IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.5, May 2022 E-ISSN: 2347-2693

A Hybrid Model for Enhanced Medical Intelligence Process using Ontology Based and Virtual Data Integration Technique

Asogwa E.C.[1], Ejiofor V.E.[2], Amanze B.C.[3*], Agbakwuru A.O.[4], Belonwu T.S.[5]

[1,2,5]Department of Computer Science, Faculty of Physical Sciences, Nnmadi Azikiwe University, Awka. Nigeria
[3,4]Department of Computer Science, Faculty of Physical Sciences, Imo State University, Owerri, Imo State, Nigeria

Corresponding Author: amanzebethran@yahoo.com, +2347062523438

DOI: https://doi.org/10.26438/ijcse/v10i5.3642 | Available online at: www.ijcseonline.org

Received: 27/Apr/2022, Accepted: 11/May/2022, Published: 31/May/2022

***Abstract*—**The business model developed focused on expert system for health sector that uses intelligent agent to guide doctors in accurately carrying out disease control procedures. The objective of the paper is to create ontology-based data integration (OBDI) system process model that can uses intelligent agent to guide doctors in accurately carrying out disease control procedures. The system developed used to manage a disease registry that consists of the concepts of the domain, the attributes characterizing each disease, the different symptoms, and treatments. A model for enhanced medical intelligence process using ontology based technique 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) 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. Ontology-based data integration technique for disease control procedure achieved 95% accuracy in predicting the disease control procedure.

***Keywords*—**Patients, production rule, OOADM, OBDI technique and Expert System

1. INTRODUCTION

Medical intelligence is the ability to detect and cure an ailment on time with minimal effort. It requires vast knowledge on disease symptoms, cure, and this can only be achieved by having a data warehouse build from the knowledge of medical experts. To apply medical intelligence effectively, the healthcare condition of the patients ascertained. Ontology for data integration is as a data model that consists of parts such as classes, properties and relationships between them. It says the term ontology refers to a machine-readable representation of knowledge, particularly for automated inference. .The health-care condition of a patient defined as all the past and current medical and social information about the patient that may affect the professional immediate and short-term management of that patient. In this research, this information corresponds to all the diseases, syndromes and social issues that are diagnosed for the patient, the signs and symptoms (including family medical history), the problem assessments performed (i.e., medical, social, cognitive, and mobility tests), and the current interventions, either pharmacological, rehabilitative, nurse care, social care, counseling, and special medical services. In healthcare system all over the world, the amount of patient oriented data is constantly growing. More hospitals are opening up with various departments / units. For example, the Intensive Care Unit (ICU) is an extremely data intensive environment where large volumes of data from patient monitoring and observations are recorded continuously. Physicians and nurses could generate such patient oriented data from medical devices, laboratory results, electronic prescriptions, therapeutic decisions, and clinical observed values. These data disintegrated and access to the data done by requesting for it through the various hospital units. This is not only time consuming but obsolete. The following are the existing challenges of the medical system;

1. Lack of fast, accurate, reliable and intelligent software solutions that can help healthcare practitioners make decisions that would solve urgent, and in some cases, complex medical problems in real-time.
2. Cost of processing and analyzing large volumes of data in a medical environment is high most especially in terms of time consumption.

Ontology-Based Data Integration (OBDI) Technique
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 [1]. [2] 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 explicitly defined. Ontologies allow more complete and precise domain models. Ontologies intended shared and reused, and the approach perceived to be beneficial. Ontology-based design has an advantage of being syntactically correct and semantically consistent as a model.

[3] 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 classified as Global-As-View (GAV). The effectiveness of ontology based data integration closely tied to the consistency and expressivity of the ontology used in the integration process. Inter-application interoperability 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) [4]

Single ontology approach: A single ontology used as a global reference model in the system. This is the simplest approach 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 modeling an individual data source used in combination for integration. However, 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 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 intended to be share 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 provides retrieval of additional rules possibly not recognized during the design phase. 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 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 mapped to enabling integration across multiple business intelligence, data sources and domains. 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.

