Ontology -Based Data Integration Methods: Framework for Medical Process

[1]Duru Rosetta Okwuchi., [2]Agbakwuru A. Onyekachi.,[2]Amanze Bethran C., [2]Agbasonu V. Chinedum
[1]Department of Computer Engineering, Abia State Polytechnic, Aba, Abia State, Nigeria
[2]Department of Computer Science, Imo State University, Owerri, Imo State, Nigeria
amanzebethran@yahoo.com

D.O.I: 10.56201/ijasmt.v8.no4.2022.pg91.101

Abstract

The research addressed the problems, by having a hybrid of ontologies and virtual data integration in order to enhance medical intelligence process. A hybrid model for enhance medical intelligence process using ontology based and virtual data integration technique were developed. The design provided for a database system for storing medical records, software for enhanced Medical Intelligence Process that would be more user-friendly, flexible, adaptive, intelligent, agile and automatic in integrating and analyzing medical data thereby helping medical practitioners at various levels to make realistic intelligent and real-time decision on critical health issues. Object Oriented Analysis and Design Methodology (OOADM) were adopted in the design of the system. The system achieved integration of various patients medical records from different hospitals using ontology based and virtual data integration technique that will allow clinic data of one patient collected together to form a Combinational resource, and could be accessed by physician if authority is assigned to the physician

Keywords: *Patients, Medical Records, OOADM, Hospitals*

Introduction

Ontologies are one of the most successful ways of representing actionable knowledge in biomedicine (Rosse, 2019). Two of the reasons for this success are their ability to capture biomedical knowledge in a formal but simple, powerful and incremental manner, and their easy application in the reasoning processes performed by medical decision support systems (Ouwens, 2015).In health care, the most common, complex and resource-consuming clinical cases to deal with correspond to chronically ill patients, who are a kind of patients that deserve long term and simultaneous assistance provided by several sorts of professionals, as for example family doctors, specialists, nurses, or social workers. In order to deal with this highly variable kind of patients, we need mechanisms to personalize the knowledge describing both the condition of these patients (each individual patient is a potential different case with specific diseases, syndromes, social needs, signs and symptoms), and the intervention plan for these patients (the actions to be followed for different patients are eventually very varied). But we also need mechanisms to assess whether the decisions and recommendations on these patients are correct or not in part because the possibilities of over- and under-treat these kinds of patients can be very high. Also, integration becomes a crucial challenge as heterogeneous data is generated by various healthcare systems. This has to do with how to integrate various types of data including patient demographics and environmental data, clinical monitoring systems, pathology and radiology imaging data, textual data from clinical reports etc. Bianchi et al. (2019) describe the integration of clinical, environmental and genetic data and point out with examples how ontologies are used to normalize data from various disparate systems. This thesis examines how to utilize modern technologies available and design a system that assists health workers in decision making. The system will allow for easy and fast reach to information. One of the core objectives is to support physicians, nursing, laboratory, and radiology staffs in various hospitals to exchange data and information electronically. This is due to importance of time that helps decision makers implement necessary and most appropriate procedure for patient, especially if provided by internet. Since most care givers or clinicians do non-resident consultations, and some have hospitals at different locations, most of the data or records concerning patients are either incomplete or totally out of reach; and that is why incorporating patients' records into an integrated platform will allow the different service providers to the healthcare system share in the networked records (Heldt, 2019). In order to support the Ontology content accessing, this thesis presents a hybrid model to locate and get hospital data which are stored in heterogeneous hospital information systems. In the proposed method, clinic data of patient is defined as resource with unique URL address. Related clinic data of one patient is collected together to form a Combinational resource, and could be accessed by physician if authority is assigned to the physician. A wrapper system integrating this two personalization and the decision support tool was also implemented in the new system. With this system, the ontology can be directly maintained (and extended) by health-care professionals without any intervention of information technology specialists. As the ontology grows up with new medical concepts and properties, the diversity of patients that our system is able to deal with increases, the possible intervention plans can be more detailed and accurate, and the decision support system automatically becomes more powerful. Therefore, the health-care utility of the system is exclusively dependent on the incorporation of new knowledge in the case ontology. In computer science, ontology is a controlled vocabulary that describes objects and the relations between them in a formal way. Ontologies provide a sound basis for sharing domain knowledge between human and computer programs, or between computer programs. An ontology normally defines concepts (or classes), individuals (or instances), properties, relationships and their constraints. Logical formalization of ontology language ensures semantic interpretation, i.e. inference, by computer programs. Ontology is a major instrument toward realization of the Semantic Web vision (Rosse, 2019). Helena and Lidia (2019) defined ontology as a formal, explicit specification of a shared conceptualization and further defined conceptualization to be an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Ontologies allow more complete and precise domain models. Ontologies are intended to be shared and reused, and the approach is perceived to be beneficial. Ontology-based design has an advantage of being syntactically correct and semantically consistent as a model.

