



To whom it concerns

FIT CVUT Prague

Referee report of the PhD thesis
Linked Data Based Knowledge Provisioning
By Ing. Milan Dojchinovski

The thesis under report deals with a challenge – how to help developers to find suitable API's for a software project (and also a more general problem with large portion of the information on the Web which is available only in an unstructured format).

The defense process was interrupted and I obtained some comments of candidate. Parts of my report concerned by candidate are in italic and numbered with Roman numbers. Where I require further comments (again italic and sometimes bold), listed with capitals or bold in the text. **In general I would like to see a record of a session with and user using the system.**

Of course, there are tens of thousands API's on the web, many of them on Programmable Web. Nevertheless, there is a problem – description is in HTML, it is not clear how much effort needs their integration. Problem is attacked by author with weapons of semantic web – first semantization, then knowledge and finally personalization (extraction like Wikipedia → DBPedia is a dream). Use cases are nice,

- A. *Thesis claims creation of semantized storage of API descriptions (with SPARQL endpoint) in the spirit of Linked Data.*
- B. *Knowledge is understood to be hidden in Named Entities.*

This is a hot topic, up-to-date and PhD adept is well placed in the international community.

Thesis is structured by three goals and solutions are presented in three main chapters.

Chapter 3 deals with Acquisition and Semantization of Web API Descriptions. Main contributions are the (so far) largest Linked Data dataset with semantic Web API descriptions; a light-weight ontology for modelling relevant Web API information and a *survey* on the usefulness of the dataset *aiming to* ascertain the added value and the degree of achievement.

Results are nice, especially Linked Data dataset.

- I. *It would be interesting to see, how the ontology was used (by developers) for their work. E.g. was it enough intuitive (session record-trained/untrained user)? Ontology itself has concepts and properties very general. There are some helpful examples e.g. of data formats JSON, XML, RDF, etc. protocols HTTP, SOAP, etc. examples of categories: social, video, music, etc. example of tags are: auction, mobile, shopping, etc. Something like a more detailed user study would be nice to see.*
- II. *I would like to see an example of a transformation of an API description from Programmable Web and resulting instance in Linked data repository annotated by author's ontology (true user's session record).*
- III. *Where can we find more details about the process of this transformation – is it a tool which does this (something remembering extraction rules from Wikipedia to DBPedia) how it was trained? This is the most interesting part of the semantization process.*

Chapter 4 Knowledge Extraction and Integration with Salient Linked Entities is oriented **more general to the web**. It considers semantization of texts (from Linked TV EU project e.g. transcripts from automated speech recognition). Chapter lists following contributions: an open-source system for named entity recognition and entity linking; a crowdsourcing based methodology for learning entity salience; ground-truth datasets for named entity recognition and entity salience; an evaluation framework for named entity recognition and experimental evaluations.

IV. *It is a lot of work hidden in this chapter to bring this pipeline functioning. I am interested how and who constructed pre-defined JAPE grammars in this solution (do untrained users understand these?).*

V. *Can entity salience be personalized? see requirement of a session record.*

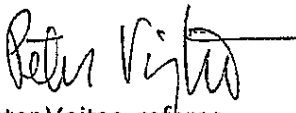
Chapter 5 Personalized Access to Linked Data comes back to linked API dataset. Contributions consist of a method for personalized access to Linked Data; a Linked Data resource similarity and relatedness metrics; resource recommendation use case for the Web services domain and experimental evaluations and results from the evaluations. As noted in the Introduction “huge and diverse amount of information published as Linked Data, the actual access to a piece of information in the LOD cloud demands significant amount of effort” this is also true for the API search domain – personalization is an important step towards reducing these efforts.

VI. *My question is – is personalization more influenced by developer's history or rather by the customer requirements (session records, user studies)?*

I value the thesis and author's work very high – he proved to be a valuable member of many international excellence teams and has an excellent publication record. Dissertation objectives were met, methods used in solution are modern, up-to-date-although some details of usage (user's session record) would be helpful.

I do recommend the thesis for the presentation and for defense with the aim of receiving the PhD degree.

Prague, September 10th, 2018


Peter Vojtas, referee



Professor
Dr. Harith Alani, BSc MSc PhD

Knowledge Media Institute

The Open University
Walton Hall
Milton Keynes MK7 6HF
United Kingdom
Phone: +44 1908 659399
Fax: +44 1908 653169
Email: h.alani@open.ac.uk

26 November 2018

Thesis Review

Linked Data Based Knowledge Provisioning

By Milan Dojchinovski

The candidate has significantly improved his thesis and addressed all my comments from the first review very well. In particular, the chapters are now well connected and follow nicely from each other, and the problematic and irrelevant material fixed or removed. The research questions are now better articulated, more strongly related to the work in the thesis, and well connected to each other.

Evaluations have been extended and are now far more appropriate for the assessment of the work in this thesis, and cover the whole work done in the various chapters. Research outcomes are now compared against appropriate baselines and related works, thus positioning the work well from the state of the art.

In the attached PDF, I have highlighted and added notes to a number of very minor corrections (typos and grammatical errors). These are mainly in the new text that was added in this last version of the thesis.

Based on the above, I believe the thesis is ready to pass the examination. In accordance with par. 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act) I **DO** recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Harith Alani', written in a cursive style.

Prof Harith Alani
Knowledge Media institute, The Open University

Ruben Verborgh
Ghent University – imec



Dear Dean Jiřina,

Subject: Review PhD dissertation Milan Dojćinovski

12 September 2018

With this letter, I am providing a review of the PhD dissertation of Mr. Dojćinovski. Given that this is my second review, I will respond briefly to the topics.

Up-to-datedness of the dissertation

The material of the thesis is sufficiently up to date.

Formal structure and organization of the dissertation

The organization has vastly improved since my previous reading.

Completion of the dissertation objectives

The objectives are clear and I consider them completed.

Assessment of the methods used in the dissertation

The comparisons are adequate to validate the research questions.

Evaluation of the results and contributions of the dissertation

The candidate has made excellent contributions to the field.

Remarks, objections, notes, and questions for the defense

I would like to thank Milan for taking my previous comments into consideration. I understand that this has cost significant effort, but I think we will be in agreement that this strongly reflects on the quality of the end result.

The overall evaluation of the dissertation

Solid work with considerable depth, pleasant to read.

Given the above, I conclude that the author of the dissertation proved the ability to conduct research and achieve scientific results. In accordance with par. 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act); I do recommend the thesis for the presentation and defense with the aim of receiving the PhD degree.

Sincerely,

A handwritten signature in black ink, appearing to be "R. Verborgh", with a long horizontal line extending to the right.

Prof. Ruben Verborgh, PhD