Personalizable Information Interchange Architecture for EducationalInstitutions

Ting Tin Tin
Information Systems Engineering Research Group, Universiti Sains Malaysia
ruthting.cod06@student.usm.my
Khoo Kay Teong
Information Systems Engineering Research Group, Universiti Sains Malaysia
vkhoo@cs.usm.my

Abstract

This paper proposes a personalizable service oriented architecture for the interchange of information between educational institutions through an intermediate party. The issues pertaining to standards-personalization convergence are first discussed, followed by a description of the proposed architecture, and an evaluation of its performance at three levels of granularity. Although Web services is an emerging technology for integrating service enablers in diverse platforms, there is a lack of widely accepted flexible standards in formats, contents and processes. Among the existing business-to-business standards, it is observed that they have been focusing only on service integration at the expense of flexibility during the execution of information interchange. From the perspective of this weakness, this paper makes an attempt to introduce possible personalizable capabilities into a set of information interchange standards in the education industry.

1.Introduction

The more progressive business organizations are now moving towards standardized trade documents interchange for global supply chain management. Among the standards consortiums such as RosettaNet [20], Open Application Group [19], and Postsecondary Electronic Standards Council [17], it has been claimed that business interactions efficiency could be increased through standardized business processes, and documents structure, format and content. However, as in many standards, users have to pre-define the business processes and document format and content. Such pre-definition in the standards have imposed some restrictions in allowing individual changes during the interactions. In particular, the receivers of information have no control over what they receive. Therefore, there is a need for online personalization in the execution of standards.

Web services have been adopted by many multi-national companies such as Intel and IBM to ensure business agility [5], [4], [16]. In this paper, the proposed architecture attempts to facilitate Web services technology for information interchange with personalization features among the smaller organizations such as the educational institutions.
Section 2 introduces an educational standard called Postsecondary Electronic Standards Council, followed by a discussion on the convergence of information interchange standards and personalizable capabilities in section 3. In the following section, it presents a personalizable service oriented architecture with an evaluation of its performance at three levels. Section 5 concludes the study with some foreseeable future work.

2. Research Background

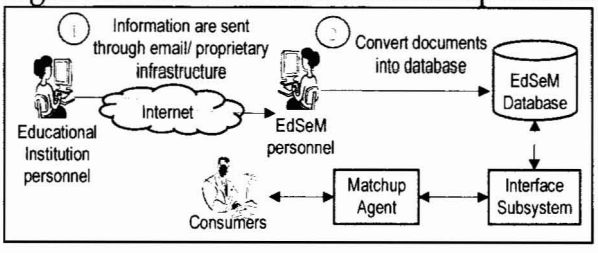
2.1.Postsecondary Electronic Standards Council (PESC)

PESC is a non-profit community-based association of the higher education industry. The primary activity is to focus on the establishment of Web service-based data standards definition and adoption which will help to ease the information interchange between the higher educational institutions [17]. Among the standards which have been approved by the consortium are High School Transcripts, College Transcripts, Academic Records, etc. The standards use XML schema which is compatible with the Web services standards. Although the accepted main standards are reusable in this study, the PESC standards interchange documents in Electronic Data Interchange (EDI) environment require fairly complicated infrastructure development. Moreover, the pre-defined standard schema could hardly be changed online as in the case of RosettaNet. As such, more personalizable capabilities are required.

2.2. Education Services Mashup (EdSeM)

Education Services Mashup (EdSeM) comprises a suite of software tools for combining information from various educational institutions into a single standardized format stored at a server farm. Interested students can then search for the appropriate institutions through only one Website. It also facilitates the educational institutions to share and interchange information via some standard procedures. It is felt that more personalizable capabilities could be incorporated into the architecture shown in Figure 1 below:

Figure 1. Education Services Mashup architecture



3.Personalization versus Standardization

3.1.Personalization

Personalization is widely adopted in B2C e-commerce to individualize products or services that match the customer needs [1]. Riemer and Totz [7] sub-divide the opportunities of B2C e-commerce personalization [15] (or individualization) into nine sub-layers from the customer perspective as shown in Figure 2. A product/service personalization is complemented by additional website and communication layers which take care of the customer interaction/communication with suppliers.