According to Vipul, et al. (2018), Ontologies provide a common language to express the shared semantics and consensus knowledge developed in a domain. This research will explore this in its ontology-based integration technique phase. 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 expressivity of the ontology used in the integration process. Inter-application interoperability has been long seen as schema mapping and data integration problem. In this manner integration requires mapping systems and integration systems that uses those mappings to answer queries or translate data across data sources. There are three different categories of ontology-based integration approaches; single ontology approaches (SOIA), multiply ontology approaches (MOIA), hybrid ontology approaches (HOIA) (Bostjan and Vili, 2018)

Single ontology approach: A single ontology is used as a global reference model in the system. This is the simplest approach as it can be simulated by other approaches. The SIMS (Search in Multiple Sources) system is a prominent example of this approach. The Structured Knowledge Source Integration component of Research Cyc is another prominent example of this approach. (Title = Harnessing Cyc to Answer Clinical Researchers' Ad Hoc Queries)

Multiple ontologies: Multiple ontologies, each modelling an individual data source, are used in combination for integration. Though, this approach is more flexible than the single ontology approach, it requires creation of mappings between the multiple ontologies. Ontology mapping is a challenging issue and is focus of large number of research efforts in computer science. The OBSERVER (Ontology Based System Enhanced with Relationship for Vocabulary heterogeneity Resolution) system is an example of this approach.

Hybrid approaches: The hybrid approach involves the use of multiple ontologies that subscribe to a common, top-level vocabulary. The top-level vocabulary defines the basic terms of the domain. Thus, the hybrid approach makes it easier to use multiple ontologies for integration in presence of the common vocabulary.

Ontologies enable the unambiguous identification of entities in heterogeneous information systems and assertion of applicable named relationships that connect these entities together. Specifically, ontologies play the following roles;

a. Content Explication: The ontology enables accurate interpretation of data from multiple sources through the explicit definition of terms and relationships in the ontology.
b. Query Model: In some systems like SIMS, the query is formulated using the ontology as a global query schema.
c. Verification: The ontology verifies the mappings used to integrate data from multiple sources. These mappings may either be user specified or generated by a system.

