

# Flexible database transformation for XML-RDBMS

A. Zailani<sup>1</sup> & M. Mustafa<sup>2</sup>

<sup>1</sup>*Department of Computer Science,  
University Malaysia Terengganu, Malaysia*

<sup>2</sup>*Faculty of Information Technology & Multimedia,  
University Tun Hussein Onn Malaysia, Malaysia*

## Abstract

Transformation of original dataset into predefined data format always becomes a main concern in pre-processing of text mining or information retrieval. Unfortunately most of the available tools and databases in the market are not capable to incorporate implicit flexibility of two-way XML-RDBMS transformation. Thus, the development of flexible database transformation is absolutely significant in order to overcome these drawbacks. Here, the exiting tools for database conversion are examined in an attempt to determine their strengths and weaknesses. In this paper, we develop a model using Relational Unified process (RUP) via C# programming language, based on .NET framework for database transformation. It shows that, the database conversion process can be easily performed and flexible.

*Keywords: transformation, text mining, database, conversion, flexible.*

## 1 Introduction

Flexi-DB Transformation (F-DBT) is a tool for transforming XML into Relational Database Management Systems (RDBMS) or vice-versa. These two types of databases have been employed widely especially in the web applications. Nowadays, there are huge of data available in both formats and always need further processing. However, not all databases can afford to provide a good solution for advance processing or data interchanging.

In text mining, pre-processing phase usually consumes a lot of resources in term of efforts and man-hour. One of the main concern is to reduce irrelevant



features via dimensionality reduction process. Too many dimensions are not good for further analysis and sometime may produce unexpected result. Typically, calculation of dimensionality reductions such as Chi-Square, Mutual Information and Information Gain are very difficult in XML format compare to normal relational (row-column) format.

Currently, there are quite a number of database transformation tools such as Access2MySQL Pro 5, dbQwikSync Pro, dataPro etc. However, one of the major problems of these tools is, database transformation is only applicable in the one way direction. Besides that, functions for selecting the users' preferable attributes are also not available. Realizing the advantages offered by XML, a few major RDBMS vendors have extended its basic RDBMS functionality to include XML data management capabilities named XML-enabled. By using XML-enabled databases, data can be transferred from relational tables to XML documents in both ways [1]. However, the limitation is, the XML-enable database has better performance only for a small database size (number of records  $\leq 1,000$ ) compare to native XML-database.

Reuters-21578 [2] is a classic example of dataset in SGML / XML format for text mining. Numerous researches have used this dataset for further analysis. However, the constraint is these data must be converted first into certain formats for example in table (row-column) format. Otherwise, it is very difficult to analyze and apply some formula in XML format. Thus, development of database conversion tool from or to XML-Relational database is very significant for further data analyzing and processing in the area such as text mining, information retrieval, etc. This paper proposes a flexible model using RUP based on .NET engine to transform XML to or from Relational database.

The reminder of the paper is organized as follows. Section 2 introduces all the basic concepts and some related works. Section 3 focuses on methodology used to accomplish this project. Section 4 describes the tool being developed and overall findings. Finally, Section 5 concludes the paper with a short summary.

## 2 Literature review

Two tools have been evaluated for benchmarking purposes; Access2MySQL Pro 5 and dbQwikSync Pro. Access2MySQL Pro 5 [3] is one of the database conversion tool which can convert the relational database in Microsoft Access, into MySQL format, and then converted into XML format. dbQwikSync Pro [4] is another database conversion tool which can convert directly relational database in Microsoft Access or Microsoft Dbase format, into XML format. Although this tool provides step-by-step instruction for the user during the whole conversion process, but it focuses more on experienced users with some basic knowledge in databases.

XML [5], or 'Extensible Markup Language', is a simple text format for a semistructured information document. XMS has growth very fast as a dominant database for representing the hypertext level of a website such as logical composition of web pages and the navigation structure [6–8]. Recently, there are large amounts of data in both financial and business data, that is, most of the data



in today's web-driven world are being continuously converted to XML [9]. XML has basic features including storage the data using XML documents, define schemes through XML schemes and DTD [10], query languages function [11] by XQL and XPath, programming interface capability via DOM and JDOM, and many more.

