WFMS-based Data Integration for e-Learning

Jianming Yong[1] Jun Yan[2] Xiaodi Huang[3]
[1]*Department of information Systems*
Faculty of Business, University of Southern Queensland, AUSTRALIA
yongj@usq.edu.au
[2]*School of Information Technology and Computer Science*
University of Wollongong, AUSTRALIA
jyan@uow.edu.au
[3]*School of Mathematics, Statistics and Computer Science,*
Faculty of the Science, University of New England, AUSTRALIA
xhuang@mcs.une.edu.au

Abstract

As more and more organisations and institutions are moving towards the e-learning strategy, more and more disparate data are distributed by different e-learning systems. How to effectively use this vast amount of distributed data becomes a big challenge. This paper addresses this challenge and works out a new mechanism to implement data integration for e-learning. A workflow management system based (WFMS-based) data integration model is contributed to the e-learning.

Keywords: WFMS, Data Integration, E-learning

1. Introduction

E-learning is becoming one of most important educational means. As more and more educational organisations are moving into e-learning, more and more useful disparate data are distributed over the Internet. How to form a universal e-learning environment for all e-learning users becomes a challenge job. Like in the United States, an academic member is called a faculty member, sometimes directly called a faculty, while in the United Kingdom and Australia, a faculty has a totally different meaning, which is an academic organisation under the university. Also sometimes one same thing can be represented by different words, like in USA, an assistant professor has the same meaning as the British word, lecturer. These differences have caused a big confusion for the users of e-learning. In order to solve this problem, an effective mechanism of data integration for e-learning systems is definitely needed. WFMS-based data integration for e-learning is suggested and analysed to meet this challenge.

The reminder of the paper is organised as follows. Section 2 discusses WFMS-based e-learning. Section 3 addresses general approaches for data integration. Section 4 shows the architecture of data integration for e-learning and WFMS-based data integration is discussed. Section 5 gives the conclusion remarks.

2. WFMS-based e-learning

Workflow [1] has been used in many large organisations to control their business processes and business re-engineering. According to Workflow Management Coalition (WfMC)[2] , workflow focuses on handing business processes. It is concerned with the automation of procedures where information and tasks are passed between participants according to a defined set of rules to achieve, or to contribute to, an overall business goal. It is often associated with business process re-engineering, which is concerned with the assessment, modeling, definition and subsequent operational implementation of the core business process of an organization (or other business entity). In order to implement an effective workflow system, WfMC has published its reference model of the workflow system. In April 2000 Object Management Group (OMG) [3] also published its workflow management facility specification in order to use its CORBA and relevant technologies to implement workflow systems. The workflow mechanism has been used in e-learning design [9]. In [9], the e-learning system is modelled into four sub-systems, teaching workflow system, learning workflow system, admin workflow system, and infrastructure workflow system. These four sub-workflow systems need seamlessly work together to form an efficient e-learning environment. Furthermore a new e-learning system needs good connections with "legacy systems" which are used prior to an e-learning platform. With the extension of e-learning, more and more e-learning systems need more

co-operations. All these connections and co-operations need involving in using diverse data sources [10]. It is essential to have a mechanism of data integration for e-learning systems.

3. Data integration

The vast diversity of data sources for e-learning systems has put forward a big challenge for data integration. In [4], semantic and syntactic interoperability is very important for data integration. Semantic interoperability is the knowledge-level interoperability that provides cooperating businesses with the ability to bridge semantic conflicts arising from differences in implicit meanings, perspectives, and assumptions, thus creating a semantically compatible information environment based on the agreed concepts between different business entities. Syntactic interoperability is the application-level interoperability that allows multiple software components to cooperate even though their implementation languages, interfaces, and execution platforms are different. Syntactic interoperability gives a technology solution, while semantic interoperability provides semiotic, linguistic, philosophical, and social solutions. For aged care information systems, the focus will be put on the semantic interoperability because the aged care organisations definitely need accurately understanding all the semantic aspect of all the data sources, aged care records.

Previous research in semantic interoperability can be categorized into three broad areas: mapping-based 5], intermediary-based [6, 7], and query-oriented approaches [8].

- Mapping-based approach: The mapping-based approach attempts to construct mappings between semantically related information sources. It is usually accomplished by constructing a federated (or global) schema and by establishing mappings between the federated (or global) schema and the participating local schemas.

- Intermediary-based approach: This approach depends on the use of intermediary mechanisms (e.g., mediators, agents, ontologies, etc.) to achieve interoperability.

- Query-oriented approach: The query-oriented approach is based on interoperable languages, most of which are either declarative logic-based languages or extended SQL. They are capable of formulating queries spanning several databases.

