A Survey: Approaches to Enhance the Data Integration using Ontology and Methodologies to Improve Data quality

Pravithra CNS[1] Hema MS[2] Sumathi VP[3]
[1,2,3]Department of Computer Science & Engineering
[1,2,3]Kumaraguru College of Technology, India

Abstract— In the today's world, the amount of the data is been increasing tremendously. In order to analyse the data and make decisions on the data, data residing at the different sources are integrated by providing users. The data integration is the solution for above said problem. The complete data integration solution encompasses the discovering, monitoring, transforming and delivery of the data from a variety of sources. This paper deals with various approaches of the data integration to resolve the semantic heterogeneity using ontology. Various levels of ontology based data integration techniques are reviewed and the issues are summarized. Different metrics and the approaches are also discussed to improve the data quality using biomedical database.

Key words: Ontology, Data Integration, Data Quality, Semantic Heterogeneity

I. Introduction

Data integration is an approach to integrate data from multiple data sources. The three different approaches for the data integration are the data consolidation, the data propagation and the data federation. Data federation creates a virtual view and returns the data. Data consolidation is the process of summarizing large quantities of information into a single database. Data Propagation is the distribution of data from one or more source data warehouses to one or more local access databases, according to propagation rules. It gains more popularity than data consolation and data propagation. Ontology is the main formal and explicitly defined specification of the shared conceptualization. The different approaches for integrating the data and methodologies to improve the data quality are discussed in this survey paper.

A. Introduction To Data Integration:

Data Integration defines the integration of data were the data's are integrated from various heterogeneous data sources. Different types of data integration are,

- Data federation creates a virtual view and returns the data.
- Data consolidation is the process of summarizing large quantities of information into a single database.
- Data Propagation is the distribution of data from one or more source data warehouses to one or more local access databases, according to propagation rules.

B. Heterogeneities:

Fundamental characteristic of services which results in variation from one service to another, or variation in the same service from day-to-day or from customer-to-customer. Heterogeneity makes it hard for a firm to standardize the quality of its services. In systematic reviews heterogeneity refers to variability or differences between studies in the estimates of effects. A distinction should made between "statistical heterogeneity" (differences in the reported effects), "methodological heterogeneity" (differences in study design) and "clinical heterogeneity" (differences between studies in key characteristics of the participants, interventions or outcome measures). Where there are large differences in clinical or methodological nature between studies where the heterogeneity is known to exist.

1) Semantic Heterogeneity:

Semantic heterogeneity is the database schema or datasets for the same domain are developed by independent parties, resulting in differences in meaning and interpretation of data values. Beyond structured data, the problem of semantic heterogeneity is compounded due to the flexibility of semi-structured data and various tagging methods applied to documents or unstructured data. Semantic heterogeneity is one of the more important sources of differences in heterogeneous datasets.

2) Structural Heterogeneity:

Structural heterogeneity is the database schema that deals with the different data models and its hierarchical representations. It actually defines the native model or structure to store data differs in data sources leading to structural heterogeneity. Schematic heterogeneity that particularly appears in structured databases is also an aspect of structural heterogeneity.

3) Syntactic Heterogeneity:

Syntactic Heterogeneity is a result of different format of data representation. Syntactical heterogeneity deals with different languages with the data representations by syntactical analysis.

4) System Heterogeneity:

System Heterogeneity is the use of different operating system and hardware platforms.

C. Introduction To Data Quality:

Data quality is a perception or an assessment of data's fitness to serve its purpose in a given context. Data quality is defined as "they are fit for their intended uses in operations, decision making and planning." In computing, data quality is the reliability and application efficiency of data, particularly when kept in a data warehouse. Data quality assurance (DQA) is the process of verifying the reliability and efficiency of data.

D. Virtual View Creation:

1) Local-as-view (LAV) Model

In this model, the source database is modelled as a set of views over an underlying global schema. The advantage of this model is that new sources can be added easily when compared to GAV. However the query rewriting process is complex because the system has to choose from a set of choices to determine the best possible rewrite [1].

2) Global-as-view (GAV) Model

In this model, the global schema is defined by having one or more views over the source schemas for each class. In this approach, changes in information sources or adding a new information source requires mapping between the global and source schemas [1].