Ontology allows more complete and precise domain models. They are intended to be shared and reused and one of the main advantages of its design is that it has syntactically correct and semantically consistent model and reasoning over them also provides retrieval of additional rules that were possibly not recognized during the design phase. The issue of structural difference in the data that data warehouse integrates and stores (relational databases, Object databases, unstructured data, etc) as well as when sources have the same structure but with data integration problem of synonyms and homonyms, ontologies are used to describe them as truly equivalent despite their appearing in different databases, forms and names (i.e. they describe the same objects. (Helena and Lidia, 2019). In any domain such as that of business intelligence systems, ontology play the role of providing a common language to express the shared semantics and consensus knowledge developed in such domain. The shared semantics are typically captured in the form of various domain specific ontologies and classifications. The concepts provide the shared semantics to which various data objects and data interpretations can be mapped to enabling integration across multiple business intelligence data sources and domains. (Vipul, et al, 2018), After the 80s the massive adoption of database systems inside organizations leads to the need to integrate different data repositories with possibly incompatible data schemata. The process of integrating different data residing at different sources to provide a unified view of this information is known as data integration problem. As early stated in this research, data warehousing is one of the approaches to data integration problem solution. Here data originates from different sources and are submitted to a process called ETL (Extraction, Transformation and Loading) and the stored into a new database with a single and usually de-normalized schema. The resultant database is often structured to store various aggregations of the sources' data in order to speedup query processing. Architecturally, data warehousing can be seen as a tightly coupled approach because the integrated data reside in a single place at query time. Recent approaches to data integration as would be found in our proposed research are sometimes "loosely coupled". (Letizia, 2016). A data integration system provides a uniform interface to distributed and heterogeneous sources. These sources can be databases as well as unstructured information such as files, HTML pages, etc. One of the most important problems within data integration is the semantic heterogeneity, which analyzes the meaning of terms included in the different information sources. As earlier stated in this research, Data integration is concerned with unifying data that share some common semantics but originate from unrelated sources.

Ontology-Based Data Integration Methods

Data integration methods generally depends on which integration architecture the developer is quite familiar with, how much is known about heterogeneity of data sources, etc. In the models shown in figure 1 selected criteria for analyzing is used, ontologies and data sources are represented as classes and semantic relationships between data source content and ontology as links and links between local ontologies as mappings. The data integration task automatically performs data extraction and integration from both structural and semi-structured data sources. The proposed model integrates and reuses data using ontologies by relevant criteria. As earlier stated, ontology is a data model which consists of classes, properties and relationships between them as its parts. Its power lies in the ability to represent relationships between the classes. The benefits of the model mainly are its runtime interpretation, an open-world assumption and its ability to clearly interpret disseminating of knowledge between people and applications. It supports integration task as it describes the exact content and semantics of data sources more explicitly.

Also, in contrast to database schemas which are static, ontology schemas are highly dynamic and are an evolving object. (Viriginija and Rimantus, 2018)

