

РЕЦЕНЗИЯ

върху дисертационния труд за придобиване на научната степен
„доктор на науките”
по научното направление 4.6 Информатика и компютърни науки

Автор на дисертационния труд: доц. д-р Стоян Милков Михов от секция „Лингвистично моделиране и обработка на знания“ на ИИКТ към БАН.

Тема на дисертационния труд: „Теория и имплементация на крайни автомати, преобразуватели и бимашини” (“Finite-State Automata, Transducers and Bimachines: Algorithmic Constructions and Implementations”).

Член на научното жури: акад. проф. д-мн Веселин Стоянов Дренски от ИМИ към БАН.

Дисертационният труд е в област, която е гранична за математиката и теоретичната информатика. В съществена част от труда се изследват математически проблеми от теорията на формалните езици и теория на автоматите, които се реализират програмно на специално създаден за целта език, а след това се прилагат за решаване на конкретни задачи с практическо приложение. Трудът е написан на английски език и изложен на 226 страници. Състои се от кратък увод на три страници, 8 глави, авторска справка с приносите, списък на използваната литература и списък на публикациите по дисертацията. Изложението е илюстрирано с 38 фигури. То следва буквално първите 8 глави от книгата, публикувана от дисертанта през миналата година в съавторство с Клаус Шулц. Но за разлика от книгата, в дисертацията не са включени кратките резюмета и упражненията в края на всяка глава.

1. Актуалност на разработвания проблем. Теорията на крайните автомати и свързаната с нея теория на формалните езици са важни области на съвременните математика и теоретична информатика с многобройни приложения в други математически дисциплини (например в алгебрата, математическата логика и комбинаториката) и в други области на знанието, включително с приложения за решаване на практически задачи. В частност, тези теории се прилагат широко в лингвистиката за решаване на трудни задачи в обработката на текстове и естествени езици, обработката на реч, търсене на шаблони, приближено търсене, корекции на текстове и други. Представената дисертация е мотивирана именно от такива задачи. В нея се комбинират два подхода. От една страна крайните автомати и свързаните с тях математически обекти като крайни преобразуватели и бимашини се изучават от чисто математическа гледна точка. Но след това получените резултати се използват за създаване на работещи имплементации с конкретно реализирани програмни продукти. В частност, специално за приложенията е създаден езикът $S(M)$, на който са написани програмите. Накрая на дисертацията се дават конкретни приложения за важни задачи от областта. Считаю, че в дисертацията се изучават и решават важни и актуални задачи, които водят началото си от математическата лингвистика, но представляват и самостоятелен математически интерес. Освен това се дава лингвистична интерпретация на повечето от получените резултати. При това се използват известни и се развиват нови разнообразни алгебрични, комбинаторни и изчислителни техники, като се преодоляват сериозни принципни и технически трудности.

2. Научни приноси. Първа глава „Формално въведение“ има уводен характер. В нея се въвеждат много от основните математически понятия, които се използват по-нататък –

азбуки, думи и езици. Основно внимание се обръща на понятието „моноид“ (т.е. полугрупа с 1).

Оригиналните изследвания на автора започват от втора глава „Моноидни крайни автомати“. В класическата теория на крайните автомати и свързаната с нея теория на регулярните езици обикновено се разглеждат автомати и езици на базата на свободни моноиди, когато изписаните по различен начин думи са различни елементи на моноида. Централно място в теорията заема теоремата на Клини, в която се доказва, че класът на регулярните езици съвпада с този на езиците, реализирани чрез крайни автомати. Дисертантът разглежда крайни автомати и регулярни езици, които използват произволен крайно-породен моноид. Той доказва, че така получените крайни автомати и техните езици са хомоморфни образи на класическите и доказва съответно обобщение на теоремата на Клини. След това авторът представя методи за опростяване на автоматната структура, което води до по-проста проверка дали една дума принадлежи на автоматния език. Лично на мене би ми било интересно да се види как определящите съотношения и алгоритмичните свойства на моноида влияят на алгоритмичните свойства на автомата и автоматния език, но това е въпрос на друго изследване, което не е в духа на въпросите, разглеждани в дисертацията.

В трета глава „Класически крайни автомати и регулярни езици“ се разработват методи за детерминизация и минимизация на класическите крайни автомати. Дава се конструкция, която заменя произволен краен автомат над свободен моноид с еквивалентен на него детерминиран такъв. С алгебрични средства се доказва съществуване на минимален детерминиран автомат за даден език и се дава съответната конструкция. След това се въвежда понятието оцветен детерминиран краен автомат. Тези автомати са подходящи за решаване на задачи, при които множеството от думи на езика трябва да се раздели на няколко групи. Доказва се теорема, която дава еквивалентност на оцветените детерминирани крайни автомати с минимален нецветен такъв. Оказва се, че за детерминирани крайни автомати над произволен моноид не са в сила редица свойства, типични за автоматите над свободен моноид. Поради тази причина дисертантът въвежда по-слабото понятие на псевдо-детерминиран моноиден краен автомат и доказва, че с тези автомати могат да се реализират същите езици.