E. Ontology Based Data Integration:

Ontology based Data Integration involves the use of ontology's to effectively combine data or information from multiple heterogeneous sources. It is one of the multiple data integration approaches and may be classified as Global-As-View (GAV). The effectiveness of ontology based data integration is closely tied to the consistency and expressivities of the ontology used in the integration process. There are several methods created to address the problem of dealing with different concepts and interpretations. Use of ontology is used as one of the methods to resolve heterogeneities. The Ontology is defined as the process of exposing formal explicit specification of the shared conceptualization and Interpretations. Three types of ontology are used to carry out data integration. They are single ontology, multiple ontology and hybrid ontology [6, 2]. A comparative study is made among various ontology approaches which are shown in the Table 1.

Approaches used	Implementation	Semantic heterogeneity	Adding or removing of sources
Single ontology	Straight forward	Similarity in domain	Need for the some adaption in the global ontology
Multiple ontology	Costly	Supports heterogeneity	Providing a new source Ontology that is related to the other
Hybrid ontology	Reasonable	Supports heterogeneity	Providing a new source ontology

Table 1: Various Ontology Approaches

II. LITERATURE REVIEW COMPARISONS

A. Ontology Based Data Integration:

The Integration of data is carried out by various integration approaches [3]. They various integration approaches are as follows

1) Integration by Manual

Users directly interact with all relevant information systems and manually integrate selected data. That is, the users have to deal with differently used interfaces and query languages.

2) Integration by application

This approach integrates the application that accesses various data sources and return integrated results to the user. This solution is possible for a small number of component systems

3) Uniform Data Access

The logical way of data integration is successfully accomplished at the data access level. The global applications are accompanied with the unified global view of physically distributed data, though only the virtual data that is available on this level. This global view of physically integrated data can be time consuming since the data access, the homogenization and the integration have to be done at run time.

4) Integration by Middleware

Middleware mainly provides the usable functionality that is generally used to solve the dedicated aspects of the integration problem, example: solved by the SQL middleware.

5) Common User Interface

The uniform interface provided to the endured to have uniform look and feel. The data from the relevant information systems is still separately presented and executed on the users feel, so that the homogenization and the integration of data must be done by the users.

6) Common Data Storage

The physical data integration is carried out by transferring data to the new data storage; so that the local sources can either be retired or remain operational. During the integration of informational data from various data sources, resolving heterogeneities remains as a challenging task. The heterogeneities available in the data bases make data integration a tougher task [2]. The various heterogeneities are as follows.

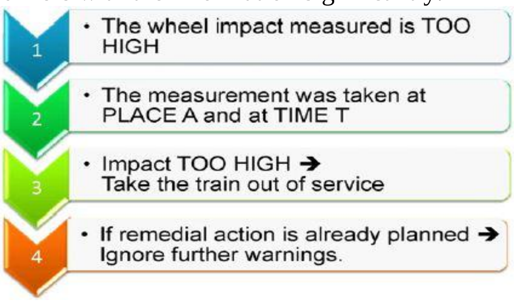
- Structural heterogeneity deals with the different data models and its hierarchical representations.
- Systematic heterogeneity deals with hardware and operating system via systematic approach.
- Syntactical heterogeneity deals with different languages with the data representations by syntactical analysis.
- Semantic heterogeneity deals with different concepts and interpretations relating the conflicts.

Semantic heterogeneity deals with the three types of concepts are as follows,

- Semantically equivalent concepts with the different analysis models.
- Semantically related concepts with same analysis models.
- Semantically unrelated concepts with different analysis models.

The three main causes for semantic heterogeneity [12]:

- Confounding conflicts occurs when information items seem to have concluded the same meaning, which differs in reality, e.g. different temporal contexts.
- Scaling conflicts occurs when different reference systems are used to measure a value. Examples are different currencies.
- Naming conflicts occurs when the naming schemes differs with the information significantly.



To implement the semantic heterogeneity using ontology OWL language is used with different levels of expressive power. This was motivated by three sub languages, each of them varying in their trade-off between expressiveness and its inferential complexity. They are, in order of increasing expressiveness:

– OWL Lite: It supports the classification hierarchies and simple constraint features.
– OWL DL: The OWL Description Logics provides a subset providing the great expressiveness without losing the computational completeness and decidability with specific features.
– OWL Full: It supports the maximum expressiveness and the freedom of syntactical analysis however without the computational guarantees and its features.