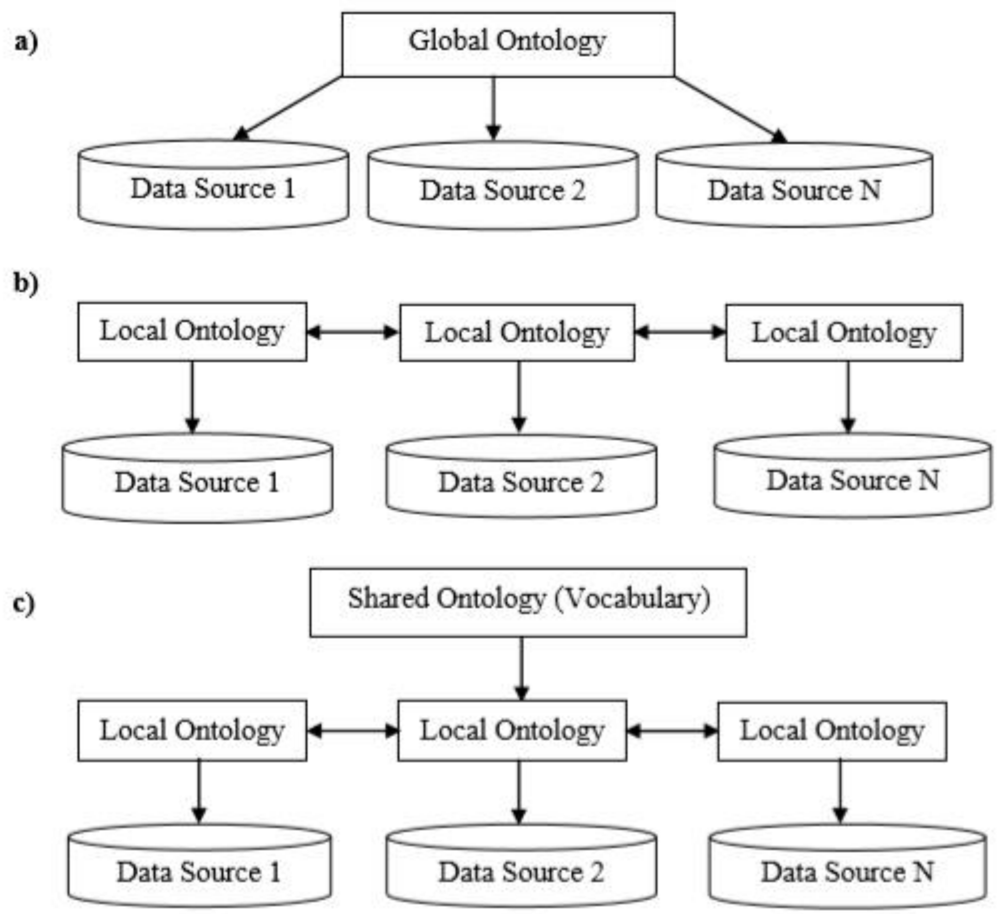

Figure 1: The Methods: a) Global, b) Multiple c) Hybrid.   (Virginija and Rimantas, 2011)

The global ontology as shown in figure 1a) uses a single ontology which as a single main stage of building global ontology by domain expert that knows the semantics for all data sources. It can also be a combination of several specialized ontologies. It describes data from heterogeneous data sources from which query is executed via the main ontology. Every data source in this method is related. It is applied to integration solutions where all data sources to be integrated provide the same view on a domain. The b) of figure 1 is the multiple ontology method which uses local ontologies and mapping rules between them. Each data source is described by its own ontology and the mapping rules can be modified according to the dynamic change of data source. It has two main stages; building local ontologies and defining mappings. This method local ontology describes data from heterogeneous data sources such that integrated query is executed via the local ontologies. The essential feature of this method is that the ontologies for individual data source could be developed or changed without respect to other semantic relations, data sources or their ontologies. In the case of the hybrid ontology method c) of figure 1, it uses a vocabulary of a domain to represent a shared ontology, a local ontology and mapping rules between them. The specification of a vocabulary includes definitions of classes, relations, functions, and other objects. The mapping rules can be modified according to the dynamic change in data source. That is, it integrates users' profiles, data warehouse, online Analytical Processing, Data Mining and underlying Enterprise Integration System (EIS). It aims to develop mechanisms for smooth mapping from user-defined keywords to meta-data items in data warehouse or physical attributes and entities dispersing in business system service or office system service (BSS/OSS). This architecture requires ontological engineering techniques for effective operation, such as building ontological name space and semantic relationships for organizing items in the domains; mechanisms for ontological transformation and mapping intra or inter domains; services for ontological query and search in the warehousing (Longbing, et al, 2015). The advantage of this hybrid method which this research will implement is that new data sources can easily be added without the need to modify the mappings or the shared vocabulary. So this architecture gives more autonomy to data sources. The use of shared vocabulary makes the source ontologies comparable and avoids the disadvantages of single or multiple ontology methods. Table 1 gives a summary of the benefits and drawbacks of the ontology-based integration methods.

