Ontological Mapping for Semantic Search in Shodhganga: A National Repository of Electronic Theses and Dissertations (ETDs)

Manoj Kumar K *Nirmal Chand* *Savita Gandhi*

Abstract

Creation of research results and its publications is extremely important for further research in the same area. Research institutions are already making their best effort to digitize their research output and keeping it in an archive as repository for internal access and sharing among peer research. Open access model has paved way for publishing the research result on public access mode. Shodhganga is a National Repository of Electronic Theses and Dissertations (ETDs) awarded in Indian Universities. Search and discovery of content in a semantic approach is paramount for a researcher to navigate to the proper content in ETDs. Creation of meta-data, organizing it in a logically linked relation is proposed for a semantic visual browser. Use of ontology for the explicit specification of conceptualization of the important meta-data elements by using latest semantic tools such as RDF/XML, OWL, Protégé, NeOn toolkit are discussed in this paper. A conceptual framework for ontologies based on the standard and advanced ontological model and its creation and implementation with reference to ETDs is also discussed.

Keywords: ETD, Shodhganga, Semantic Web, Ontology

1. Introduction

The World Wide Web is experiencing a transmutation from the web of interlinked static documents to the web of dynamic meaningful data. The goal of Semantic Web, i.e. the web of data, is to provide a general framework which allows to reuse and share data across different applications or enterprises.

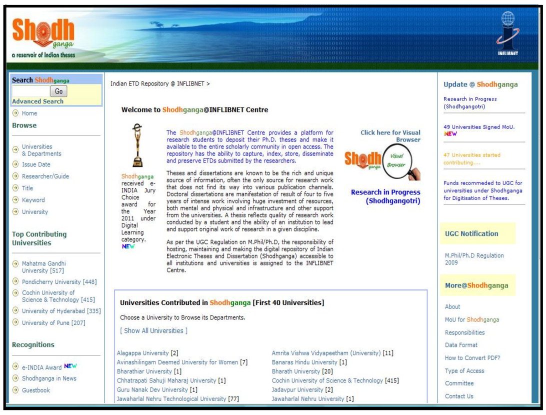
The Semantic Web is an extension of the current web providing reasoning, inference and linking to the resources. According to M. Alexander and et al. [1] ontology is related to provide machine interpretable data for the Semantic web. The ontological applications in semantic web are based on XML which provides machine interpretation and reasoning to heterogeneous systems and are platform independent. By incorporating automatic inference in properly organized metadata, the discovery of knowledge is possible.

Ontology provides a detailed description of design criteria for any domain in an explicit manner. According to B. Chandrasekaran and et al. [2], for any given domain, domain ontology forms the heart of any system of knowledge representation. Ontology created for one domain by an enterprise can be reused for the same domain by the other organization. Ontology merging between different domain results in the creation of a new ontology. An approach towards making a machine interpretable intelligent repository for Shodhganga i.e. a National repository of ETDs is proposed in this paper.

8th Convention PLANNER-2012
Sikkim University, Gangtok, March 01-03, 2012
© INFLIBNET Centre, Ahmedabad

2. ETD and Shodhganga

The primary purpose of the ETD repository is to provide access to one of the most important intellectual products of the university i.e. doctoral theses, dissertations etc. UNESCO ETD guide defines ETD as a document which explains the research or scholarship of researcher and ETDs consists of theses and dissertations which are submitted or archived by the Institutions on the web or on internal network. Theses and dissertations are known to be rich and unique source of information. ETD stored in repositories are information stored, the repositories lack the reasoning and interpretation. Shodhganga is the name coined to denote digital repository of Indian Electronic Theses and Dissertations set-up by the INFLIBNET Centre. The word “Shodh” originates from Sanskrit and stands for research and discovery. The ‘Ganga’ is the holiest, largest and longest of all rivers in Indian subcontinent. Hence, Shodhganga stands for the reservoir of Indian intellectual output stored in a repository hosted and maintained by INFLIBNET Centre.