Beside the most common approach using description logics or frame based ontology languages, several approaches exist that represent knowledge about the information to be integrated in a different way [5]. These approaches often also refer to these models as ontology's,

Formal Concept Analysis is one of the approaches that issued for the integration of information based on the calculation of a common concept hierarchy for different information sources

Object Languages with very different scopes and structures are frequently used by information integration systems.

Annotated Logics are sometimes used in order to resolve conflicts. Thereby, values of confidence or belief act as a basis for the calculation of a most promising fact to include into a common model.

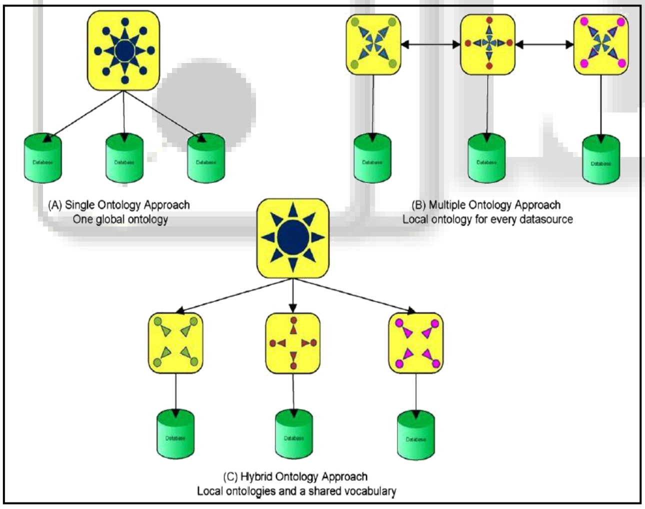


Fig. 1: Approaches

Well-known and the good methodologies exist to construct ontology; there are five widely accepted stages for building ontology

– Specification: Defines the purpose and the scope of the ontology are identified.
– Conceptualization: The specified conceptual model of the ontology is constructed which consists of the different concepts, relations and properties that can occur in the domain.
– Formalization: Here the conceptual model is translated into a formal model for example by adding axioms that restrict the possible interpretations of the model.
– Implementation: The formal model defined above is implemented in a knowledge representation language, for example OWL.
– Maintenance: The above implemented ontology has to be constantly evaluated, updated and corrected.

As the conceptual analysis, the Prostate Cancer Ontology (PCO) [6] is created for the development of the Prostate Cancer Information System (PCIS). PCIS is applied to demonstrate how the ontology is utilized to solve the semantic heterogeneity problem from the integration of two prostate cancer related database systems. As a result of the integration process, the SPARQL semantic query language is applied to perform the integrated queries across the two database systems based on PCO.

III. DATA MODELING AND ITS ANALYSIS OF DIFFERENT SYSTEMS

The data integration systems are characterized by an architecture based on global schema and a set of available sources. The available source that's present contains the real data while the global schema provides a reconciled and the integrated virtual view of the underlying sources.

In this survey paper, we have presented the analysis and comparisons of seven systems that use ontology to solve the problems involved in data integration. In order to do so, a conceptual framework with three main categories is created. They are semantic heterogeneity, architecture, and query resolution. The seven systems defined are as follows.

A. OBSERVER [Ontology Based System Enhanced with Relationship for Vocabulary heterogeneity Resolution]

In [7], OBSERVER is an approach that proposes managing multiple information sources through ontology's. OBSERVER uses the concept of data repository, which might be seen as a set of entity types and attributes. The architecture is based on wrappers, ontology servers and an IRM (Inter- ontology Relationship Manager). OBSERVER is classified as multiple ontology approach. In this system, each information source is represented through single ontology, thus a modification or addition of information to some source will only impact on the related ontology and on the IRM. Users can use any language based on description logics such as CLASSIC or Loom. The query construction is carried out by the user. It is followed by the access to underlying data and the controlled query expansion to new ontology's steps.

B. SIMS [Search in Multiple Sources]

In [8], the authors Knoblock, Arens. Y, Hsu. C has Discussed that the semantic heterogeneity, architecture, and the query resolution of the SIMS system. SIMS was created assuming dynamic information sources, i.e. changing information sources, availability of new information, etc…The sources can be databases and information sources such as HTML pages. The architecture is based on the wrapper/mediator. Wrapper is used to translate the data set description into a query, which is submitted to the source. Mediator is used to retrieve and process data. A global ontology approach is used in the SIMS. The ontology is represented in the Loom language. Users make a query in terms of the global ontology without knowing the terms or languages used by the underlying information sources.