Figure 2. Nine sub-layers of personalization [7]

Product, services	
Additional services	Individualized services.
Offers	Recommendation, individualized product bundling, individualized pricing.
Website	
Website content, features	Individualization content modules, customizable website features.
Layout, look, and feel	Individualization of layout and look and feel of single web pages or whole website.
Interface, navigation	Individualized website interaction and navigational page orders.
Communication	
Communication Subject	Individualized communication subjects and contents of (one-to-one) messages.
Communication Channel	Selection and customization of communication channels and media.
Communication Attributes	Individualization of communication intensity, frequency, time and individualized address.

This study focuses on possible personalization of the *Communication Layer* in a B2B transaction. Communication in a B2B transaction includes the sending and receiving of messages or documents which contain important information about a transaction. For example, a PESC Request for Transcript process defines the student records (personal details, academic achievements, etc). This is a *Communication Subject sub-layer.* The document can be interchanged through different media/channel (*Communication Channel sub-layer*), e.g. email, EDI, infrastructure such as RosettaNet Implementation Framework [13]. In addition, the communication process attributes, such as response time and non-repudiation requirements are covered by the *Communication Attributes sub-layer.*

PESC standards define document content, structure and format (*Communication Subject*), and EDI implementation guidelines (*Communication Channel*). RosettaNet standards define business processes (*Communication Attributes*) other than the other two sub-layers. In this context, communication sub-layers are inter-related. Any alteration to a sub-layer will affect the other sub-layer(s). For example, to add or delete new elements into a standard schema requires infrastructure modification and vice versa. Personalization is restricted because of the concerns that it may violate the defined standard, jeopardize the automated process, interrupt the interaction, and reduce the inherent data quality.

Proprietary infrastructures are more flexible and personalizable to each user, but it is not suitable for global interactions due to the problems of interoperability between different applications. The connections between different proprietary infrastructures of trading partners can be calculated as n*(n-1), where n equals the number of trading partner, in which a connection is a set of relationship/ application/ interface [3], [18]. The complexity of integration increases as the trading partner increases.

Limited online alteration is allowed in standardized procedures. This study attempts to explore the possibility of introducing flexibility into the standardized *Communication Subject sub-layer.* The challenge is to adopt B2B standard documents while incorporating some degrees of personalization. Thus, personalization in this context is defined as alteration to the pre-defined standard schema. In order to narrow down the study, the personalization capabilities are confined only to online changes of pre-defined standard schema in terms of eliminating the undesired elements without modifying the Communication Channels and Attributes.

3.2.Research Challenges

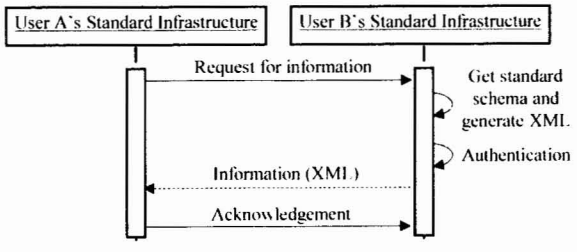
B2B standards are important to ensure that the interchanged documents can be processed by the backend system automatically [8], [11].However, as discussed in the previous section, the personalization features are limited in the standards. A modified architecture is required to adopt standardization while embracing personalization. The architecture must ensure that online changes made to the standard schema will not affect the communication channel and attributes.

The architecture may require the private processes to transform data extracted from database into XML document via the Web services. This is an enhancement for standards such as RosettaNet and PESC standards which only focus on the public processes. Therefore, Web service portability [12] is required whereby a set of *generic Web services* is necessary to interfere with different users' database management systems.

3.3. Conventional 'Push' Models

The existing standards information interchange procedure (e.g. PESC, RosettaNet) use a 'Push' model. This model requires complicated infrastructure development at each user's end, as in the case of RosettaNet Implementation Framework (RNIF) [13] and PESC EDI. Pre-defined standard schema is stored at each user's server. The interchanged document is generated based on the standard schema. Changes to the standard schema involved infrastructure design alteration at both the user ends. All other users who adopt the same standard schema must change accordingly. Therefore, the opportunity for personalization is very much restricted.