Relational database is a database that conforms to the relational model, and refers to a database's data and schema. This database can be considered as a dominant database and being employed widely to store content of a web site. The term relational database was first defined and coined by E.F. Codd [12]. The term relation stems from the mathematical theory which derived from his database model. Conceptually, all the data are stored in tables (row-column) format.

Compared to the XML database, relational database possess its own advantages too. Some advantages [13] include control of data redundancy, data consistency, sharing of data, and improved security and integrity. Among the most popular relational databases are Microsoft Access, Microsoft SQL Server, IBM DB2, Oracle and MySQL. Currently, there are several top vendors in database industry already incorporated the function of transforming data into XML. For example, IBM's DB2 introduces DB2 XML Extender [14,15], Microsoft SQLServer 2000 [16] with OPENXML row set and Oracle has Oracle XML SQL Utility [17].

### 3 Materials and methods

Relational Unified Process (RUP) methodology has been chosen for developing F-DBT tool. RUP is an iterative software development process framework created by the Rational Software Corporation, a division of IBM since 2003. RUP lifecycle is an implementation of the spiral model. Four main phases involved in RUP are inception, elaboration, construction and transition.

#### 3.1 Inception phase

During this phase, business case is established. A basic use-case diagram is designed to provide main functionalities offered by the F-DBT tool. Figure 1 shows a use case diagram consist of two (2) actors named "User" (the final users of the tool) and "Printer" (to output information on the screen), and five (5) main use-cases which are "Retrieve important elements from XML database", "Create and insert data into RDBMS", "Retrieve important attributes from RDBMS", "Create and insert data into XML database" and "Show output records".

#### 3.2 Elaboration phase

The elaboration phase is where the project starts to take shape. Analysis on how the conversion processes between both XML and MS Access are performed. The basic database characteristics and restrictions which may occur along the database conversion tool's development will take into consideration. All related



diagrams such as activity diagrams, relational entity classes' diagrams, and basic GUI are produced.

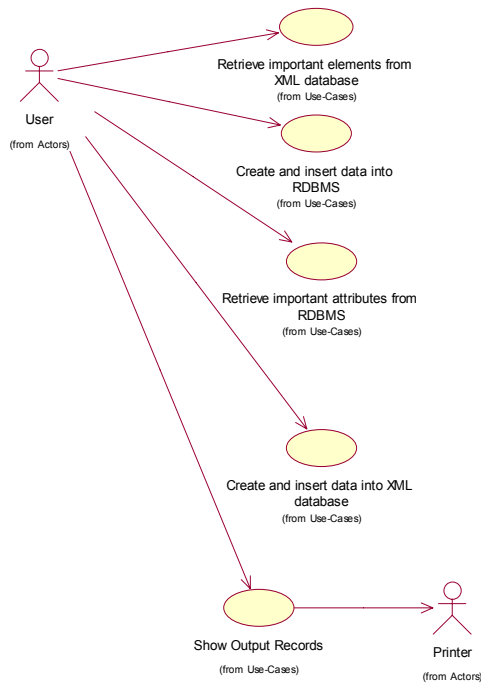


Figure 1: A use-case diagram.

### 3.3 Construction phase

All codes are built in Microsoft C#. During the development, the complete model to show the basic workflow is derived. From the model, one can clearly see how the data and processes in the database conversion tool relate and how it interacts with each other via Microsoft's C# programming language. Figure 2 shows overall model of the database conversion tool with .Net framework.

In the Figure 2, both XML and RDBMS (Microsoft Access) databases' format are connected via the C# programming language, based on .NET Framework. The conversion goes in two (2) ways, which fulfils the concept of flexi-transformation in the database conversion tool.

### 3.4 Transition phase

At this phase, the testing processes of the main functions will be performed several times to ensure that it is free of errors and bugs. If any errors or bugs are

