Dr Georg Grossmann
School of IT & Mathematical Sciences
University of South Australia
Mawson Lakes Campus
Mawson Lakes, SA 5095
AUSTRALIA
Phone: +61 8 8302 3194
Fax: +61 8 8302 3988
Email: georg.grossmann@unisa.edu.au

Prof. RNDr. Jan Kratochvíl
Studijní oddelení-doktorské studium
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
Ke Karlovu 3, 121 16 Praha 2

August 26, 2013

Dear Prof. RNDr. Kratochvíl:
Please find enclosed my examiner's report for Jakub Malý's PhD thesis.

If you have any questions, please do not hesitate to contact me.

Sincerely,

Dr. Georg Grossmann
Lecturer

Examiner's Report
on the dissertation entitled

XML Document Adaptation and Integrity Constraints in XML
By Jakub Malý

Topic

Computer systems are constantly changing and evolving over time. The main reason driving changes are performance improvements and adaptation to new functionality, new business strategies or external factors such as law which leads to changes in interfaces and data structures. With the dominance of the World Wide Web as the standard communication infrastructure, interfaces in form of (Web) services and the Extensible Markup Language (XML) for data structures became the new standard. Fabio Casati and Boualem Benatallah pointed already out in 2007 that the main challenge behind dynamic evolution of Web services is an efficient management of protocol evolution and the handling of instances running under old protocol when it has been changed. This challenge became more important than ever with new technologies in the area of Big Data and Cloud Computing which require handling the adaptation of not only a large amount of data but also a large diversity of data. The thesis "XML Document Adaptation and Integrity Constraints in XML" written by Jakub Malý is therefore a valuable contribution to this challenge and proposes a unique approach that combines state-of-the-art technologies with a new innovative approach that simplifies and increases the automation of XML document adaptation.

The scientific contribution of Jakub Malý is recognised in the community by an extraordinary large number of publications. They include three top international journal articles published in Information Systems Frontiers, Journal of Systems and Software (JSS) and Data & Knowledge Engineering (DKE) and eight conferences, including top-ranked International Conference on Web Services (ICWS) and Information Systems Development (ISD) as well as Advances in Databases and Information Systems (ADBIS) and Asia-Pacific Conference on Conceptual Modelling (APCCM). Further his work has been recently awarded with the Best Paper Award at the Ninth Asia-Pacific Conference on Conceptual Modelling (APCCM 2013).

Research challenge and objectives

Jakub addresses multiple important aspects in his thesis: evolution of XML schemas, handling multiple versions of schemas and documents, adaptation of documents after schemas have been changed, and dealing with complex integrity constraints within documents - an essential requirement when dealing with complex data structures. He proposes an interesting and sophisticated approach based on Model Driven Engineering principles. This approach lifts the evolution of XML schemas and documents to a conceptual level and allows users to concentrate on the actual evolution of data structures

rather than on underlying XML implementation which can easily complicate changes and distract from the actual task.

The formalisation of platform specific and platform specific model level as well as the formalisation of the mapping between the levels and the operators that execute evolutionary changes, allow a substantial increase of automation. Rather than performing changes manually and adapting documents individually, the framework provides a comprehensive list of change operators, a change detection algorithm and adaptation approach which will reduce the laborious and error prone tasks of XML document adaptation dramatically. Further it allows the analysis of impact of changes prior to the adaptation of documents. This is a valuable contribution in identifying costs and effort of changes at an early stage of evolution.

