prof. dr hab. inż. Krzysztof Goczyła

Politechnika Gdańska

Wydział Elektroniki, Telekomunikacji
i Informatyki

kris@eti.pg.edu.pl

Gdańsk, 31.03.2022

Reviewer's opinion

on Ph.D. dissertation authored by

Dawid Wiśniewski

entitled:

AUTOMATING COMPETENCY QUESTIONS HANDLING IN ONTOLOGY DEVELOPMENT PROCESSES

1. Problem and its impact

The subject of the dissertation is the ontology engineering – the field of computer science dealing with the development of real world models based on first-order logic. So, this area falls into the branch of software engineering called knowledge-based systems engineering. It is an important area for modern computer science due to the fact that the modern world, subject to constant and inevitable progressive computerization, is becoming more and more complicated and very difficult to precisely describe, which is crucially important in the development of usable information systems.

The Author has set itself the goal of developing methods and tools to support the knowledge engineer, i.e. the person who creates ontologies, in the process of creation, verification and validation of ontologies. This task is ambitious for several reasons. First, the process of creating an ontology is an advanced synthesis process, not a simple analysis of existing knowledge about a fragment of real world. Secondly, it requires the knowledge engineer to delve into the modeled domain, and this is not an easy task due to the inherent impedance mismatch between the computer scientist and the domain expert, which results mainly from the rather hermetic modeling language. Thirdly, there are no widely recognized and accepted knowledge-based systems construction processes, as is the case in the construction of "classical" information systems.

The Author concentrates on the approach based on so-called competency questions, that is exemplary questions to which the ontology (the knowledge base) should deliver appropriate answers (in the form of knowledge). The aim of the dissertation was formulated by the Author in the form of five research problems (research questions). This approach has its advantages and disadvantages. Typically, in doctoral dissertations, a general thesis (or research hypothesis) is formulated, which is the core of the dissertation, and research goals are subordinate to this thesis. In the reviewed dissertation, in principle, research problems correspond to these goals, but there is no general thesis that would link these goals into one consistent research direction. I will return to this problem later in the review. On the other hand, the advantage of this approach is that makes the dissertation more

specific and its goals become easier to prove to be achieved than a general thesis, which usually requires verification and validation in an IT industry setting.

To sum up this part of the review - the Author of the dissertation took up the issue of knowledge-based systems engineering, important from the point of view of modern computer science, which is of practical importance for the emerging need to organize knowledge about the contemporary complex real world. Undoubtedly, this issue is of a scientific nature and the Author attacked this by formulating adequate research problems.

2. Contribution

As it was mentioned before, the Author focused his work on five research questions. These questions are original and the analysis of the research tasks related to these questions is itself an original contribution to the field of knowledge engineering. The author attacked these tasks and related problems in a comprehensive manner and achieved original and interesting results, mainly from the practical, software engineering point of view.

1) (RQ1) Examining whether there are recurring patterns among CQs, SPARQL-OWL queries, and between CQs and SPARQL-OWL queries, in Chapter 6 of the dissertation the Author analyzed a set of 234 competency questions relating to five real ontologies. Based on the analysis of the syntactic and semantic structure of these questions, he constructed domain-independent patterns, which were then related to templates formulated in the SPARQL-OWL language, creating the extensive set called CQ2SPARQLOWL. He then used this set in the SeeQuery tool, presented in Chapter 9 of the dissertation.
2) (RQ2) By examining automatic methods of extracting terms from competency questions, in Chapter 7 the Author developed and analyzed two methods - a method based on machine learning and a method based on rules. As a result of experiments, it turned out that the better method that shows better precision and recall factors is the rule-based method, which eventually was implemented by the Author within the rule-based tagger.
3) In the next research question (RQ3) elaborated upon in Chapter 8, the Author posed the issue of the automatic creation of pairs: competency question - SPARQL-OWL query, on the basis of axioms contained in the ontology. For this purpose, the Author proposed a method of verbalizing ontological axioms and their subsequent linguistic transformation into generic pairs of a competence question – an SPARQL-OWL query. The Author also analyzed the scope of possible linguistic constructions that can be transformed in this way. As a result of this work, he created BigCQ - an extensive set of patterns of competency questions related to SPARQL-OWL query templates, to be used then in the SeeQuery tool.
4) The objectives of the next two research questions (RQ4, RQ5) were to determine how the developed methods can assist the ontology engineer in developing and testing ontologies. The fourth research question suggests that it is possible to automatically recommend the form of a SPARQL-OWL query based on a competency question asked in a natural language. Six consecutive steps were proposed, based on the results of work done to answer previous research questions. The whole procedure has been implemented in the SeeQuery tool. It is important that the Author analyzed the applicability of the proposed procedure, examining carefully cases in which SeeQuery was unable to cope with some competence questions that did not fit exactly into templates.
5) In the fifth research question, the Author posed the problem of integrating the proposed methods, obtained results and developed tools with the method of ontology engineering based

on testing (Test-Driven Development). As a result, a scheme of such an integration was proposed, including additionally the so-called presupposition tests.

