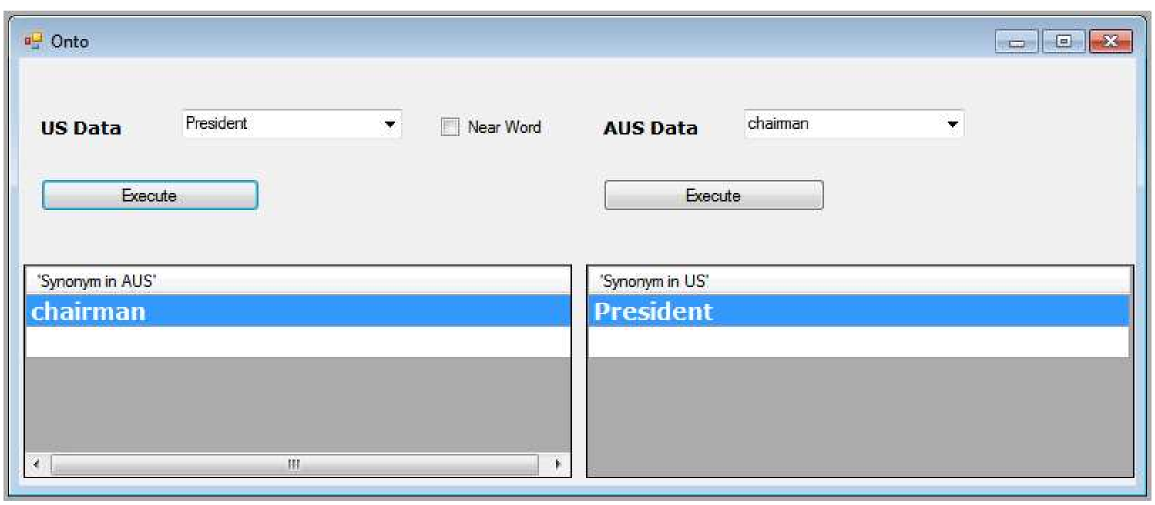
Snapshot 5.16: Clerk in US smilar to Steno in Australia & steno in Aus similar to clerk and computer operator in US.



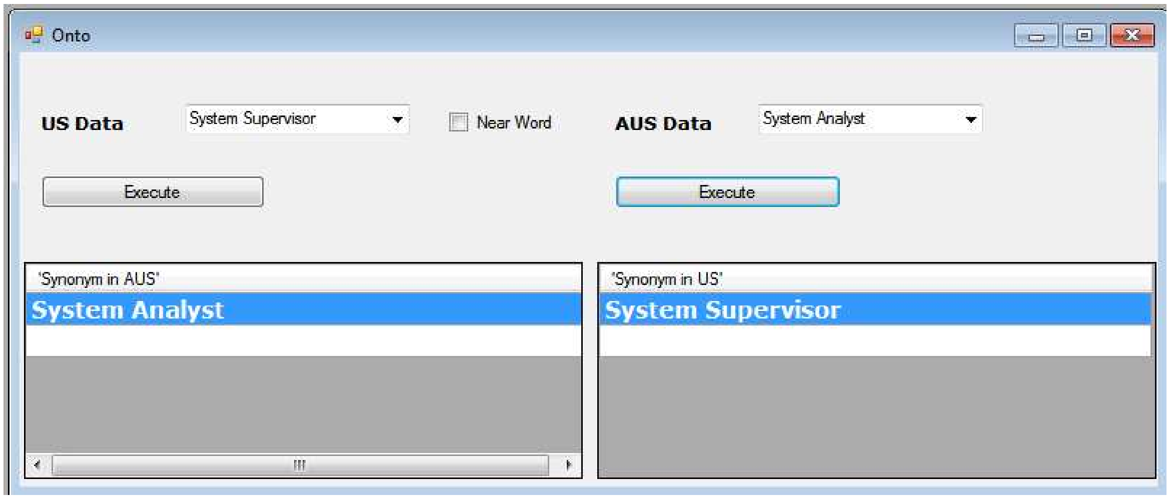
Snapshot 5.17: Principle in US smilar to Dean in Australia