Table 1: **Advantages and Disadvantages of ontology-based integration methods (Virginija and Rimantas, 2018)**

	Ontology-based Architectures		
Criterions	**Global**	**Multiple**	**Hybrid**
Evaluation of Semantic Heterogeneity	Useful for Systems which have the same view on a domain	Useful for Systems which have the same on a domain	Useful for Systems which have different view on a domain.
Appending new data sources	Some modification is necessary in the global ontology	Supports an opportunity to append the new data source with some adaption in other ontologies	New data sources can easily be added without the need of modification.
Elimination of data sources	Some modification is necessary in the global ontology	Supports an opportunity to remove the data source with some adaption in other ontologies (need to remove relation between ontologies)	Data sources can easily be removed without the need of modification

		Data	
Comparison of multiple ontologies	Impossible	Difficult because of lack of a common vocabulary	Simple because ontologies use a global shared vocabulary

Ontology-based data integration is an effective method to cope with the heterogeneous data. This solution is based on the idea of decoupling information semantics from the data sources. Moreover, ontologies support dynamic domains better. For this reason, it is necessary to analyze data source elements: data, schema, schema elements and content, values, entities and attributes, query result classes. It is known that ontology-based search system gives the user more meaningful query results than the normal search system, which queries data with syntactic parameters. The query result is based on data retrieval methods. (Virginija and Rimantas, 2018). Ontology Integration is similar to Ontology merging only that in ontology integration the integrated ontology is created reusing parts of source ontologies as they are. Both has a key task of in consistency checking process which must ensure the absence of unforeseen or wrong implications into the integrated ontology. The ontology is created by using Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), DARPA Mark-up Language (DAML) + Ontology Interchange Layer (OIL) and Ontology Web language (OWL). Among these, the OWL is more powerful than others. The OWL has well defined semantics and highly optimized implementation system. The data quality is often defined as "fitness for use". Data is fit for use whenever a user, (1) is able to get information, (2) is able to understand it, (3) finds it applicable to a specific domain and purpose of interest and (4) believes it to be credible. The Key measures of data quality are data completeness, data consistency and data accuracy. Completeness is defined as the extent to which data are of sufficient granularity for the task at hand. Data consistency expresses the degree to which a set of data satisfies a set of integrity constraints. Data accuracy is defined as the closeness between the given value and the correct representation of the same in real life phenomenon. A unified view is created to resolve the semantic conflict among different heterogeneous databases by using ontology. This view is used by the user for shopping and business analysts for decision support. (Hema and Chandramathi, 2018). This ontology technique will provide quick access to documents and information with the help of taxonomy created from the concepts, called a concept map, with the incorporation of ontology array indexing. The system will be adaptive as it will retrieve the most relevant information as well as documents which are close to the user's queries and the array-indexing is key to this concept as it helps in obtaining an inter-relation between the documents and information. It also, ensures the redundancy of concepts or document in the database. Business intelligence requires the acquisition and aggregation of key pieces of knowledge from multiple sources in order to provide valuable information to customers or feed statistical business intelligence models and tools. Since the analyst of business intelligence has lots (massive) of information to extract and process, there is the challenge as par its acquisition and use of its semantic information, ontology-based data integration approach is applied in this research as an effective way to resolve the challenge. It implies the process of identifying in text or other sources relevant concepts, properties, and relations expressed in ontology. (Maynard, et al, 2018). As an intelligent system like humans which implies having the ability to take in data from their environment, understand the meaning and significance of the information and then act appropriately, business intelligent and analytic has to do with integrating all information streams produced by a firm into a single, coherent enterprise-wide set of data, and then, using modelling, statistical analysis tools (normal distribution, correlation and regression analysis, chi square analysis, forecasting and cluster analysis), and data mining tools (pattern discovery and machine learning), to make sense out of all the massive data so managers can make better decisions and better plans, or at least know quickly when their firms are failing to meet planned targets.

The tools that use ontologies for data integration can be compared on the following ways;

1. The use of ontologies which implies the different roles ontologies play in the data integration process in terms of making the context explicit and in the expression semantics or as a global query model.
2. Ontology representation which has to do with representational capabilities of the used ontologies that can vary among the different tools.
3. Use of Mapping has to do with linking the ontologies to the actual information and to each other if multiple ontologies are used.
4. Ontology engineering that deals with the building and re-use of ontologies in different manners.

Class Diagram

Class diagrams are one of the most useful types of diagrams in UML as they clearly map out the structure of a particular system by modeling its classes, attributes, operations, and relationships between objects.

