REVIEW

by Prof. Dr. Tinko Velichkov Tinchev

on a dissertation for awarding the scientific degree „Doctor of Sciences“ area of higher education: 4. Natural Sciences, Mathematics and Informatics, professional field 4.6 Informatics and Computer Sciences

Author of the dissertation: Assoc. Prof. Dr. Stoyan Milkov Mihov
Topic: Finite-State Automata, Transducers and Bimachines: Algorithmic Constructions and Implementations

For the submitted documents

The dissertation presented by Assoc. Prof. Mihov is in English in a volume of 225 pages and is structured in 8 chapters (210 pages), content, introduction, conclusion and bibliography containing 48 cited sources. The conclusion includes a declaration of originality and contributions - scientific and scientific-applied. The dissertation is based on a joint monograph with Klaus Schultz, published by Cambridge University Press in the series Cambridge Tracts in Theoretical Computer Science in 2019, a chapter from a book in the series Trends in Linguistics-Studies and Monographs and 11 articles. Three of these articles are in reputable journals with an impact factor - two in Computational Linguistics (Q1) and one in Theoretical Computer Science (Q3); another seven are in journals and collections with the SJR factor. According to Scopus, they are cited 227 times (excluding selfcitations). All citations of scientific works of Assoc. Prof. Mihov according to Scopus are over 300 (excluding selfcitations). Therefore, the scientometric indicators of Assoc. Prof. Mihov far exceed the minimum national requirements. An abstract in Bulgarian (58 pages) and an abstract in English (57 pages) are presented, which accurately reflect the dissertation. All other requirements of the Law on the Development of the Academics in the Republic of Bulgaria (LDARB) and the regulations for its implementation are also presented. I will not list all these documents, but I will only mention one of them - a declaration by Prof. Klaus Schultz that in each of the chapters of their joint monograph the contribution of Dr. Stoyan Mihov is greater than his, as the chapters relating to to the C (M) language and the C (M) programs included in the monograph are entirely the contribution of Dr. Stoyan Mihov.

For the area of the dissertation

In the early 1970s, it was widely believed that the theory of finite automata was a mathematically mature and largely complete theory with the theorems of Kleene, Rabin-Scott, Myhilll-Nerode, McNotton, Salomaa, and others.; on the agenda are more expressive models such as pushdown automata, context-free grammars, etc.

Now, as we well know, the truth has turned out to be different - the rapid growth of computing systems has caused a hunger for efficient algorithms and finite structures for the knowledge representation and processing of knowledge. In particular, a number of variants of the finite state transducers have been developed, often due to specific applications or for the technical convenience of the explication. With this, the scales seemed to lean to the side of algebra and logic. The proposed dissertation offers a reasonable general algebraic approach with a clear algorithmic colour. He is in an actively developed field, both as mathematics for theoretical computer science and as applications in computational linguistics and artificial intelligence.

Review and evaluation of the dissertation content

The dissertation is a single comprehensive scientific and applied research, following a clear and appropriate methodology and contains all the necessary concepts for its tracking, and, roughly speaking, has two naturally related and interdependent parts.
The first 6 chapters contain the mathematical theory of the considered transducers with a finite number of states - finite state machines, different types of transducers and bimachines, here are the scientific theoretical contributions of Assoc. Prof. Mihov: a number of effective algorithmic constructions are given, for which correctness is proved and complexity is investigated. In the second part, they are implemented. This part contains significant scientific and applied contributions of Assoc. Prof. Mihov - a target programming language C (M) (Chapter 7), which is convenient for mathematicians, has been developed. In Chapter 8 the algorithms from the first part are implemented using C(M), as well as four examples of applications of the developed technique to specific practically important non-trivial tasks. Here we must immediately say that for this programming language Assoc. Prof. Mihov has created a compiler that translates programs into efficient C code, which can be executed after compilation. Thus, the dissertation contains correct executable code for a number of algorithms.

Chapter 1 is introductory and introduces the basic concepts necessary for further exposition - words, n-tuples of words, monoidal words, languages, monoidal languages and the classical operations with words and languages. At the same time, the chosen monoidal approach is motivated in order to generalize, naturally encompassing classical concepts such as words, n-tuples of words, languages, automata and n-tape automata.

