An Integration Method of Digital Campus Resources based on Heterogeneous Data Integration

Manman Cao[1], Mian Wang[2]

1 Jining University
Qufu 273155, China
2 Institute of Scientific and Technical Information of Jining
Jining 272000, China

***Abstract* — In this paper we study how to effectively integrate digital campus resources through a heterogeneous data integration algorithm. A framework of heterogeneous data integration based on working process design is proposed. In the proposed framework, in order to integrate different data resources, we call on the corresponding local Web service query exploiting XML documents. Then XML based Heterogeneous data integration algorithm is designed by converting the different data resources integration problem into a multiple XML tree merging problem. Next, inspired by the proposed heterogeneous data integration algorithm, digital campus resource integration scheme is developed in three main layers of data layer, middle layer and application layer. We then consider seven sub-systems in digital campus construction. To do performance evaluation, experiments are carried out to testify the effectiveness of the proposed design. The results demonstrate that exploiting the proposed algorithm various information systems can be effectively integrated into the digital campus resource sharing platform.**

Keywords - Digital campus, Resource integration, Heterogeneous data integration, XML document

I. INTRODUCTION

The construction of digital campus infrastructure is a kind of integrated project, which can combine the network system, the computer system, and the storage system together [1]. This paper aims to propose a design of digital campus resource integration system. As is well known that the construction of digital campus infrastructure needs to give a high performance data center switch network to transmit data of digital campus applications [2]. In particular, the data transmission process should be completed in time. Moreover, the construction of digital campus infrastructure also should give a high performance calculating module by separating virtual computers and real computers [3][4]. The implement of all the digital campus applications should be realized effectively and all programs can be realized to cluster, backup or move. In general, the construction of digital campus infrastructure should provide a high-band and large scale data storage, data backup and data disaster recovery system as well.

Considering the stage characteristics of IT technique development, there are a large number of heterogeneous database in the network, and this causes inconvenience in data exchange and integration. The existing method of information integration no longer satisfies humans' requirements [5]. In recent years, more and more attentions have been paid on heterogeneous database integration, and isolated systems should be effectively connected together and information sharing level should be enhanced.

The technology of heterogeneous information integration aims to tackle the problem of convenient and flexible information accessing between heterogeneous database systems. In the process of developing heterogeneous information integration, some traditional integration technologies cannot meet our requirements [6]. Therefore, in this paper, we concentrate on the problem of how to fully utilize the heterogeneous data integration technology in digital campus resource integration.

The rest of this paper is organized as follows. The next section gives the related literature about heterogeneous data integration. Section 3 provides the problem of digital campus resource integration. In section 4, XML based Heterogeneous data integration algorithm is illustrated. Section 5 presents digital campus resource integration scheme. Afterwards, experiments are given in section 6, and the conclusions are illustrated in the section 7.

II. RELATED WORKS

In recent years, there are many studies about heterogeneous data integration, which has been widely applied to a variety of research areas [7]. Before proposing our design scheme, in this section, we introduce the related works about the applications of heterogeneous data integration technology.

Bale et al. developed a scheme to simulate the diffusion of energy innovations on a heterogeneous social network. Moreover, this paper describes the development of the model to incorporate heterogeneous parameters and, therefore, become more like a real social system [16].

Nguyen et al. we presented an alternative conception for local data integration, which is named Biological Integration and Retrieval Data, which defines 4 important terms: 1) a hybrid flat file and relational database architecture, 2) a generic data model, 3) configuration rules, and 4) a simple,

declarative query language. Particular, in this paper, generic design allows the integration of diverse data formats in a searchable database with high-level functionalities according to the specific scientific context[8].

Liu et al. presented a detailed survey on the state-of-the-art siRNAs design, concentrating on some important problems, such as 1) inconsistencies among the proposed guidelines for siRNAs design and the incomplete list of siRNAs features, 2) improper integration of the heterogeneous cross-platform siRNAs data, 3) inadequate consideration of the binding specificity of the target mRNAs and 4) reduction in the 'off-target' influence in the above design [9].