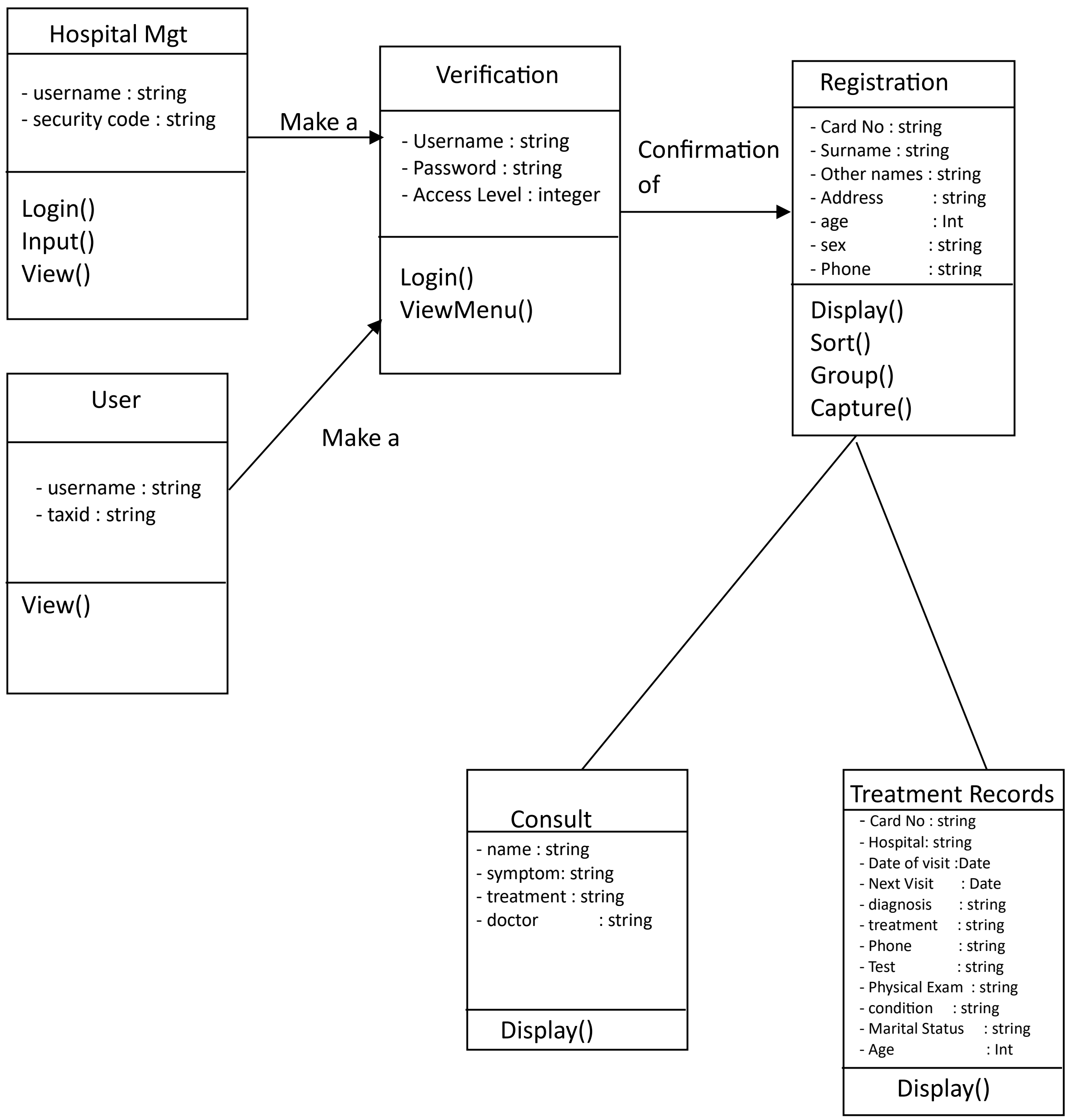

Figure 2: Class Diagram of the proposed system

Figure 2 shows various information sources that is needed at various stages of the data integration processes.