In Chapter 2 monoidal finite-state automata are defined and studied. The closedness of the class of monoidal automaton languages under monoid homomorphisms and regular operations union, monoidal product and monoidal Kleene-star is shown. The notions monoidal regular language and monoidal regular expression are defined. Kleene's classical theorem for monoid regular languages and monoid regular expressions is proved. This chapter clearly shows the main purpose of the chosen generalized approach - not just to constructively prove the existence of an object with a given property, but to give an efficient direct construction of an object that is proven to have the desired property. Usually the elegant and economical from the point of view of the explanation proof is in a remark after the proof that the constructed object has the desired property. The last paragraph of Chapter 2 shows the removal of unaccessible states of a monoidal finite-state automaton and two ways to remove e-transitions.

In Chapter 3, the focus is on determinization and minimization, mainly studying the classical finite-state automata (the monoid is of the words in the finite alphabet). Efficient constructions are shown, proving the closure of regular languages with respect to set-theoretical union, intersection and set-theoretical difference, as well as inverse of words. The reason for the restriction to classical finite automata is shown in Section 7 and is rooted in the fact that set-theoretical complement of a monoidal regular language is not always regular language. A reasonable option are the notions pseudo-deterministic and pseudo-minimal monoidal finite-state automata - the corresponding free companion to be deterministic, respectively minimal. It is shown how they are obtained on the basis of the constructions developed in the first 5 paragraphs for determinization and minimization of classical finite-state automata. It is worth noting the in-depth study of the Myhill-Nerode equivalence relation in sections 4 and 5, which further became a powerful tool for studying similar minimization problems of various types of tranducers. A direct application of the minimization technique presented is to the coloured deterministic finite-state automata introduced in paragraph 6, which are used essentially in the pseudo-minimization of bimachines in Chapter 6.

Chapter 4 examines multi-tape automata as a special kind of monoidal automata. Constructions are given showing the class of the languages recognized by monoidal multi-tape automata is closed with respect to Cartesian products and projections, and in the case of two-tape automata also about the inverse operation. The construction giving the relational composition for two-tape letter finite-state automata is described, which gives a construction for the composition of classical n-tape automata. An essential part of the chapter is devoted to the monoidal finite-state transducers and their functionality in the case when the second tape is also a free monoid (classical finite-state transducers). The aim is to obtain an efficient method for recognizing the property functionality of classical finite-state transducers. The presented construction is based on a non-trivial technique using a number of special kind transducers and the so-called function of the admissible balance. The obtained method is necessary for the study of the determinization of finite-state transducers in the next chapter, due to Schützenberger's theorem.

Chapter 5 is devoted to deterministic finite-state transducers and is particularly interesting because of the importance in terms of applications of subsequential transducers (functional and deterministic on the input tape) in computational linguistics, for example, for speech recognition and text processing. Section 2 essentially describes the construction of Roche & Schabes for determinization: an iterative procedure which, starting from a classical real-time (no ε-transitions on the second tape) finite-state trimmed transducer T, consecutively constructs finite-state transducers as it finish after finite number of steps, the result is a classical subsequential transducer T ', equivalent to T. Moreover, if the initial converter T has a bounded variation, then the procedure ends after a finite number of steps. Hence immediately follows the characterization of the regular functions between words represented by classical subsequetial transducers: they are precisely those regular functions which have a bounded variation. In Section 3, Assoc. Prof. Mihov finds a suitable elegant and far from obvious form of observed by Béal et al. relationship between the property of T to be of bounded variation and the validity of an elementary inequality related to the number of states of T, the longest transition on the second tape of T, and the admissible state progress of the squared automaton associated with T. This allows it to embed in the determinization algorithm the recognition of the property in question, with which the algorithm acquires integrity. Sections 4 and 5 examine the minimization of finite-state subsequential transducers based on an analogue of the Myhill-Nerode equivalence relation and give an appropriate construction that is based on an initial transformation into the so-called canonical form. It is in this algorithm for conversion into a canonical normal form, with little complexity and using purely automata techniques, is one of the serious theoretical contributions of Assoc. Prof. Mihov. The last section of this chapter demonstrates the transfer of results for determinization, recognition of functionality, recognition of bounded variation property and minimization obtained for classical subsequential transducers, for monoidal subsequential transducers over the additive monoid of natural numbers.

