Институт по информационни и
комуникационни технологии-БАН
Вх. № 439 / 28.05 2020 г.

РЕЦЕНЗИЯ

на дисертационен труд
за придобиване на научната степен "Доктор на науките"

Област 4 "Природни науки, математика и информатика"

Научно направление 4.6 "Информатика"

Тема: "Крайни автомати, преобразуватели и бимашини:
алгоритмични конструкции и имплементации"

Автор: доц. д-р Стоян Милков Михов

Тема на дисертационния труд

Представеният дисертационен труд е посветен на теорията на крайните автомати, преобразуватели и бимашини. Целта на дисертанта е да представи конструкции за устройства с краен брой състояния, доказателства за коректност на тези конструкции, както и техни работещи имплементации. Разглежданите абстрактни машини имат важни приложения при съхраняването и обработката на текстове на естествен език. Дисертацията предствлява определен напредък на концептуално ниво за някои видове абстрактни устройства, използвани за обработка на естествен език.

Работата обобщава изследванията на автора през последните 15 години. По-важните задачи, разгледани в дисертационния труд са следните:

(1) Да се представи единно и затворено описание на основите на теория на крайните автомати, преобразуватели и бимашини.

(2) Да се дадат формални доказателства за коректност на представените конструкции и за съществуващите връзки между различните абстрактни устройства.

(3) Да се разработи език за описание на предствените абстратни изчислителни устройства.

(4) Да се разработят имплементации заедно с документиран код за представените конструкции.

(5) Да се очертаят възможни приложения на разработените техники.

Литературен обзор

Общото ми впечатление е, че дисертантът познава отлично съвременното състояние на разглежданите проблеми. Голяма част от изследванията му са върху един кръг от задачи от теория на автоматите, считани като значими, както в теоретичен план, така и за приложенията. Дисертантът демонстрира дълбоко познаване на областта на изследванията и възможности творчески да прилага знанията си.

Методика

В изследванията си дисертантът използва основно езика на теория на множествата и методи на доказване, типични за теория на автоматите. Дисертацията има допирни точки с някои области на абстрактната алгебра (теория на групите).

Съдържание и резултати на дисертационния труд

Дисертационният труд е в обем от 219 нестандартни машинописни страници и се състои от увод, осем глави, заключение и списък на използваната литература, включващ 48 заглавия. По-долу ще изложа накратко съдържанието на отделните глави от дисертациията.

В глава 1 са представени дефиниции на основни математически понятия, използвани в дисертацията. Те включват множества, функции и релации (раздел 1.1), продължаване на функции върху подмножества (раздел 1.2), езици над произволни азбуки (раздел 1.3), низове и релации върху низове (раздел 1.4). В раздел 1.5 са въведени моноиди като множество с бинарна операция, която е асоциативна и има единичен елемент. След това се дефинират езици върху моноиди и операции върху моноиди.

В глава 2 са въведени моноидни крайни автомати като обобщение на класическите крайни автомати. С всяка дума $a_1 \ldots a_k$ от езика, разпознаван от автомата се свързва етикет $a_1 \circ \ldots \circ a_k$, където изразът се изчислява в моноида M. Класическите автомати се получават като частен случай на моноидни автомати над свободния моноид. При тези условия са доказани теорема 2.1.22, че всеки моноиден автомат е хомоморфен образ на класически автомат, и твърдение 2.2.1, че класът на езиците, разпознавани от моноидни автомати, са затворени относно операциите обединение, моноидно произведение, моноидна звезда на Клини, както и по отношение на моноидни хомоморфизми.

В раздел 2.3 се въвеждат монидни регулярни езици, аналогично на класическите регулярни езици. Разликата се състои в изискването $L_1 \circ L_2$ да е моноиден език. Тази дефиниция е по-обща: класическите регулярни езици се получават като специален случай на монидните езици. По-нататък се въвеждат монидни регуларни изрази и, очаквано, се доказва, че моноидните регулярни езици са точно езиците разпознавани моноидни крайни автомати (теорема 2.4.2).