Heterogeneity classified into four categories: (1) structural heterogeneity, involving different data models; (2) syntactical heterogeneity, involving different languages and data representations; (3) systemic heterogeneity, involving hardware and operating systems; and (4) semantics heterogeneity, involving different concepts and their interpretations. The semantic heterogeneity deals with three types of concepts: the semantically equivalent concepts, the semantically unrelated concepts, and the semantically related concepts. In the first case – semantically equivalent concepts – a model uses different terms to refer the same concept, e.g. synonymous, or some properties modeled differently by different systems, for example, the concept length may be “meter” in one system and “mile” in one another. In the second case – semantically unrelated concepts – the same term may be used by different systems to denote completely different concepts; and in the last case – semantically related concepts – different classifications may be performed, for example one system classifies “person” as “male” and “female” and other system as “student” and “professor”.

2. RELATED WORKS

Table 1: Summary of Related works

Author	Techniques	Work done	Limitations
[5]	Ontology-based	clinical reminder system that link clinical guideline knowledge with patient registries	The paper didn’t integrate electronic health record (EHR) standards
[6]	open data integration platform	facilitates centralization of data assets	Lacks analytics, data visualization, monitoring and reporting functionalities for clinical decision support
[7]	machine learning	facilitating personal health care, reducing costs of health care, and improving outcomes	The work was carried out using a single hospital and may not represent the facts when you broaden the scope.
[8]	Adaptative Neuro-Fuzzy System (ANFIS)	The obtained simulation results demonstrate the efficiency of using ANFIS model in the identification of heart attacks	The intelligent system was limited to heart attacks only
[9]	Survey	A major finding of the survey is that although significant advances have been made in introducing AI technology in critical care, successful examples of fielded systems are still few and far between	Theoretical review. No practical implementation
[10]	cloud computing	Addressed the challenge of sharing medical data	Didn’t incorporate medical intelligence
[11]	Expert System	help a great deal in identifying those diseases and describing methods of treatment to be carried	Lacks data integration
[12]	Expert System	Able to give appropriate diagnosis and treatment for two heart diseases namely; angina pectoris and infarction	Is limited to two heart diseases
[13]	Fuzzy	The analysis clearly shows the effectiveness and accuracy in the system performance through false result elimination	Lacks data integration
[14]	Multi Agent Enhanced Business Intelligence	Results indicate that the pMAEBI managed stores performed better (in terms of profit) than the comparison stores	Narrowed to product pricing
[15]	ontology-based personalization	Helps health-care professionals to detect anomalous circumstances such as wrong diagnoses, missing information, unobserved related diseases, or preventive actions	Needs specialized knowledge to operate
[16]	Framework for Comparison	This survey describes seven systems and three proposals for ontology-based data integration.	It is a survey
[17]	ontology based	bridges the gap between ontology based integration and service oriented architecture by enabling dynamic and transparent integration of information which is provided by services	The problem of splitting the query into static and dynamic query was not addressed fully.
[18]	Virtual Data Integration	They developed a Virtual – Data Integration Framework (V-DIF) that meets most of the users’ expectations	concentrate mainly on data integration process and avoid or ignore the other two processes (inconsistency detection and resolution)
[19]	Mapping Approach	Provides a linkage between the fundamental components required to provide accurate and unambiguous answers to the users’ queries from the integration system	Cannot use the sources of the data to resolve the duplicate through source preferences.
[20]	Ontology based	By using Internet of Things will help us to cure the patient in a short period of time	The paper didn’t integrate electronic health record (EHR) standards

[21]	Intelligent-Knowledge Authoring Tool (I-KAT)	Developed technologically integrated healthcare system	Increased complexity
[22]	case based reasoning	Medical dataset	Existence of many problems without solutions
[23]	KNN	The performance of CBR applications was enhanced	The missing data values of some attributes have been handled while others are not treated
[24]	Decision tree	Prediction of presence and absence of diabetes	It is a review of techniques and no model was developed
[25]	Support Vector Machine (SVM) and Artificial Neural Network (ANN)	Diagnosis heart disease	The accuracy is low when compared to other research
[26]	Fuzzy logic	Modeled clinical practice guidelines	Is not a dynamic systems