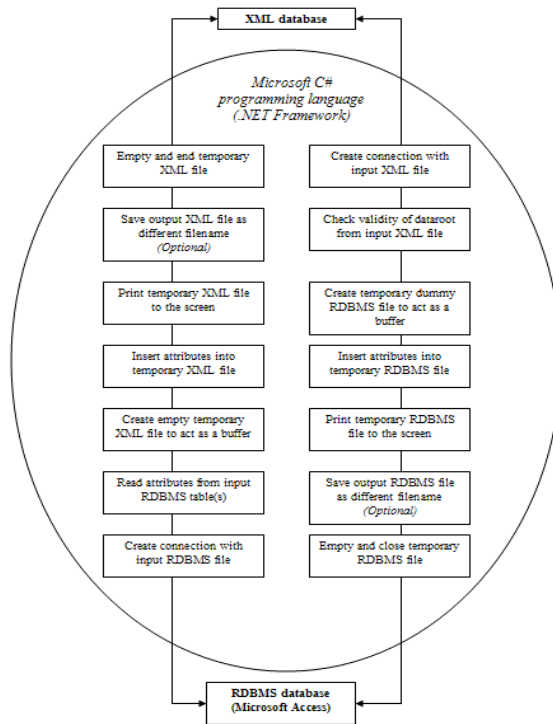


Figure 2: An overview model.

detected during testing, there will be code fixing until all errors have been eliminated. However, if there is a need to add any extra minor functions into the database conversion tool, the updating process of the tool will be performed again via the iterative process concept. Not only functional attributes which are being tested, but also user interfaces may also be commented and changed according to suitability.

## 4 Result and discussion

This section presents the main functions and output generated from the tool.

### 4.1 Conversion from RDBMS to XML

When this option is chosen, user has to select the location of the input Microsoft Access database file first by browsing through the computer.

After the input database file is selected, a tree-view of the RDBMS input database will appear on the left side of the window as at Figure 3. From there, user can choose which table(s) or attribute(s) to be converted and then select them into the conversion area on the right side of the window. If the user want to control any attributes by converting them all into NULL (or empty) values, the

user just need to select the checkbox(s) in the conversion area on the right side of the window. When everything is done, user need to click on the ‘Convert’ button to perform the conversion, and after confirming in the pop-up confirmation box, the conversion process will begin.

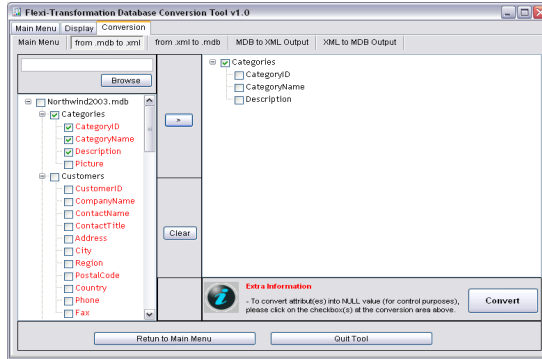


Figure 3: Conversion from RDBMS to XML window.

After the conversion process succeeds, an output window which shows the output database file in the form of XML will appear on the screen as at Figure 4. From here, user can select whether to save the output file, perform another conversion, return to main menu of the tool, or simply exit from the tool.

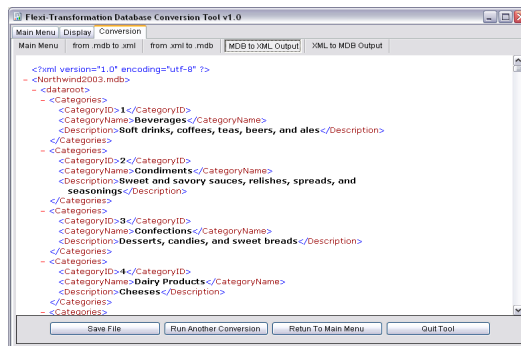


Figure 4: Output window of conversion from RDBMS to XML.

## 4.2 Conversion from XML to RDBMS

Alternatively, when this option is chosen, user has to select the location of the input XML database file first by browsing through the computer. Do take note that most of the steps in this conversion way are similar to the conversion from RDBMS to XML.

After the input database file is selected, a tree-view of the XML input database will appear on the left side of the window as at Figure 5.

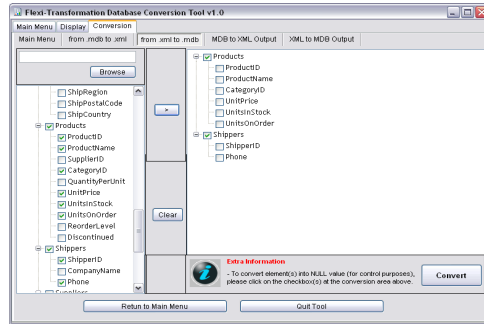


Figure 5: Conversion from XML to RDBMS window.