В последния раздел на тази глава (2.5) се разглеждат някои упростявания на разглежданите моноидни крайни автомати. Първото от тях е т.нар. тримоване, което се състои в премахване на "излишните" състояния: всяко състояние трябва да е върху път, по който стига до крайно състояние. Друга операция е премахване на e-преходите (преходи, които имат за етикет единичния елемент на моноида). Доказано е, че за всеки моноиден краен автомат съществува еквивалентен моноиден краен автомат без e-преходи (твърдение 2.5.4).

Глава 3 е посветена на детерминирани крайни автомати като най-естествения и най-важен клас крайни автомати. В раздел 3.1 са въведени детерминираните крайни автомати и са изложени някои добре известни свойства, като например, че всеки краен език е автоматен. В раздел 3.2 е описана конструкция, чрез която от всеки моноиден краен автомат се получава детерминиран краен автомат с тотална функция на преходите.

В раздел 3.3 се раглеждат някои своиства на класическите крайни автомати. Тук централни резултати са твърдения 3.3.1, 3.3.2 и 3.3.4. В първото от тях при зададен автоматен език $\mathcal{A}$ се показва съществуването на автомати за допълнителния език; в твърдение 3.3.2 се доказва аналогичен резултат за сечение и разлика на автоматни езици, а твдение 3.3.4 се отнася до посимволно обръщане на краен автомат.

Раздели 3.4 и 3.5 са посветени на минимизация крайни автомати. Въвеждат се понятията дясно инвариантна релация и съвместимост на релация и език. Доказано е, че ако език е съвместим с дясно инвариантна релация от краен индекс, то езикът се разпознава от детерминиран краен автомат (твърдение 3.4.4). Освен това, ако $R_{\mathcal{A}}$ е релация, включваща всички двойки думи, по които се достига от q_0 до едно и също състояние, то $R_{\mathcal{A}}$ е дясно инвариантна (твърдение 3.4.7). По нататък за прозволен език L над азбуката Σ е въведена релацията на Майхил–Нерод R_L и са доказни няколко резултата, по-важните от които са следните: теорема 3.4.13, съгласно която за всеки (класически) краен автомат съществува единствен еквивалентен детерминиран краен автомат, който е минимален относно броя на състоянията; теорема 3.4.14, съгласно която един език L е автоматен тогава и само тогава, когато индексът на R_L е краен; твърдение 3.4.17, съгласно което един автомат е минимален тогава и само тогава, когато няма различни еквивалентни състояния. В раздел 3.5 е представена конструкция на построяване на еквивалентен минимален автомат чрез последователно идентифициране на всички еквивалентни състояния.

В раздел 3.6 се разглеждат оцветени детерминирани крайни автомати. Това се автомати чиито крайни състояния са "оцветени" в няколко цвята (т.е. зададено е разбиване на крайните състояния, което от своя страна индуцира разбиване върху разпознаваните думи). Тук централни са теореми 3.6.12 и 3.6.14. Според първата от тях за всеки оцветен детерминиран краен автомат съществува единствен еквивалентен детерминиран краен автомат, който е минимален относно броя на състоянията. Във втората теорема се твърди, че един детерминиран краен автомат е минимален тогава и само тогава, когато няма различни еквивалентни състояния.

Накрая в раздел 3.7. се разглежда псевдо-детерминизация и псевдо-минимизация на моноидни крайни автомати. Един моноиден автомат е псевдодетерминиран, ако чрез всеки елемент на моноида може да се достигне най-много до едно състояние на автомата и е псвдоминимален, ако е хомоморфен образ на минимален автомат. В твърдение 3.7.2 се доказва, че за всеки моноиден краен автомат може да се построи еквивалентен псевдодетерминиран краен автомат, а в твърдение 3.7.4 – че за всеки моноиден краен автомат съществува еквивалентен псевдоминмален краен автомат.