3. ANALYSIS OF THE SYSTEM

The system is designed to bring together the benefits of ontology-based data integration technique as it relates to seamless transition, solving inconsistencies in semantics and accuracy issues in syntactic and that of virtual data integration technique as it relates to hiding of technical jargons from users and unifying integrated and reconciling views of data residing at different sources as well as at different location for users. These would be implemented in a Business Intelligence environment using health sector, thereby enhancing the existing models that adopted exclusively either of the two techniques. The interoperable regional healthcare system lies in the maximum usage of medical and health resources by integrating important medical technologies and sharing the medical resource and information. Since the information of healthcare domain is diversified and dynamic because of the heterogeneous and distributed information resources and the large amount of daily updated data produced by many information systems from medical institution, the key problem for constructing integrated healthcare system is the efficient management of medical workflow and the effective integration of mass data resources. Therefore the modularization design and expansibility with the hierarchy structure are essential to this system. In the system, the healthcare centers are responsible for the medical data from each regional data center to monitor disease. Each regional information integration platform is composed of hospital, clinics and community healthcare center by which different levels of medical institutions are able to share the data from data centers and realize two-way referral and medical record lending. The information sharing among different medical systems not only needs to provide full accessibility to the data but also requires the interoperability among these systems. The main problems caused by bringing together heterogeneous and distributed computer systems are summarized as semantic heterogeneity and structural heterogeneity as shown in Figure 3.2. Semantic heterogeneity refers to the variation of semantic meaning in medical information resources which will lead to the semantic conflicts and complication for data integration. Structural heterogeneity means that the same data will be described in different structures by different systems because of various application systems, DBMS and operating systems.

