OPINION

on

the dissertation of **Viktor Ernestov Senderov** on
"The Open Biodiversity Knowledge Management System in Scholarly Publishing"
for acquiring PhD degree in Informatics and Computer Sciences

by

Assoc. Prof. Dr. **Gennady Pavlovich Agre**
Institute of Information and Communication Technologies at Bulgarian Academy of Sciences

According to the order № 86/30.04.2019 of the Director of the Institute of Information and Communication Technologies - Bulgarian Academy of Sciences (IICT-BAS), I have been approved as a member of the Scientific Jury in connection with the procedure for acquiring educational and scientific degree "doctor" in the doctoral program "Informatics" in professional field 4.6 "Informatics and Computer Sciences" by Viktor Ernestov Senderov with a dissertation entitled "The Open Biodiversity Knowledge Management System in Scholarly Publishing" with the Scientific Supervisor Prof. Lubomir Penev (Pensoft Co.) and Scientific Adviser Assoc. Prof. Kiril Simov.(IICT-BAS)

As a member of the jury I received:

1. A copy of the dissertation for acquiring educational and scientific degree "doctor" in professional field 4.6 "Informatics and Computer Sciences" (in English)
2. An abstract of the dissertation (in Bulgarian)
3. Professional Curriculum Vitae (in English)
4. Copies of the author's publications related to the dissertation.

The dissertation consists of 113 pages, structured in: Introduction, 8 chapters, Conclusion, List of Abbreviation and Bibliography.

1. Actuality of the dissertation topic and relevance of its objectives and tasks

The dissertation is related to an actual topic from the field of semantic technologies – creation and management of the link open data (LOD) with application to a concrete problem domain – scholarly publishing of biodiversity data. The main objective is to propose a formal semantic model of the domain that can be applied for solving the task of developing an open system for managing the LOD The stated objectives and tasks are of a great importance for the domain under consideration since they allow systematizing the domain knowledge by ontological description of concepts, unification of terms as well as constructing a digital store of the linked open domain data.

2. Familiarizing with the state-of-the-art in the field

The PhD candidate showed a very good knowledge of the state-of–the art in the field of semantic technologies related to the management of the linked open data as well as of the most important specifics of the problem domain – scholarly publishing of biodiversity data. The bibliography consists of 163 references from which only 3 are written in Bulgarian. .

3. Methodology of the study

The research methodology chosen by the author evolves from 6 tasks stated in the dissertation that aim at achieving the dissertation objectives. The methodology includes the motivated

choice of technology for creating a biodiversity database, the selection of main information resources describing the problem domain, the choice of program environment for implementing the system, as well as a decision to use the open-source approach in order to provide the ability to check and reuse the achieved results.

4. Short analytical resume of the dissertation

The Introduction begins with the motivation of the study and continues with an analytical review of the related works in the problem domain under consideration. After analyzing the existing approaches and the achieved results, .the author formulates the objectives of the dissertation, the tasks to be solved in order to achieve the objectives, as well as the research methodology needed for solving the tasks. The first chapter describes the system architecture that consists of three main blocks – the semantic graph database (the ontology and LOD), software packages for managing the database and the web portal for accessing the database. The second chapter (which, in my opinion, is the central chapter of the dissertation) describes the created formal model of the problem domain - OpenBiodiv-O ontology. This ontology provides a conceptual model for scholar publishing in the domain of biodiversity by introducing specific domain classes and relations as well as providing semantic modelling of the biological nomenclature and taxonomic concepts. The third chapter is intended to the linked open data created by the system by means of acquiring data about the biodiversity from scholarly articles published by Pensoft and Plazi publishing houses based on the proposed ontology. Some results of the tests measuring the behaviour and processing time of different amounts of data are also presented. The forth chapter describes the functionality of RDF4R – a software package developed by the author for working with RDF. The comparison of the RDF4R with other packages with similar functionality is also presented. Chapter 5 considers two automated workflows for exchange of biodiversity data developed as a part of OpenBiodiv system: the automatic import of specimen records into manuscripts, and the automatic generation of data paper manuscripts from Ecological Metadata Language metadata. Chapter 6 discusses the website that was prepared to serve on top of OpenBiodiv-LOD and its applications. Chapters 7 and 8 present a part of the system source code listings as well as some details of the webinar accompanying Chapter 5. In my opinion these chapters are not significant for the dissertation and should be presented as annexes. The conclusion summarizes the results of the dissertation, presents the author's contributions, his publication activity and information about his participation in various projects and conferences.