Глава 4 е посветена на моноидни многолентови крайни автомати и крайни преобразуватели. В раздел 4.1 се въвеждат многолентови моноидни крайни автомати. При тях входната азбука е директно произведение на n моноида. Така езикът разпознаван от n-лентов автомат е n-местна релация. Моноидните многолентови автомати притежават някои очаквани свойства, които са описани в твърдение 4.2.1. Оттук следва, че класът на моноидните многолентови крайни релации е затворен относно декартово произведение, проекции и обратна релация. В раздел 4.3 са дефинирани класическите n-лентови крайни автомати (като n-лентови моноидни автомати над свободния моноид) както и n-лентови посимволни крайни автомати. В раздели 4.4 и 4.5 се въвеждат съответно моноидни крайни преобразуватели и класически крайни преобразуватели като двулентови крайни автомати над азбука във вида $\mathcal{M}_1 \times \mathcal{M}_2$, където $\mathcal{M}_1$ е свободен моноид (съответно $\mathcal{M}_1$ и $\mathcal{M}_2$ са свободни моноиди). Накрая в раздел 4.6 са доказани няколко резултата, които водят до процедура за разрешаване на функционалността на класически краен преобразувател. По-нататък това е програмно реализирано в глава 8.

В глава 5 се изследват детерминирани крайни преобразуватели. Най-напред в раздел 5.1 се въвеждат моноидни подпоследователни преобразуватели и класически подпоследователни преобразуватели. Дефинирани са важните понятия секвенциално разстояние между две думи и функция с ограничена вариация като функция в която "близки" думи се изобразяват в "близки" думи. По-нататък целта е да се изследват класически крайни преобразуватели с ограничена вариация. В раздел 5.2 е показано, че всеки функционален класически преобразувател с ограничена вариация може да се преобразува в еквивалентен класически подпоследователен преобразувател. По-нататък в раздел 5.6 е показано, как тези резултати могат да се обобщат за моноиди, различни от свободния (за моноида на естествените числа по отношение на събирането).

В раздел 5.4 е дефинирана релация, подобна на релацията на Майхил-Нерод за подпоследователни преобразуватели, с помощта на която може да се получи минимален подпоследователен преобразувател. В раздел 5.5 е представена процедура за минимизация на подпоследователни крайни преобразуватели.

В глава 6 се изследват бимашини, дефинирани като абстрактно изчислитлно устроиство състоящо се от моноидна азбука, два детерминирани крайни автомата с обща азбука и частична функция, наречена функция на изходите.

В раздел 6.2 е изследвана връзката между регулярните функции върху думи и бимашините. Доказано е, че за всяка моноидна бимашина $\mathcal{B}$ съществува моноиден краен преобразивател $\mathcal{A}$, за който $O_{\mathcal{B}} = L(\mathcal{A})$ (Твърдение 6.2.1). Обратно, за всеки тримован функционлен преобразуvател с изход в моноида $\mathcal{M}$ съществува моноидна бимашина $\mathcal{B}$, за която $L(\mathcal{T}) = O_{\mathcal{B}}$. В раздел 6.3 е въведено понятието псевдоминимизация на моноидна бимашина и е представена конструкция за псевдоминимална бимашина. В раздел 6.4 е представена и изследвана т. нар. директна композиция за класически бимашини.

Глава 7 съдържа описание на езика $C(M)$, който се използва в следващата глава 8за реализиране на представените в дисертацията алгоритми. Разработен е компилатор, който компилира програма на $C(M)$ до код на С. В последната глава са представени имплементации на алгоритми за различни абстрактни машини: крайни автомати, класически крайни преобразуватели, детерминирани крайни пробразуватели и бимашини.

Приноси на дисертационния труд

По-важните приноси в дисертационния труд се свеждат до следното:

(1) Изложени са теоретичните основи на крайните автомати, преобразуватели и бимашини. Дадени са доказателства за коректност на представените конструкции, както и за връзките между тях. С това дисертацията придобива и значителна дидактична стойност.

(2) Разработен е метод за тестване на ограничената вариация на класически краен преобразувател.

(3) Представена е нова конструкция за канонизация на подпоследователни преобразуватели, чиято сложност е полиномиална.

(4) Разработен е език за програмиране $C(M)$ за реализиране на представените в дисертацията алгоритми. Разработен е компилатор, който компилира програми на $C(M)$ до С код.

(5) Представени са имплементации на алгоритми за различни абстрактни машини: крайни автомати, класически крайни преобразуватели, детерминирани крайни пробразуватели и бимашини.