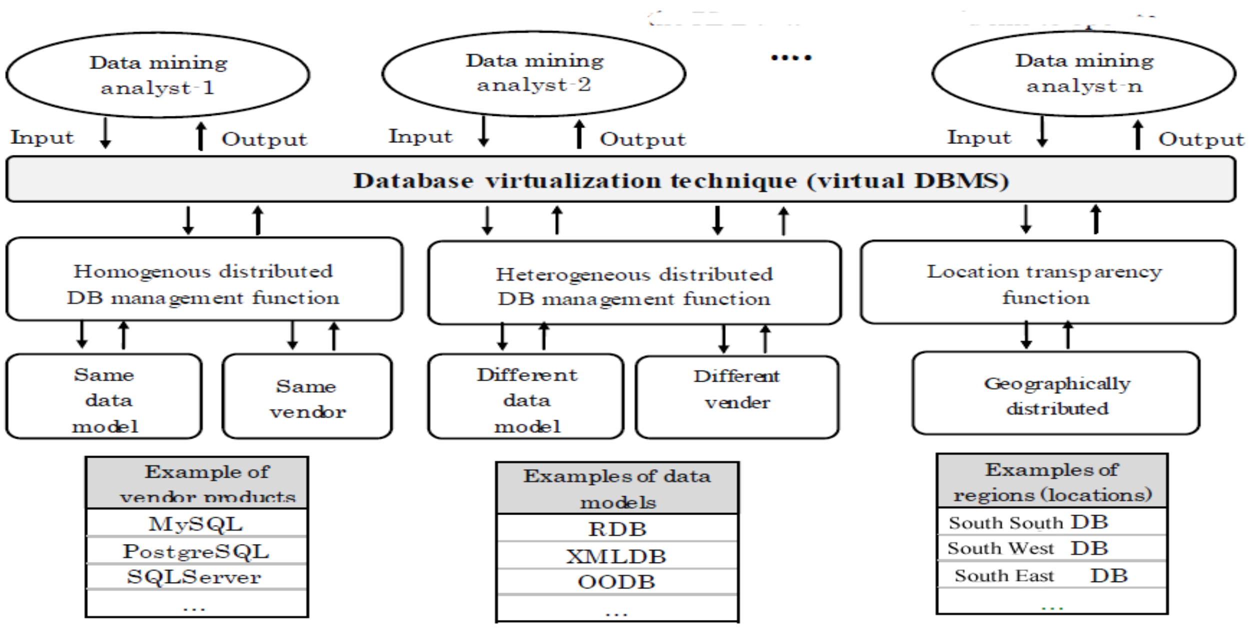

Fig. 3.2: Database virtualization technique

The system will track patient's visits and will generate a reminder for patient or physician of the needed follow-up or preventive care against a discovered disease. The typical work flow begins with the physician reviewing a patient's information from a computer terminal to view the information online. The physician can also retrieve patients list with similar disease. The richer the medical registry's data set, the greater the possibilities for examining subgroups of patients with related cases and how it was treated. The registries include standard reports and permit user to query the system for specific date ranges and interventions or patient status indicator.

In the system, physicians can use the patient's information to develop a strategy for treatment of each patient. It can allow the physician to contact the patient via SMS to advice the patient on the need to come for follow up or take certain medication. During this process, physician become aware of problems with registry information such as finding that the patient is no longer coming for treatment or has changed health service provider. The disease registry can generate report with different views of aggregating information about the process or outcome of health care management. It can show the population of patients suffering from a particular disease, the number treated successfully, and a feedback to physicians about the status of their patients, and the possible treatment to apply.

Data Flow Diagram (DFD) of the System

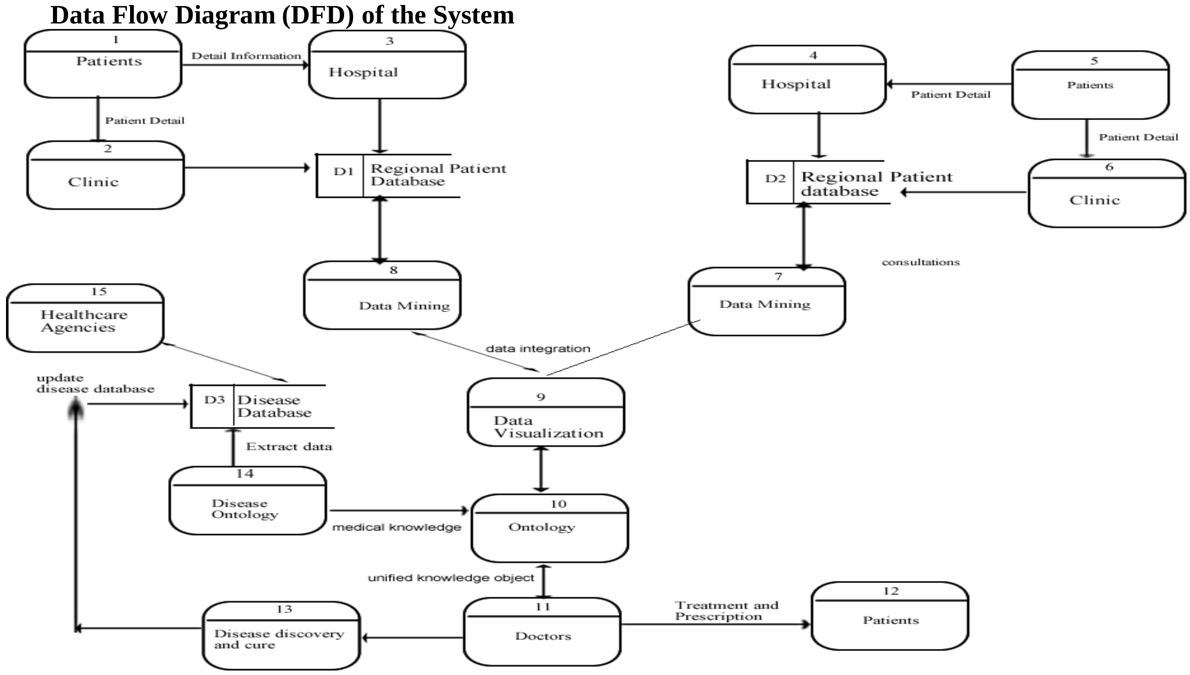

Fig. 3.4: Data Flow of the System

As shown in figure 3.4, the health sector is responsible for the medical data from each regional data center to monitor disease. Each regional information integration platform is composed of tertiary hospital, clinics and community healthcare center by which different levels of medical institutions in a region are able to share the data from data centers and realize two-way referral and medical record lending.

