Review of Doctoral Thesis

Author: Tomáš Knap
Title: Towards Trustworthy Linked Data Integration and Consumption
Reviewer: RNDr. Irena Holubová, Ph.D. (thesis supervisor)

General Description

The thesis consists of 7 chapters covering 196 pages altogether. The first chapter provides an excellent introduction and motivation for the problems solved in the rest of thesis and explains precisely which aspects are solved in which parts of the thesis and respective papers. Author's proposals are described in the following five chapters, whose research value and contributions are confirmed by the respective international publications. The seventh section sums up the achievements of the thesis and outlines future work.
The structure of thesis conforms to principles and requirements on the structure of a scientific thesis. The author has studied and used appropriate number of bibliography sources used and quoted in the thesis. It is the evidence of the deep theoretical knowledge and very good orientation in the problems discussed in the text.
The word processing of the thesis is adequate. The usage of different fonts and structure of the text is proper and helps the reader in better orientation in the text. In the text there are also numerous figures, listings and examples which help the reader to understand the ideas. The thesis fulfils the general formal requirements at a very good level.
In general the author proved the ability to prepare a sound and explanatory research text describing the solutions of the selected problem with all appropriate parts.

The Topic of the Thesis
The area of Linked Data is currently a very popular and still open topic which brings numerous novel research challenges to data management and processing, as well as many new applications and areas of usage. The author has focussed on an important task of data provenance and trust which forms a basic and key aspect necessary for most Linked Data applications. On the other hand, the problem is nontrivial and so are the respective proposed solutions. The author performed an extensive analysis of existing solutions and focused on their several further improvements.
The author has performed good orientation and wide knowledge of different parts of the provenance and trust problem both from practical and theoretical point of view. He has studied the area from several points of view and within several research teams and projects, namely the *XML and Web Engineering Research Group*, the GAUK project *Efficient Processing of Linked Data*, the GACR project *Handling XML Data in Heterogeneous and Dynamic Environments*, the TACR project *Intelligent library INTLIB*, the *DERI institute* (within his 3-month research stay), the FP7 project *LOD2*, the *SoSIReCR* project funded by the European Social Fund, and the *OpenData.cz* initiative.

Author's Contributions
Most of the author's contributions are related to the *ODCleanStore* tool which was implemented as a student SW project at the Faculty of Mathematics and Physics of the Charles University in Prague under author's supervision and on the basis of his proposals (mainly the W3P provenance model, quality assessment, strategies to data cleansing and linking etc.). The tool serves as a proof of the concepts outlined in the thesis and related

papers. Its impact is proven also by its usage within the OpenData.cz initiative and, especially, the FP7 project LOD2.
The tool also depicts author's teaching abilities, which are further proved via supervision of 3 Bachelor theses, 2 Master theses and 2 SW projects.

Author's Publications
The publications covered in the thesis involve 1 journal paper with impact factor, 7 full papers published at refereed international conferences and workshops, and 6 demo and positional papers published at refereed international conferences and workshops. The results are more than sufficient with regard to the respective research level. The author published his results at several acknowledged events, such as a VLDB workshop, the ESWC and WISE international conferences, international summer schools (namely the *Indian-summer school on Linked Data* and the *ESWC Summer School*) etc.

Conclusion
In my opinion, the thesis of Tomáš Knap undoubtedly fulfils all the conditions for gaining the Ph.D. degree in Computer Science; therefore it is recommended.

In Prague, April 28, 2013

RNDr. Irena Holubová, Ph.D.

Department of Software Engineering
Faculty of Mathematics and Physics
Charles University in Prague
Malostranské nám. 25
118 00 Praha 1
Czech Republic