Shodh ganga
a reservoir of Indian theses
INFLIBNET
Search Shodhganga
Go
Advanced Search
Home
Browse
Universities & Departments
Issue Date
Researcher/Guide
Title
Keyword
University
Top Contributing Universities
Mahatma Gandhi University [517]
Pondicherry University [448]
Cochin University of Science & Technology [415]
University of Hyderabad [335]
University of Pune [207]
Recognitions
e-INDIA Award NEW
Shodhganga in News
Guestbook
Indian ETD Repository @ INFLIBNET >
Click here for Visual Browser
Shodh ganga Visual Browser
Research in Progress (Shodhgangotri)
Update @ Shodhganga
Research in Progress (Shodhgangotri)
49 Universities Signed MoU. NEW
47 Universities started contributing....
Funds recommeded to UGC for universities under Shodhganga for Digitisation of Theses.
UGC Notification
M.Phil/Ph.D Regulation 2009
More@Shodhganga
About
MoU for Shodhganga
Responsibilities
Data Format
How to Convert PDF?
Type of Access
Committee
Contact Us

Welcome to Shodhganga@INFLIBNET Centre

Shodhganga received e-INDIA Jury Choice award for the Year 2011 under Digital Learning category. NEW

The Shodhganga@INFLIBNET Centre provides a platform for research students to deposit their Ph.D. theses and make it available to the entire scholarly community in open access. The repository has the ability to capture, index, store, disseminate and preserve ETDs submitted by the researchers.

Theses and dissertations are known to be the rich and unique source of information, often the only source for research work that does not find its way into various publication channels. Doctoral dissertations are manifestation of result of four to five years of intense work involving huge investment of resources, both mental and physical and infrastructure and other support from the universities. A thesis reflects quality of research work conducted by a student and the ability of an institution to lead and support original work of research in a given discipline.

As per the UGC Regulation on M.Phil/Ph.D, the responsibility of hosting, maintaining and making the digital repository of Indian Electronic Theses and Dissertation (Shodhganga) accessible to all institutions and universities is assigned to the INFLIBNET Centre.

Universities Contributed in Shodhganga [First 40 Universities]

Choose a University to Browse its Departments.

[Show All Universities]

Alagappa University [2]	Amrita Vishwa Vidyapeetham (University) [11]
Avinashilingam Deemed University for Women [7]	Banaras Hindu University [1]
Bharathiar University [1]	Bharath University [20]
Chhatrapati Sahuji Maharaj University [1]	Cochin University of Science & Technology [415]
Guru Nanak Dev University [1]	Jadavpur University [2]
Jawaharlal Nehru Technological University [77]	Jawaharlal Nehru University [1]

Figure 1: Home page of Shodhganga

Ontological structure for the ETD brings the semantic sense to the repository and also helps in easy information retrieval of the data which will be machine interpretable. Moreover, now a days, digital copies of theses and dissertation are stored in a repository to have multiple access simultaneously. An easily categorization of the ETD based on the subject of their domain can be easily done with ontological mapping. As Shodhganga stands for the reservoir of Indian intellectual output stored in a repository hosted and maintained by the INFLIBNET Centre, attempt is to be made to access it in a meaningful semantic way.

3. Methontology for Shodhganga

Ontological structure and mapping for ETD are created using Methontology proposed by Fernandez (1997). Ontology structure works on the backend so that the complexity is not seen by the normal user.

The Shodhganga@INFLIBNET is set-up using an open source digital repository software called DSpace developed by MIT (Massachusetts Institute of Technology) in partnership between Hewlett-Packard (HP). The DSpace uses internationally recognized protocols and interoperability standards. The repository has the ability to capture, index, store, disseminate and preserve ETDs (Electronic Theses and Dissertations) submitted by the researchers. The data about the theses and dissertations stored in a database are fetched and is structured according to the ontology designed for it. Before it appears on the user’s screen it is reasoned with the help of a reasoner.

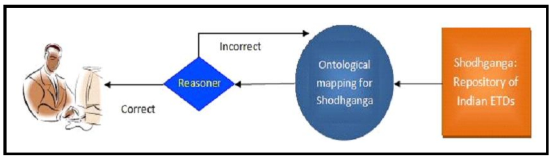

Incorrect
Reasoner
Correct
Ontological mapping for Shodhganga
Shodhganga: Repository of Indian ETDs

Figure 2: Ontology Work Flow

According to F. Noy and et al. [3] Ontology helps to share common understanding of the structure of information among people or software agent, to enable reuse the domain knowledge, to separate domain knowledge from the operation knowledge. The goal of ontological engineering is to develop theories, methodologies and tools suitable to elicit and organize domain knowledge in a reusable and transparent way by G.Nicola[4].

