OPINION

of Prof. Ludmila Petrova Dimitrova-Rashkova, Ph.D.,
Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences,
regarding Alexander Nikolaev Popov's PhD dissertation titled "**Modeling Lexical Knowledge
for Natural Language Processing**" in domain 4. Science, Mathematics and Informatics;
Professional Area: 4.6 Informatics and Computer Science, PhD Program (Research specialty)
Informatics

I present this opinion as a member of the Scientific Jury appointed by Order № 127/12.07.2018
of the Director of IICT-BAS.

This opinion is prepared in accordance with the requirements of the Decree for the Development
of Academic Staff in the Republic of Bulgaria (Publ. 3 April 2018) and its Implementing
Regulations (Publ. 6 July 2018) at the time the procedure was launched.

1) Methodology of the study

The study methodology used by the PhD candidate is clearly described. The purpose of the
study, the tasks and the scope of the study are clearly outlined. The PhD candidate demonstrates
literacy on the research methodology and demonstrates capacity to use in practice, to conduct his
own research with subsequent analysis and summary of the results obtained.

2) Knowledge the state of the art

The PhD candidate demonstrates an understanding of the studied problem, the systematic and
creative approach, and uses modern methods and techniques for researching and solving the
assigned tasks. The dissertation is written clearly; the presentation is concise and consistent; the
conclusions are justified.

3) General description of the dissertation work

The PhD thesis is written in English and consists of 147 pages, divided into Introduction, 7
chapters (2 – 8), Summary and Outlook (Chapter 9), Declaration of Originality, Bibliography (a
list of references, incl. 14 publications by the PhD candidate – author of 4, co-author of 10, incl.
one, where he is listed as first), 23 tables, 8 figures, and list of abbreviations.

The thesis focuses on models for representing lexical knowledge for the purposes of natural
language processing (NLP). Lexical knowledge is considered broadly to include traditional
aspects of the meaning of lexical semantics, supplemented with morpho-syntactic information,
patterns of collocation, knowledge of the world, etc. The presented work studies ways of
presenting lexical knowledge with symbolic, statistical or hybrid means in order to apply them to
specific tasks of NLP. Two deep learning systems for performing knowledge-based word sense
disambiguation and context embedding are developed. The algorithms used to solve the task rely
not so much on statistical knowledge learned directly from data but on the information encoded
in a graph model of the lexicon. The advantage of knowledge-based method for word sense
disambiguation is that algorithms have full coverage over the meaning of the lexicon used. The

work is initiated in connection with a project to construct a Bulgarian version of the WordNet lexical resource and for its use in performing word sense disambiguation in running text. The work is also used in a project to build machine translation (MT) systems. Since the MT systems under development cover both the Bulgarian-to-English and English-to-Bulgarian directions of translation, the research was focused on both languages in parallel.

4) Analysis of the scientific and applied achievements of the PhD Thesis

Firstly, in Chapter 2, definitions of the work problems are provided. Each task serves as a testing ground for the hypothesis, explored in the dissertation. The problems are firstly described in a non-formal way by means of examples for illustration, and then formal definitions are given. Part-of-Speech tagging, Word sense disambiguation, Word similarity and relatedness are widely discussed. Next, some related models for automatic lexical analysis in NLP, relevant to the thesis, are presented in Chapter 3. The overview is based on the set of publications, cited in the Bibliography, including six publications by the PHD candidate. The main results are described in details in Chapters 4 – 8 of the dissertation. Recurrent neural networks for Part-of-Speech tagging are described in Chapter 4. The results of experiments with three models of neural networks that differ in the type and number of input words are compared and discussed. Chapter 5 studies knowledge-based methods for word sense disambiguation that rely on a graph model of the lexicon. Some of the most popular WSD variants are grouped under graph-based methods. In the chapter 6 the author introduces some work on generating distributed representations of lexical units: words, lemmas and meanings. The chapter 7 presents supervised neural network architectures for WSD. The Chapter 8 (final) presents original research aiming to explore the possibilities of doing WSD in parallel with other NLP tasks.

5) Assessment of the contributions

The PhD candidate's contributions in the dissertation and in the autoreferat correctly reflect the achieved results. The presented in the dissertation results are used in two European Commission's FP7 projects: *QTLeap:Quality Translation by Deep Language Engineering Approaches* (this project has funded some of the thesis-related work in the period 2014-2016) and *EUCases - EUropean and National CASE Law and Legislation Linked in Open Data Stack* (the project has funded some of the thesis-related work in the period 2014-2015), and in the project *Deep Models of Semantic Knowledge (DemoSem)*, funded by the Bulgarian National Science Fund (it has supported some of the thesis-related work in 2017 and partially in 2018). Besides the aforementioned 3 projects, the PhD candidate has made a set of presentations at international conferences, workshops, and national events.

6) Evaluation of the PhD candidate's Publications

The results of the dissertation are published in 14 publications. All original results described in the thesis have been presented at international conferences and workshops and subsequently published in conference proceedings. There is one publication from 2018 in a journal with SJR (0.204 for 2017).

In accordance with Databases Scopus and Web-of-Science, 5 citations in articles in Scopus of 4 author's papers (excluding self- and co-author citations), are matched. 11 citations of 14 publications were found.

7) Evaluation of the PhD candidate's autoreferat

The dissertation's autoreferat consists of 46 pages. The relevance of the topic under consideration is justified, the objectives are formulated, and the research tasks are set. The main and most important results obtained in the dissertation are presented. So are the contributions and publications of the author, and his participation giving talks at relevant international conferences and seminars at home and abroad. I believe that the autoreferat, as well as the presented summary, correctly reflect the content of the dissertation, results and scientific contributions.

8) Critical remarks

I have no critical remarks. I do not know the PhD candidate personally, but his work reflects an erudite and promising young scientist in the field of computer science.

9) Conclusion

The presented dissertation by Alexander Nikolayev Popov on "Modeling Lexical Knowledge for Natural Language Processing" meets all the requirements of the Decree for the Development of Academic Staff in the Republic of Bulgaria (Publ. 3 April 2018) and it's Implementing Regulations (Publ. 6 July 2018). The candidate has demonstrated competence, academic maturity, and ability for independent research. Therefore, I firmly recommend to the respected academic jury to award Alexander Nikolaev Popov the educational and academic degree "doctor" in the field of higher education 4. Natural Sciences, Mathematics and Informatics professional sub-field 4.6. Informatics and Computer Science, Doctoral Program of Informatics.

Sofia, 15.10.2016