A Framework for Personalized Information Integration in Higher Education Institutes

Bahram Amini
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, Malaysia

Roliana Ibrahim
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, Malaysia

Mohd Shahizan Othman
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, Malaysia

ABSTRACT

Information integration plays an important role in academic settings since it provides a comprehensive view of educations and enables managers to analyze and evaluate the effectiveness of education processes. However, the diversity and huge amount of information available in different academic sources overwhelms the education administrators and debilitates decision makings. This paper proposes a service-oriented framework which augments recommendation approach with components of semantic-based information integration and provides interactive and contextual-based information integration for decision makers in Higher Education Institutes. The underlying semantic web technology facilitates on-demand integration of information from internal sources as well as the Web and provides web service discovery and invocation for effective information analysis. In addition, the framework enables the users to analyze instances of student's information and to receive recommendation of new information sources as well as appropriate analytical services based on the students' status. Service orientation paradigm provides dynamic and flexible means of communication for service interoperability among the framework components.

General Terms

Framework, Personalization, Decision Making

Keywords

Personalization, Semantic Web, Information Integration, Service-Oriented Architecture, Higher Education Institute

1. INTRODUCTION

Data integration has emerged as a natural requirement of organizations from the early life of software systems. With the advent of new information technologies such as Internet and intranet, the need for information integration from a number of independent and heterogeneous sources is rapidly growing. Although organizations such as Higher Education Institutes (HEI) began to deploy large and complex information systems to support their educational activities, the need for comprehensive information which provides the global view of education process based on different occasion is still required. There is no need to say that HEI often needs to combine various types of educational data such as student records, schedules, scholarship, and alumni information altogether, preferably in a flexible, scalable and personalized manner, in order to make a global insight into current state of individual education processes including students. Therefore, information integration plays an important role in the educational environment since it provides a comprehensive view of education data and enables the managers to analyze and evaluate the performance of education processes and make effective decisions.

However, in spite of many research on information integration in the recent years, the need for flexible, scalable, and personalized information integration encourages scholars [1]. This paper proposes a framework which combines suitable components of semantic based technologies. It provides interactive and flexible information integration for decision making in higher education institutes such as universities and colleges. This framework aims to integrate information sources from internal education database and combines them with relevant information from the Web. In addition, the system enables users to analyze and evaluate instances of information with various analytical tools in a personalized fashion and to incorporate new sources of information on demand.

In the following sections, the aspect of higher education institutes is described in Section 2 and challenges and motivations in Section 3, respectively. Personalization and technology supports in terms of information integration are described in Section 4, and the proposed framework to personalized information integration in Section 5. A working example is presented in Section 6 and finally, the summarization of the framework and contribution is discussed in Section 7.

2. ASPECTS of HIGHER EDUCATION INSTITUES

In general, HEI from the information system perspective are academic organizations combining legacy and modern systems of different technologies. Basically, there is a need to have a unified access to information systems in order to monitor and analysis important aspects of the education processes. This situation

requires the application of distributed computing and information analysis solutions. Moreover, higher education institutions need to improve the education and professional practice through information analysis and visualization. Information integration and visualization are extremely accepted by diverse communities including academic and educational institutes due to its potential ability to add values to the available information [2]. It is therefore crucial for HEIs to analyze and explore hidden relationships between different educational objects such as students, courses, scholarship, publications, and relevant information from the Web to ensure the effectiveness of the education processes.

For example, information integration and analysis allows the academic administrations to take action for course scheduling based on the students' requirements, to identify those who may likely to attend in alumni programs, to identify how students learn better, what subjects are often taken together by students, which lecturers are mostly successful in specific course, and compare student progresses with standard education metrics taken from the Web [3]. Besides, higher education sectors are responsible to communicate with external parties including research institutes to exchange and share academic information such as grants and lectures, etc.

3. CHALLENGES and MOTIVATIONS

Bringing together information from different sources that are in different platforms such as hardware, software, and technology burden many problems and challenges [4]. Although there are many tools and technologies focused on information integration, new problems still exist. Laura et al, [5] addressed several challenges of enterprise information integration (EII) as follows: Firstly, in dynamic environments such as educational environment, a challenge is to provide on-the-fly data integration when new data are required. For example, a particular information analysis might require additional statistical data from other institute's database to enrich analysis of current student information. Thus, a run-time demand to incorporate new sources of data is required. Secondly, one emerging research challenge is to incorporate several analytic tools together to perform a personalized information analysis [6].