Florido et al. presented a novel framework to be used in Artificial Neural Networks for the problem of forecasting functional relationships between proteins via the integration of evidences from heterogeneous database. This paper chooses smaller representative/non-random subsets from the original dataset for Artificial Neural Networks optimization process [10].

Emery et al. presented a case study conducted at the Port Mayaca aquifer through porosity logs at two wells and high-resolution crosswell seismic measurements of P-wave impedance. The main innovation of this paper is that the model parameters are inferred and checked against the data. Moreover, multiple realizations of porosity can then be constructed conditionally to the impedance information in the interwell region [11].

Leida et al. presented a novel definition of a declarative mapping language, which is able to map precisely and unambiguously the semantics of a domain conceptualization into queries to a set of data sources. The mapping model is design based an ontology, and the data can be shared with other information systems [12].

Vincini et al. concentrated on how to implement data integration systems manage syntactic and semantic heterogeneities, which is of great importance. In particular, the authors designed a Data Transformation System which supports data transformation functions and produces query translation to push down to the sources the execution [13].

Lin et al. proposed an effective and efficient computational approach to integrate heterogeneous data sources based on advanced technologies. The main innovations of this paper are that the authors developed a Bayesian classifier for human PPI predictions from model organisms [14].

Smith et al. presented an open-source data warehouse system to facilitate the building of databases with complex data integration requirements and a requirement for a fast customizable query facility. Experimental results of this paper show that large biological databases can be obtained from a specific heterogeneous database, and the extensible data model permits to combine new data types [15].

III. Problem Statement

The digital campus resource information integrated service business platform uses the idea of integrated business workflow, and then designs the query and update operation of heterogeneous data sources. According to Fig.1, in our proposed heterogeneous data source workflow integration framework, different users login this platform with different identities, and then submit business processing requirements to information systems. Afterwards, data integration engine is used to construct business processes, and service analysis. Next, a global business process is defined in accordance with the table query data set.

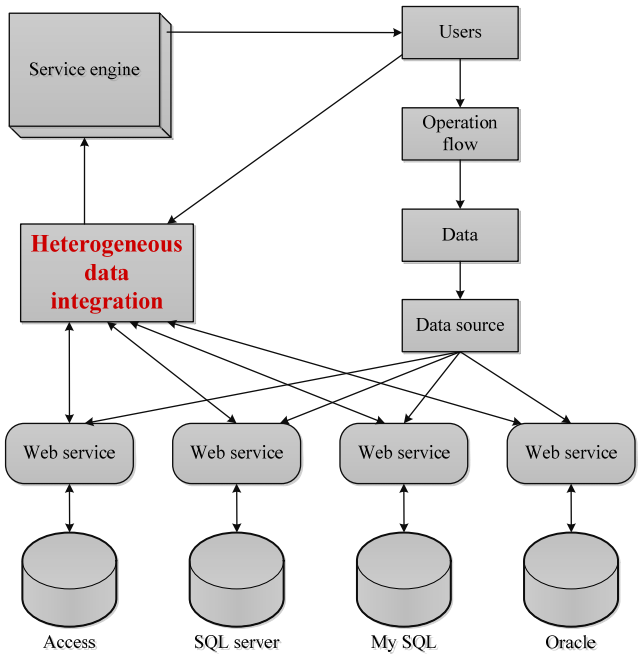

Figure 1. Framework of the heterogeneous data integration based working process design

To integrate different data resource, we call the corresponding local Web service query utilizing XML documents, and then the results are transmitted to the buffer service engine. In the end, the service integration program is implemented by mapping rules and field mapping table in terms of the matching rule. The main idea of this paper is that, the temporary data mode is converted by exploiting a XML document.

Based on the idea heterogeneous data source workflow integration design, the realization process of heterogeneous database integration becomes very simple. As is shown in Fig.1, the integrated digital campus resource integration platform can be built up based on the existing data integration model. In the next section, we will show how to design the heterogeneous data integration algorithm.