4. Conceptualization of ETD Ontology

According to L. Fernandez[5] , no methodology is mature enough still that can be used to develop a common ontology, A comparison is done with IEEE 1074-1995 standard for software development process in which Methontology is emerged as a better option than other methods. Based on the Methontology, the building of ETD ontology covers the process of specification, conceptualization, formalization, implementation and maintenance. In order to create ontologies, it is desirable to build a complete Glossary of Terms (GT) first. Based on the GT, taxonomy of the concept is to be made systematically and an ad hoc binary relation diagram has to be prepared M. Fernandez and et al [6]. The basic and essentials terms to be included on the ontology are the synonyms, meronyms (i.e. part-of), hyponyms (i.e. superclass-of), hypernyms (i.e. subclass-of) of the terms and metadata that are used to define a thesis in Shodhganga. Setting semantic relation among the subjects and keywords in Shodhganga for interlinking theses from one area to the similar area will be based on these concepts.

Meronymy (Part-of) in knowledge representation means part of a whole domain. E.g. Wheels are part of Vehicle. Holonymy (Has-a) is used to define the relationship between a term denoting a whole and a term denoting a part of the whole. E.g. 'tree' is a holonym of 'bark', of 'trunk' and of 'limb. Such relations are to be defined in ETDs also for keywords and subject. Hyponymy (Superclass-of) is a less familiar term to most people than either synonymy or antonym, but it refers to a much more important sense relation. It describes what happens when we say 'An X is a kind of Y' or A daffodil is a kind of flower, or simply, A daffodil is a flower." House is a hyponym of the subordinate building, but building is in turn, a hyponym of the subordinate structure, and, in its turn, structure is a hyponym of the subordinate thing. A subordinate at a given level can itself be a hyponym at a higher level. Hypernymy (Subclass-of) is a word with a general meaning that has basically the same meaning of a more specific word. For example, dog is a hypernym, Labrador Retriever and Dalmatian are more specific subordinate terms. The hypernym tends to be a basic-level category that is used to denote a specific domain.

Based on the concept explained above, classifications of terms are used to build taxonomy and a taxonomy based on the terms collected is built thereafter. Any additional concept can be easily integrated since the hierarchy of the ETD ontology is manageable. Ad-hoc binary relation diagram is built on the basis of the terms collected as shown in Figure 3. This ontology mapping will result in easy, powerful and precise search result with relations mentioned.

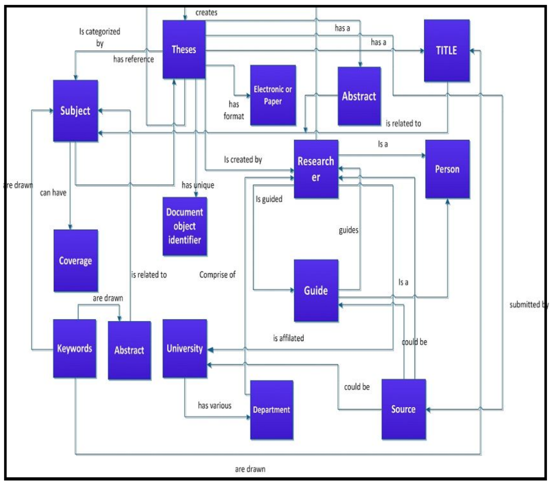


Figure 3: Binary Relation Diagram of Metadata of ETDs

5. Implementation using Ontological Tools

5.1 Semantic and OWL Tools

Based on the online survey on the analysis of the tools is carried out by M. Rahamatullah and et al [7] asking different aged people, working in different domains, some experienced while others-not to compare popular semantic tools such as Protégé, Swoop, Top braid composer, Oiled, Web ODE, Ontolingua, Internet Business Logic, Onto Track and IHMC Cmap Ontology Editor etc. Protégé was adopted by most of the users and it has more functionality and characteristics which is used in in our testing for ETD ontology. Features and characteristics of sme of the common tools are listed below:

5.1.1 Protégé

Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies. At its core, Protégé implements a rich set of knowledge-modeling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats. Protégé can be customized to provide domain-friendly support for creating knowledge models and entering data. Further, Protégé can be extended by way of a plug-in architecture and a Java-based Application Programming Interface (API) for building knowledge-based tools and applications.