These capabilities would enable organizations to make use of the wealth of information whether it is typical information (such as students' records) or supplementary information (such as grants and benchmark data) on the Web. Thus, the challenges include how to supply analytics tools as flexible as possible in a standard communication framework and how to integrate the results of queries on local information (from the university database) with relevant information from the outside of the education institute to gain comprehensive view of education performance [5].

Furthermore, the personalization and information integration have been motivated by the demand to integrate the diverse collection of Web information with organization's information to gain better insight to the academic performance and improve decision making

by exploiting personalized information analysis tools. Recent trends such as the continual improvement of Web-based applications, Web service integration, and personalization to Web applications present more interesting topic of research in the field of information integration [7].

4. TECHNOLOGY SUPPORT

Fortunately, Semantic Web Technologies (SWT) and interoperability paradigms such as Service Oriented Architecture (SOA) truly support the aforementioned requirements for personalized information integration in a distributed environment such as the Internet [8]. Today, Semantic Web and Web services (WS) are the basic building blocks of information sharing and information integration [9]. A Web service is a platform-independent, loosely coupled program that can be executed over the network such as Internet and is accessible through an interface which specifies the physical address as well as the messages that a client can consume it. The actual consumption of the Web service is realized by exchanging of XML data over the network via SOAP protocol which is a service access protocol for which invokes a Web Service. Transport protocols over the internet such as HTTP, FTP, SMTP, and other transport protocols may be used with SOAP with different purposes. Alonso, et al. [10] argued that SOA allows developers to overcome many challenges in distributed computing including enterprise information integration.

Web services technologies and SOA interconnectivity paradigm has been adapted by several standardization and technologies including: 1) Semantic-based Web Service Description Language (WSDL-S) which specifies adequate information on how clients can access to the web services, 2) SOAP, a messaging technology for exchanging XML data over the Internet, and 3) UDDI (Universal Description, Discovery and Integration Protocol) which is a registry technology for registering and publishing Web services over the Internet.

Semantic Web Service technologies (SWS) provide appropriate techniques for automatic detection and consumption of Web services over the network for solving a particular client request within the SOA architecture. To achieve this goal, ontology is employed as the underlying data model in SWS approaches to specify formal knowledge for the Semantic Web Services. In fact, ontologies provide a pure data model for the domain of discourse (domain problem) by enriching information with description of data items and bridge the gap between real world and the information source [11]. In essence, they support integration of heterogeneous data sources by defining mappings between ontologies and the date model of information sources.

Moreover, Semantic Web provides the following supports to the integration problem: 1) Formal ontology languages for describing information sources [12], 2) Ontology management technologies for ontology engineering task such as development and maintenance [12], 3) Techniques for supporting ontology-based
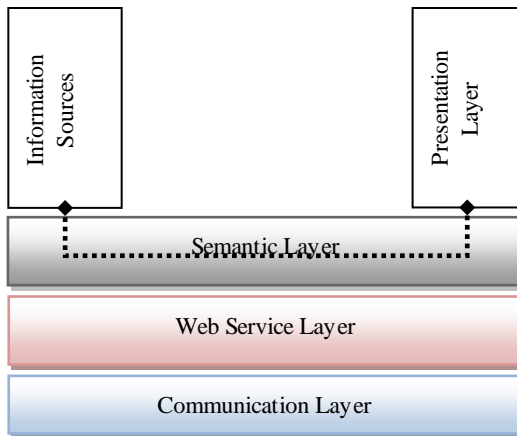
Fig. 1: Conceptual structure of the framework

data integration [13], and 4) Processing capability of Semantic Web technologies such as Web services [14].

In addition, personalization technologies have been extensively employed in different domains such as e-learning to tailor more interesting objects to individual user's preferences [15]. User profile maintains user's preferences or items the user already rated. Personalization approaches then recommend those unseen objects that are most similar with the items in the user profile. User profiles are usually made up based on the user score of items rated during the user interaction with the system.