From Figure 5, user can choose which table(s) or element(s) to be converted and then select them into the conversion area on the right side of the window. If the user want to control any elements by converting them all into NULL (or empty) values, the user just need to select the checkbox(s) in the conversion area on the right side of the window. When everything is done, user need to click on the 'Convert' button to perform the conversion, and after confirming in the pop-up confirmation box, the conversion process will begin.

After the conversion process from XML to RDBMS succeeds, an output window which shows the output database file in the form of Microsoft Access's style will appear on the screen as at Figure 6. From here, user can select whether to save the output file, perform another conversion, return to main menu of the tool, or simply exit from the tool.

### 4.3 Display input database

User can also select the option to only display the input database on the screen without the need to perform any conversion in the first place. User can select whether to display input database in the format of RDBMS (Microsoft Access format) or XML format, by first selecting the location of the input database in the computer. After that, the tool will display the input database in the screen, as at Figures 7 and 8.

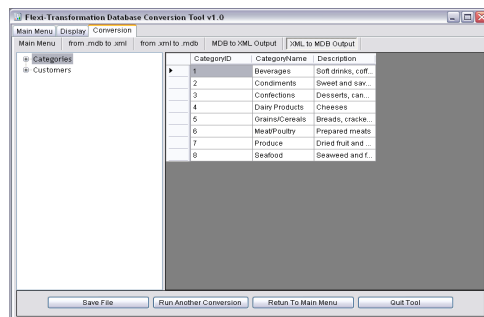


Figure 6: Output window of conversion from XML to RDBMS.



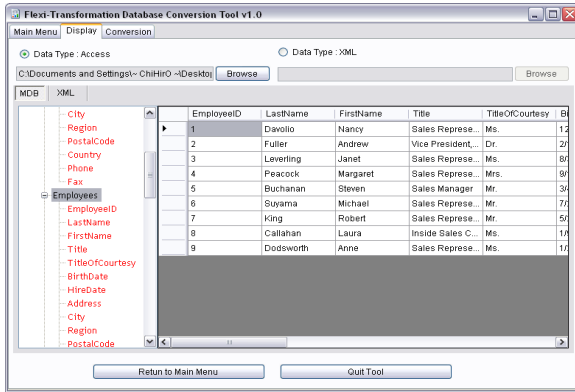


Figure 7: Display window of RDBMS (Microsoft Access) input database.

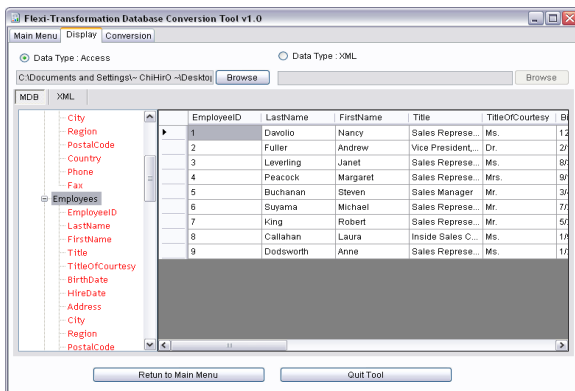


Figure 8: Display window of RDBMS (Microsoft Access) input database.

#### 4.4 Advantages of F-DBT tool

Further from the issues discussed in Literature Review, all the specified problems has been catered for and solved in the development of this database conversion tool.

*Support for double-way conversion:* F-DBT tool is flexible for two-way conversion, which is to convert from RDBMS to XML database, and from XML database back to RDBMS. Besides that, the original structure, integrity, and original data from the input database, stays intact. Minor errors in the conversion process will be significantly reduced with increasing number of error checkers and loops at critical stages along the whole database conversion process.

*Possess function to choose preferable attributes from input database:* Besides having the option to choose the whole table or elements to be converted, it also provides a feature to choose a number of needed attributes or elements for conversion. This will greatly save storage space, as only needed attributes and tables will be converted, and the tool's response time will be greatly shorten as



only selected attributes will be converted and not the whole table with all attributes.

*Greater user-friendliness:* The database conversion tool will emphasize more on the user-friendliness aspect by preparing a much simpler and appropriate interface, without sacrificing its original functionality in the process. Besides providing functions in the tool using easy to understand buttons, technical jargons are also kept to the minimum and only relevant inputs will be asked, thus increasing user's satisfaction towards using this tool.

