

STATEMENT

from

Assoc. Prof. Dr. **Dimitar Ivanov Vassilev**,
Department of Computer Informatics, Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski "

concerning:

the defense of the PhD thesis of Iliyan Nedkov Mihailov,
full-time doctoral student at the Department of Information Technology (IT),
Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski "

on the topic:

**"Intelligent information systems in bioinformatics: semantic
integration, analysis and classification of bio-medical data "**

for

awarding with the educational and scientific degree "**Doctor**"
in the field of higher education 4. Natural sciences, mathematics and informatics,
professional field 4.6. Informatics and Computer Science, doctoral program "Information
Technology - Bio- and Medical Informatics",
Faculty of Mathematics and Informatics, Department of Information Technology (IT),
Sofia University "St. Kliment Ohridski ".

By an ordinance of the Rector of the Sofia University "St. Kliment Ohridski" RD 38-255 / 02.06.2021 I was appointed as a member of the scientific jury to prepare a statement concerning the submitted PhD thesis of Iliyan Nedkov Mihailov for the official defense of the thesis and awarding with the educational and scientific degree "Doctor", in the domain 4 of higher education: Natural sciences, Mathematics and Informatics, professional subdomain 4.6. Informatics and Computer Science, doctoral program "Information Technology - Bio- and Medical Informatics", Department of Information Technology, Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski "

The statement prepared by me complies with all the requirements of: the Law on the Development of the Academic Staff in the Republic of Bulgaria (LASRB) and the relevant regulations for its implementation (RRILASRB), as well as the regulations on the structure and activity of Sofia University and the relevant regulations for acquiring scientific degrees and holding academic positions at Sofia University and respectively at the Faculty of Mathematics and Informatics of Sofia University "St. Kliment Ohridski ".

Research, tuition and applied activities of the doctoral student

Iliyan Nedkov Mihailov was enrolled as a full-time doctoral student at the Department of Information Technology by an ordinance of the Rector of Sofia University, RD 20-359 / 08.02.2018 and was expelled with the right to defend the thesis by an ordinance of the Rector of Sofia University RD 20-376 / 12 / 02.2021. The supervisor of the doctoral student is Assoc. Prof. Dr. Dimitar Vassilev from the Department of Computer Informatics, Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski ". The dissertation work, together with the abstract written in Bulgarian and English are presented according to the requirements in time, as well as the accompanying documents of the thesis.

The research and scientific-applied work of doctoral student Iliyan Mihailov are in the domain of bioinformatics, comprising various aspects of informatics, methods of machine learning, statistics. The listed areas represent a new, integrative and interdisciplinary field of research - bioinformatics, which requires in-depth knowledge and skills in the above-mentioned knowledge domains. During his doctoral studies, Ilian Mihailov led exercises in the courses "Linux and programming languages in bioinformatics" and "Database applications in bioinformatics" in the Master's program Bio- and medical Informatics at the Faculty of Mathematics and Informatics at Sofia University 'St. Kliment Ohridski'.

Submitted materials

Ilian Mihailov presented a well-written dissertation on 189 pages, with 49 figures, 21 tables, a list of references (151 titles), a glossary of terms, and a list of abbreviations. The paper is divided into 5 chapters: Chapter 1 - introduction to the problem area, Chapter 2 - literature review and comments on the theoretical justification, Chapter 3 - methodological aspects of the tasks in the dissertation, Chapter 4 - results, discussion and software implementation, Chapter 5 - contributions, perspectives, conclusions

The doctoral student also presented 9 publications and 1 patent in connection with the topic of the dissertation, which significantly exceeds the required minimum and are an excellent certificate for the work of Iliyan Mihailov. All publications are in journals and proceedings referred by Scopus and Web of Science, with impact factor, there have the significant number of 32 citations with h-index = 3.

Review on the research and applied achievements of the thesis.

The main subject of the dissertation are problems of semantic integration of bio-medical data, as well as the related various innovative models for analysis of these data. In general, the topic related to data integration occupies a very important place in modern computer science theory and application, which is especially evident in various aspects of the digitalization of bio-medical research and practices. At present, digitalization in biology and medicine is a major priority of modern bioinformatics, which in turn leads to a significant improvement in diagnosis and related follow-up therapy, especially in the context of preventive and personalized medicine. Separately, biology is now seen as both an experimental and a digitally related domain of human knowledge, especially after the advent of genome sequencing technologies. Of particular relevance are the publications of Ilian Mihailov, related to models for semantic integration of bio-medical data on cancer studies such as laboratory, patient, clinical, and different *-omix* data. The origin of these *-omix* data is the main end product of the so-called parallel sequencing or as it sounds the most widely used name - *next generation sequencing (NGS)*, a new and widespread technology that has created great prerequisites for increasing the relevance of bioinformatics as big data generation, new methods, algorithms, procedures and a huge variety of software implementations. This extremely intensive development of bioinformatics is related to new methods of analysis, storage, visualization, integration and protection of such data, as well as the use of new areas of computer science, applied mathematics, methods of artificial intelligence and machine learning, graph theory, discrete mathematics, logic, object-oriented programming, etc. In this sense, the significance of Ilian Mihailov's dissertation can be highly appreciated.