Algorithm

The algorithm for the proposed model is as follows
Start
Algorithm to set up the disease control dataset
Enter the disease symptoms
Enter the procedure for treatment
Create a dataset
Store in database
Stop

Algorithm for disease control
Start
Enter the symptoms
Query the disease db
Integrate all the dataset found
Match the symptoms entered with the one in the database
Use intelligent agent to filter the database
Apply disease ontology
Search for best matching case
Call Virtual data integration algorithm
Search the knowledge base
Is the disease a new case?
If yes then store in the disease dictionary else
Find the matching case
Is similar case found?
If yes search for the disease control procedure
Otherwise search other dataset from global view

Extract the suggested disease control procedure that matched the disease found from disease ontology
Display the suggested disease control procedure
Stop.

IV. RESULT

Table 2: **Confusion matrix applied to test dataset of Disease Control**

		Observed	
		True	**False**
	True	18	0
Predicted	**False**	1	1

Table 2 shows that out of 20 transactions, 18 diagnoses are True Positive and was predicted correctly. One diagnosis detected to be False Negative while it is not. Finally, one False Positive detected. A model of performance metrics derived from the confusion matrix as show in equation 3, which show the accuracy of the system.

Substituting the values, we have

AC = (18+1) / (18+0+1+1)

AC = 0.95 i.e. 95% accuracy in predicting the disease control procedure

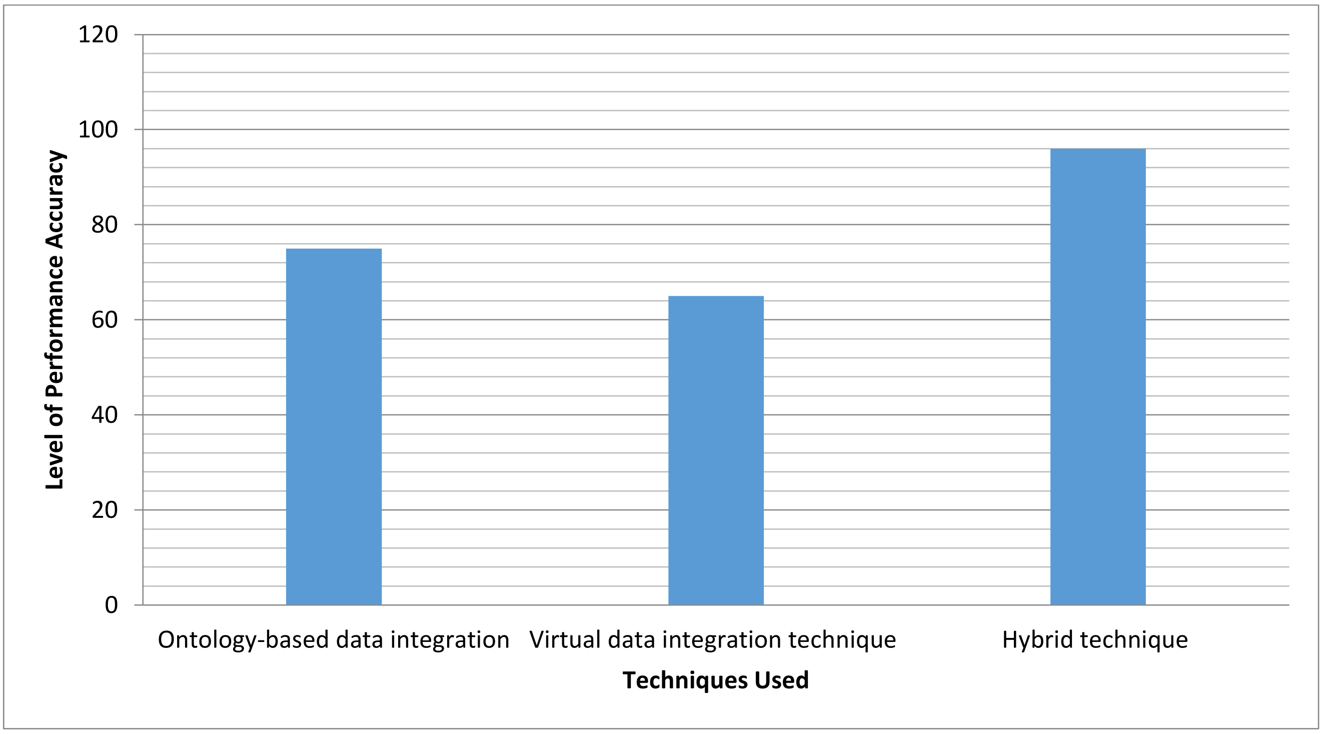

Figure 6: Comparison of level of prediction accuracy using various techniques

From Figure 4.17, one can see that the Ontology-based data integration technique for disease control procedure has 75% accuracy in predicting the disease control procedure; Virtual data integration technique for disease control procedure has 65% accuracy in predicting the disease control procedure; while Hybrid technique using both Ontology-based data integration and virtual bydata integration technique for disease control procedure has 95% accuracy in predicting the disease control procedure. This shows that the Hybrid technique outperforms the existing techniques with (95 – 75) = 20%, i.e. there is 20% improvement from the existing technique.

V. CONCLUSION

In developing nations like Nigeria, 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 thesis.