All three approaches depend on the schemes of databases. For e-learning systems, the data sources are far broader than databases. It is not enough to focus only on the database scheme integration.

The available data sources include databases, knowledge bases, traditional files, Web page files, email messages, maps and images, videos and audios. With the growing number of information sources available, the problem of exploiting and integrating distributed and heterogeneous data sources is becoming more and more critical. Data gathering and integration is the fundamental requirement for e-learning systems. Usually different organisations have different data sources. Even some organisations possibly have the same data sources, like the same databases, the format of these data sources might be in significantly different formats. It is essential to have an effective approach of data integration for the Internet-based e-learning systems. In the following section, we will demonstrate an effective architecture of data integration for e-learning systems.

4. Architecture of data integration for e-learning system

It is hard to describe the very details of data integration for e-learning systems. In order to understand the essentials of data integration for e-learning systems, the following subsections are presented to address some issues of data integration for e-learning systems.

4.1. Data formats of e-learning system

In order to work out an effective way to integrate the data of e-learning systems, it is important to identify the data formats of e-learning systems. Generally speaking, all data are categorised into three categories: structured data, semi-structured data, unstructured data.

Structured data: Many systems use structured data to manage their large data sets, like relational database systems, object-oriented database systems. Currently Oracle, Sybase, MySQL, PostgreSQL, MS Access are popular DBMS systems for large systems or individual computers. Structured data is managed by technology that allows for querying and reporting against predetermined data types and understood relationships. Many large systems use structured data to manage their large amount of data. Most e-learning systems use structured data to manage their large volume data, like student records, administration information, course information, staff record, financial data, etc. These structure data can be well managed by its operating languages, like SQL.

Semi-structured data: Semi-structured data is that the information is usually associated with a schema and contained within the data [11], like HTML files and XML files. E-learning systems are using more and more

semi-structured data, Web pages and files, to communicate between instructors/lecturers, students/learners, administration staff and technical support people [9].

Unstructured data: Except for structured and semi-structured data, some information is totally non-structure. Usually unstructured data have two categories, map files and text files. Map files are non-language based, such as image, video or audio files. Text files are based on a written or printed language, such as Microsoft Word documents, e-mails or Microsoft Excel spreadsheets. These unstructured data are widely used in e-learning systems by all participants.

4.2. Data integration between new e-learning systems and "legacy" systems

E-learning systems have been developed only in recent years. Many education organisations or institutions used their old systems, usually called "legacy systems", to manage their learning, teaching, and administration for many years prior to adopting modern e-learning systems. It is vital to seamlessly integrate useful data from "legacy" systems into new e-learning systems. There are two ways to do data integration: once-all and on-call. Once-all is an approach that all previous data transfer to new formats which are the same as e-learning systems. After once-all transfer, the old data of "legacy" systems can be discarded. On-call is an approach that only requested data are presented to new systems based on a mediated data integration tool, while all other data keep intact. On-call can make new e-learning systems and "legacy" systems running independently if necessary or needed. First approach is rarely used because of too much cost. The second approach is often used by organisations. Figure 1 shows the architecture of data integration for new e-learning systems and its "legacy" systems.

Mediated data integration tool (Mdtt) is a key component for data integration between e-learning and "legacy" systems. Mdtt has the capacity to process the structured data, semi-structured data and unstructured data respectively. It is a bridge across two systems to implement data integration.

4.3. Data integration between different e-learning systems

From the learners' perspective, e-learning should be pervasive, whenever and wherever they want to access e-learning systems, they can get required and expected results in a consistent way. One of most important aspects is to achieve data integration across different e-learning systems. Currently most e-learning systems are connected by the Internet. In the networking environment, a term, architecture, can be used effectively to describe how heterogeneous systems work together while the technical details do not have to be addressed. Figure 2 shows an overall diagram of e-learning systems with a module of data integration. The data integration module is in charge of full data integration of all different data sources which belong to different e-learning systems.