Chapter 6 is devoted to the finite-state transducers introduced by Schützenberger, called bimachines, and contains, in my opinion, the most important theoretical contributions of Assoc. Prof. Mihov: pseudo-minimization and direct construction of the composition of bimachines. The efficiency of these operations largely determines the practical usability of bimachines. Here the approach is completely original and is based on monoidal finite-state transducers. A suitable monoidal transducer is constructed for each monoidal bimachine. And the more interesting and important construction is in the opposite direction, with every classic real-time functional transducer (with minimal restrictions) a monoidal bimachine with a suitable output function is associated. Thus, on the one hand, it is shown that the regular functions on words, which maps the empty word to the empty word, are exactly the functions representable by bimachines. On the other hand, this construction results in a bimachine, generally speaking, with a significantly smaller number of states.

The applied part of the dissertation - chapters 7 and 8 - is organically connected with the theoretical part. Chapter 7 presents in a solid way the declarative functional language C (M) developed by Assoc. Prof. Mihov, intended both for mathematicians who want to write efficient programs, but cannot write in C, and for specialists in the field of computer technologies. Undoubtedly, the language will be useful for students and PhD students studying or using the techniques of finite-state transducers. The design, implementation and testing of C (M) in itself is quite non-trivial, both ideologically and technically, a huge amount of work. Mihov's compiler provides efficient C code, including memory management. This has allowed in Chapter 8 to be described a number of non-trivial algorithms in a completely understandable way. An additional convenience provided by the compiler is the ability to compile to a LaTeX file containing a beautifully arranged pseudocode.

In Chapter 8 are given implementations on C(M) of the main constructions presented in the dissertation. They are all accompanied by meaningful explanations and comments. There are also 4 interesting examples of application programs written in C(M) and using the presented algorithms. All of them are meaningful, useful and not easy for those unfamiliar with the technologies of finite-state transducers.
Hints for modifications are given that encourage the reader to adapt them for their own purposes. For the first group of algorithms related to finite-state automata, example 8.1.16 is for a finite deterministic automaton recognizing valid dates from the Gregorian calendar in a fixed format. For the second group of algorithms related to classical finite-state transducers,

example 8.2.13 is for a program that realizes the functionality of a fully-fledged spell checker. It implements a function to test a given word to be in a dictionary and a function that retrieves dictionary words close to the given word in terms of the Levenshtein distance.

The third group of algorithms, relating to deterministic finite-state transducers, is illustrated by a program, Example 8.3.10, giving functionality that is used in speech synthesis. The minimal subsequential transducer is constructed that converts a number written in a decimal number system into its English phonetics. The fourth group of algorithms related to bimachines is example 8.4.4 for the implementation of basic arithmetic operations with unlimited natural numbers using bimachines.

Overall impression

The applicant has presented all the necessary documents required by the Law for Development of the Academics in the Republic of Bulgaria and the Regulations for its implementation. The extended summaries of the dissertation in Bulgarian and in English correctly reflect the contents of the dissertation. The scientometric indicators of Assoc. Prof. Dr. Stoyan Milkov Mihov far exceed the minimum criteria described in the Regulations for application of the Law for Development of the Academics in the Republic of Bulgaria.

The author's self-evaluation correctly reflects the main scientific and applied contributions in the dissertation.

The dissertation is written in a refined well-balanced style. The figures and examples are well thought out and help for faster and profound understanding. The constructions are precisely and clearly described. All propositions are correctly proven.

The presented dissertation is an impressive example of a successful and fruitful symbiosis of in-depth high-value mathematical research and useful non-trivial practical applications.

Conclusion

I am convinced that the dissertation contains significant scientific and scientific-applied contributions along with all other qualities required by the Law for Development of the Academics in the Republic of Bulgaria and the Regulations for its implementation and I strongly recommend the Scientific Jury to award the degree of "Doctor of Science" in professional field 4.6. Informatics and Computer Science to Associate Professor Dr. Stoyan Milkov Mihov for the presented dissertation „Finite-State Automata, Transducers and Bimachines: Algorithmic Constructions and Implementations“ (“Крайни автомати, преобразуватели и бимашини: алгоритмични конструкции и имплементации”).

30.03.2020 год.
Sofia

Signature:

NOT FOR PUBLIC RELEASE

(Prof. Dr. Tinko Tinchev)