## 5 Conclusion

The development of the database conversion tool using the .NET engine will greatly broaden the horizons of the future of text mining, information retrieval, database development etc. With the usage of this database conversion tool, the database conversion process would be as much easier and really flexible for user.

Further analyses in text mining processes such as dimensionality reductions can be easily performed once the output is converted into tables (row-column) format. By using F-DBT tool, it can reduce time for researchers to write a program for database conversion and they can focus directly to other advance processes.

Future work of this project is to test and evaluate the actual performance of the tool into Reuters-21578 dataset. Another potential work is to allow flexible conversion between XML to and from another type of databases such as Microsoft SQL Server, MySQL, DB2 and Oracle databases.

## References

- [1] Serna, A., Gerrikagoitia, J, K. "David & Goliath: A Comparison Of XML-Enabled And Native XML Data Management Techniques" *XML Journal*, 28 Jun 2005, [xml.sys-con.com/read/104980.htm](http://xml.sys-con.com/read/104980.htm).
- [2] Reuters-21578 Text Categorization Test Collection [www.daviddlewis.com/resources/testcollections/reuters21578/](http://www.daviddlewis.com/resources/testcollections/reuters21578/).
- [3] Acces2MySQL Pro, [accessmysqlpro.say-it-now.com/](http://accessmysqlpro.say-it-now.com/).
- [4] dbQwikSync Pro, [www.supershareware.com/download/dbqwiksync-pro.html](http://www.supershareware.com/download/dbqwiksync-pro.html).
- [5] World Wild Web Consortium, Extensible Markup Language (XML) 1.0, W3C recommendation 10-February-1998, [www.w3c.org/XML/](http://www.w3c.org/XML/).
- [6] Abiteboul, S., Buneman, P., Suciu, D.: *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufman Publishers, 2000.
- [7] Ceri, S., Fraternali, P., Paraboschi, S.: *Design Principles for Data-Intensive Web Sites*. ACM SIG-MOD Record, Vol. 24, No. 1, March 1999.
- [8] Widom, J.: *Data Management for XML – Research Directions*. IEEE Data Engineering Bulletin, Special Issue on XML, Vol 22, No. 3, September, 1999.
- [9] SenGupta A. and Mohan, S., Formal and conceptual models for XML Structures – the past, present and future, *Technical Report no. 137-1*,



- Information Series Department Working Paper Series*. Indiana University, April 2003.
- [10] Obasanjo, D. An Exploration of XML in Database Management Systems, 2001, [www.25hoursaday.com/StoringAndQueryingXML.html](http://www.25hoursaday.com/StoringAndQueryingXML.html).
- [11] Bourret, R. XML and Databases XML and Databases, 2001 [www.rpbourret.com/xml/XMLAndDatabases.html](http://www.rpbourret.com/xml/XMLAndDatabases.html).
- [12] Codd, E.F. *A Relational Model of Data for Large Shared Data Banks*. Communications of the ACM 13 (6), 1970, pp. 377–387.
- [13] Connolly, T., Begg, C. *Database Systems*. Fourth Edition, Addison-Wesley, England, 2005
- [14] Carey, M., Florescu, D, Z. Ives, Lu, Y., Shanmugasundaram, J., Shekita, E., and Subramanian, S. XPERANTO: publishing object-relational data as XML. *In Proceedings of WebDB*, Dallas, TX, May 2000.
- [15] Shanmugasundaram, J., Shekita, E., Barr, R., Carey, M., Lindsay, B., Pirahesh, H., and Reinwald, B. Efficiently Publishing Relational Data As XML Documents. *In Proceedings of VLDB*, pp. 65–76, Cairo, Egypt, September 2000.
- [16] Rys, M. Bringing The Internet To Your Database: Using Sqlserver 2000 And XML To Build Loosely-Coupled Systems. *In Proceedings of the International Conference on Data Engineering*, pp. 465–472, 2001.
- [17] Oracle XML Sql Utility For Java. Oracle Corporation, 1999. [technet.oracle.com/tech/xml/oracle\\_xsu/](http://technet.oracle.com/tech/xml/oracle_xsu/).