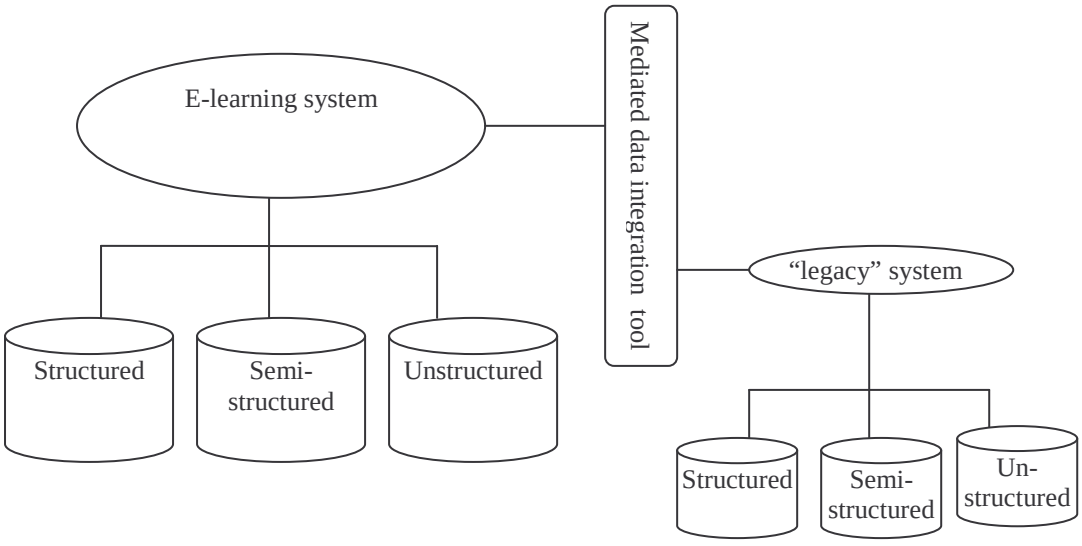

Figure 1. Architecture of data integration for e-learning and its " legacy" systems

The data integration module can handle all the different data sources, like database, knowledge base, seme-structure files, flat files, maps and images, videos and audios, etc. Based on different data sources, different approaches are used to implement the data integration for the e-learning systems. We do not want to discuss more technical details here. We just want to use an effective architecture to demonstrate the needs of data integration for the e-learning systems.

In order to describe the Internet as a whole for the e-learning, we do know there are countless data sources distributed over the Internet. Thus we use a logical architecture [10] to illustrate the overall data integrated systems over the Internet. Figure 3 shows a general view of data integration across the Internet for e-learning.

4.4. WFMS-based data integration processes for e-learning

So far data integration is still a tedious job and also involves a huge amount of labour work. In order to simplify the process of data integration for e-learning, we suggest a new approach, WFMS-based data integration for e-learning. Figure 4 shows the diagram of WFMS-based data integration for e-learning.