5. INFORMATION INTEGRATION FRAMEWORK

In the proposed framework, components interact by means of service invocation and standard message passing [16]. Each source data is wrapped by an appropriate service and services are executed in a dynamic or on-demand fashion. The architecture of the framework which is based on SOA paradigm encompasses five different layers: information source, presentation, semantic, Web service, and communication layers. A layering principle has been followed here because it appropriately helps to support strong separation of concerns, abstraction, and loose coupling that, in turn, supports flexibility and scalability [17]. It elaborated by Web Service technologies which allows interoperability among framework components by relying on semantic standards.

As shown in Figure 1, both Web services and semantic layers play the key role in the architecture because they provide an interface between information sources and presentation layer. In fact, presentation layer (user interface) access the information sources through three inter-related layers including semantic, services, and communication layers. As mentioned in previous sections, the aim is to address a flexible and scalable personalized framework for information integration. Layered structure provides sufficient flexibility, scalability, and modularity as well as scalability for the framework [17].

Moreover, layered structure consolidates the dynamic interoperability between components because components can be easily replaced and upgraded without affecting other layers. Besides, this architecture enables and encourages decision makers to interact with the system by incorporating their own analysis tools, including legacy and sophisticated tools via presentation layer. The following sections discuss the functionality of those layers in detail.

5.1. Information Sources

In most HEIs, there are many pertinent information sources from which to make decisions including internal sources such as academic records, external data source such as e-journals, and the Web information such as research portals. The flexible and efficient information integration system must enable knowledge workers to dynamically bring in and out information sources from these sources in order to make effective decision. The dynamic nature of the Web implies that some information sources are relatively temporary and therefore cannot be involved permanently into the system. Thus, they should be replaced with the new ones on-occasion by support of semantic web service approaches which wraps information source and helps resolving heterogeneities. Besides, new information may be required based on the individual student records to fulfil specific analysis.

A global ontology describes the general keywords of HEI domain including structured information such as courses, events name, graduation steps, progress report, and so on. Each information source is described by a local ontology, fine-grained concepts and their semantic relationships. A mapping between the global ontology and the local ontology has been established to resolve concept heterogeneities among different sources [18]. An ontology engineering approach, as the basis of a service, can be used to keep track of changes in local ontologies due to source replacement and concept renewal. As an illustration, consider an enquiry on student information who will graduate in the nearby future. User can monitor the student progress and include living information to see whether living condition affects on student progress. Another example is that if a student is going to graduate, relevant job opportunities from the Web sources such as portals will be recommended to the user. Since this form of information is unstructured, information retrieval techniques can be used to extract relevant information. The global ontology helps to overcome structure inconsistency that might exist among local structured data and retrieved data from the Web. Finally, new terms of the analysis affects the "context ontology" and the recommendation source recommend relevant information sources based on the new terms such as "scholarship", "finance", and "alumni" information.
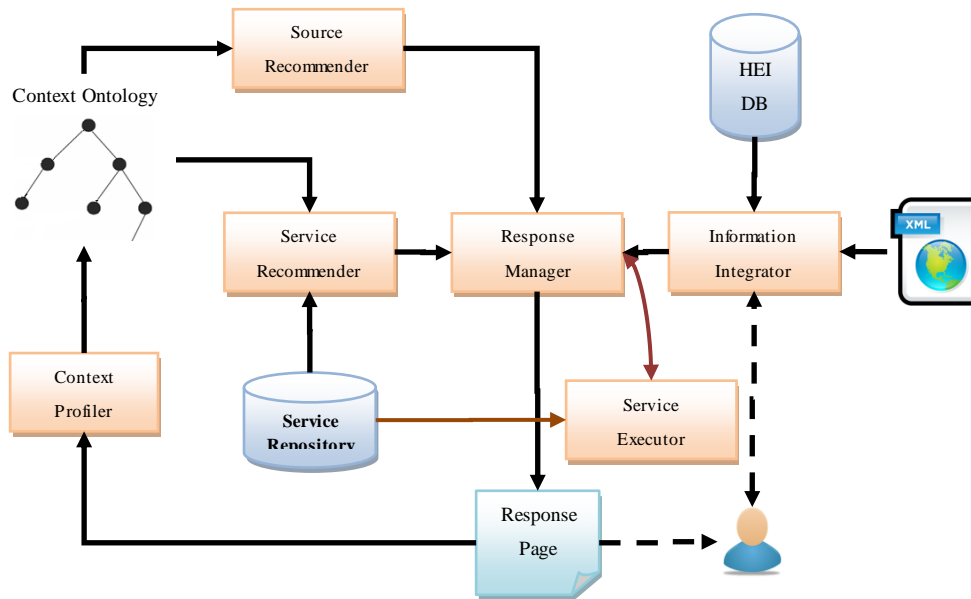
5.2. Semantic Layer

Fig. 2: Basic Components of the Web Service Layer in the framework

The Semantic layer plays three important roles in the proposed framework: 1) enriches information sources to overcome information heterogeneities, 2) facilitates discovery, composition, and execution of services, and 3) keeps updated the "context ontology" when users issues a query.