Queries are written in high level languages. The first step to answer a query is transforming it into another query expressed in terms of concepts that correspond to information sources. The four reformulation operations are as follows.

- Select-Information-Source
- Generalize- Concept
- Specialize concept
- Decompose relation

SIMS uses the Semantic Query Optimization that can speedup database query answering by using knowledge intensive reformulation.

C. DOME [Domain Ontology Management Environment]

In [9], DOME is focused on ontology development by using software reverse engineering techniques. The most important architectural components are wrappers, a set of tools for extracting and defining ontology's and mappings between them, the mapping server and the ontology server. The DOME system uses the multiple ontology approach. DOME uses XRA as a tool to generate ontology's.

D. COIN [Context Interchange]

In [10], COIN system is with a goal of achieving semantics interoperability among heterogeneous information sources. It has mediator based architecture. This COIN technology uses a hybrid ontology approach.

E. KRAFT [Knowledge Reuse and Fusion/Transformation]

In [11], KRAFT was conceived to support configuration design of applications among multiple organizations with heterogeneous knowledge and data models. It uses the concept of "Knowledge fusion" to denote the combination of knowledge from different sources in a dynamic way.

KRAFT offers two methods for building ontology's: the building of shared ontology's and extracting of source ontology's. The development of shared ontology's are (a) ontology scoping, (b) domain analysis,(c) ontology formalization, (d) top-level-ontology. The minimal scope is a set of terms that is necessary to support the communication within the KRAFT network. The development of extracting ontology's introduces bottom-up approach to extract ontology from existing shared ontology's. This extraction process has two-steps. The first step is a syntactic translation from the KRAFT Exportable view of the resource into the KRAFT-schema. The second step is the ontological upgrade, a semi-automatic translation plus knowledge-based enhancement, where local ontology adds knowledge and further relationships between the entities in the translated schema.

F. Garlic

It addresses large scale multimedia information systems by considering specialized component systems to store and search for particular data types like image management systems. Garlic provides an object-oriented schema to applications, interprets object queries, creates execution plans for sending pieces of queries to appropriate data servers, and assembles query results for delivery back to

The applications even changes of capabilities do not affect the mediator. Garlic requires quite powerful wrappers, since query execution depends on a interactive communication between mediator and wrappers about the component's capabilities.

G. Shoe

In [6], SHOE is an ontology based knowledge representation language designed for the Web. SHOE uses knowledge-oriented elements and associates meaning with content by making each Web page commit to one or more ontology's. A real methodology proposing how to create an ontology does not exist (SHOE: Simple HTML Ontology Extension) but, one can define categories, relations and other components in an ontology. The SHOE-annotated web pages are uploaded on the web, the Expose tool has the task to update the repositories with the knowledge from this pages. This includes it list of pages to be visited and an identification of all hypertext links, category instances, and relation arguments within the page. The tool then stores the new information in the PARKA knowledge base.

H. DWQ

In [6],Further development within the DWQ project lead to a tool called i·com. i·com is a Supporting tool for the conceptual design phase. This tool uses an extended entity relationship conceptual (EER) data model and enriches it with aggregations and inter-schema constraints. i·com does not provide a methodology nor is it an annotation tool, it serves mainly for intelligent conceptual modeling.

Table 2 gives the comparison of different systems used in ontology based data integration.

Systems	Information sources	Architecture	Ontology use	Languages	Query
OBSERVER	HTML Pages, databases and files.	Wrappers, ontology servers and IRM	Multiple ontology approach	CLASSIC or Loom	Cost based query optimization
SIMS	HTML pages	Wrapper/mediator	Single ontology	Loom	Query subsumption
DOME	Structured and semi- structured data sources	Wrappers, mapping server and ontology server	Multiple ontology approach	CLASSIC	Cost based query optimization
COIN	traditional databases and semi structured sources	Mediator based architecture	Hybrid ontology approach	F- Logics	Cost based query optimization
KRAFT	Knowledge	Wrappers,	Hybrid	Classical frame	Constraint

	bases	mediators, facilitators and user agents	ontology approach	based representational language	based query
SHOE	PARKA knowledge base and HTML based	Wrapper/mediator based server	Multiple Ontology	CLASSIC	Constraint based query
DWQ	Knowledge bases	i-com based server	Multiple Ontology	CLASSIC or Loom	Constraint based query