IV. XML Based Heterogeneous Data Integration Algorithm

Before illustrating the algorithm, two important definitions should be defined in advance.

(Definition 1) Supposing that $T=(V,E)$ refers to a directed data tree, in which V and E denote node set and edge set respectively. Moreover, $v \in V$ is allocated with a label, which is represented as $l(v)$.

In general, we define a sub-tree T^* is established by indicating node v which refers to the root node of T^*. In an XML data tree, we define a function π to allocate the node type as follows:

1) Θ_1: membership degree node
2) Θ_2: possibility node
3) Θ_3: deterministic XML node

Function γ is designed to allocate membership degrees to nodes of the XML data tree, and root of the XML tree is defined as a possibility node.

(Definition 2) We define an XML data tree by a quadruple $\Omega=\{V,E,\pi,\gamma\}$, where V,E denote node set and edge set. Function π is used to map a node v to the set $\{\Theta_1,\Theta_2,\Theta_3\}$. Furthermore, function γ is designed to map a node to local membership degree.

Based on the above two definitions, XML based Heterogeneous data integration algorithm is given as follows.

Algorithm 1: XML based Heterogeneous data integration algorithm

Input: XML tree set $\{T_1,T_2,\cdots,T_n\}$, query set Q

Output: New XML tree ($\tilde{T}$) obtained from the heterogeneous data Integration process

(1) Set $\tilde{T}=\varnothing$

(2) For each tree T_i in $\{T_1,T_2,\cdots,T_n\}$

(3) Conduct query q ($q \in Q$)

(4) If T_i is a new tree which has not been integrated

(5) Integrate T_i to the integrated tree $\tilde{T}$ by the XML tree linking method

(6) End If

(7)Return $\tilde{T}$

(8)End For

V. DIGITAL CAMPUS RESOURCE INTEGRATION SCHEME

In this paper, we aim to develop a digital campus resource integration system which is utilized in comprehensive colleges and universities. Considering the restriction by the different historical periods and the planning and construction of the long-term idea, there are many various functional departments and several separate information systems, such as 1) Personnel management system, 2) Educational administration management system, 3) Scientific research management system, 4) Library management system, 5) Financial management system 6) Logistics management system, and 7) Student management system. However, the existing information systems cannot operate with each other, and the work efficiency is very low. It is urgent to realize the data sharing and integration of the digital campus information integration platform based on the existing heterogeneous application systems. Overview of the digital campus resource integration system is shown in Fig.2.

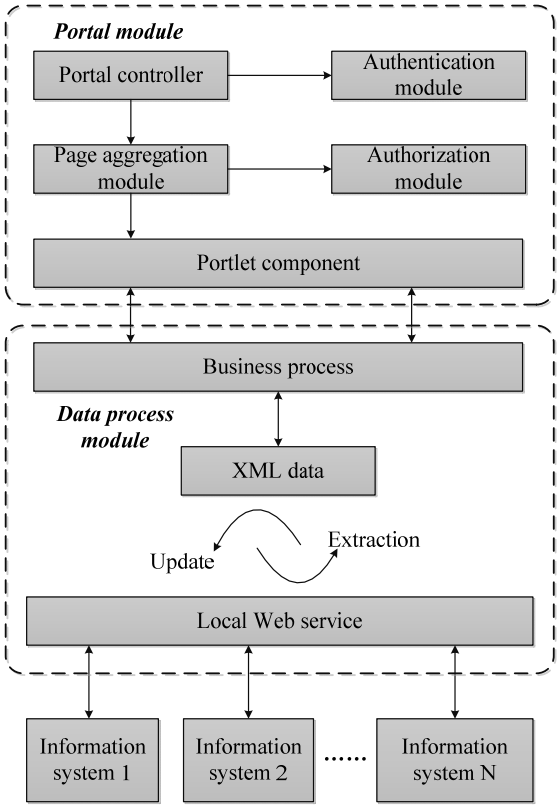

Figure 2. Overview of the digital campus resource integration system

Based on the above analysis, digital campus information integration platform should share with each other, and the goal is to achieve data sharing in business process integration. As is illustrated in Fig. 3, architecture of the digital campus information sharing platform is developed by integrating the business processes of each application system, and establishing the sharing and exchange model based on the business process of system integration. Particularly, as is shown in Fig.3, our proposed digital campus information sharing platform is composed of portal module, Portlet module, data processing module, and service driven engine.