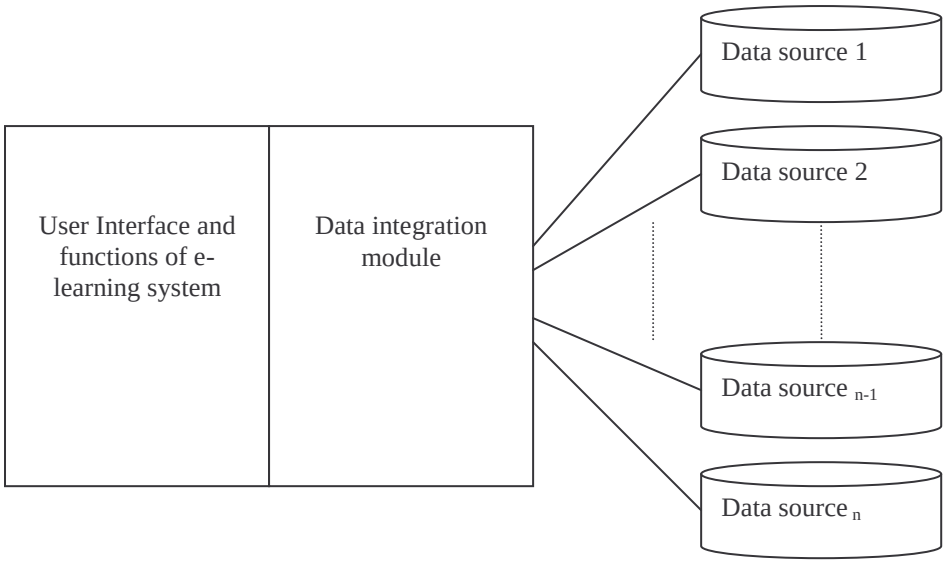

Figure 2. Architecture of data integration for e-learning systems

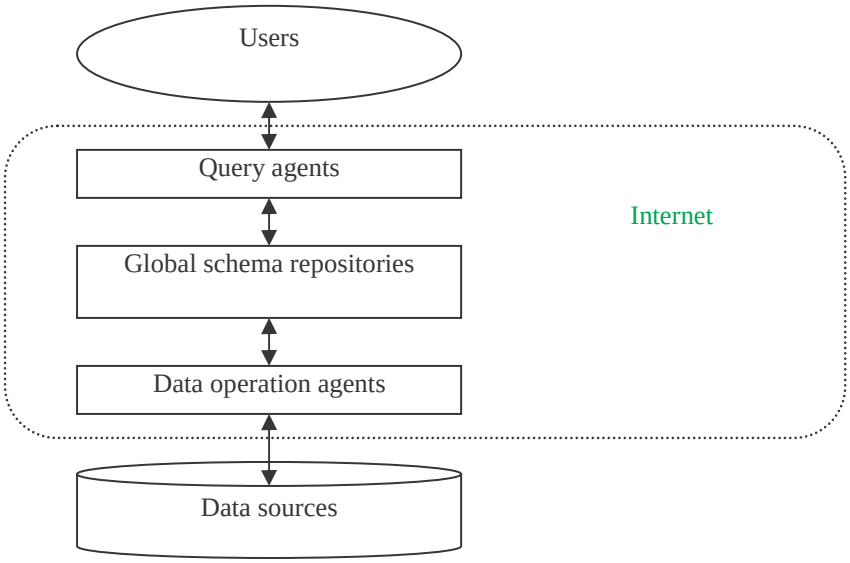

Figure 3. Data operation architecture through a data integrated system over the Internet for e-learning

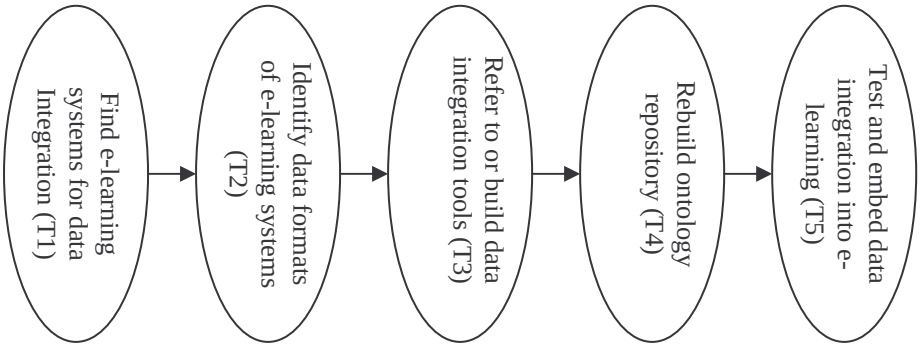

Figure 4. Diagram of WFMS-based data integration for e-learning

Find e-learning systems for data Integration (Task 1): Task 1 is first step towards data integration for e-learning. At this step, the main objective is to establish a set (S) of e-learning systems for data integration. This is a dynamic process for individual systems. Prior to the start of data integration process, S is an empty set. With the more processes of task 1, more and more specific e-learning systems are added into the set, S. This process represents by the following models.

$S \leftarrow \phi$ Prior to data integration,

$S \leftarrow S1$ System S1 looking for data integration as an initiator,

$S \leftarrow S \cup Si \,\&\, Si \notin S$ More and more e-learning systems are added into the data integrated system.

Identify data formats of e-learning systems (Task 2): After the completeness of Task 1, we assume an e-learning system, Si, is added for data integration, Task 2 needs staring to identify data formats (*df*) of Si. Si.*df* can have the following options.

Si.*df*=only one of {structured, semi-structured, unstructured}, or
Si.*df*=only two of {structured, semi-structured, unstructured}, or
Si.*df*=all of {structured, semi-structured, unstructured}

Refer to or build data integration tools (Task 3): After the completeness of Task 2, Task 3 needs to build data integration tool for Si.*df* if data integration tool for Si.*df* does not exist yet. Otherwise if data integration tool for Si.*df* exists at data integrated e-learning system, task3 needs referring to existing tools.

Rebuild ontology repository (Task 4): Now the data ontology of e-learning system, Si, needs to be reflected in the central ontology repository (*OR*) for a unified e-learning system. This process can be modelled as the follows.

$OR \leftarrow \phi$ Prior to data integration,

$OR \leftarrow S1.df.terms$ System S1' terms as an initial ontology ,

$OR \leftarrow OR \cup Si.df.terms \,\&\, Si.df.terms \notin OR$ More and more ontologies mapping are added into the central ontology repository.

For an example, there is an e-learning system in the United States, a course instructor can be an assistant professor. At another e-learning system in Australia, a course instructor usually is called a lecturer. In two systems, assistant professor and lecturer actually have the same meaning. If the mapping of these two terms has existed in *OR*, *OR* does not need to be rebuilt. Otherwise this new mapping has to be added into the central repository for data integration.

Test and embed data integration into e-learning (Task 5): At this stage, the built functions of data integration for e-learning are needed testing and embedding into a universal e-learning platform. An overall e-learning architecture is shown in Figure 5.