Table 2: Comparison of different systems for ontology based data integration

IV. ONTOLOGY – CENTRIC DATA MANAGEMENT SYSTEMS

There are several methods for managing data through ontology concepts and interpretation. The need for effective data management is, in a large part, due to the fact that the massive amounts of digital data are being generated by modern instruments. The ability to provide a data management service is that, they can manage large quantities of heterogeneous data in multiple formats and structure the data with effective search, query and dissemination using technologies and processes. At the past, Database systems have traditionally been used successfully to manage research data in which database schemas are used as domain models. So the ontology language OWL is being widely used as a modelling language in a number of domains.

The PODD (Phonemics' Ontology Driven Data Management) [12] has been enhanced as an extensible, domain-agnostic architecture for scientific data management that uses an ontology-centric approach. This architecture is very much supportive to object oriented approach making PODD highly extensible. PODD repository is used in a large scale as database for maintaining structured and precise metadata around the raw data distributed in a reusable fashion.

A. Scientific Data and Resource Management Systems and Tools:

1) Fedora Commons is an open-source digital resource management System based on the principles of modularity, interoperability and extensibility were abstract concepts are defined as models. Data in Fedora Commons repositories are represented as objects, which contain data streams that store either Metadata or data.
2) Apache Jackrabbit is an open-source implementation of the Content Repository for Java Technology (JCR) API. In JCR; data is stored in a tree of nodes which holds the properties of arbitrary values.
3) Bioinformatics Resource Manager (BRM) is one example of client–server style data management software for bioinformatics research. The client software is installed on users' computers to access resources stored on a BRM server in a Postgre SQL relational database. The BRM server supports data acquisition from external sources.
4) Array Express is a public database of microarray gene expression data. Its three components, the Repository, the Warehouse and the Atlas which contains a large amount of data, summaries and analytical tools across many experiments and biological conditions.

B. Domain Modeling in Scientific Research:

1) CCLRC

Central Laboratory of the Research Councils
CCLRC Scientific Metadata Model that models data holdings in scientific activities in free text.

2) OBI

The Ontology for Biomedical Investigations (OBI) is an ongoing effort aimed at developing an integrative ontology for biological and clinical investigations (top-down approach).

3) FuGe

The Functional Genomics Experiment Model (FuGe) is an extensible modelling framework for high-throughput functional genomics experiments for the molecular biology research community.

For all these there are in need of some Requirements of data management systems,

– Data storage and management
– Data contextualization
– Data security
– Data identification and longevity
– Data reuse and integration
– Model extensibility

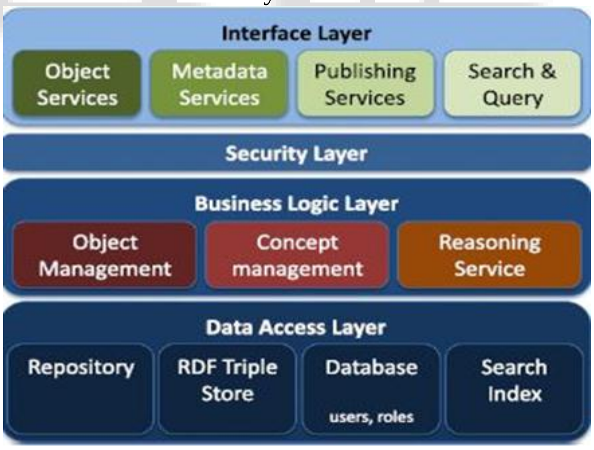


Fig. 2: Layered Architecture

V. METHODOLOGIES INITIATED FOR ASSESSMENT OF DATA QUALITY

The main goal of this survey paper is to provide a systematic and the comparative description of different methodologies of assessing the data quality. The dimensional classifications of quality are provided. By analyzing these classifications it is possible to define the basic set of quality dimensions, including the accuracy, the completeness, the consistency and the timeliness. The different methodologies are as follows.

A. The TDQM (Total Data Quality Management) Methodology

In [13], the TQDM methodology was the first general methodology published in the data quality literature. The objective of TDQM is to extend the data quality, the principle of Total Quality Management (TQM). TDQM proposes a language for the description of information production (IP) processes, called IP-MAP. IP-MAP has been variously extended, towards UML and also to support organizational design. TDQM's goal is to support the entire end-to-end quality improvement process, from requirement analysis to implementation. TDQM cycle consists of four phases that implement a continuous quality improvement process: definition, measurement, analysis and improvement.