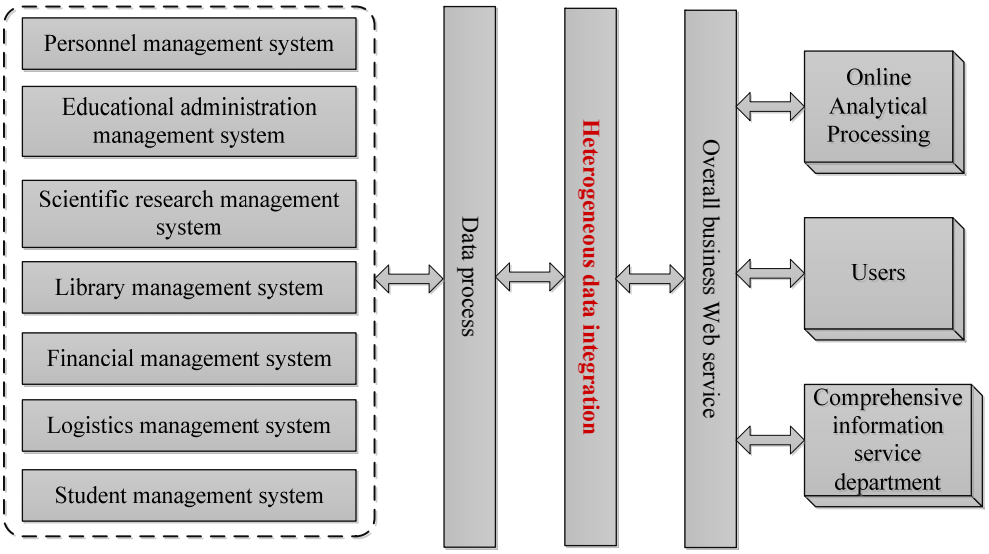

Figure 3. Architecture of the digital campus information sharing platform

The data processing module is composed of XML document template library, rule matching library, and business process integration component. Using portal components, according to user identity, the Portlet component is used to form the integrated platform in the implementation of business processes. Moreover, to extract or update data from a heterogeneous application system, data processing components are responsible to build up a data set using and XML documents. Next, the model of digital campus resource integration system is described in Fig.4. In our proposed digital campus resource integration system, seven sub-systems are combined together, and seven databases are constructed. After extracting data from each database, XML documents are provided to implement the business activity synthesis using Business activity synthesis.