5.1.2 JENA

Jena is a Java framework for building Semantic Web applications. Jena provides a collection of tools and Java libraries to help you to develop semantic web and linked-data apps, tools and servers.

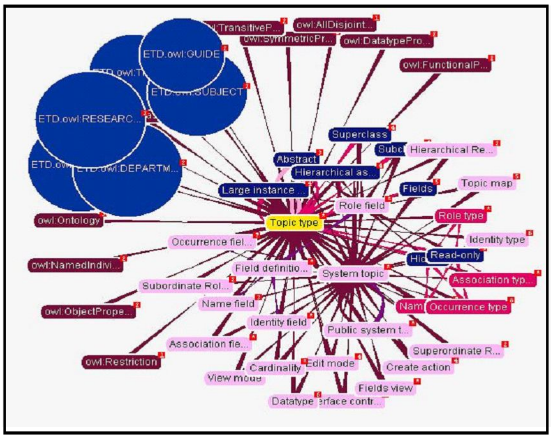


Figure 4: Topic Map to Show Relation of OWL Developed in Protégé for Shodhganga

The Jena Framework includes: an API for reading, processing and writing RDF data in XML, N-triples and Turtle formats, an ontology API for handling OWL and RDFS ontologies, a rule-based inference engine for reasoning with RDF and OWL sources, stores to allow large numbers of RDF triples to be efficiently stored on disk, a query engine compliant with data the latest SPARQL specification, servers to allow RDF data to be published to other applications using a variety of protocols, including SPARQL.

5.1.3 Sesame

Sesame is a de-facto standard framework for processing RDF data. This includes parsing, storing, inferencing and querying of/over such data. It offers an easy-to-use API that can be connected to all leading RDF storage solutions.

Sesame has been designed with flexibility in mind. It can be deployed on top of a variety of storage systems (relational databases, in-memory, file systems, keyword indexers, etc.), and offers a large scale of tools to developers to leverage the power of RDF and related standards. Sesame fully supports the SPARQL query language for expressive querying and offers transparent access to remote RDF repositories using the exact same API as for local access. Finally, Sesame supports all main stream RDF file formats, including RDF/XML, Turtle, N-Triples, TriG and TriX.

5.1.4 NeOn

The NeOn toolkit is a state-of-the-art, open source multi-platform ontology engineering environment, which provides comprehensive support for the ontology engineering life-cycle. The toolkit is based on the Eclipse platform, a leading development environment, and provides an extensive set of plug-ins (currently 45 plug-ins are available) covering a variety of ontology engineering activities which includes Ontology Dynamics, Ontology Evaluation, Ontology Matching, Reasoning and Inference, and Reuse.

5.2 Reasoning tool

Knowledge systems require to do something which has not been explicitly told how to do it. The system should be able to apply its own reasoning to inference from the available source of data and give result based on the inference. There are many reasoning tools for OWL. Some of them are explained below:

5.2.1 Pellet

Pellet is an OWL 2 reasoner. For applications that need to represent and reason about information using OWL, Pellet is the leading choice for systems where sound-and-complete OWL DL reasoning is essential. Pellet includes support for OWL 2 profiles including OWL 2 EL. It incorporates optimizations for nominal, conjunctive query answering, and incremental reasoning. According to E. Sirin and et al. [8] Pellet has been the first reasoner to support all of OWL-DL, i.e. the Description Logic (DL).

5.2.2 FaCT++

FaCT++ is the new generation of the well-known FaCT OWL-DL reasoner. FaCT++ uses the established FaCT algorithms, but with a different internal architecture. Additionally, FaCT++ is implemented using C++ in order to create a more efficient software tool, and to maximize portability.

5.2.3 HermiT

HermiT is reasoner for ontologies written using the Web Ontology Language (OWL). Given an OWL file, HermiT can determine whether or not the ontology is consistent, identify subsumption relationships between classes, and much more. HermiT is the first publicly-available OWL reasoner based on a novel “hypertableau” calculus which provides much more efficient reasoning than any previously-known algorithm. Ontologies which previously required minutes or hours to classify can often by classified in seconds by HermiT, and

HermiT is the first reasoner able to classify a number of ontologies which had previously proven too complex for any available system to handle uses direct semantics and passes all OWL 2 conformance tests for direct semantics reasoner.