Diversity through information sources imposes many heterogeneities problems such as the lack of common semantic among data items. Semantic Web technologies such as ontology provide a global conceptualization (common terms) of information sources. Since information sources are not mandatory and replaced during the analysis, an ontology engineering techniques are employed which enable users to make local source ontology at run-time.

To facilitate this process, a hybrid approach [18] for domain conceptualization is employed. In the hybrid approach, the semantics of each source is described by a particular ontology.

Furthermore, in order to make the sources ontologies compatible and comparable to each other, they can be constructed based on a shared vocabulary. The shared ontology encompasses basic terms of the domain. Since each term of the local ontology is based on the basic terms, the terms of involving ontologies can be easily mapped to each other [19]. In fact, the domain ontology is a partial refinement of the terms of global ontology. Since domain ontologies only use terms of the global ontology, they remain compatible and easily comparable [20]. The advantage of using hybrid approach in the framework is that it allows new information sources easily replaced without the need of changing the mappings between data source and the concepts of shared ontology.

However, the drawback of hybrid approach is that existing ontology cannot be easily reused in other domains and also required developing from scratch, because all ontology terms refer to the shared ontology's terms. Thus, some parts of ontology mapping, i.e. mapping between concepts of the information sources and the global ontology, may require user modification.

In addition to this role, the Semantic layer facilitates automated Web service discovery, composition, and execution of Semantic Web services [13]. All functionalities in this layer are ontology development and maintenance. Information query and analysis are supported by the Semantic layer. This layer provides required flexibility to deal with dynamic aspects of the framework in which information source can be easily added or replaced, for example, regarding the students' status. It provides facilities for better support of detection and consumption of Web services in the framework. Therefore, semantic web service uses ontologies as the underlying data model to specify the knowledge model of the Web services. All services in the framework can be described by WSDL-S language which augments semantic tags to WSDL data types as well as the messaging type and operations. WSDL-S also provides efficient discovery and composition of services, since it augments semantic features and functionality with service specification, allowing for better semantic search, efficient service mediation and automatic service composition [9].

The semantic layer also maintains the context ontology entity to provide final version of context vocabulary that user might need to know more about them. The context ontology supports personalized query and service invocation by enriching query terms and service discovery.

5.3. Web Service Layer
In the framework, the foundation of communication model is SOA and the underlying processing components are Semantic Web Services. Therefore, it supports better service interoperability and facilitates easier service integration and better automation of service discovery, composition, and execution [13].

As illustrated in Figure 2, this layer includes fundamental services for supporting information integration tasks, ontology management, and basic run-time services for service registration, discovery, automatic composition and execution. It also

5.4. Communication Layer
Communication layer provides a medium that allows Web services exchange messages by means of SOAP protocol. We separated the Web service layer from the communication layer in order to emphasis on flexibility in the framework. The basic elements of Web service technology such as WSDL-S, SOAP, and UDDI enables the provision of Web services usage (service discovery and invocation) over the Internet [20]. The components of the framework which are a collection of Web services by means of these technologies can be efficiently and easily run over the distributed platform such as the Internet. The proposed framework relies on the Internet as a stable communication platform because it enhances the traditional communication methods for information integration. Moreover, depending on the integration situations, the Web services could be originated from one source or multiple sources interconnected over the

encompasses source and service recommenders [21], context profiler, and information integrator. This layer allows diverse kind of system component including legacy tools for information analysis, wrapped to form a semantic web service and, therefore, can be incorporated in the architecture at run time. These services exploit semantic description provided by the semantic layer in order to provide transparent and dynamic service discovery and invocation in the framework.