(6) Очертани са приложения на резултатите към важни практически проблеми, свързани с обработката на естествен език.

Забележки и коментари по дисертационния труд

Във връзка с дисертационния труд имам следните забележки, въпроси и коментари:

(1) Авторефератът е написан в конспективен стил и преставлява набор от дефиниции и резултати. На места липсват факти, които правят невъзможно разбирането без справка с основния текст. Така при дефиницията на секвенционалното разстояние не е дефинирано $u \wedge v$; в твърдение 3.3.3 не ясно, какво е ρ и т.н.

(2) В дисертацията се забелязват известни езикови неблагополучия, които се обясняват с това, че в българския език няма установена терминология: "изфиняване на функция", "повдигане на множество", "последователност" (редица) и т.н.

(3) Някои от резултатите са наречени твърдения а други – теореми. Каква е разликата между твърдения и теореми в дисертационния труд?

Публикации по дисертационния труд

Изложението в тази дисертация следва монографията на автора и К. Шулц "Finite-State Techniques: Automata, Transducers and Bimachines", издадена от Cambridge University Press през 2019 г.

Резултатите от дисертационния труд са публикувани в 12 статии и една глава от книга. Три от статиите са в списания с импакт-фактор:

- Computational Linguistics (Q1:2000,2004) – 1.657
- Theoretical Computer Science (Q3: 2011) – 0.667

От останалите седем статии са в списания и сборници със SJR: International J. of Document Analysis and Recognition, ACM Transactions on Speech and Language Processing, Lecture Notes in Computer Science и др. Изполвана е глава от монографията Finite State Techniques: Automata, Transducers and Bimachines от поредицата Cambridge Tracts in Theoretical Computer Science, Cambridge University Press. Представените публикации напълно отговарят на минималните национални изисквания по т. 7 (показател 7).

Авторство на получените резултати

Шест от статиите на Стоян Михов с с един съавтор, пет са с двама и една – с трима съавтори. Представено е писмо от един от съвторите на общия монографичен труд, в което се декларира, че приносът на докторанта е поне равностоен на този на другия съавтор. От писмото става ясно, че изследванията в главите, отнасящи се до езика $C(M)$ и имплементациите на крайни автомати са изключително дело на дисертанта. За останалите съвместни работи приемам, че приносът на дисртанта е равностоен с този на останалите автори.

Цитирания на публикациите от дисертационния труд

Дисертантът е приложил списък на 227 цитирания на статиите, в които са публикувани резултатите от дисертационния труд. В материалите към дисертационния труд е направено описание на местата, където се появяват тези цитирания. Приемам, че резултатите са известни и добре приети в професионалната общност.

Автореферат и авторска справка

Авторефератът и авторската справка са направени съгласно изискванията и отразяват правилно резултатите и приносите в дисертационния труд.

Заключение

Дисертацията е посветена на един кръг от задачи от теория на крайните автомати, които имат и определено практическо значение, особено тези, свързани с важни задачи за търсене в и обработка на текстове, обработка на естествен език, правописна корекция, фонетизация. Разработени са нови алгоритми за конструиране и изследване на специални автомати, с което е постигнат съществен напредък в актуална област на информатиката.

Считам, че представеният дисертационен труд "Крайни автомати,преобразуватели и бимашини: алгоритмични конструкции и имплементации" с автор Стоян Милков Михов съдържа резултати, които представляват оригинален принос в теория на крайните автомати. Дисертантът показва задълбочени теоретични знания в областта на абстракните изчислителни устройстава. С това той отговаря на изискванията на "Закона за развитие на академичния състав" за даване на научната степен "Доктор на науките". Гореизложеното ми дава основание да дам **положителна оценка** на представения дисертационен труд и да препоръчам на Уважаемото Жури да присъди на Стоян Милков Михов научната степен "Доктор на науките" в област 4 "Природни науки, математика и информатика", научно направление 4.6 "Информатика".

София, 28.03.2020 г. Рецензент:

(проф. д.м.н. Иван Ланджев)

NOT FOR PUBLIC RELEASE