Figure 3. Sequence diagram for 'Push' model



4. Proposed 'Push'/'Pull' Models and Architecture

The Education Services Mashup (EdSeM) aims to provide a set of standardized information interchange procedures for the educational transactions, e.g., request for High School Transcript, Academic Records and others. In addition to the existing 'Push' model, this study proposes two alternative models namely, the 'Pull' model and 'Push and Pull' model as shown Figure 4. The proposed models are based on the following observations [13], [14]:

- Most of the small and medium-sized industries (SMIs) do not adopt B2B standards.
- SMIs have limited fund to invest in B2B standards adoption.
- SMIs are more concerned about information fraud in the interchange process.

Since the educational institutions belong to the SMI, the objective of the proposed models is to encourage a wider adoption of the more personalizable standards among the educational institutions.

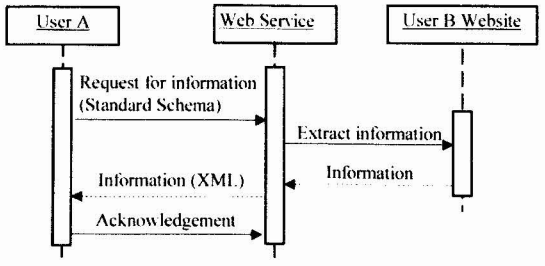
Figure 4. Proposed 'Push'/ 'Pull' models

		Push Model (Existing model)	Pull Model	Push and Pull Model
Technology		e.g. RNIF, EDI	Website extraction Web services	Database extraction Web services
Data update: schema changed?	No	Sender 'push' the latest data	Automated by software agent	Sender detects latest data and 'push' data to receiver
	Yes	Redefine standard schema and modify infrastructure	Receiver updates standard schema and 'pull' from website	• Receiver updates standard schema and 'pull' data • Sender detects latest data and 'push' data (based on latest standard schema) to receiver

4.1 'Pull' Model

The 'Pull' model is proposed in this study to add possible personalization capabilities for altering the content of standard documents. The users define the standard schema and generate DataMap (via Web services) to match websites with standard schema. Information is extracted from websites based on the defined DataMap and standard schema via the Web services. A receiver (User A) can edit the standard schema as long as the DataMap matches the schema. However, whatever changes made by User A will not affect User B.

Figure 5. Sequence diagram for 'Pull' model



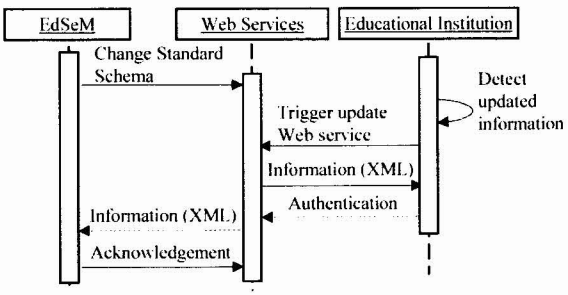
This model has its own disadvantages. The software agent involved is required to periodically navigate through the institutions websites to capture the most current data, otherwise the contents of the mashup database will be outdated. Furthermore, the websites will not provide complete information compared to direct database access.

4.2. 'Push and Pull' Model

The 'Push and Pull' model also utilizes Web services to map educational institutions' databases with the standard schema and generate the XML document for information interchange. If there is any new information, the Web services can be invoked by the

institutions to 'push' the information to the mashup server based on the standard schema. Before any information is 'pushed' out, it must be authorized by the respective educational institutions to prevent possible information fraud. The standard schema can be altered by the receiver by eliminating the undesired data fields as and when required. This will not affect the updating task because the software agent will generate the XML document based on the latest schema. This alteration also does not require any infrastructure modification affecting other parties.