B. TQMIS (Total Quality Information Management System) Approach

In [14], TIQM methodology has been proposed to support data warehouse projects. The methodology assumes the consolidation of operational data sources into a unique integrated database, used in all types of aggregations performed to build the data warehouse. The goal is to improve the data quality level.

C. CDQ [Complete Data Quality]

In [15], CDQ follows an approach similar to ISTAT with more emphasis on the autonomy of organizations in the cooperative system. In fact, the resolution of heterogeneities proposed as best practices are performed through record linkage on a very thin layer of data, namely the identifiers.

D. AIMQ (A Methodology for Information Quality Assessment)

In [16], the AIQM methodology is the only information quality methodology focusing on benchmarking which is an objective and domain independent technique for quality evaluation. Gap Analysis Technique is advocated as a standard approach to conduct benchmarking and interpret results.

E. AMEQ (Activity-based Measuring and Evaluating of Product information Quality) methodology

The main goal of AMEQ methodology is to provide the rigorous basis for Product Information Quality (PIQ) assessment and improvement in compliance with organizational goals [17].

F. IQM (Information Quality Measurement)

In [18], the fundamental objective of the IQM methodology is to provide an information quality framework tailored to the Web data. In particular, IQM helps the quality based selection and personalization of the tools that support webmasters in creating, managing and maintaining websites.

G. DaQuinCIS (Data Quality in Cooperative Information System) Methodology

In DaQuinCIS, instance-level heterogeneities among different data sources are dealt with by the DQ broker. Different copies of the same data received as responses to the request are reconciled by the DQ broker, and a best quality value is selected [19].

Methodologies	**Data quality dimension**	**Type of data**	**Extensible to other dimensions and metrics**
TDQM	Timeliness, security	Monolithic, distributed	Fixed
TIQM	Concurrency of redundant data	Focused on monolithic and distributed	Fixed
CDQ	Syntactic/ semantic accuracy	Monolithic, distributed	Open
AIMQ	Freedom from errors	Monolithic	Fixed
AMEQ	Unambiguity, consistency	Monolithic	Open
IQM	Accuracy, interactivity	Strongly focused on web	Open
DaQuinCIS	Consistency, currency	Monolithic, distributed	Open

Table 3: Comparison of Different Methodologies of Data Quality

VI. SEMANTIC SIMILARITY IN BIOMEDICINE

The determination of semantic similarity between word pairs is an important component of text understanding that enables the processing with classification and structuring of the textual resources. In the past, the several approaches are proposed for assessing word similarity by exploiting different knowledge sources. Some of these measures have been adapted to the biomedical field by incorporating domain information extracted from clinical data or from medical ontology's. Some of the approaches are introduced and analysed with respect to the considered knowledge bases. After that, a new measure is exposed based on the exploitation of the taxonomical structure of a biomedical ontology is proposed by Using SNOMED CT [5] as the input ontology. As the digital libraries have become valuable resources for clinical and translational research, the Semantic technologies play an important role in this context enabling proper interpretation of this information.

The determination of the semantic similarity between words being successfully applied in many natural language processing tasks such as word sense disambiguation, document categorization or clustering, word spelling correction, automatic language translation, ontology learning or information retrieval. Semantic similarity makes understandable by computing the likeness between words as the degree of taxonomical proximity. For example, similarity between bronchitis and flu, because both are disorders of the respiratory system.

According to the type of domain knowledge exploited, number of relevant biomedical ontology's, knowledge repositories and structured vocabularies that model and organize concepts in a comprehensive manner. The Well-known examples are follows as,

- MeSH (Medical Subject Headings) for the indexing literature.
- ICD taxonomy (International Classification of Diseases) for recording death causes and diseases.

– SNOMED CT (Systematized Nomenclature of Medicine, Clinical Terms) using SNOMED CT as knowledge source, evaluating them over particular datasets or in the context of a concrete application.

From the Similarity measures based on the previous structure of taxonomy, the path-based measures are the most adequately used and implemented. As they exploit the geometrical model of the ontology, no need of pre-calculus or pre-processing, this makes them more computationally efficient.