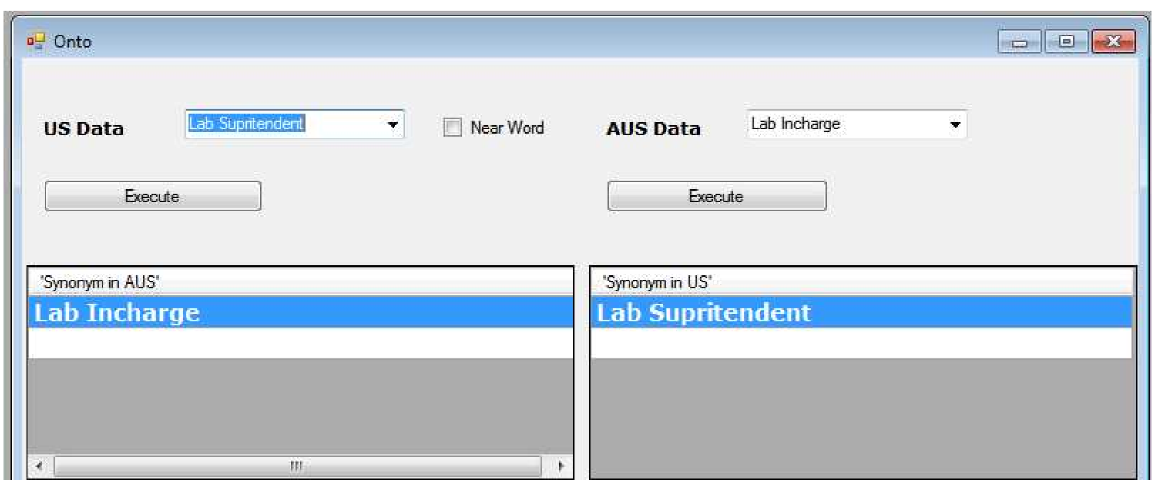
Snapshot 5.18:2[nd] year in US smilar to Undergraduate courses in Australia



Snapshot 5.19:President in US smilar to Chairman in Australia



Snapshot 5.20:System Supervisor in US smilar to System Analyst in Australia



Snapshot 5.21: Lab Supritendent in US smilar to lab Incharge in Australia

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In the nut shell we can conclude that the main obstacle that we face in data interoperability is data heterogeneity, where similar source data is represented differently using different naming conventions and structures. This thesis addresses the problem of heterogeneity between Ontologies in the context of the Semantic Web. Domain Ontologies have become an integral part of the Semantic Web and as their usage increases, the need for resolving semantic differences among them becomes very important. Ontologies do not solve interoperability problems, because the conceptualizations they represent are not commonly shared. In order to realize the Semantic Web vision, disparate ontological representations must be connected through mappings, Ontology Mapping is the process of establishing a common ground for interoperability between domain ontologies. The purpose of Ontology mapping is to look for corresponding relations among concepts and design relevant mapping rules.

Ontology Mapping is only one way to achieve compatibility among ontologies. However, current methods suffer from a number of problems. Most techniques do not provide formal semantics for mapping data structures. They also rely heavily on Graph-based and Grid-based, which often times fail to produce accurate mappings.

Although many proposed solutions for Ontology Mapping exist, but they are either dependent on manual input or demand more resources which results in time overrun or sometimes lack in required solution. It is therefore very difficult to conclude as which technique is the best. As mentioned previously, most techniques are theoretical models that use syntactic and machine learning matching algorithms.

It can be further substantiated, in Literature survey on current Ontology mediation research we had find that initiatives such as InfoSleuth, SMART, CHIMERA, PROMPT and CAIMAN among others, revealed that Ontology mapping is to a large extent based on mainly syntactic schema matching that supported binary schema matches (1:1) only.
Only PROMPT and the Cooperative Framework use taxonomic features and synonym detection of concepts. However, the rigid criterion of having exact attributes in the latter makes the model less flexible. Similarly the Works of Doan [34] and Calvanese [29] seem to be very labor intensive and assume that the Ontology engineer understands the domain, formalisms and mapping rules.

Therefore an over-arching approach is needed to resolve heterogeneity amongst ontologies which we had tried to develop in this thesis. The goal of this thesis is to develop a novel approach to provide the solution for heterogeneity and semantic interoperability by using Ontology mapping in which we construct the local Ontology by referring to vocabulary of the shared domain based on word senses such as synonyms, hyponyms, etc.

6.2 Future scope

Ontology is a latest buzz word within which research and application on Ontology Mapping are going at fast pace. It is hailed as the 'next big thing' and according to researchers, will change the world for good. But still there are many problems that need to be further researched and worked on, to provide a holistic solution to current and future problems. Present problem faced in this research is that there is no uniform definition of lifecycle and criterion of Ontology and also there is common ontological developing methodology and technology at universal level. Therefore, the need of the hour is to build systematic and concrete methods in this area, which will provide an integral and universal solution to problems at global level and will help in better resource utilization.

[1]Siegfried Handschuh, Raphael Volz and Steffen Staab "Annotation for the Deep Web". IEEE Intelligent Systems Published by the IEEE Computer Society, 2003.

[2]C. Sherman and G. Price. The Invisible Web: Uncovering Information Sources Search Engines Can't See. CyberAge Books, 2001.

[3]Updegrove A., 2005. The SW: an Interview with Tim Berners-Lee. Available at: http://www.consortiuminfo.org/bulletins/semanticweb.php

[4]Tatarinov, I., Ives, Z., Halevy, A., Weld, D. (2001), "Updating XML". *In: Proceedings of ACM SIGMOD International Conference on Management of Data*, Santa Barbara, CA. pp. 413-424. New York, NY: ACM Press.