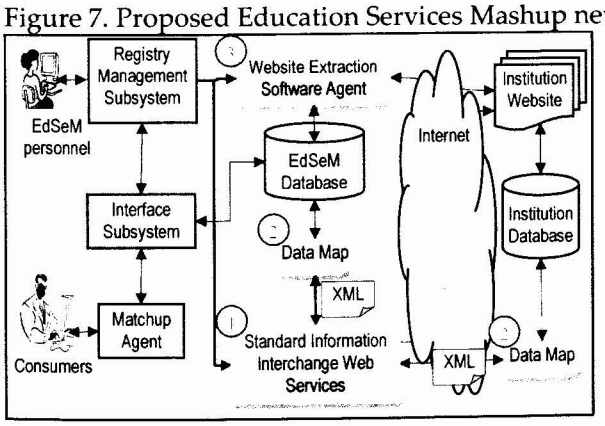
Figure 6. Sequence diagram for 'Push and Pull' model



4.3.Proposed Education Services Mashup Architecture

This section presents an architecture designed for the 'Push'/'Pull' models proposed earlier. The architecture adapts the Chen et al. architectural design [10] and the Huang et al. framework [9] with the incorporation of additional personalization capabilities together with the B2B standards component.

Figure 7. Proposed Education Services Mashup new architecture



The main subsystems of Education Services Mashup are Information Interchange Web Services, Website Extraction Software Agent, and DataMap. The design adopts a three-tier design approach, namely Presentation Layer, Business Logic Layer and Database Layer [4], [10]. The User Interface Subsystem represents the Presentation Layer, the Web Services reside at Business Logic Layer [4], [10] whereas UDDI is at the Database Layer.

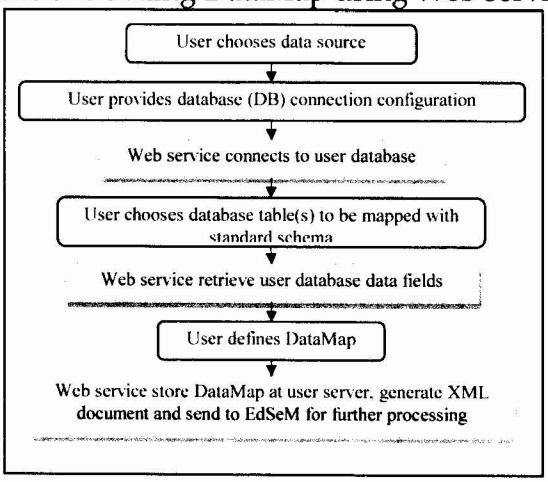
4.3.1. Education Services Mashup Personalizable Standard Information Interchange Web Services

The semantic, format and content of interchanged document (XML format) are defined based on a pre-defined standard schema. However, the content of the standard schema can be changed by the receivers so long as new elements are not added. In other words, a receiver has the flexibility of retrieving smaller size of document than those pre-defined. Generic Web services are defined to extract data from different database management systems (e.g. Microsoft Access, SQL, MySQL, Oracle, etc).

4.3.2. DataMap

The DataMap attempts to map the standard schema with the user's database for preparing the XML document. The Web services provide a graphical user interface tool for the users to generate the DataMap (Figure 8).

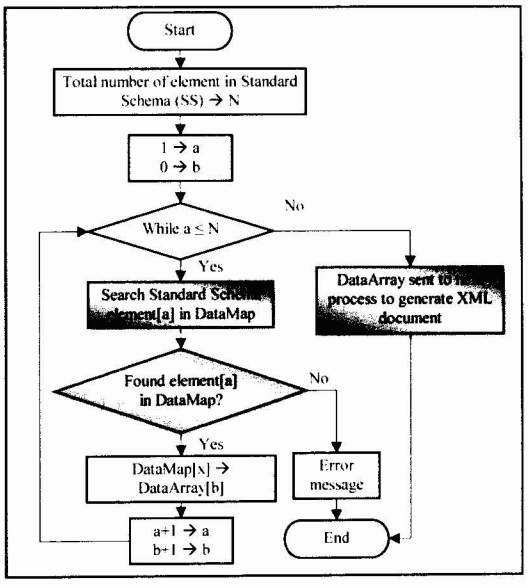
Figure 8. Defining DataMap using Web services



'Push and Pull' model allows the users to eliminate unwanted data fields from the standard schema (

Figure 9). The XML documents are generated based on the available elements in the standard schema. With the assistance of the DataMap, data are extracted from the database. Error will only occur when new data fields are added to the standard schema which is not found in the Data Map.