В четвърта глава „Моноидни многолентови крайни автомати и крайни преобразуватели“ се изучават многолентовите крайни автомати, които са естествено обобщение на класическите крайни автомати. Тъй като при тях съответният моноид е декартово произведение на няколко моноида, те дават възможност да представят релации. Дисертантът разглежда тези обекти като частен случай на моноидните крайни автомати. Оказва се, че много от свойствата на класическите крайни автомати са в сила и за многолентовите. Освен това в главата се изучават свойствата на крайните преобразуватели, при които моноидът е декартово произведение на два моноида, единият от които е свободен, и на класическите крайни преобразуватели, когато моноидът е декартово произведение на два свободни моноида. Дава се и процедура, която решава въпроса за функционалността на класическите крайни преобразуватели.

Петта глава „Детерминирани крайни преобразуватели“ е посветена на свойството детерминираност по входната лента, което се разглежда в класа на функционалните преобразуватели. Такива преобразуватели се използват широко за обработка на текстове и разпознаване на реч. Като следствие на получените резултати се дава характеризация на регулярните функции от думи с ограничена вариация на езика на класическите

подпоследователни преобразуватели и се дава ефективно условие за липса на ограничена вариация. Дава се и процедура за минимизация на подпоследователните крайни преобразуватели. Накрая на главата се показва, че получените резултати са в сила и за случая на изходен моноид на естествените числа относно операцията събиране.

Шеста глава „Бимашини“ изучава класа на бимашините. Това е клас от детерминирани машини с краен брой състояния, които представят класа на регулярните функции между думи. Съществува стандартна конструкция на бимашина, започвайки от функционален преобразувател. В главата се предлага нова конструкция, която превръща преобразувателя директно в бимашина. Оказва се, че за редица класове от преобразуватели конструираната бимашина има експоненциално по-малко състояния в сравнение със стандартната конструкция.

Седма глава „Програмният език $C(M)$ “ представя разработения нов език за програмиране $C(M)$, на който са реализирани редица алгоритми, използвани в първите глави на дисертацията. Новият език има предимството, че за реализацията на решението на дадена задача формално се описва видът на математическия обект, който трябва да се получи. Това позволява пишещият програмата да се съсредоточи върху математическите стъпки на абстрактно ниво, вместо да се грижи за детайлите на ниско ниво. Компиляторът на езика $C(M)$ е достъпен свободно на страницата на дисертанта в интернет.

В последната осма глава на дисертацията „Имплементация на крайни автомати, преобразуватели и бимашини на $C(M)$ “ са представени реализации на $C(M)$ на автоматните конструкции от предходните глави. Във всеки от четирите параграфа е даден по едно илюстративно приложение: (1) Програма, която определя валидността на дадена дата; (2) Програма, която проверява дали дадена дума влиза в даден речник, при което дава набор от думи от речника, близки до дадената (относно разстоянието на Левенщайн); (3) Преобразуване на число в неговата фонетизация на английски език; (4) Използване на бимашини за реализиране на аритметични операции с неограничени естествени числа.

3. Преценка на публикациите по дисертационния труд. Дисертацията се базира на една монография, публикувана през 2019 г. в авторитетното издателство Cambridge University Press, 11 статии и една глава от книга, от които 7 са публикувани в периода 2000 – 2008 г., 4 са от периода 2011 – 2017 г. и една е препринт в arXiv.org от 2017 г. Всички статии са публикувани в авторитетни списания или трудове на конференции: в Computational Linguistics – 2, в Theoretical Computer Science – 1, International Journal of Document Analysis and Recognition – 1, ACM Transactions on Speech and Language Processing – 1, LNCS – 3, Trends in Linguistics - Studies and Monographs – 1, в трудове на други конференции – 2. От публикациите 3 са с импакт фактор (две от тях, в Computational Linguistics, попадат в квартила Q1, а една, в Theoretical Computer Science, е в квартила Q3). Освен тези три статии, още 4 статии са с SJR фактор. Като пример за качеството на публикуваните статии ще отбележа, че за конференцията, където е представена статия с номер 1, е имало над 70 постъпили статии, от които първите 19 са публикувани в том на серията LNAI на Springer, а следващите 19, сред които е и тази статия, са публикувани в допълнение и са достъпни от страницата на CEUR-WS. Според мене, 12 статии с качество като тези, представени към дисертацията, е скромно постижение, но то напълно се компенсира с наличието на монография, издадена в авторитетно издателство. Дисертантът е указал, че материалите по дисертацията са цитирани 227 пъти съгласно SCOPUS, но би било добре той да представи към

