OPINION

on the thesis of Stoyan Milkov Mihov "Finite-State Automata, Transducers and Bimachines: Algorithmic Constructions and Implementations"

presented for awarding of the scientific degree "Doctor of Sciences" in

Professional field 4.6. Informatics and Computer Science, Subfield 01.01.12. Informatics

written by Prof. Dr.Sc. Galia Angelova, Institute of Information and Communication Technologies (IICT), Bulgarian Academy of Sciences (BAS)

The thesis of Dr. Mihov considers one of the fundamental structures for discrete representation and processing of information – the finite-state machines. The author presents the theory of finite-state automata, transducers and bimachines from an algebraic point of view and describes the practical tools he developed for applying finite-state machines in real computational problems: the programming language C(M) and a software library of 45 C(M) programs for automata constructions and applications in practice. Given the importance of the tasks of approximate search as well as text and speech processing, which Assoc. Prof. Mihov has been solving for over 20 years by applying the formalism of finite-state automata, we should assess the topic of the dissertation as very relevant and challenging, especially in the arriving big data era. My personal opinion is that the need to find effective practical solutions to modern large-scale natural language processing challenges has given an incentive to Dr. Mihov to consider the theory of finite-state machines in an original and application-tailored manner.

According to the requirements of the Regulations for awarding the scientific degree "Doctor of Science" at IICT-BAS, applicants have to submit a thesis accompanied by evidence that the results are published in scientific papers which bring scores covering the minimal requirements of NACID: in the group of indicators "Г" at least 100 points and in the group of indicators "Д" - also at least 100 points. These scientific articles should not be used in previous competitions for awarding scientific degrees or employments at academic positions. The reference for fulfillment of the minimum requirements, presented by Dr. Mihov, shows that the attached 7 publications on the dissertation bring 142 points in indicators "Г", and the 53 citations in publications indexed by Scopus carry 318 points for indicators "Д". Thus, the points corresponding to the presented papers and citations of Assoc. Prof. Mihov easily exceed the requirements of the IICT-BAS Regulations, which are higher than the minimum national requirements of NACID. Thus, the formal requirements of the IICT Regulations for obtaining the degree of "Doctor of Science" are met and even significantly exceeded with the number of citations.

Thesis content

The thesis text contains 226 pages in English and consists of an introduction, eight chapters, a conclusion and bibliography with 48 titles. The main body of the dissertation covers the first eight chapters of the monograph *Mihov, S. and Schulz, K. (2019). Finite-State Techniques: Automata, Transducers and Bimachines, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press*. A letter sent by Prof. Klaus Schultz from the Ludwig Maximilian University in Munich is presented, confirming Dr. Mihov's authorship of the first eight chapters of the jointly co-authored monograph.

Chapter 1 presents some basic mathematical concepts including the notion of monoid. Chapter 2 introduces the monoidal finite-state automata as a generalization of the classical automata with a finite number of states and studies the properties of the monoidal languages. Chapter 3 considers the classical finite-state automata and regular languages from the perspective of the monoidal automata, with focus on the basic properties (determinization, minimization).

In chapter 4 the basic features of the monoidal multi-tape automata are presented. This chapter defines the monoidal finite-state transducers as monoidal 2-tape automata. A procedure for deciding the functionality of a given classical finite-state transducer is presented.

Chapter 5 seems to be central in the thesis. It introduces the deterministic finite-state transducers, more especially the ones that are deterministic on the input tape (called sequential or subsequential transducers). Several statements are proved that allow to construct a determinization procedure for functional transducers with the bounded variation property. An effective procedure for deciding the bounded variation property for a classical real-time transducer is constructed. In addition, an effective way for testing the termination of the inductive determinization construction is presented. It is shown how to minimize classical subsequential finite-state transducers, and it is shown that essentially the same technique can be used for other target monoids that satisfy some additional conditions. Dr. Mihov's expertise in automata theory is clearly demonstrated in this chapter as he proposes alternatives to Mehryar Mohri' ideas (M. Mohri is Professor in Computer science at the Courant Institute of Mathematical Sciences, New York University). The suggested "partial" determinization of the initial part of an automaton is interesting too (remark 5.5.16). It shows that the search for effective practical applications is an important motivation for the development of Dr. Mihov's theoretical considerations.