The establishment of local ontologies must represent the vocabulary used in the domain in order to recognize the synonyms relation and the hierarchical relations between the concepts. On this basis, the ontology matching becomes less time-consuming than the global schema matching as the method aims to reduce the amount of integration decisions and the number of rules.

Advances in intelligent systems, e.g., "Intelligent Information Agents" for the Internet, will help doctors in accurately carrying out disease control procedures. Emerging and more mature standards such as "Extensible Markup Language" (XML), "Ontology Web Language" (OWL) and Web Services based on "Simple Object Access Protocol" (SOAP), "Universal, Description, Discovery, and Integration" (UDDI) and "Web Service Description Language" (WSDL), will also help to resolve many software-level interoperability problems. The application developed in this thesis relies on these Web interoperability standards in order to integrate information dynamically.

REFERENCES

[1] Rosse, C. and Mejino, J.L.V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy**.** Journal Biomedical Information, **3(6), 478-500, 2013.**

[2] Helena, V. and Lidia, R. (2021) "Ontology-based Data Warehouse Development Process" Proceedings of the ITI 2009 31st International Conference. On Information Technology Interfaces, June 2225, 2009, Cavtat, Croatia, retrieved on June, **2021**

[3] 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.**

[4] Bostjan, G. & Vili, P. Automating ontology based information integration using service orientation" WSEAS Transactions on Computers Bostjan Grasic, Vili Podgorelec **9(6), 12-26, 2021.**

[5] Marut, B. (2016). Ontology-based Clinical Reminder System to Support Chronic Disease Healthcare. Article in IEICE Transactions on Information and Systems. DOI: 10.1587/transinf.E94.D.432 · Source: DBLP. March **2016**,

[6] Madhura, J., Dinithi, N., Daswin, S., Damminda, A., Brian, D., Kate, E .W. A data integration platform for patient-centered e-healthcare and clinical decision support. Research Center for Data Analytics and Cognition, La Trobe University, Victoria, Australia b School of Allied Health, La Trobe University. Victoria, Australia, **2020**.

[7] Chih-Lin, C. Medical decision support systems based on machine learning. PhD (Doctor of Philosophy) thesis, University of Iowa, https://doi.org/10.17077/etd.o5gmwvxk, **2019.**

[8] Sozan, S. M. Intelligent System for Identification Heart Diseases. A thesis submitted to the graduate school of applied sciences of near East University, **2019**.