документацията и списък на цитатите, с което би облекчил работата на Научното жури и би дал допълнителна възможност на колегията да оцени неговите постижения. Същото се отнася и до липсата на по-конкретни данни за импакт фактора и SJR фактора. Всички публикации са съвместни: монографията и 7 от статиите са с един съавтор, 4 от статиите са с двама съавтори и 1 е с трима съавтори. Тъй като в документацията няма декларации от съавторите за съвместните статии, приемам, че участието на съавторите е равноправно. Съавторът на монографията Клаус Шулц е декларирал, че Стоян Михов има водеща роля в монографията. Съавторите са Клаус Шулц от Германия (съавтор на монографията и на 6 от статиите), Стефан Герджиков от ФМИ на СУ, ученик на дисертанта (в 4 от статиите), Петър Митанкин от ФМИ на СУ (в 2 публикации) и Гая Ангелова от ИИКТ, Христо Ганчев от ФМИ на СУ, Ян Дачук от Полща, Брус Уотсън и Ричард Уотсън от Южна Африка, Дени Морел от Франция и Кристоф Рингълщетер от Канада - съавтори в една публикация. В списъка от статии с импакт фактор и SJR фактор също има само съвместни публикации. Наличието на съвместни статии е типично явление за областта, защото изследванията носят интердисциплинарен характер. Освен това, съвместните публикации говорят за умение за работа в екип, а това е качество, което лично аз ценя високо. Въпреки това, бих пожелал на дисертанта да публикува и самостоятелни работи, на което той според мене е напълно способен.

4. Аprobация на резултатите. Много от резултатите, включени в дисертацията, са докладвани на национални конференции по математическа логика в България, на семинар в Университета в Мюнхен и на международни конференции в България, Швеция, Франция, Испания и САЩ.

5. Мнения, препоръки и бележки. Нямам особени критични бележки по изложението. Би било добре, ако авторът бе дал малко повече информация кое негово постижение къде е публикувано като например отбележи коя теорема в коя статия се съдържа. Такава информация не липсва в дисертацията, но според рецензента тя е твърде оскъдна. Това, че дисертацията следва почти буквално 8 от 11 глави на монографията на автора и Шулц има своите предимства. Това означава, че текстът вече е минал сериозно рецензиране преди публикуването на книгата. Освен това, една от целите на книгата е да послужи за навлизане с областта, което ще позволи дисертацията да може успешно да служи за навлизане в тематиката на магистри и докторанти, което несъмнено е нейно допълнително достойнство. От друга страна, следвайки изложението в книгата има и недостатъци. Както се отбелязва в предговора на книгата, примерите и програмите имат „педагогически“ характер в смисъл, че се фокусира главно върху прозрачността на стъпките и простотата на програмите. Докато това е преимущество на една книга, според мене в една дисертация би трябвало да има по-сериозни приложения. Това само е загатнато в няколко забележки, в които се казва, че авторът има подобрения на алгоритмите и програмите, като се цитират работи, върху които се базира дисертацията. Смятам, че дисертантът добре познава литературата в областта. В три от статиите, на които се базира дисертацията, има над 30 цитирани източника, а в една от тях цитираните източници са над 40. Но списъкът на литературата в дисертацията се състои само от 48 заглавия, като са включени източници от 1961 г. досега. Трябва да отбележим, че няма нито един литературен източник след 2010 г., в който дисертантът не е съавтор. Тъй като дисертацията ще се чете отделно от статиите, върху които е написана, би било разумно списъкът на литературата да е малко по-голям. Би било добре, ако към документацията по дисертацията има и автобиография на кандидата, както и списък на

всички публикации, а за тези с импакт фактор и SJR фактор се укаже и стойността на фактора.

6. Авторефератът и справката за приносите са написани достатъчно подробно и дават ясна и адекватна представа за съдържанието и основните резултати на дисертацията.

Заключение: Представеният дисертационен труд е в актуална област. Той е на високо научно ниво и удовлетворява всички изисквания, поставени пред един дисертационен труд в областта на информатиката и компютърните науки. Препоръчвам на почитаемото Научно жури да присъди на доц. д-р Стоян Милков Михов научната степен „доктор на науките” в научното направление 4.6 Информатика и компютърни науки.



София, 27 май 2020 г.

Подпис: /

(акад. дмн В. Дренски)