Figure 9. Flowchart for matching standard schema data fields with DataMap to generate XML document



4.3.3. Website Data Extraction Software Agent

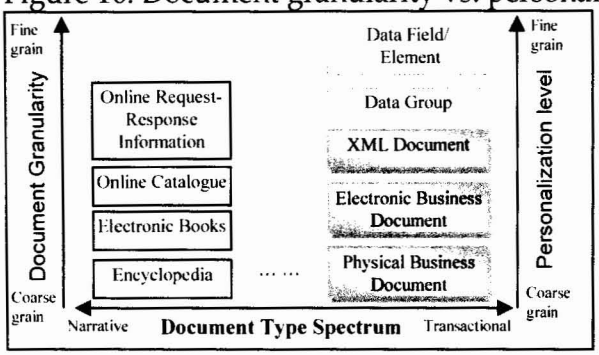
Besides the Information Interchange Web services, the website extraction software agent would realize the 'Pull' model. Data is extracted from the educational institutions' websites in the 'Pull' model. A software agent is scheduled to periodically detect possible changes at the web pages and extract the modified data, if any. The receivers will decide when to 'pull' data from the websites. However, this 'Pull' model might miss out the most current data if the schedule is not properly planned.

4.4. 'Push'/ 'Pull' Models Evaluation

The evaluation is focused on the level of personalization capabilities in the standardized procedures. Three levels of content personalization intensity are applied, i.e., coarse grain, medium grain and fine grain. The intensity is based on the document granularity model defined by Glushko and McGrath [6] for the narrative document granularity from coarse grain to fine grain. The model reflects transactional documents granularity. The finer the granularity of a document, the finer the granularity of data is accessible to the user. In this case, the finest grain of data is a data field, e.g., StudentName. Personalization level is proportional to document granularity. A document with fine granularity can be more

personalized (Figure 10). With a personalization level of fine grain, the users are allowed to modify the schema in units of a single data field, compared to the medium grain where the users can only modify groups of data fields.

Figure 10. Document granularity vs. personalization level



The levels of personalization among the models are summarized in Figure 11. In the original 'Push' model, the Standard Schema can only be changed together with *Communication Channel and Attributes*. It is not alterable during online execution. The 'Pull' model allows the receiver to modify the standard schema. However, due to the incomplete information provided in websites, granularity of personalization is quite limited. The 'Push and Pull' model allows the users to drop the undesired data elements in a standard schema.

Figure 11. 'Push' / 'Pull' models and level of personalization

	Push	**Pull**	**Push and Pull**
Personalization level	Coarse grain	Medium grain	Fine grain

In order to further justify the adoption of personalization and standardization, the following guidelines are proposed:

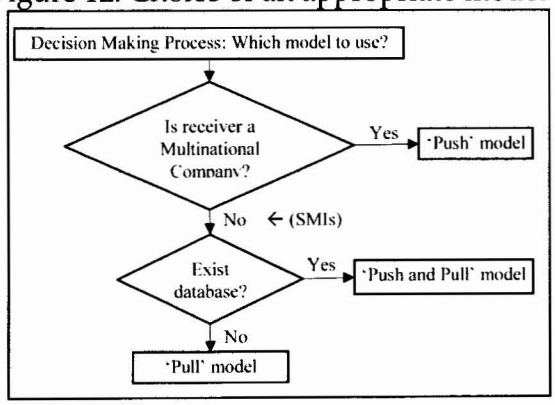
- The standard schema can be altered online by eliminating the pre-defined data elements but not adding new elements.
- The receivers can alter the standard schema online but not the sender.

Although the 'Push and Pull' model appears to be a better alternative compared to the others, each model has its own usability. How each model could be chosen is illustrated in the following flowchart (

Figure 12).