I consider the above-mentioned original contribution of the Author to be valuable for computer science in the area of knowledge management. I have no objections to the approach to solving the problems posed in the research questions. However, I have some remarks related to the lack of a clearly formulated thesis (research hypothesis) in the dissertation. Constructing a dissertation in the form of answers to research questions makes the results presented in individual chapters seem like individual blocks, from which it is not known what can be put together. The only fragment of the dissertation in which the methods proposed in the dissertation are placed in a more general process one can find in Chapter 10, Fig. 10.2. This chapter, however, is somewhat on the sidelines of the whole argument and the reader gets the impression that it is not an essential part of the dissertation.

I believe that the entire dissertation, and in particular the valuable results obtained by the Author, would gain more value if it presented the overall process of building and testing an ontology, where the developed methods and tools would be used. This process could be either an extension of a known state-of-the-art process (methodology) or a proposal for a new one. Perhaps this problem should be left for further studies and work, although a sketch of such a holistic approach would be highly desirable. What was presented in Chapter 10 is just a slight touch of the problem.

My second substantive remark, related to the first one, concerns the fact that the Author's methods and tools have not been practically validated through the process of creating and verifying a real, maybe not very complicated, ontology. It is known that any engineering tools should be checked at work. Usually, their creators are not able to see all of their possible shortcomings until someone else tries to apply them. I am aware of the difficulties of such a validation task, and it is quite a common difficulty in PhD dissertations in software engineering. However, taking into account the vision of practical application of the proposed approach, it is worth thinking about carrying out such a verification and validation process, even in a purely academic settings.

3. Correctness

The dissertation is written very well, both in terms of editing and technical point of view. It is written in very good English. In the entire, quite extensive text I found only a few typos (e.g. missing "2" in "CQ2SPAQLOWL"). I have no detailed substantive remarks to the original arguments presented in Chapters from 6 to 10. Some general remarks I have presented above. Nevertheless, while studying the dissertation, some debatable comments came to my mind.

1) In Section 2.3. The author discusses the differences between the Open World Assumption (OWA) and Closed World Assumption (CWA). From the comment under Table 2.1. one could get the impression that OWL reasoners work in the three-valued logic: *Yes*, *No*, *Do-not-know*. This is not true, since DL is a subset of (two-valued) first-order logic. Actually, in the world of OWA, in the absence of knowledge, the DL reasoner answers *No* to both questions of type: *Can you prove that* [*this*] *is true*? and *Can you prove that* [*this*] *is false?*. In this context, it would be reasonable to refer to the epistemic operator **K** "closing" the world (the knowledge base) and compare the "classic" (CWA) SPARQL with the SPARQL-OWL. This would not increase the size of an already extensive dissertation; for instance Chapter 3 could be considerably shortened, so as it turned out that machine learning methods were eventually not used in the developed methods.

2) In several places of the dissertation, the Author verifies his methods and states that in some cases errors may occur in the automatic interpretation of competence questions. This is understandable due to the fact that these questions are formulated in natural language. However, it may happen that one error in the ontology is enough to make the entire ontology useless (e.g., inconsistent). It would be worth discussing this problem.
3) In the dissertation, the Author consistently uses the XML-based OWL notation for DL class expressions. In some places it makes the dissertation difficult to read, especially in Chapter 8, where the so called axiom shapes are presented. In such places, OWL expressions could have been accompanied by expressions with DL operators.

As I stated before, the above remarks may be, among others, topics for discussion during the public defense of the dissertation and they do not contradict with my high opinion on the original contribution of it.

4. Knowledge of the candidate

In my opinion, the Author of the dissertation showed very good, deep knowledge in the field of Semantic Internet and ontology engineering as well as the OWL and SPARQL-OWL languages that constitute the basis for the construction and querying DL knowledge bases. In the initial chapters of the dissertation – Chapters 2 to 4 - he presented the current state of knowledge in this field, pointing to some difficulties in practical ontology engineering, shortcomings of existing methods and processes and the need to support an ontology engineers, particularly in testing completeness of ontologies by means of competency questions. He confirmed his knowledge with a wide, exhaustive selection of literature on the subject. It is also of great importance that the Author has published the results of his work in notable journals and at international conferences on knowledge management.

In general, I state that the candidate has a general knowledge in the **Information and Communication Technology** discipline of Computer Science sufficient to conduct independent research in this discipline in the areas related to modern knowledge management.

5. Conclusion

Mr. Dawid Wiśniewski, MSc, has achieved original and valuable scientific results in the field of knowledge management in the area of modern IT research. Therefore, I recommend for the reviewed dissertation to be transferred to the next stages of the doctoral process.

Moreover, taking into account what I have presented above and the requirements imposed by Article 187 of the A*ct of 20 July 2018 - The Law on Higher Education and Science* (with amendments)[1], my evaluation of the dissertation according to the three basic criteria is the following:

A. Does the dissertation present an original solution to a scientific problem? (the selected option is marked with **X**)

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

[1] http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276

B. After reading the dissertation, would you agree that the candidate has general theoretical knowledge and understanding of the discipline of **Information and Communication Technology**, and particularly the area of **Knowledge Management**?

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

C. Does the dissertation support the claim that the candidate is able to conduct scientific work?

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

Finally, taking into account the scientific quality of the dissertation and the outstanding publication achievements of the candidate in the area of **Information and Communication Technology**, strictly connected with the subject of the dissertation, **I recommend to distinguish the dissertation**.

Signature