Chapter 6 introduces and studies the bimachines – deterministic finite-state devices that represent exactly the class of all regular string functions. The construction of a pseudo-minimized bimachine is shown as well as the direct composition of classical bimachines.

Chapter 7 presents the programming language C(M), an original functional language built over declarative representations of formal constructions in a set-theoretical language. Tasks are solved by posing formal descriptions of the desired kind of mathematical objects the user wants to obtain as results. The C(M) compiler, which is a piece of software with open access, constructs executable C-code using the input formal descriptions. Obviously, the C(M) language is very suitable for mathematicians who want to make experiments with processing of discrete structures. Unfortunately, the thesis lacks details of how many people use C(M) and whether they easily learn the ways to formulate expressions that lead to more efficient performance of the operations.

Chapter 8 describes a software library of 45 C(M) programs. As a toolkit for construction of automata and their applications, it shows that Dr. Mihov's group actively uses C(M) in solving various practical tasks.

Some negative remarks can be made concerning the document package accompanying the thesis (e.g. one lacks any hint where each result is published originally), but they do not reduce the scientific value of the dissertation and the presented results.

Original achievements of the thesis

Assoc. Prof. Stoyan Mihov is a world-famous specialist in approximate search using finite-state machines. His expertise has been accumulated for more than 20 years. In 2000, in his doctoral

dissertation, he proposed an algorithm for directly constructing a minimal acyclic finite-state automaton from a dictionary of words given in a lexicographically ordered list, and published this result in the Journal of Computational Linguistics as an article cited to this day. His research on different types of automata, the constant pursuit of combining theoretical perfection with practical applications allow him to offer today a comprehensive exposition of the theory of finite-state automata from an abstract algebraic point of view, designed to introduce and study computationally efficient structures. I accept the scientific contributions of Dr. Mihov as formulated on page 211 of the thesis (a procedure for deciding the bounded variation property of a finite-state transducers has been developed, which can be integrated in the sequentialization construction; a new construction with polynomial complexity for canonization of a subsequential transducer is presented; a new construction has been developed for constructing a bimachine from a finite-state transducer; a construction together with correctness proof was obtained for direct composition of bimachines).

I highly appreciate the development of the C(M) language and the library of 45 programs for constructions of finite-state automata, transducers and bimachines, realization of operations as well as construction of applications in practical tasks. The C(M) language and the library prove that the software toolkit is actively used to solve real problems and that C(M) is something like a "lingua franca" (common language of communication) for the group of experts working on Dr. Mihov's projects.

It is also worth noting that the co-authorship in the presented monograph and publications does not diminish the importance of the achievements of Assoc. Prof. Mihov, but rather emphasizes the importance of his position as a valuable and sought-after collaborator, partner and leader.

Other Achievements

Dr. Mihov is the leader of the team that created the Wallbreaker program, which won a world competition for approximate search in 2013. (https://www2.informatik.hu-berlin.de/~leser/searchjoincompetition2013/Results.html). In the last two years, under his leadership, with a lot of work, patience and persistence, a prototype of a speech-to-text system for Bulgarian speech was created.

Conclusion

I believe that Assoc. Prof. Stoyan Mihov is a rare example of a talented mathematician who is deeply interested in creating real systems and is ready to work almost as a professional programmer for their construction. The quantity and quality of his research papers and citations prove that he is recognized by the international scientific community as a leading expert with in-depth knowledge in automata theory. The projects funded by the H-Tech company prove Dr. Mihov's ability to implement the generated ideas and create software for industrial applications. **I strongly support the award of the scientific degree "Doctor of Sciences" to Assoc. Prof. Stoyan Mihov and I propose to all members of the Scientific Jury to vote unanimously in support of such a decision.**

27 май 2020 г.

Sofia

Member of the Scientific Jury:

Prof. DrSc Galia Angelova