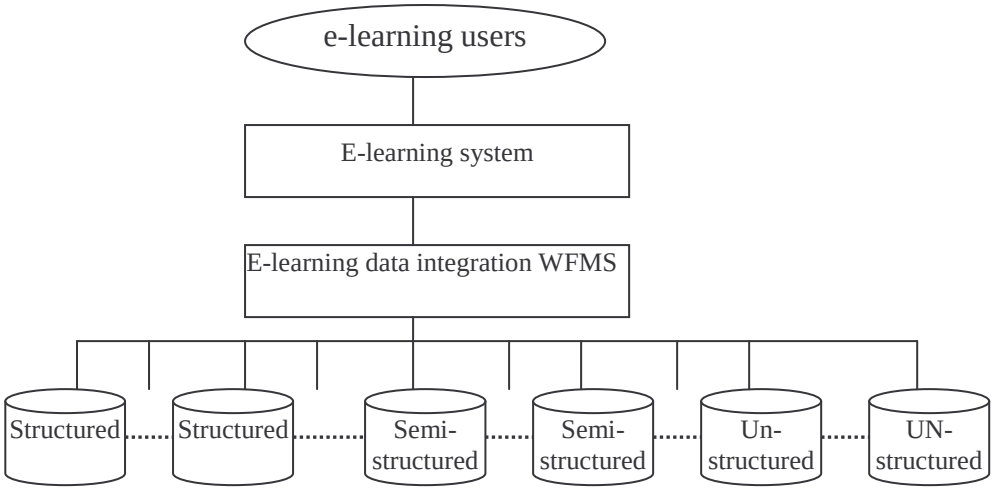

Figure 5. Overall architecture of e-learning with the support of data integration

5. Conclusion remarks

This paper contributes a new modelled concept to use WFMS principles to simplify data integration for e-learning. As more and more organisations are moving into e-learning and most disparate data are presented through different e-learning systems, data integration becomes a new challenge for establishing a universal e-learning environment. Through e-learning as an applicable application, data integration is well addressed and analysed. The principles can be applied in all arenas, like banking, health, aviation, etc. In the near future we will do some further implementations of data integration for e-learning and other applications.

Acknowledgement

This paper is written during my academic stay at Purdue University. Prof. Elisa Bertino gives me a valuable support and advice. The paper is also partially supported by USQ's early career research grant.

References

1 J. Yong, and Y. Yang. Modeling and integration for Internet-based workflow systems. The IASTED Int. Conf. on Internet and Multimedia Systems and Applications (IMSA2001). 2001. Hawaii,USA, pages:345-350.

2 D. Hollingsworth, "The Workflow Reference Model," vol. http://www.aiim.org/wfmc/standards/docs/tc003v11.pdf: WfMC-TC-1003, Version 1.1, 1995.

3 OMG, "Workflow Management Facility Specification, V1.2," http://www.omg.org/docs/formal/00-05-02.pdf, 2003.

4 J. Park and S. Ram, Information systems interoperability: What lies beneath? ACM Transactions on Information Systems (TOIS), Volume 22 Issue 4, October 2004, Pages 595-632

5 M. P. Reddy, B. E. Prasad, P. G. Reddy, and A. Gupta, A methodology for integration of heterogeneous databases. IEEE Trans. Knowl. Data Eng. 6, 6, 920–933, 1994.

6 A. M.Ouksel and M. Klusch, A framework for a scalable agent architecture of cooperating heterogeneous knowledge sources. In Intelligent Information Agents: Agent-Based Information Discovery and Management on the Internet, Springer, Berlin, Germany, 100–124, 1999.

7 J. Kahng and D. Mcleod, Dynamic classificational ontologies: Mediation of information sharing in cooperative federated database systems. In Cooperative Information Systems: Trends and Directions, Academic Press, San Diego, CA, 179–203, 1998.

8 L. V. S. Lakshmanan, F. Sadri, and I. N. Subramanian, Logic and algebraic languages for interoperability in multidatabase systems. J. Logic Programm. 33, 2, 101–149, 1997.

9 J. Yong, Workflow-based e-learning platform design, The 9th International Conference on Computer Supported Cooperative Design, Coventy, UK, 24-26 May 2005, pp1002-1007

10 J. Yong and Y. Yang, Data integration over the Internet for e-commerce, The Proc. of IASTED International Conference on Communications Systems and Applications (CSA'03), pp253-258, Banff, Alberta, Canada, July 2003.

11 P. Buneman, Semestructured Data, PODS' 97, Tucson, Arizona, USA, pp117-121.