The standard class diagram is composed of three sections:

- **Upper section:** Contains the name of the class. This section is always required, whether you are talking about the classifier or an object.
- **Middle section:** Contains the attributes of the class. Use this section to describe the qualities of the class. This is only required when describing a specific instance of a class.
- **Bottom section:** Includes class operations (methods). Displayed in list format, each operation takes up its own line. The operations describe how a class interacts with data.

Conclusion

In a decentralized system, data are rarely collected and stored at a single entry point especially in the health sector. Integration from multiple heterogeneous sources is a prerequisite step for many applications, e.g., decision aids, data/information fusion and data mining. It is also a prevailing task by many organizations in order to improve their knowledge sharing as well as the efficiency of their operations. This will be of immense benefit to physicians who are in need of these vast amounts of knowledge for their daily life saving operations. Utilizing ontology-based data integration and virtual data integration is an attractive avenue as it is also a key factor for enabling interoperability. However, integrating vast amount of information from different sources is a difficult, complex and demanding task. The use of ontology-based data integration systems and virtual data integration tools to automate partly the data integration task and reduce this effort has been achieved in this paper.

References

Rosse, C. ,Mejino,J.L.V. (2013) A reference ontology for biomedical informatics: the Foundational Model of AnatomyJ Biomed Inform, 36 (2003), pp. 478-500 Article Download PDFView Record in Scopus Google Scholar

Ouwens, M. ,Wollersheim, H. , Hermens, R. , Hulscher, M. , Grol,R. (2015) Integrated care programmes for chronically ill patients: a review of systematic reviews. Int J Qual Health-Care, 17 (2) (2005), pp. 141-146

Bianchi,S. et al., (2019) “Biomedical data integration - capturing similarities while preserving disparities,” Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2019, pp. 4654–4657, 2019.

Heldt,T. and Verghese, G. C. (2010) “Model-based data integration in clinical environments,” Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Helena, V.& Lidia, R. (2009) “Ontology-based Data Warehouse Development Process” Proceedings of the ITI 2009 31st Int. Conf. on Information Technology Interfaces, June 2225, 2009, Cavtat, Croatia, retrieved on June, 2021.

Vipul, K., Kei-Hoi, C., Donald, D., Matthias, S., Scott, M., Joanne, L., Susie, S., Ivan, H. and Raymond, H. (2008) "Chapter 5 An Application Of Ontology-Based Data Integration For Biomedical Research" retrieved on May, 2016

Virginija, U.& Rimantas, B. (2011) "Ontology-based Foundations for Data Integration" BUSTECH 2011: The First International Conference on Business Intelligence and Technology Copyright (c) IARIA, 2011. ISBN: 978-1-61208-160-1 retrieved on October, 2014

Bostjan, G., Vili, P. (2014) Automating ontology based information integration using service orientation. Faculty of electrical engineering and computer science University of Maribor Smetanova ulica 17, 2014 Maribor SLOVENIA

Letizia, T. C. (2016) "An Ontology-Based Data Integration System: Solving Semantic Inconsistencies" Relatore: Letizia Tanca Correlatore: Ing. Carlo Curino Tesi di Laurea di: Giorgio ORSI matr. 674222 Anno Accademico 2015-2016, retrieved on October, 2020

Longbing, C., Chengqi, Z., & Jiming, L., (2015) "Ontology-Based Integration of Business Intelligence" Retrieved from www.w639.pdf/ on Dec.5, 2017

Hema, M. S. & Chandramathi, S., (2013) "Quality Aware Service Oriented Ontology Based Data Integration" WSEAS Transactions on Computers E-ISSN: 2224-2872 Issue 12, Volume 12, December 2013, retrieved on November, 2015