

# A Novel Approach for Semantic Integration of Data using Ontology

Ghonia Pooja Mansukhlal\*, C. Malathy and U. Uma Pratheebha

Department of Computer Science and Engineering, SRM University, Chennai - 603203, Tamil Nadu, India;  
ghoniapooja@gmail.com, malathyc03@gmail.com, Pratheebha1994@gmail.com

## Abstract

Integration process of data is recently recognized like a significant visualization of the Semantic Web explore for which researchers focus on numerous areas, such as integration of information, ontologies and databases. **Objectives:** Users generally requires an incorporated analysis of information accessible from various data and it was proposed to grant users with this view of data. **Methods:** The Meta data is created from different data sets like excel data set, RDF data set and XML data set. **Findings:** In this work a survey was made for integrating databases using ontology and a new approach for integration of databases is devised for finding correspondence between ontologies. **Applications:** View and Search functionality was provided by filtering data from Meta data. This provides easy access to the integrated data.

**Keywords:** Database Integration, Ontologies, Schema, Semantic Web, Semantic Integration

## 1. Introduction

Nowadays the sources for data and data providers are increasing voluminously every day. However, users are having requirement of incorporated analysis of the information presented from various data sources. Hence, spotlight was made on integration of heterogeneous databases which was found to be an issue. Data incorporation defines to merge information in such a way that a uniform and constant visualization is accessible by users<sup>1</sup>.

Two types of database heterogeneity are available: Data heterogeneity and Semantic heterogeneity.

Data heterogeneity defines the difference among restricted definitions. Semantics defines meaning which is contrast type, formats, or correctness and semantic heterogeneity refers to difference and comparison in consequence of limited information. Schema fundamentals in two confined information source may have matching projected significance, but may be having diverse names<sup>2</sup>. Hence, it must be appreciate that two essentials really submit to the identical idea through procedure of integration. Schema fundamentals in data sources may be defined similarly, and their individual meanings were unequal. So, these fundamentals are

supposed to be validated as different effects through integration.

The technique of Ontology based concept ladder extraction of Web information helps to take out concept ladder efficient technique for ontology construction<sup>3</sup>.

Semantics is the confirmation given by people for the attribute to data based on their understanding and this different understanding of information cause semantic heterogeneity. In database part, semantics defines to the implication of schemas basics<sup>4</sup>.

A schema specifies organization of information. It gives structure to database plan period. When integrating information, all databases provide a description of the data. The plan of the combination procedure used to build up of an inclusive schema. Global schema adds with combine local schema so that clients are visualized with a consistent and accurate breakdown of global database. Main issue here in related semantic heterogeneity. This is to be addressed properly and should be resolved.

Defined and prescribed meaning of semantics for heterogeneous databases has motivated several researchers to affect recognized ontologies as possible clarification for semantic heterogeneity. Formal ontology contains reasonable axiom which communicates significance of conditions for exacting community.

\* Author for correspondence

Permission for ontological meanings in the middle of members of a society is significant dissimilarity among conceptual models and ontologies<sup>5</sup>.

The development of ontology has the need to explain some points linked to the Formalisms, Tools and Languages to be worn. Ontologies can be confidential according to the conceptualization topic (content)<sup>6</sup>. The Semantic Web technologies represent data as information with a machine-readable configure and a standardized representation framework so that the data integration and management of data can be performed<sup>7</sup>.

## 2. Materials and Methods

### 2.1 Ontology

Semantics describes to significance which is difference to grammar that relates to arrangement. In database part, semantics could be considered as human understands of information and schema objects as relating to their thoughtful of world in certain background<sup>8</sup>. For information incorporation, kind of semantic measured is usually actual world semantics. The concern in actual world semantics is mapping of substance in model or globe into the actual world. Also it considers issue which involves individual understanding, utilize of statistics and data.

Distinction in interpretation of the related representation or information thing among information supplier and information consumer escort to semantic heterogeneity. The main aim is to conquer semantic heterogeneity by specifying proposed actual world semantics of information and schema elements<sup>9</sup>. But it was difficult toward totally explain what an information and schema element means. Some means must be provided to give enough unambiguous semantics toward understand information at all times constantly and unambiguously<sup>10</sup>.

Ontologies can be single means to characterize open, recognized semantics. It gives an open requirement of concepts. Ontology is open, official explanation of concept and relationships that be current in a convinced world of dialogue and gives a communal language to pass on to these concept. In comparison with further categorization scheme, such as taxonomy and keywords, ontologies permit additional absolute and precise province models<sup>11</sup>.

In region of integration incorporation, ontologies may be useful to build certain semantic relationships among data source. Due to common understanding, risk of semantic heterogeneity may be reduced<sup>12</sup>. To keep away

from harms comparable to solitary universal schemas, no solitary global ontology can be information suitable used for conclusion making tasks.