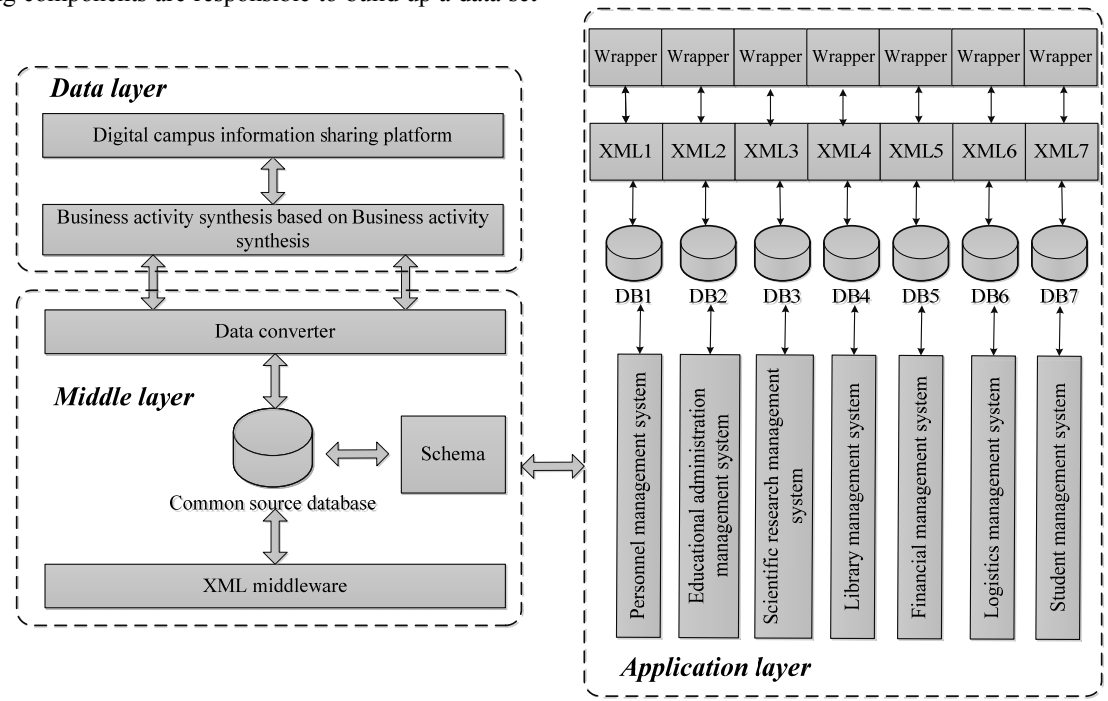

Figure 4. Model of the digital campus resource integration system

Fig. 4 shows that there three various layers in our proposed model. The middle layer is important to integrate the heterogeneous data. In particular, XML document is able to integrate heterogeneous data sources. All information of the digital campus system is stored in common source database by XML format, and we construct

a map between heterogeneous data and common source database.

Through constructing mapping relationship, the data model can be represented by XML Schema. In the application layer, we can implement the data integration process. Moreover, user's operation to the middle layer can be diverted to the data layer, afterwards, information can be transmitted from the application layer to the data layer.

VI. Performance Evaluation

We investigate the effectiveness of our proposed heterogeneous data integration method by trying to integrate the data in the above seven sub-systems (shown in Fig. 4). Three tasks are designed in this experiment: 1) querying student information (T1), 2) querying library information (T2) and 3) querying financial information (T3). Next, we test the performance of our algorithm by two experiments.

In experiment 1, we test the query precision and recall with different tasks.

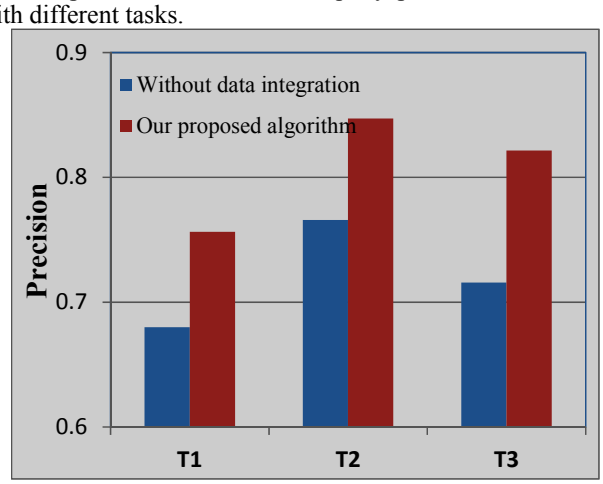

Figure 5. Precision for different tasks.

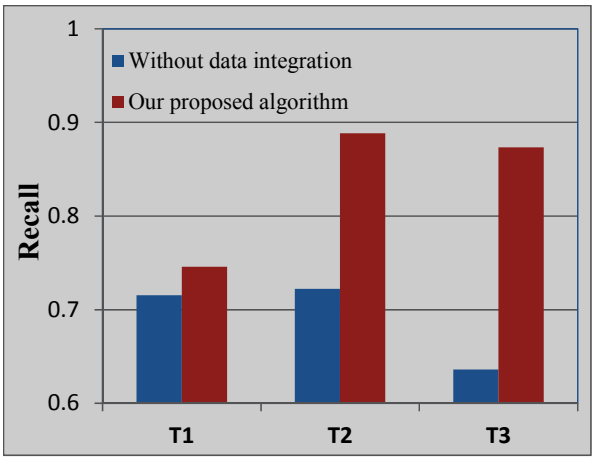

Figure 6. Recall for different tasks.