An important component of automated service composition and execution is the discovery of the required services [9]. The run-time service management in the framework deals with the service discovery, composition, and execution. The functionality and the semantic specifications of all Web services in the service repositories can be formally described by WSDL-S language which supports semantic search techniques for more precise Web service discovery. Semantic matchmaking of functional descriptions helps the discovery of the available Web services and detection of suitable candidates for an analysis or a particular presentation service. The usability of the discovered services can be determined by selection and ranking them either to select one of the candidates or recommend a priority list for user's selection. If a single service satisfies the user request, then the executor calls the respected Web service in order to accomplish the user's request. Otherwise, the composer combines a series of suitable Web services for solving the user request.

Context Profiler (CP) in the framework processes "response page" and employs information retrieval approaches such as Natural Languages Processing (NLP) and statistical analysis to determine important vocabulary of information that user is analyzing and updates context ontology. In fact, CP identifies those important terms that best describe the student status and helps the two recommenders in the framework, i.e. source and service recommender, to propose most relevant source of information and service (for analysis) to the user. In other words, CP assists knowledge worker to analyze the retrieved student information efficiently in terms of accuracy and spending time.

communication network. Thus, the Internet plays two important roles in our framework; it provides Web services for users as well as interconnects Web service components for exchanging message and communication.

5.5. Presentation Layer
This layer deals with data visualization which engages a wide variety of tools for presentation the information contained in the information sources. Efficient visualizations can make complex relationships among piece of information easily understandable [22]. This layer helps all aspect of education information including trend analysis, performance of teaching and supervision, and other academic information easily analyzed in a service-based manner. All of these analyzes can be accomplished by means of third party tools which are wrapped as a Web Service. This layer is simply a Web browser which enables users to pose queries and navigate

through the information and access the Web service repository through the links offered by the recommenders in the web service layer. Users also are enabled to invoke appropriate web services semi-automatically to accomplish specific analysis through semantic search of simple keyword in UDDI and WSDL-S description.

6. WORKING EXAMPLE

The proposed framework provides sophisticated search capabilities over the educational data as well as the Web information for decision making and information analysis. User simply poses keyword-based queries to Information Integrator (II) interface. Response Manager receives the result of the query from II and invokes both Source Manager and Service Manager for their recommendation and combines the recommendation altogether as a single Response Page (RP), observed by the user. After analyzing the RP, user may either recall additional services or request additional information source for further analysis. At run-time, Context Profiler (CP) exploits RP to extract relevant new vocabulary which best describes the status of students information and updates the context ontology.

For instance, if some students failed in specific course, CP may extract new terms such as "prerequisite course", "complementary course", and "grading system" from the current RP. Source Recommender and Service Recommender components use context ontology and extra knowledge to improve the recommendation process. In the next turn of analysis, user may retrieve more personalized information due to updated context ontology. User also may search Service Repository for appropriate services manually which can be applied to the current retrieved information. For example, he may wish to draw a graph of student grades in a course and compare them with grades of the past semesters. It can be accomplished by searching a tool for drawing the graph and run it as a web service.

7. CONCLUSION and CONTRIBUTION

This paper proposes a service-oriented framework which provides personalized information integration for higher education institutes. Also, the principles of semantic Web technologies such as ontology, Web Service, and service oriented architecture in the framework have been described. These technologies address the dynamic combination of simultaneous information integration namely on-demand integration, as well as personalized and dynamic information analysis. Technical aspects of the framework such as dynamic information source and personalized service composition were examined by realization of run-time hybrid ontology development and contextual ontology. To overcome the

traditional challenges of information integration, the usage of a flexible and open interoperation environment such as Internet is proposed. Working example clearly shows that the framework can assist HEI managers to analyze a wide variety of relevant information and perform sophisticated analysis based on the student's status. This research framework is also different from traditional recommender systems because it recommends objects (information sources and web services) based on the situation of student information but not based on the user preferences. Personalization therefore assists the academic managers to analysis only those aspects of information which affects the actual state of the educational process by proposing relevant information and services.

8. REFERENCES

[1] Jingtao Z., Mingwei W., Han Z. 2006. Enterprise Information Integration: State of the Art and Technical Challenges. In International Federation for Information Processing (IFIP), Vol. 207, Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing, and Management, eds. K. Wang, Kovacs G., Wozny M., Fang M., pp. 847-852, Springer Boston

[2] Vernadat F. B. 2010. Technical, semantic and organizational issues of enterprise interoperability and networking. Annual Reviews in Control, Vol. 34, pp. 139–144, Elsevier

[3] Serban A., Luan, J. 2002. Knowledge Management: Building a Competitive Advantage in Higher Education: New Directions for Institutional Research. Data Mining and Its Applications in Higher Education, pp. 1-13, Jossey Bass, San Francisco

[4] D.F. Barrero, Maria. D. R-Moreno. 2010. Information Integration in Searchy: An Ontology and Web Services Based Approach. International Journal of Computer Science and Applications (IJCSA), Vol. 7, (2010), pp. 14-29.