5. Contributions of the dissertation

The author states three contributions – one scientific and two – applied-scientific. The scientific contribution is the creation of an ontology and a formal model of the domain of biodiversity knowledge publication. The applied-scientific contributions are the analysis of information sources and the creation of OpenBiodiv-LOD. I think that these main contributions of the author are stated correctly.

6. Assessment of publications related to the dissertation

The author presented 8 publications related to the dissertation. I think that these publications properly reproduce the content of the dissertation and represent the original contributions of the dissertation's author.

According to the "Rules for specific requirements for acquiring scientific degrees and for holding academic positions in the Institute of Information and Communication Technologies – BAS", a PhD candidate should have at least 30 points collected from his publications related to the dissertation. Only publications indexed in some well-known reference databases (like Web of Science (WoS), Scopus, etc.) have to be evaluated. The evaluation points are counted according to the Rules approved by the Decision dated to 20.05.2019 of the 48th meeting of the General Assembly of Bulgarian Academy of Sciences.

Among the eight publications presented by the PhD candidate only three are indexed in WoS and/or Scopus (publications №8, №7 и №3). Publications №8, №7 were published in journals having Impact Factor in WoS (both belongs to the third quartile – Q3), while publication №3 was published in a journal having SJR rank in Scopus. That is why, according to the accepted evaluation procedure these three publications accumulate $2x30 + 1x20 = 80$ points, which is significantly exceeds both national and IICT-BAS minimal requirements for acquiring PhD degree.

In addition, it should be noted that up to now, the mentioned above three publications are have been cited 15 times (excluding self citations) in Scopus.

7. Personal contribution of the PhD candidate

Although the PhD candidate has no any one-man publications related to the dissertation's topic up to now, he is pointed as the first author of publication № 8 having the highest IF. This has shown that the personal contribution of the PhD candidate is significant.

The analysis of the dissertations and all publication provided by the PhD candidate has allowed me to conclude that there is no any plagiarism in the dissertation.

8. Abstract of the dissertation

The abstract of the dissertation correctly presents the objectives, tasks and results of the dissertation.

9. Remarks

I have some remarks to the dissertation structure – in my opinion Chapters 7 and 8 should be presented as Annexes rather than chapters. I have also some concerns to Bulgarian translation of certain English terms, for example "графични" бази данни (graph database), „буквални" стойности (literal values) etc. However, these remarks do not decrease the scientific value of the dissertation.

10. Conclusion

All mentioned above allow me to conclude that all requirements of the Law on the Development of the Academic Staff in the Republic of Bulgaria, the Rules for applying the Low, as well as of the Rules for specific requirements for acquiring scientific degrees and for holding academic positions in the Institute of Information and Communication Technologies – BAS are completely satisfied. I am confident that the dissertation of Viktor Ernestov Senderov has all characteristics required by a PhD Thesis. The results presented in the

dissertation are significant and contribute to wider applications of the semantic technologies to various problem domains and especially to the domain of scholarly publishing of biodiversity data.

All this facts give me a reason for a positive evaluation of the dissertation and I strongly recommend the honoured Scientific Jury to award the educational and scientific degree "doctor" in professional field 4.6 "Informatics and Computer Sciences" to Viktor Ernestov Senderov

June 10, 2019

Soifa