[5]Sean B. Palmer, "The Semantic Web: An Introduction", viewed February 2007, <http://infomesh.net/2001/swintro>

[6]OWL Web Ontology Language, viewed February 2007, <http://www.w3.org/TR/owl features>

[7]Tramullas, J. 2003. Material from Profesor Jesús Tramullas Saz, Department of Documentation, Universidad de Zaragoza, Spain. Available at: http://www/tramullas.com

[8]Sure Y. , Erdmann M, Angele J., Staab S., Studer R., Wenke D. 2002. OntoEdit: Collaborative Ontology Development for the Semantic Web. In Proceedings of the 1st International Semantic Web Conference – SWC2002, Springer, LNCS.

[9]Giorgos Flouris, "On Belief Change and Ontology Evolution."

[10]Ram, S., J. Park, and D. Lee, Digital Libraries for the Next Millennium: Challenges and Research Directions. Information System Frontiers, 1999. pg. 75-94.

[11]Noy, N.F. and M.C.A. Klein. A component-based framework for Ontology evolution. in IJCAI. 2003. Acapulco, Mexico.

[12]Park, J. and S. Ram, Information Systems Interoperability: What Lies Beneath? ACM Transactions on Information Systems (TOIS), 2004. **22**(4): p. 595-632.

[13]Natalya F. Noy and Deborah L. McGuinness "Ontology Development: A Guide to Creating Your First Ontology.

[14] Marta Sabou, "Building Web Service Ontologies."

[15]T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, May 2001.

[16]From Data Federation Pyramid to the Semantic Web 'Birthday Cake', http://www.mkbergman.com.

[17]http://www.ibm.com/developerworks/grid/library/gr-semgrid/

[18]Stuckenschmidt, H. and F. v. Harmelen (2005). Information sharing on the semantic web, Springer.

[19]Patel, M., T. Koch, et al. (2004). Semantic Interoperability in Digital Library Systems.

[20]Fensel D., 2001. Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer.

[21]Giorgos Flouris, "On Belief Change and Ontology Evolution."

[22]M. Fern´and´ez, A. G´omez-P´erez, and N. Juristo. MethOntology: From ontological arts towards ontological engineering. In Proceedings of the AAAI'97 Spring Symposium Series on Ontological Engineering, Stanford, USA, March 1997.

[23]Guarino, N. editor 1998: Formal Ontology in Information Systems. IOS Press, Amsterdam.

[24]Ontology and interoperability, 2001; R.Laurini

[25]Ontoweb deliverable 1.4: A survey on methodologies for developing, maintaining, evaluating and reengineering ontologies, 2003

[26]Gruber, T. (1993). "A Translation Approach to Portable Ontology Specifications." Knowledge Acquisition 5(2): 199-220.

[27]Jorge Cardoso, University of Madeira, Portugal, "Semantic Web Services: Theory, Tools, and Applications."

[28]Madhavan, J., P. Bernstein, et al. (2002). Representing and Reasoning About Mappings between Domain Models. Proc. 18th National Conference on Artificial Intelligence (AAAI 2002),Edmonton (CA).

[29]Calvanese, D., G. D. Giacomo, et al. (2001). Ontology of Integration and Integration of Ontologies. Description Logics.

[30]Ehrig, M. and S. Staab (2004). QOM: Quick Ontology Mapping. Proceedings of the 3rd International Semantic Web Conference (ISWC).

[31]M. Ehrig ,Y. Sure .Ontology Mapping - An Integrated Approach”, In: Proceedings of the 1st European Semantic Web Symposium, Heraklion,Greece, Springer, LNCS,2004,pp.10-12.
[32]Kiryakov, A., K.I. Simov, and M. Dimitrov. OntoMap: Portal for Upper-Level Ontologies. In Proceedings of the 2nd Conference on Formal Ontology in Information Systems 2001. Ogunquit, Maine, USA: ACM Press
[33]Kent, R.E. The Information Flow Foundation for Conceptual Knowledge Organization. In Proceedings of the Sixth International ISKO Conference. 2000. Toronto, Canada.
[34]Doan, A., et al. Ontology Matching: A Machine Learning Approach. in Proceedings of the 11th international conference on World Wide Web. 2002. Honolulu, Hawaii, USA.
[35]Lacher, M.S. and G. Groh. Facilitating the exchange of explicit knowledge through Ontology mappings. in 14th International FLAIRS conference. 2001. Key West Florida, USA: AAAI Press.
[36] Deutsch, A., Fernandez, M., Florescu, D., Levy, A. & Suciu, D. (1998). XML-QL: A Query Language for XML Submission to the World Wide Web Consortium 19-August-1998 Cambridge, MA: W3C. Available at: http://www.w3.org/TR/NOTE-xml-ql/
[37] Fowler, J, et al., Agent-Based Semantic Interoperability in Infosluth. ACM SIGMOD Record, 1999. : pg. 60-67.

LIST OF PUBLICATION

Paper Communicated

Komal Saluja, Seema Bawa and Shalini Batra, “A Noval Approach in Ontology Mapping” communicated to “The 15th IASTED International Conference on Software Engineering and Applications 2011”.