VII. Conclusion

Data Integration is the hectic task in ontology via semantic web. Resolving of the semantic heterogeneity is the challenging task in data integration. In this paper we have discussed several systems and methodologies that use ontology to solve the problem involved in data integration. Different methodologies are also used to improve the integrated data quality that can be used in biomedical field.

References

[1] Maurizio Lenzerini, "Data Integration: A Theoretical Perspective", Proceedings of PODS, page no (233-246), December 2002.
[2] Stijn V, Femke O , Leanneke L , Frederik V , Pieter D, Bart D, Tom D, Filip D T, " Efficient data integration in the railway domain through an ontology-based methodology", Transportation Research Part C: Emerging Technologies Volume -19, Issue - 4, Page no (617–643) ,August 2011.
[3] Batini C, Lenzerini M, Navathe S B, "A comparative analysis of methodologies for schema integration", Journal of ACM Computing Surveys, (CSUR 1986), Volume-18, Issue-4, page no (2-6), January 1986.
[4] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner, "Ontology-based Integration of Information - A Survey of Existing Approaches," In: Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, 2001, Vol. pp. (108-117).
[5] Montserrat B, David S, Aida V, "An ontology-based measure to compute semantic similarity in biomedicine", Journal of Biomedical Informatics archive Volume - 44 Issue- 1, Page no (118-125), February 2011.
[6] Hua M, Frank J. Manion, Elizabeth G, Yu-Ning W, Eric Ross, J. Robert Beck, " Integration of prostate cancer clinical data using an ontology", Journal of Biomedical Informatics Volume - 42, page no (1035–1045), 2009.
[7] Mena, E., Kashyap, V., Sheth, A. and Illarramendi, A. "Observer: An approach for query processing in global information systems based on interoperation across pre-existing ontologies", Kluwer AcademicPublishers, Boston. http://citeseer.nj.nec.com/mena96observer.html,1-49, 2000.
[8] Arens, Y., Hsu, C., Knoblock, C.A, "Query processing in the SIMS Information Mediator". Advanced Planning Technology, Austin Tate (Ed.), AAAI Press, Menlo Park, CA, page no (61-69), 1996.
[9] Cui, Z. and O'Brien, P. "Domain Ontology Management Environment". In Proceedings of 33rd Hawaii International Conference on System Sciences, 2000.
[10] Goh, C.H., Bressan, S., Siegel, M. and Madnick, S. E. "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information". ACM Transactions on Information Systems, Vol. 17(3), pp (270–293), 1999.
[11] Gray, P.M.D, Preece, A., Fiddian, N.J. and colab, "KRAFT: Knowledge Fusion from Distributed Databases and Knowledge Bases", Proceedings of 8th International Workshop on database and Expert Systems Applications (DEXA'97), 1997.
[12] Yuan-Fang L, Gavin K, Faith N, Philip W, Jane H ,"An ontology-centric architecture for extensible scientific data management systems", Future Volume - 29, Issue - 2, , Page no (641–653) ,February 2013.
[13] Wand, R. "A product perspective on total data quality management". Comm. ACM 41, 2, 1998.
[14] English, L. "Improving Data Warehouse and Business Information Quality". Wiley & Sons, 1999.
[15] M.J. Carey, L.M. Haas, P.M. Schwarz, M. Arya, W.F. Cody, R.. II,J.H. Williams, and E.L. Wimmers, "Towards heterogeneous multimedia information systems: The Garlic approach", IBMAlmaden Research Center, San Jose, CA, 1996.
[16] Lee, Y.W., Strong, D. M., Kahn, B. K., Andwang, R. Y. "AIMQ: A methodology for information quality assessment". Inform. Manage. 40, 2, pp (133–460), 2002.
[17] Su, Y. and Jin, Z. 2004. "A methodology for information quality assessment in the designing and manufacturing processes of mechanical products", In Proceedings of the 9th International Conference on Information Quality (ICIQ). Page no (447–465), December 2004.
[18] Eppler, M. and Munzenmaier, P, "Measuring information quality in the Web context: A survey of state-of-the-art instruments and an application methodology". In Proceedings of the 7th International Conference on Information Systems (ICIQ).ISTAT, 2002.
[19] Scannapieco, M., M.Virgillito, Marchetti, M., Mecella, M., and Maldoni, R.. "The DaQuinCIS architecture: a platform for exchanging and improving data quality in Cooperative Information Systems", Inform. Syst. 29, 7, pp (551–582) ,January 2004.