5.2.4 RacerPro

RACER stands for Renamed ABox and Concept Expression Reasoner. RacerPro can also be used as a system for managing semantic web ontologies based on OWL (e.g., it can be used as a reasoning engine for ontology editors such as Protégé). However, RacerPro can also be seen as a semantic web information repository with optimized retrieval engine because it can handle large sets of data descriptions (e.g., defined using RDF). Last but not least, the system can also be used for modal logics such as Km.

5.3 Ontology Created in Protege

ETD Ontology is created and tested using Protégé 4.1 application. Figure 5 shows the class hierarchy using the OWL visualization which shows various classes and sub –classes used in ontology. For e.g. Researchers and Guides are the subclasses of 'Person' in Shodhganga. Shodhganga is the main domain concept to define all other sub classes. i.e. super-class of all sub-classes. An IRI is provided (Internationalized Resource Identifier), which helps in easy linking of the resources that helps in the easily information retrieval.

As shown in figure 6, the browser output is generated in protégé, which shows class Guide in the class hierarchy, its super classes disjoints and its usage. Usage means how it is related to other concepts, while disjoint means it is not part or included in other concept. Ontology developed is based on advanced OWL-DL(Web Ontology Language Description Logic). Mainly, two types of properties are used for building the relation ie. object property and data property. Object properties are used to link between concepts, individuals while the data property is used to define value. In case of ETD ontology for linking the value of the document object identifier the data property is used. Moreover the characteristics functionality is used for this data property since the Handle number of any thesis is going to be unique value only. Similarly the characteristics are applied to the object property.

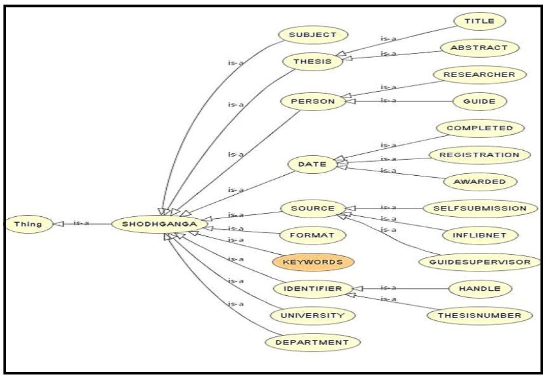
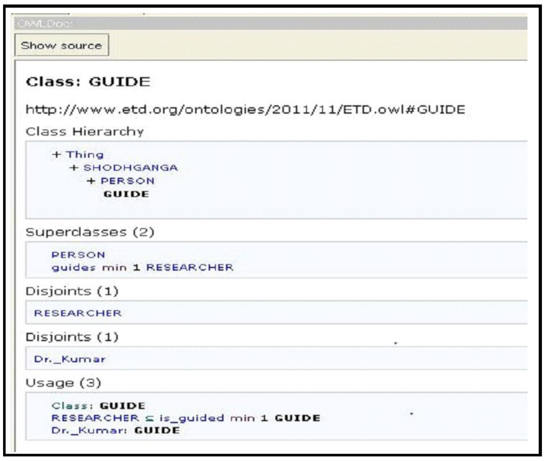


Figure 5: OWL Visualization

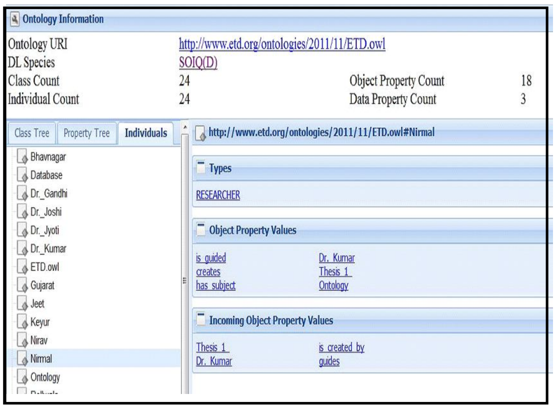
Class hierarchy and superclass and its dependence on other classes with properties are shown in Figure 6 which is created in Protege 4.1.



```
OWLDoc
Show source

Class: GUIDE

http://www.etd.org/ontologies/2011/11/ETD.owl#GUIDE
Class Hierarchy
    + Thing
       + SHODHGANGA
          + PERSON
             GUIDE

Superclasses (2)
    PERSON
    guides min 1 RESEARCHER
Disjoints (1)
  RESEARCHER
Disjoints (1)
  Dr._Kumar
Usage (3)
    Class: GUIDE
    RESEARCHER ⊆ is_guided min 1 GUIDE
    Dr._Kumar: GUIDE
```