It is believed that with the personalized models, it will encourage small and medium-sized industry to adopt B2B standards in information interchange process. This is necessary due to the high cost and time involved in implementation and inflexibility to alter standard schema in conventional model. Any changes made to any layer of communication protocol require all trading partners to modify the communication subject/infrastructure/attributes accordingly. This is an expensive, time consuming and complex practice.

Figure 12. Choice of an appropriate model



5. Conclusion and Future Work

This paper has presented the 'Pull' model and 'Push and Pull' model for a more personalizable Web service oriented information interchange architecture. It is found that in most B2B document interchange standards, personalization is restricted by the pre-defined standard schema. Any subsequent schema alternation requires the Communication Channel and Attributes to be modified accordingly. In the proposed architecture, the receivers have the control to 'pull' the desired data by altering the standard schema without affecting the Communication Channel and Attributes.

The 'Pull' model utilizes a software agent to extract information from the websites and update the database periodically based on the pre-defined standard schema, whereas the 'Push and Pull' model prepares generic Web services to interchange information extracted from the databases. Since each of the models has its own strengths, guidelines are provided for a more appropriate choice of each.

The more immediate future work is to carry out experiments to further verify the actual performances of the models, followed by the research into tasks categorization in the private processes and possible incorporation of some self-learning capabilities into the Web services and/or software agents, either maintained at the central server farm and/or at the individual institutions.

6. References

[1] W.A. Hanson, and K. Kalyanam, Internet Marketing & e-Commerce, Neil Marquardt, Canada, 2007.
[2] T.T. Ting, and K.T. Khoo, "Receiver-Oriented 'Pull' Model for ResettaNet Trade Documents Interchanges", National Conference on Software Engineering & Computer Systems, Malaysia, 2007.
[3] P. Strong, "Enterprise Grid Computing", Enterprise Distributed Computing, 2005.
[4] A. Coenen, and C. Harding, "An SOA Case Study: Agility in Practice", SOA Magazine, 2006.
[5] A.W. Brown, M. Delbaere, P. Eeles, S. Johnston, and R. Weaver, "Realizing Service-Oriented Solutions with the IBM Rational Software Development Platform", IBM Systems Journal, 2005, pp. 727-752.
[6] R.J. Glushko, and T. McGrath, Document Engineering - Analyzing and Designing Documents for Business Informatics & Web Services, The MIT Press, London, 2005.
[7] K. Riemer, and C. Totz, The Customer Centric Enterprise: Advances in Mass Customization and Personalization, Springer Verlag, New York/Berlin, 2003.
[8] S.C.W. Kong, H. Li, T.P.L. Hung, J.W.Z. Shi, D. Castro-Lacouture, and M. Skibniewski, "Enabling Information Sharing between E-commerce Systems for Construction Material Procurement", Automation in Construction, 2003, pp. 261-276.
[9] Y. Huang, and J.-Y. Chung, "A Web services-based framework for business integration solutions", Electronic Commerce Research and Applications, 2003, pp. 15-26
[10] M. Chen, and M.J. Meixell, "Web-Services Enabled Procurement in the Extended Enterprise: An Architectural Design and Implementation", Journal of Electronic Commerce Research. 2003, pp. 140-155.
[11] E. Söderström, "Challenges in the Field of B2B Standardisation", The Conference on System Integration, 2003.
[12] P. Clements, R. Kazman, M. Klein, Evaluating Software Architectures: Methods & Case Studies. Addison Wesley, 2001.
[13] RosettaNet, RosettaNet Technical Overview, RosettaNet Engineering Team, Penang, 2007.
[14] T.T. Ting, RosettaNet Engineering Team Interview, Penang, Malaysia, 2007.
[15] C. Belz, et al., Erfolgreiche Leistungs – systeme, Stuttgart, Schäffer, 1991.
[16] SOA Transforms Intel's e-Commerce,
http://www.intel.com/it/pdf/soa_supplychain.pdf.
[17] PESC, http://www.pesc.org.
[18] Migrating to a Service-Oriented Architecture, Part 1,
http://www-128.ibm.com/developerworks/library/ws-migratesoa.

[19] OAGI, http://www.openapplications.org.
[20] RosettaNet Malaysia, http://www.rosettanet.org.my.