To overcome the problems faced by solitary global schemas, no solitary global ontology could be determined for likely client group. This type of move toward will power user to regulate to one solitary conceptualization of world<sup>13</sup>. So, a correct move toward to information integration could hold up dissimilar ontologies so dissimilar society definite semantics could be utilized in parallel.

### 2.2 Semantic Web

Semantic Web<sup>14</sup> contains the data similar to the one as considered in relational database, XML credentials, spreadsheets, RDF documents and text layout data files, which supports for files to contain access as enormous database. Semantic web is not concerning for marking accessible HTML documents for processor understand what they say<sup>15</sup> and also it is not concerning the replicated intellect areas of mechanism natural language understanding.

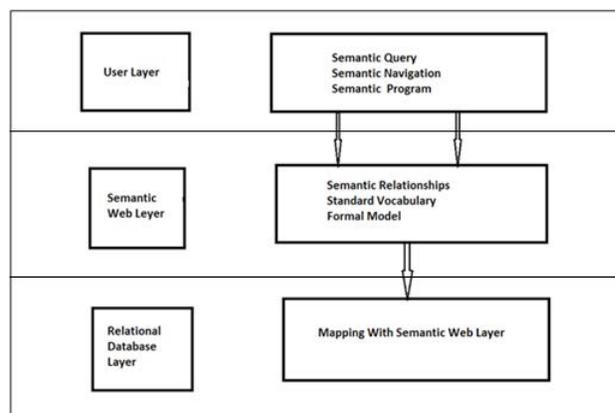


Figure 1. Semantic web of relational databases.

Vision of Semantic Web of databases in Figure 1 is,

- To communicate disseminated sited legacy database with comfortable semantics
- To comprise added deductive capability for top to expand usability and reusability of information
- To allow ontology-related uncertainty, explore and direction-finding as substantial distributed database

### 2.3 Structure of OLAP Systems

OLAP has linked through data warehouse, following these three layers:

- Data sources layer consists of all prospective information of every environment (relational, semi-structured, textual and object oriented) which may help to perform scrutiny goal.
- Integration (combination) layer converts and clean the information retrieved from the sources. It supplies them in a suitable arrangement for the following analysis.
- Analysis layer consists an amount of apparatus for extorting data and data from integrated facts and present it to analysts.

The combination representation of traditional OLAP organization is based on global representation which is considered as an observation greater than the original data resource schemas. This is generally recognized as universal view<sup>16</sup>. In combination model, uncertainty answering is easy. Outside data source are unspecified to recognized in progress since are the customer wants guide plan of the global schema. For such cases, more artificial combination representation is desirable. In exacting, combination of outside information schemas in requirements of global schema has been studied<sup>17</sup>. From global schema, local schemas are derived. The local schemas were considered like additional dedicated observation of united ordinary global schema. The ensuing combination model (Local as View) is thus greatly extensible, at the expenditure of significantly more complex query answering. Therefore, Exploratory OLAP systems are strongly the incorporated data into a DW<sup>18,19</sup>. It can be founded related to ETQ processes.

OLAP systems are using a unique data model, the multidimensional information model for the combination layer. Accurate information is retrieved from the different information source layer. Layer should be articulated into conditions of statistical actions and categorical hierarchical proportions<sup>20</sup>. The semantics of combination model consists of on behalf of any fascinating inspection of the sphere in its situation. The characteristic procedure in accuse of converting information from data sources layer to the combination layer. This process is defined as ETL processes. Sometimes, It is difficult to allow quick filling and searching at expenditure of sending just partially cleanse and imprecise information. Initially arrangement of cases are switch which leads to guide to Extract, Load, Transform (ELT) and Extract, Transform, Load and Transform (ETLT). Here, conversion are somewhat deferred to present clean other than fewer advanced information. Other alternative is to contain Extract,

Transform, Query (ETQ) by stoppage transformation for previous diminutive and ration information honestly to user on request. An ETQ process straightforwardly extracts the compulsory data from information source and converts it to fit captivated SW technology that can assist in all OLAP layer in place to hold Semantic OLAP and Exploratory OLAP systems. In data source layer, OLAP can assist in taking accurate semantics of source. In the integration layer, OLAP is worn to identify transformations and confine information extraction. In analysis level, OLAP could facilitate specifying semantics of the existing OLAP outcome<sup>21</sup>.

ETQ procedures are suitable for the stage analysis that holds exterior data available in the network. So they typically contract through semi-structured, pour out and lively information sources<sup>22</sup>. ETQ procedures can act together with the data warehouse current. So, ETQ procedure may take clean data from the ETL performance area, merge it with all exterior integrated data, and finally bring the consequences to the analytical tools. ETQ procedure may also exist separately from traditional DW/OLAP, keep away from thus require of filling in sequence and analysis about it.