The implementation of the framework is presented in Chapter 7 followed by a discussion of related work in Chapter 8. Both chapters provide a decent overview but could be elaborate in more detail for the evaluation of the presented framework. For example, the adaptation process described in Chapter 4 mentions necessary user involvement in some situations. It would be interesting to present a comparison of how much user effort is reduced by this approach, maybe based on the provided example. Similar to the chapter on related work, it provides a good overview but a comparison in detail, e.g., an evaluation against comparison criteria would make it more obvious that this work exceeds existing approaches. It might be useful to compare this work also with the following publications:

- Semantic Constraint-Based XML Updating. Md. Sumon Shahriar, Jixue Liu. FGIT-DTA/BSBT 2010: 100-109

- Towards the preservation of functional dependency in XML data transformation. Md. Sumon Shahriar, Jixue Liu. International Journal of Intelligent Information and Database Systems (IJIIDS), Vol. 4, No. 5, 2010

- Towards Evolving Constraints in Data Transformation for XML Data Warehousing. Md. Sumon Shahriar, Jixue Liu. ADBIS (Workshops) 2009: 79-86

Advantages and benefits

The advantages provided by the presented XML document adaptation framework can be summarised by four distinct benefits:

1) *Applicability*: The approach is directly applicable in the real world. It focuses on XML which is the de-facto standard for structuring content and documents on the Web. Jakub described in detail a formal framework for the evolution and adaptation of XML documents using an example throughout the thesis and in the appendix. He also addresses one aspect that is usually ignored by related research but often required in the real world: the support for multiple versions in parallel and backwards compatibility of changes which are addressed in Section 4. Further the thesis presents an implementation in Section 7 which demonstrates its applicability.

2) *Simplification and re-usability*: The model driven engineering approach explained in Section 2 places XML-specific on the platform specific level (PSM level) and lifts data structures to the conceptual level on the platform independent level (PIM level). By doing so, data structures are abstracted and simplified on a higher level which allows the user to focus on the actual data structure rather than dealing with XML-specific constraints that distract from the evolution tasks. Lifting XML schema specifications to the PIM level will also provide a higher re-usability of the evolution approach. All changes that are performed on this level can be re-used for different XML languages by generating XML expressions as demonstrated for XSLT in Section 4, and increases the re-usability of the approach for different implementations.

3) *Increased automation*: The formalisation of the adaptation framework, in particular the change predicates and the adaptation described in Section 4, enable to automate the changes of schemas and the adaptation of XML documents to a higher degree. Based on the identified guidelines for each change operator, it is known which changes can be adapted automatically and which require user intervention. There exists no approach for the evolution of XML schema and document that describe the steps in such detail as in this thesis.
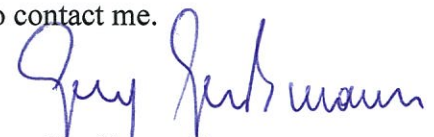
4) *Increased analysis capabilities*: The formalisation does not only increase the automation, it also builds the basis for further analysis such as impact analysis of schema changes. Taken into account the adaptation requirements for each changes, allows the prediction of impact by changes and the effort and costs of adapting existing documents.

Summary of evaluation

Overall, the conducted research presented in this thesis is *highly innovative* and *directly applicable* to real-world scenarios. It explains the approach that goes *beyond state-of-the-art* in detail that is useful for practitioners and further research. The thesis provides *original contributions* to the field of XML document evolution and adaptation, applies *appropriate methods* and shows the *deep understanding* of Jakub Malý in the area of XML and applied conceptual modeling techniques. *This thesis proves Jakub Malý's ability for creative scientific work.* Part of his work has also been supported by an *exceptional number of high-quality peer-reviewed publications* in journals, conferences and workshop proceedings which also provides adequate proof of the quality of his research.

Should you require further information, please do not hesitate to contact me.

Adelaide, 26 August 2013

Dr Georg Grossmann
Lecturer

Minor remarks (printed version):

Page 14, Figure 2.1: Provide a legend that explains different types of arrows.

Page 16: "the conceptual perspective models the semantics of the XML format in terms **of** the PIM schema.": the "of" is missing.

Page 21: I think there should be N'\timesN' instead of N'xN'

Page 37: *though* instead of *thought*

Page 43: There is no reference to Figure 4.2.

Page 47: Reference to Fig. 4.2 instead of 4.3?

Page 51: *made* instead of *mad*

Page 108: Remove "CONSIDER:" remark.