DELOS Research Exchange Program Scientific Report

Franca Debole
Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo"
Consiglio Nazionale delle Ricerche
Area della Ricerca CNR di Pisa
Via G. Moruzzi, 1 - 56124 PISA - Italy
franca.debole@isti.cnr.it

Abstract

This report summarizes on the scientific activity performed by the author during the visit to TU-Wien for the period ranging from 2^{nd} November 2005 to 30^{th} November 2005. The goal of this activity is to both test the resulting framework in large and real-world environments with different preservation requirements, and to make the resulting generic preservation testbed framework and decision support modules more comprehensive.

1 Introduction

This exchange program is related to the activity performed under the DELOS WP6 Digital preservation. The objective of this cluster is to lay the foundation for testbeds and necessary metrics and tools for assessing preservation strategies, to raise the profile of digital preservation issues within the Digital Library Community, to collaborate with other international bodies to ensure consistencies of digital repository standards, to ensure access to file format information and to establish the relationship between a typology of file formats and preservation strategies, to enable the definition of attributes and functionalities that need to be represented, and ensure that system development methodologies reflect preservation analysis and design issues. The main goal of this program is the study of a metric for testing and evaluating digital preservation strategies, which is part of the research done in Task 1 Digital Preservation Testbed Forum. The Receiving partner of this Exchange Program is *Department of Software Technology and Interactive Systems Vienna University of Technology* in the Information & Software Engineering Group with Prof. Andreas Rauber. While the sending organization is the Institute of Information Science and Technologies in particular the Networked Multimedia Information Systems Laboratory with Dott. Giuseppe Amato.

This report is organized as follows: Section 2 describes briefly the activities performed in the sending and receiving organization and how these have been integrated; in Section finally 3 some details of the integration, Section 4 reports plans for future cooperations.

2 Concepts

The team of Prof. Rauber has realized a framework for testing and evaluating digital preservation strategies [5]. The framework is based on the Utility Analysis [6], which is adapted for preservation scenarios. The followings are the steps of the analysis:

1. Defining project objectives
2. Making objectives measurable
3. Listing alternatives strategies
4. Measuring the strategies' performance
5. Transformation into comparable numbers
6. Defining the importance of objectives
7. Aggregation of values
8. Ranking the alternatives

The core element, the objective tree, is made in the first step. The objective tree captures the preservation requirements or objectives, describing the characteristics of the digital object and the process environment. The researcher Debole has developed and is working on an improved version of a high-performance, native XML database system with special features for DL applications [2, 3]. This XML database can store and retrieve any valid XML document. Once an arbitrary XML document has been inserted in the database it can be immediately retrieved using XQuery. The work done during this visit aimed to improve the usability of the preservation framework and to evaluate the obtained results in practice in a real-world preservation scenarios like a database world.

3 Works

The tool realized by Rauber's team can be applied with some modifications to preservation solutions as well. As practical examples to evaluate the migration strategies they are used an audio collection of the Austrian Phonogrammarchiv and an journal's digital library in MS Word format. On the contrary, in this cooperation we have used a collection of data of a database system to evaluate the results of tool. The idea of team of TU is to preserve the digital object in one of the standard formats which are available. With regard to the databases the SQL is the standard widely accepted, but also XML-based files. Just this last standard has been used, in this collaboration, to do database preservation. XML is an excellent choice of format for the long term preservation of databases. It can be used to specify the context, content and structure of databases. Conversion to XML can be seen as an intermediate step in a migration between a present day and a future database, for instance an native XML database. The first step of this work was to define the project as whole and its goals, this was made by exploiting the experience on relational and XML database of Franca and the knowledge on preservation of the TU group [4, 5]. The definition of project goals was realized constructing an objective tree. An objective tree is a formal representation of the various goals, which are collected and put in relation to each other to gain a certain structure. In a generic objective tree, the main goal is detailed into three subgoals to preserve the objects characteristics, to optimize the preservation process and to keep costs at a reasonable level.

From the various meeting the two groups had identified the main objectives in the field of long-term preservation of relational databases. A solution for the archiving of relational database:

- has to enable permanent retention of the original data structure, the referential data integrity, and all technical and non-technical meta-information needed to keep the data technically accessible over the long term;
- must be able to completely detach databases from its proprietary database management software, hardware, and operating system environments;
- shall easily enable the reload to current and future relational database management system products, and thereby allow queries as in the original system from which the database was archived from.

Results of the various brainstorming meetings is an objective tree of ?? nodes. Furthermore, the author looked for a tool which converts data from almost any RDBMS in XML, enabling complete detachment of databases from their vendor-specific environment. The author has examined a lot of tools for the conversion of a generic databases to XML documents. Between the software analyzed the author chose an open source java based for mapping from database to XML the *xmlizer configurator* [1].
This tool have two modules:

1. Simple DB Export
2. Database to XML

The first one exports databases, or only tables, in two different formats (simple and verbose), While the second module can be used for all mapping purposes where you want to generate XML instances on demand, only using an XML Schema (or a DTD). Using the first module the author has made a simple conversion of data contained in a mysql database, which a member of TU group has placed at her disposal. The conversion using the second module is not terminated yet. When also the second type of conversion will be completed, it will be possible to test and evaluate the alternatives according to the criteria of the objective tree. After the TU group will proceed with the other step of utility analysis framework.

4 Future Work

This report summarizes on the scientific activity performed by the PhD student Franca Debole. After her experience, the author is convinced on the effectiveness of the exchange of researchers between the DELOS organizations as one of the most effective ways to achieve interesting collaboration. The two groups planned to continue to work together on a number of topics, first of all to finish the test work and to evaluate according to the criteria of the objective tree, the performance of the different approach of conversion.

Acknowledgments

This Exchange Program is funded by the DELOS Network of Excellence for Digital Libraries. Moreover, Franca would like to thanks Dr. Giuseppe Amato and Prof. Andreas Rauber to give her the opportunity to have this research experience.

References

[1] The xmlizer configurator. http://www.e-freak.ch/xmlizer.

[2] Giuseppe Amato and Franca Debole. A native xml database supporting approximate match search. In *ECDL*, pages 69–80, 2005.

[3] Giuseppe Amato, Claudio Gennaro, Fausto Rabitti, and Pasquale Savino. Milos: A multimedia content management system for digital library applications. In *ECDL*, pages 14–25, 2004.

[4] C. Rauch, F. Pavuza, S. Strodl, and Rauber A. Evaluating preservation strategies for audio and video files. In *Proceedings of the DELOS Workshop on Digital Repositories: Interoperability and Common Services*, 2005.

[5] Carl Rauch and Andreas Rauber. Preserving digital media: Towards a preservation solution evaluation metric. In *Proceedings of the 7th International Conference on Asian Digital Libraries (ICADL 2004)*, pages 203–212, Shanghai, China, December 13-17 2004. Springer.

[6] Paul Weirich. *Decision Space: Multidimensional Utility Analysis*. Cambridge University Press, 2001.