At last, SW technology can supply as a suitable basis for important ETQ procedure, while the majority exterior data is at the present being available as related<sup>23</sup>. The quantity of come close to that utilizes software technology in first two layers is enormous. There are in particular modest proposal that concern team to third layer and ETQ procedure.

### 3. Results and Discussion

The System architecture shown in Figure 2 proposes semantic integration of data sets and how the Meta data is created. Search functionality is provided from all integrated data sets.

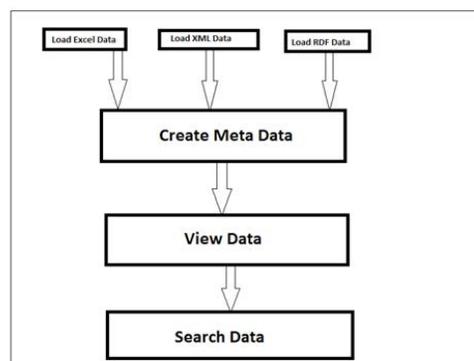


Figure 2. Architecture for semantic integration.

**Table 1.** Information of meta data

Meta Data				
ID	URL	Title	Keywords	Description
1	www.jntu.ac.in	Jntu Hyderabad	Jntu, Hyderabad	Jntu, Hyderabad
2	www.nit.ac.in	NIT	NIT, Hyderabad	NIT, Hyderabad
3	www.pune.ac.in	Pune University	Pune, Pune University	Pune, Pune University
4	www.jntua.ac.in	Jntu anathapur	Jntu, anathapur	Jntu, anathapur
5	www.jntuk.ac.in	Jntu kakinada	Jntu, kakainada	Jntu, kakainada
6	www.srmuniv.ac.in	SRM University	SRM University	SRM University

There are four modules in the proposed system. They are:

- Load Data: Load Data allows for loading different data sets like Excel File, XML file and RDF file from location.
- Create Meta Data: Create Meta Data performs merging of three different data sets and generate one common Meta data for different data sets.
- View Data: View Data shows the Meta data information which is performed in Create Meta module.
- Search Data: Search Data allows for searching information from Meta data.

Meta Data is generated from Excel file, XML file and RDF file. The Table 1 shows information of Meta data.

### 3.1 Methodology for Semantic Integration

The intelligence built in Business is intended for information congregation, transforming and shortening available information from accessible source for decision-making tasks. The grouping of data warehousing and Online Analytical Processing (OLAP) technologies and the multidimensional data representation was widely used in BI. OLAP technology is used for analysis OLAP is used in a glowing restricted congested world state, where place of information sources is quite stationary, and coherent information. It is occasionally overloaded in bunch form .information can be proficiently people and machine for an enormous mixture of tasks.

OLAP is applying heavy purification transformation. However, the explosion of XML and another wealthier semi-structured format like RDF and OWL has open positive a great deal more various and close scenario than individuals of such conventional internal DW application<sup>24</sup>. The chance and significance is of through unstructured and semi structured data in the decision making procedure, Websites and connected to oceanic

protection area. Few are associated Open Data initiative are flattering source of huge amount of valuable semi structured information. At present no single question requires of addition all information to fixed corporate analysis processes<sup>25</sup>.

An important quantity of data and information, should be establish in interested data sources like network portal, community media, amorphous data and some structured data stores such as invention appraisal, user complaint, e-mails<sup>25</sup>. Enterprises are having current to appear addicted to such wealthy in order source to add their income and look up goods and services<sup>26</sup>. So, occupying a business statement that show result of an invention movement in defined time stage may need merging information from chronological, structured information like invention transaction and customer information, reside in a data warehouse, with sentiment derived from big Data linking to products promote by the individual campaign<sup>27,28</sup>. So, companies desire to discover all these fresh data chance and contain their OLAP analysis, important to new type of OLAP: Exploratory OLAP. Main dissimilarity of Exploratory OLAP and Traditional OLAP is obviously issue of examination of latest information source and new habits of verifying data, of original presentation of giving information collectively, of new ways of querying information. While Traditional OLAP is executed in closed world situation based simply on interior information, necessary element of investigative OLAP is to find out, obtain, add and analytically query original external information. The Semantic Web has imagined as resources to assemble semantic gaps above web distributed contents so that web data can be recovered and routed by the semantic web can show to be an outstanding for accessing of different approaches and an ordinary position for analyzing with mixture approaches.

There are few instances which gives a collection of

citizens who wants to watch if fish grab being land in diverse countries and find restrictions set by the oceanic defense scheme and how they are interested in watching and examining the information by time, place where every of these scope should be prepared into ladder of level. Statistical catch data with environmental data about oceanic protection areas, fish inhabitants information from various research databases are to be integrated, in a huge number of format starting with comma divided files to data and lastly with ontology information recounting hierarchies. Analysis is to be done to carry out the composite integration and determine conflict. In general, software technologies are dominant to model diverse kind of information and gives required reasoning capability on peak.