[5] Laura, M. Haas, Aya Soffer. 2009. New Challenges in Information Integration. DaWaK '09 Proceedings of the 11th International Conference on Data Warehousing and Knowledge Discovery, LNCS 5691, pp. 1–8, Springer-Verlag Berlin, Heidelberg

[6] Stocks K. I, Condit C, Qian X, Brewin P.E., Gupta A. 2009. Bringing together an ocean of Information: An extensible data integration framework for biological oceanography. Deep Sea Research Part II: Topical Studies in Oceanography, Vol. 56 (19-20), pp. 1804-1811, Elsevier

[7] Philip A. Bernstein, Laura M. Haas. 2008. Information Integration in the Enterprise. Communications of the ACM , Vol. 51, no. 9, pp. 72-79, New York

[8] Srinivasan L., Treadwell J. 2005. An Overview of Service-oriented Architecture Web Services and Grid Computing, Technical Report, HP Software Global Business Unit

[9] Stollberg M., Fensel D. 2010. Semantics for Service-Oriented Architectures. In N. Griffiths and K. M. Chao (ed.): Agent-Based Service-Oriented Computing, pp. 113-139, Springer, Berlin

[10] Alonso, G., Casati, F., Kuno, H., Machiraju, V. 2004. Web Services: Concepts, Architectures and Applications. pp. 123-149, Springer Verlag, Heidelberg

[11] Alexiev V., Breu M., Bruijn J. de, Fensel D., Lara R., Lausen H. 2005. Information Integration with Ontologies. pp. 9-36, Wiley, UK

[12] Bruijn J. de. 2006. Logics for the Semantic Web. In J. Cardoses (ed.) Semantic Web Services: Theory, Tools and Applications. Vol 16, no 2, pp. 46-53, Idea Publishing Group

[13] Hepp M., Leenheer P. de, de Moor A., Sure Y. 2008. Ontology Management: Semantic Web, Semantic Web Services, and Business Applications. Series: Semantic Web and Beyond. Vol 7, pp. 256-288, Springer-Verlag Berlin

[14] Noy N. 2004. Semantic Integration: A Survey of Ontology-based Approaches. ACM SIGMOD Record, 33(4):65–70

[15] Uchyigit G. 2009. Semantically Enhanced Web Personalization, Web Mining Appl. in E-Commerce & E-Services. SCI 172, pp. 25–44, Springer-Verlag Berlin

[16] Rao J., X. Su X. 2004. A Survey of Automated Web Service Composition Methods, in: Proceedings of First International Workshop on Semantic Web Services and Web Process Composition(SWSWPC 2004), pp. 43-54, San Diego, California, USA

[17] Tan V. V., Yi M.-J. 2010. Flexibility and Interoperability in Automation Systems by Means of Service Oriented Architecture. ICIC 2010, LNAI 6216, pp. 554–563, Springer-Verlag Berlin

[18] Tong G., Sun Y., Tang J., and Qin K. 2009. Application of Ontology-Based Information Integration on BI System. IEEE WCSE'09 World Congress on Software Engineering, pp. 171-175.

[19] Davis J., Studer R., Warren P. 2006. Semantic Web Technologies. Trends and Research in Ontology-based Systems. Wiley & Sons

[20] Kashyap V., Bussler C., Moran M. 2003. The Semantic Web: Semantic for Data and Services on the Web. In Proc of the 29th VLDB, pp. 37-58, Springer-Verlag Berlin

[21] White R. W., Bailey P., Chen L. 2009. Predicting User Interests from Contextual Information. ACM SIGIR'09, pp. 363-370, Massachusetts, USA

[22] W. Maseri Bt W. Mohd, Embong A., Zain J. M. 2010. A Framework of Dashboard System for Higher Education Using Graph-Based Visualization Technique. NDT 2010, Part I, CCIS 87, pp. 55–69, Springer-Verlag Berlin Heidelber