Figure 6: OWL Browser Output in Protégé

The output of the ontology is generated as shown in Figure 7 using OWL Sight which shows the ontology URI, the DL species used in building the ontology, class and individual count, object and data property count. Class Tree shows the class hierarchy while second tab property tree shows the properties used in the ontology, the third tab shows the individuals, only limited individuals have been taken for the sake of understanding, whiich can be populated from the database. On the side of these three tabs an URI is provided to link to the individual selected, below that the description of the individual type its relation with the object property values is shown.



Ontology Information

Ontology URI	http://www.etd.org/ontologies/2011/11/ETD.owl		
DL Species	SOIQ(D)		
Class Count	24	Object Property Count	18
Individual Count	24	Data Property Count	3

Bhavnagar
Database
Dr_Gandhi
Dr_Joshi
Dr_Jyoti
Dr_Kumar
ETD.owl
Gujarat
Jeet
Keyur
Nirav
Nirmal
Ontology

http://www.etd.org/ontologies/2011/11/ETD.owl#Nirmal

Types

RESEARCHER

Object Property Values

is guided	Dr. Kumar
creates	Thesis 1
has subject	Ontology

Incoming Object Property Values

Thesis 1	is created by
Dr. Kumar	guides

Figure 7: Ontology on ETD Created by OWL Sight

6. Conclusion

Ontology on Indian ETD is built in this paper based on the data used by Shodhganga. The concepts were classified in this domain with their bindings and slots to build the relationship between them. Ontology is built in Protégé 4.1 and was inferred using FACT++ and also Pellet- DL reasoner. Ontology is developed to provide a semantic ETD which can reason and can help in easily retrieval of information about the thesis, author, and date from the inference. The ontology developed can be reused and integrated. The ontology is developed so that the researcher can easily retrieve the information needed by them since the ontology can easily categorize the theses and dissertations based on the factors like subject, university and author and which is machine interpretable.

References

1. Alexander Maedche, Steffen Staab Learning Ontologies for the Semantic Web in Semantic Web Workshop 2001 Hongkong, China

2. B. Chandrasekaran and John R. Josephson. Richard Benjamins: What are Ontologies, and Why do we need them? in IEEE INTELLIGENT SYSTEMS 1999

3. Natalya F. Noy and Deborah L. McGuiness :Ontology Development 101: A Guide to Creating Your First Ontology at Stanford University.

4. Nicola Guarino: UNDERSTANDING, BUILDING, AND USING ONTOLOGIES at LADSEB-CNR, National Research Council.

5. Fernandez Lopez: Overview of Methodologies for Building Ontologies In Proceedings of IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) in Stockholm Sweden, August 2,1999.

6. Mariano Fernandez, Asuncion Gomez-Perez, Natalia Juristo:"METHONTOLOGY: From Ontological Art Towards Ontological Engineering" from AAAI Technical Report ss-97-06

7. M. Rahamatullah Khondoker, Paul Mueller:Comparing Ontology Development Tools Based on an Online Survey IN: Proceedings of the World Congress on Engineering 2010 Vol I

8. Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, Yarden Katz “Pellet: A Practical OWL-DL Reasoner” at University of Maryland, MIND Lab.

About Authors

Mr. Manoj Kumar K, Scientist-D (CS), INFLIBNET Centre, Ahmedabad.

Mr. Nirmal Chand, Student, M.Tech. Gujarat University, Ahmedabad.

Dr. Savita Gandhi, HOD, Dept. of Computer Science, Rollwala Computer Centre, Gujarat University, Ahmedabad.