The main results and achievements of the dissertation are presented in nine papers and one patent, where the most significant ones concern not only the semantic integration of data on cancer, but also the integration and various methods of data analysis for

antimicrobial resistance data, developing model and software implementation of an advisory system for generating recommendations for diet in patients with diabetes, as well as an innovative model for compressing *-omix* data. The submitted dissertation work is generally distinguished by presenting the developed innovative methods and models with obvious levels of complexity.

A special place in the dissertation is occupied by the software implementation of the developed and presented models, aimed mainly at semantic integration of large arrays of bio-medical data. The various methods presented in the work also aim to create a generalized model of a platform for intelligent integration of bio-medical data, which has a universal solution for working with data with different structures and formats. In this regard, doctoral student Iliyan Mihailov developed and used an innovative and non-traditional approach based on the NoSQL databases. NoSQL databases are a relatively new domain, which has recently been increasingly used in the semantic integration, storage, protection, management of large arrays of biological and medical data with different structures and formats. Of course, this important methodological and applied part of the dissertation concerning the semantic data integration of bio-medical data, which is an independent proposition, development and implementation of Iliyan Mihailov, despite the large and detailed volume (as a text) occupied in the work, could be presented more clearly and concisely, which does not reduce the value of the thesis.

The sections, including a review of the literature and the introduction to the problem area of the thesis, as well as a critical analysis of the theoretical formulation of the dissertation are written in detail. Here the author's ideas for creating a generalized approach for semantic integration of bio-medical data are manifested. Despite their markedly large volume, these chapters provide clear, albeit sometimes unnecessarily detailed, information on the setting of the doctoral student's work.

The chapter dedicated to the setting and defining of goals and tasks, is written definitely clearly, but with some insignificant inaccuracies such as the existence of one goal and many tasks, which is slightly different from the title of the work. Criticism here could be attributed to both the doctoral student and the supervisor and the preliminary discussion of the work.

Chapter 4, in my opinion, is the most important part of the dissertation, presenting the methodological ideas of the doctoral student. This chapter highlights the originality of Ilian Mihailov's work, as well as the development of applications that have not only methodological but also definitely applied sense. The main criticism here is that this chapter could be written more consistently. The main reason for such a volume of Chapter 4 is the use of all publications related to the dissertation. Another potential remark is related to some extent to the the narrative appearance of the chapter and the lack of more analytical formulas that are reflected in the articles.

Chapter 5 is devoted to the presentation of the software implementation of the developed models. This is the chapter that perfectly presents the in-depth knowledge of doctoral student Ilian Mihailov related to programming languages and technologies, software architectures, applied orientation. This chapter emphasizes once again the innovative ideas of the doctoral student and their implementation in the dissertation. I have criticisms of the text as volume and consistency, which in no way reduce the value of the work of Ilian Mihailov.

The contributions of the dissertation are well presented, and they could be clearer. The prospects for future work are interesting, and would be related in several projects in Iliyan Mihailov participates, but these projects are not reflected in the dissertation and the abstract. The presented conclusion is clearly written, has a certain research and development profile, necessary for the successful defense of a dissertation.

Personal impressions

As a supervisor of the dissertation of Ilian Mihailov, whom I have known for almost 4 years, I can honestly give the highest evaluation of his work, both in his research and teaching, and presentation of the work at various scientific forums. Iliyan is a person with innovative thinking, predisposed to work in a team, who loves discussions and does not contradict criticism. In addition to the social aspect, I would like to emphasize that during his doctoral years Iliyan managed to create a wonderful family that continues to grow.

Conclusion

The remarks and criticisms presented by me and concerning the dissertation are related to the design of the work, its structure, as well as the text itself. In the text of the thesis can be found both syntactic and stylistic errors and inaccuracies, as well as foreign words. Of course, this can be improved, but without undue rigor it can be said that in Ilian Mihailov's dissertation, there is a lot of new knowledge, a lot has been done and published and cited, which compensates for my criticisms. Undoubtedly, the PhD student Iliyan Nedkov Mihailov presented an excellent work based on very well formulated theses, developed and implemented in the dissertation.

I state with the an absolute conviction for awarding Iliyan Nedkov Mihailov with the scientific and educational degree "Doctor" in the field of higher education 4. Natural sciences, mathematics and informatics, professional field 4.6. Informatics and Computer Science, doctoral program "Information Technology - Bio- and Medical Informatics"

08/22/2021

Sofia

Prepared the statement:

Assoc. Prof. Dr. Dimitar Ivanov Vassilev