In experiment, we show the performance of our algorithm under different percentage of duplicates in the digital campus information system.

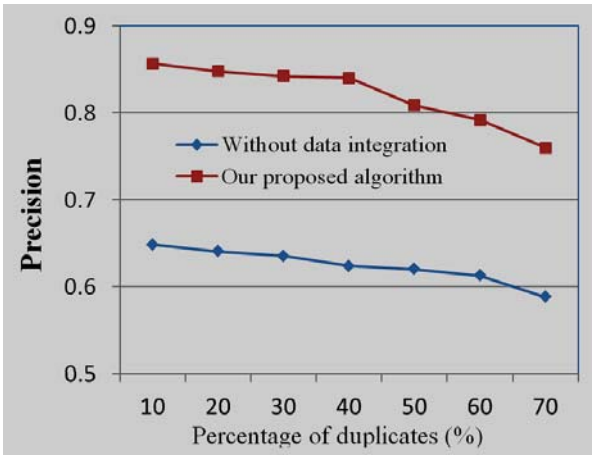

Figure 7. Precision for different percentage of duplicates

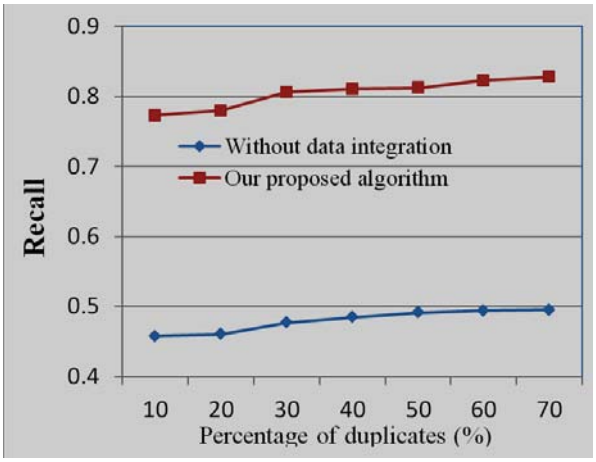

Figure 8. Precision for different percentage of duplicates

Combining the above experimental results, we can see that using our proposed algorithm the precision and recall of digital campus information query are significantly promoted. Hence, in can be concluded that our proposed algorithm can effectively the digital campus resource integration problem.

VII. Conclusion

This paper aims to effectively integrate digital campus resource through heterogeneous data integration algorithm. In order to integrate different data resource, we call the corresponding local Web service query exploiting XML documents. Then, XML based Heterogeneous data integration algorithm is designed by converting the different data resource integration problem into a multiple XML tree merging problem. Furthermore, digital campus resource integration scheme is develop by three main layers: 1) data layer, 2) middle layer and 3) application layer. Particularly, seven sub-systems in digital campus construction are adopted our design.

ACKNOWLEDGMENT

This work is supported by the Project of Science and Technology Development Plan of Shandong At 2014(No. 2014GGX109002), the Project of University-Industry Cooperation of Jining At 2014(No. 2014CXY021), and Jining University Science Foundation for Youths at 2014(2014QNKJ10).

REFERENCES

[1] Ma Nan, Yuan Mei, Cao Guoliang, "Integration of digital campus resources based on cloud computing", Lecture Notes in Electrical Engineering, vol. 163, pp. 1957-1963, 2013.

[2] Bessant, Judith, "Digital Spring New Media And New Politics on The Campus", Discourse-studies in the Cultural Politics of Education, vol. 35, No. 2, pp. 249-265, 2014.