## 4. Conclusion

Various features of ontology investigator in semantic incorporation are comparable to number that catalog and in row integration investigator have been considering. Several of methods are alike though ontology society depends extra seriously on senior animated authority of ontology language and on explanation technique. An ordinary higher ontology and position ontology is worn to resolve the integration trouble.

Most researchers have same opinion that semantic integration is one of mainly rigorous challenges for semantic web. On the other dispense, the foundation of semantic web is that work out of appliance interpretable ontologies distinct in official languages pleasant to reasoning will give subsequently production of services.

## 5. References

- Berlanga R, Romero O, Simitsis A, Nebot V, Pedersen TB, Abello A, Aramburu MJ. Semantic web technologies for business intelligence. IGI Global; 2011.
- Levy AY. The information manifold approach to data integration. *IEEE Intelligent System*. 1998; 13:12–6.
- Karthikeyan K, Karthikeyani V. Ontology based concept hierarchy extraction of web data. *Indian Journal of Science and Technology*. 2015 Mar.
- Abello A, Romero O, Bach T. Using semantic web technologies for exploratory OLAP. *IEEE Transactions on Knowledge and Data Engineering*. 2015.
- Ganter B, Wille R. Formal concept analysis: Mathematical foundations. Berlin-Heidelberg: Springer; 1999.
- Slimani T. Ontology development: A comparing study on tools, languages and formalisms. *Indian Journal of Science and Technology*. 2015 Sept.
- Jeong H, Jeong H. Ontology-based integration and refinement of evaluation-committee data from heterogeneous data sources. *Indian Journal of Science and Technology*. 2015 Sept.
- Buitelaar, Olejnik P, Sintek D. A protocol plug-in for ontology extraction from text. *Proceedings of the International Semantic Web Conference*; 2003.
- Chen, Wu H, Wang Z, Mao H. Rdf/rdfs based relational database integration. *ICDE*; 2006.
- Wely C. Ontology research. *AI Magazine*. 2003; 24(3).
- De Laborda, Conrad CP. Bringing relational data into the semantic web using SPARQL and relational OWL. *International Workshop on Semantic Web and Database at ICDE*; 2006.
- Halevy, Ives AY, Madhavan ZG, Mork J, Suciu P, Tatarinov D. The piazza peer data management system. *IEEE Transaction Knowledge Data Engineering*. 2004.
- Chen H, Wang Y, Wang H, Mao1Y, Tang J, Zhou C, Yin A, Wu Z. Towards a semantic web of relational databases. *A Practical Semantic Toolkit and an In-Use Case from Traditional Chinese Medicine*; 2012.
- Zhao H, Ram S. Combining schema and instance information for integrating heterogeneous data sources. *Data and Knowledge Engineering*. 2013.
- Ghulam M. Object-oriented approach for integration of heterogeneous databases in a multi-database system and local schemas modifications propagation. *International Journal of Computer Science and Information Security*. 2009; 6(2).
- Guarino N. Formal ontology and information systems. *Workshop on Ontologies and Problem-Solving Methods*; 1998.
- Noy NF. Semantic integration - A survey of ontology-based approach. *SIGMOD Record*. 2004; 33(4).
- Elmasri R, Navathe SB. *Fundamentals of Database System*. 3rd edition. Addison Wesley; 2000.
- McClellan S, Scotney B, Morrow P. Integrating semantically heterogeneous aggregate Schemas using multiple LI Linear Regression and SVD. *Journal Paper*. 2013
- Rahm E, Bernstein PA. Survey of approaches to automatic schema matching. *1st International Conference on Language Resources and Evaluation (LREC)*; 2001.
- Goasdoue. Answering queries using views. *A KRDB Perspective for the Semantic Web*. 2003.
- Haase, Motik P. A mapping system for the integration of OWL-DL ontologies. *Proceedings of the 1st International Workshop on Interoperability of Heterogeneous Information Systems*; 2005.
- Lambrix P, Tan H. A system for aligning and merging biomedical ontologies. *Web Semantics*. 2006.
- Richard Benjamins V, Fensel D. The ontological engineering initiative (KA). *Formal Ontology in Information Systems*. 1998.
- Kim W, Choi I, Gala S. On resolving schematic heteroge-

- neity in multidatabase systems and distributed and parallel databases. *Distributed and Parallel Databases*. 1993.
26. Kim Won, Seo J. Classifying schematic and data heterogeneity in multi database systems. *IEEE Computer*. 1991.
27. Kalfoglou Y, Schorlemmer M. An ontology mapping method based on information flow theory. *Journal on Data Semantics*. 2003.
28. Poggi A, Lembo D, Calvanese D, Giacomo GD, Lenzerini M, Rosati R. Linking data to ontologies. *J Data Semantics*. 2008.