[9] Serdar, U. Intelligent Systems in Patient Monitoring and Therapy Management. Knowledge Systems Laboratory, Stanford University, 701 Welch Road Bldg. C Palo Alto, CA 94304, USA, **2014**.

[10] Vishesh, V. Personal health record system and integration techniques with various electronic medical record systems. A Thesis Submitted to the Faculty of The College of Computer Science and Engineering in Partial Fulfillment of the requirements for the Degree of Master of Science Florida Atlantic University, Boca Raton, Florida, **2010**.

[11] Dipti, P. S., Santosh, K.P. An Expert System for Diagnosis of Human Diseases. International Journal of Computer Applications, **1(13), 23-34, 2010.**

[12] Soltan, R.A., Rashad, M. Z., El-Desouky, B. Diagnosis of Some Diseases in Medicine via computerized Experts System. International Journal of Computer Science & Information Technology, **5(5), 26-45, 2013.**

[13] Ighoyota, B. A. and Sujatha, P. Fuzzy Based Multi-Fever Symptom Classifier Diagnosis Model. International journal Information Technology and Computer Science, **6(10), 13-28, 2017.** Published Online October 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2017.10.02

[14] Alexander, P and Joseph, L. (2017). Multi Agent Enhanced Business Intelligence for Localized Automatic Pricing in Grocery Chains" A dissertation submitted in fulfillment of the requirements of the degree of Doctor of Philosophy for the School of Information Technology, Bond University. December **2017**.

[15] David, R., and Francis Real, J. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. Journal of Biomedical Informatics, **45(3), 429-446, 2020.**

[16] Augustina, B., Alejandra, C. and Nieves, R. B. (2014) "Ontology-Based Data Integration Methods: A Framework for Comparison", retrieved on July, **2016**.

[17] Bostjan, G. & Vili, P., (2010) "Automating ontology based information integration using service orientation" WSEAS Transactions on Computers Bostjan Grasic, Vili Podgorelec ISSN: 1109-2750 547, **Vol.9, Issue 6, June 2010**, retrieved on May, 2021

[18] Ali, Z., Ali, H., Osman, H. (2018) V-DIF: Virtual Data Integration Framework. International Journal of Computer Systems (ISSN: 2394-1065), **Volume 05– Issue 05, May, 2018**.

[19] Ali, Z. A Mapping Approach for Virtual Data Integration System Processes. International Journal of Advanced Computer Science and Applications, **9(12), 24-37, 2018.**

[20] Vinoth, K. Ontology based Public Healthcare System in Internet of Things. 2nd International Symposium on Big Data and Cloud Computing (ISBCC'19), **2019.**

[21] Taqdir, A., Maqbool, H., Wajahat. A., Muhammad, A., Jamil, H., Rahman, A., Waseem, H., Arif, J., Byeong, H., Sungyoung, L. Multi-model-based interactive authoring environment for creating shareable medical knowledge, *Department of Computer Science and Engineering, Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-si 446-701, Gyeonggi-do, Republic of Korea*, **2017**.

[22] Richter, M.M., Weber, R.O. Case-based reasoning. Springer-Verlag Berlin, **2016**.

[23] Jagannathan, R., Petrovic, S. Dealing with missing values in a clinical case-based reasoning system. In international conference on computer science and information technology (pp. 120-4). IEEE, **2009.**

[24] Asma, A., A. Decision Tree Discovery for Diagnosis of Type II Diabetes.", IEEE International Conference on Innovations in Information Technology (IIT), **pp.303–307, 2011.**

[25] Kapil, W., Mrudula , G. and Snehlata, D. "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network", IEEE International Conference on Computer and Communication Technology (ICCCT),**Vol.978,pp.741–745, 2010.**

[26] Nassim, D., Elpiniki, I. P., Jos, D., Hans, C. and Marie-Christine, J., Clinical Decision Support System based on Fuzzy Cognitive Maps, INSERM UMR**_S** 872, Eq 20, Medicine Faculty, Pierre and Marie Curie University, Paris 6, France, **2014**.