[3] Cheng Liang, Zhang Wen, Wang Jiechen, Li Manchun, Zhong Lishan, "Small Core, Big Network: A Comprehensive Approach to GIS Teaching Practice Based on Digital Three- Dimensional Campus Reconstruction", Journal of Geography in Higher Education, vol. 38, No. 1, pp. 119-135, 2014.

[4] Chang Tsung-Han, Hsu Shu-Chen, Wang Tien-Chin, "A Proposed Model for Measuring The Aggregative Risk Degree of Implementing an RFID Digital Campus System with The Consistent Fuzzy Preference Relations", Applied Mathematical Modelling, vol. 37, No. 5, pp. 2605-2622.

[5] Chang Chan-Chine, Chen Ruey-Shun, "Using Data Mining Technology to Solve Classification Problems - A Case Study of Campus Digital Library", Electronic Library, vol. 24, No. 3, pp. 307-321, 2006.

[6] van Schaik P, Barker P, Moukadem I, "Using A Digital Campus to Support Electronic Learning in Lebanon", Innovations in Education and Teaching International, vol. 42, No. 2, pp. 157-166, 2005.

[7] Maver A, Peterlin B, "Searching for Multiple Sclerosis Genomic Candidate Regions by Genome-Wide Integration of Heterogeneous Genomic Data Sources", European Journal of Neurology, vol. 19, pp. 736-736, 2012.

[8] Nguyen H, Michel L, Thompson J. D, Poch O, "Heterogeneous Biological Data Integration with Declarative Query Language", IBM Journal of Research and Development, vol. 58, No. 2-3, 2014, pp.45-53.

[9] Liu Qi, Zhou Han, Zhu Ruixin, Xu Ying, Cao Zhiwei, "Reconsideration of In Silico SiRNA Design from a Perspective of Heterogeneous Data Integration: Problems and Solutions", Briefings in Bioinformatics, vol. 15, No. 2, pp. 292-305, 2014.

[10] Florido J. P, Pomares H, Rojas I, Guillen A, Ortuno F. M, Urquiza J. M, An Effective, "Practical and Low Computational Cost Framework for The Integration of Heterogeneous Data to Predict Functional Associations Between Proteins by Means of Artificial Neural Networks", Neurocomputing, vol. 121, No. 64, pp. 78-85, 2013.

[11] Emery Xavier, Parra Jorge, "Integration of Crosswell Seismic Data for Simulating Porosity in A Heterogeneous Carbonate Aquifer", Journal of Applied Geophysics, vol. 98, pp. 254-264, 2013.

[12] Leida Marcello, Gusmini Alex, Davies John, "Semantics-Aware Data Integration for Heterogeneous Data Sources", Journal of Ambient Intelligence and Humanized Computing, vol. 4, No. 4, pp. 471-491, 2013.

[13] Vincini Maurizio, Beneventano Domenico, Bergamaschi Sonia, "Semantic Integration of Heterogeneous Data Sources in The MOMIS Data Transformation System", Journal of Universal Computer Science, vol. 19, No. 13, pp. 1986-2012, 2013.

[14] Lin Xiaotong, Chen Xue-wen, "Heterogeneous Data Integration by Tree-Augmented Naive Bayes for Protein-Protein Interactions Prediction", Proteomics, vol. 13, No. 2, pp. 261-268, 2013.

[15] Smith Richard N, Aleksic Jelena, Butano Daniela, Carr Adrian, Contrino Sergio, Hu Fengyuan, Lyne Mike, Lyne Rachel, Kalderimis Alex, Rutherford Kim, Stepan Radek, Sullivan Julie, Wakeling Matthew, Watkins Xavier, Micklem Gos, "InterMine: A Flexible Data Warehouse System for The Integration and Analysis of Heterogeneous Biological Data", Bioinformatics, vol. 28, No. 23, pp. 3163-3165, 2012.

[16] Bale Catherine S. E, Mccullen Nicholas J, Foxon Timothy J, Rucklidge Alastair M, Gale William F, "Modeling Diffusion of Energy Innovations on A Heterogeneous Social Network and Approaches to Integration of Real-World Data", Complexity, vol. 19, No. 6, pp. 83-94, 2014.