

Standing on the Shoulders of Giants: digital platform designing with native XPath enabled storage and retrieval on MySQL

Thomas Nicolai, Marc-Frederic Schäfer, Lars Kirchhoff

Institute for Media and Communications Management
University of St.Gallen
Blumenbergplatz 8
CH-9000 St.Gallen
{firstname.surname}@unisg.ch

Abstract: XML has become a standard for the exchange and retrieval of semi-structured data between different platforms but is not currently a standard for storing semi-structured data in relational databases. There exist a number of different approaches in transforming or mapping semi-structured documents into the highly structured relational database schema. In this paper we do not present a new approach on native XPath enabled RDBMS but instead looking at a new implementation case for the popular MySQL database system and its implication for standard in designing open source based digital communication systems as well as implications for the information retrieval scientific community.

1 Introduction & Motivation

For the information exchange between digital platforms a common standard is that XML documents are created by the underlying system from primary data stored in relational databases and translated back into their own relational data structure by the receiving agents. Therefore the XML document has to fit the relational structure of the database design or the relation structure has to be made fit. Changes in the XML document structure have mostly direct implications to the underlying relational database structure and / or to the scalability of the belonging mapping processes. Therefore the management of semi-structured data has forced another approach: native Xpath support in RDBMS. Seen from the process of designing a new medium we have to pay special attention to a growing number of agents who will request semi-structured data. Therefore most of the design requirements are influenced by the need to explore efficient ways in the architecture of such systems to overcome the scalability problem of data transformation processes. Designing a research medium [A106] with a highly adaptable user data generation process as well as focussing design requirements on Open Source Software needs an exploration design process of currently non-standard solutions to fit the future needs. We have been developing a new digital research platform and therefore have searched ways to persistently store XML documents in relational database as to overcome several limitations.

2 Design Requirements & Approach

The process of designing and the outcome of this - the artifact - is based on requirements that evolve through the whole design process. To fit this we concentrated our problem solving process on native Xpath enabled storage and retrieval enabled RDBMS solutions because of various reasons. We are aware of existing approaches in this area but they did not meet our requirements [Or06, Pg06]. Therefore we explored at first the fairly new native Xpath support of the MySQL [My06] database system as a solution to persistently store XML documents as a core basis of our information architecture.

3 Learning's & Conclusion

A number of observations were made regarding the implementation of the research platform. The most important of these are described as follows. (1) Often because of the design approach the saving of relations in XML documents can present structural problems and unwanted complexities. (2) The architecture has shown that through the selected structure and database design, independence from the saved data can be configured thus making the interoperability (layered architecture) with other systems much easier and efficient. (3) Changes to the data structure of the individual content objects require no alteration to the database structure. This abstraction has also had positive influence on the structure of the object oriented software architecture. (4) The premature assessment of an established XML schema (like for example Dublin Core) makes a later adaptation of this redundant. These eliminate adaptation or a later adjustment of the XML document and the XML data in the database. (5) Since there are currently no management tools (like PHPMyAdmin) for the support of XPath the direct manipulation of the data is highly difficult. Further developments could be undertaken to improve several limitations, e.g. the elimination of relations in XML documents to avoid structural problems. Further developments like the integration of access restriction possibilities which are embedded in the XML documents (by using XACML [21]) or the introduction of ontology in combination with used XML schemas are evaluated and perhaps implemented in the future to facilitate cross media data exchange and retrieval. Within this paper we have shown that it is possible to overcome design barriers of cross media information retrieval by using native XPath functionalities of the MySQL RDBMS.

References

- [Al06] Alexandria Research Platform (University of St.Gallen): <http://www.alexandria.unisg.ch>
- [My06] Mysql Xpath functions: <http://dev.mysql.com/tech-resources/articles/mysql-5.1-xml.html>
- [Or06] Banerjee, S. et.al.: Oracle8i-the XML enabled data management system. In Proc. of the 16th International Conference on Data Engineering, 2000.
http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=839454
- [Pg06] Dyson, T.; Gray, J.: PostgreSQL and XML.
http://www.throwingbeans.org/postgresql_and_xml.html