

СТАНОВИЩЕ

на проф. д-р Тинко Величков Тинчев,
СУ „Св. Климент Охридски“, ФМИ, катедра „Математическа логика и приложенията й“

за представената от **Стефан Владимиров Герджиков** дисертация „**Ефективен алгоритъм за приближено търсене в регулярни множества**“ за придобиване на образователната и научна степен „доктор“ в професионално направление 4.5 Математика (Математическа логика)

Представеният дисертационен труд е на английски език, има титулна страница на български език, титулна страница на английски език, съдържание, списък на фигурите, увод и основна част, която е с обем 181 страници. Основната част е структурирана в 8 глави, добавка (апендикс), заключение – резюме и библиография от 67 цитирани източника. Заключение – резюме съдържа декларация за оригиналност на научния труд и е представено отделно и на български език. Авторефератът е на български език. Намирам, че структурата на дисертацията и на автореферата отговарят на изискванията на ЗРАСРБ и на правилника за прилагането му.

Стефан Герджиков е магистър по информатика, магистърска програма „Логика и алгоритми“, ФМИ. Изпълнил е солиден индивидуален план като редовен докторант и е отчислен с право на защита.

Изследването е в активно развиваната през последните 50-ина години област на теорията на крайните автомати като модели за практически важни задачи за ефективно представяне, търсене и извличане на информация, най-често отнасяни към компютърната лингвистика поради характера на непосредствено видимите им от широка аудитория приложения. Един клас от проблеми естествено попада в тази област – при зададени формален език и мярка за близост между думи по дадена дума да се намерят близките до нея. В тази насока са изследванията на Стефан Герджиков, с които, като цяло, групата около научния му ръководител Стоян Михов, ползваща се с добър международен авторитет, прави нова забележима стъпка с видима перспектива за развитие.

Основният резултат в дисертацията на Стефан Герджиков е описанието и изследването на алгоритъм, който при предварително зададени регулярен език L над крайна азбука Σ , разстояние d от определен вид между думите от Σ и рационално число q от $(0,1)$ по дадена дума V над Σ намира всички думи U от L , удовлетворяващи условието $d(V,U) \leq q|V|$, където $|V|$ е дължината на V . Той е едно далеч отиващо нетривиално развитие на идеи на Myers, Baeza-Yates & Navarro, Mihov & Schulz и в известен смисъл отговор на въпрос на Schulz, поставен на семинара на Стоян Михов, *как можем да търсим в текст, започвайки от специфични, характерни, части на текста като постепенно ги разширяваме и какво всъщност би трябвало да означава „специфични, характерни, части на текста“*. Струва си да се отбележи удачно избраният „спираловиден“ подход на изложението: тръгва се с умело анализиран обозрим пример в глава 3, но достатъчен за да се добие идея за намерението на дисертанта, което ще видим реализирано в пълнота чак в глава 6; да почувстваме, че линейното представяне на текста (т.е. езикът е само от една дума) при Myers е съществено и структура от тип Blumer et Blumer е обещаващ заместител; а така също, че нивото на близост динамично ще се променя при рекурсивните стъпки, за разлика от подхода на Baeza-Yates & Navarro, при които езикът L е дори краен. След това, в глава 4 се излага оригинален алгебричен подход към множествата от подравнявания и списъци от разстояния, който се използва за ефективно пораждаване на кандидатите за корекция. При това в

ключовите лемми се разглеждат само специален вид разстояния, но достатъчни общи за да обхванат вече класическия случай на разстояния на Левенщайн. Общият случай е разумно отложен за параграф 6.1. В глава 5 е описан нов алгоритъм за приближено търсене в регулярен език при специалния вид разстояния, изучавани в глава 4. Това би било непостижимо разточителство, ако този алгоритъм е просто „проекция” на алгоритъма за общия вид разстояния. Ситуацията обаче е далече от горното опасение: дисертантът не само задълбочено изследва сложността на описания алгоритъм, но в случая на краен език L умело „жонглира” с изобретателни представяния на списъци от разстояния. В глава 6 е описан централният резултат – алгоритъм за приближено търсене в регулярен език при произволно разстояние (разбира се, от разглеждания вид – ортографско разстояние). Отново е отделено специално внимание на крайните езици.

В глава 7, използвайки алгебричния подход към автоматите, с помощта на подходящи поражащи функции е дадена горна граница за очакваната сложност на алгоритъма при произволен регулярен език и произволно ортографско разстояние като при достатъчно разумни предположения е показана сходимост на намерената линейна функция, задаваща горна граница. Глава 8 е особено интересна и перспективна както от теоретична, така и от практическа гледна точка с предлагания нов метод за дефиниране на близост между думи, базирана на структурата им. Най-грубо казано, задачата, която се решава е следната. Нека разполагаме с известен краен брой наблюдения за трансформации на някои думи от даден речник в думи, които ще смятаме за близки. При естествени предположения, имащи смисъл, че трансформациите са локални, т.е. не променят коренно „правописа” на думите от речника, се предлага метод, който по дадена дума дава сортиран по вероятност списък от думи от речника, близки до дадената. Типични примери са когато трансформациите са резултат от исторически промени в езика и грешно разпознати думи от OCR програма. Така, в първия случай, алгоритъмът по дадена „остаряла” дума предлага вероятни съвременни нейни варианти. Този алгоритъм е ядрото на създадената от екипа на СУ в рамките на проекта CULTURA система за нормализиране на исторически текстове на английски език от XVII век. Убеден съм, че идеите в основата на метода за дефиниране на близост са интересни и ще получат развитие в различни насоки.

Резултатите от изследванията, представени в дисертацията, са частично публикувани в една самостоятелна статия в Доклади на БАН – издание с импакт фактор, и в една статия в съавторство с научния му ръководител Стоян Михов и с Владислав Ненчев в рецензираните трудове на реномираната международна конференция ICDAR'2013. Две статии в съавторство са публикувани в трудовете на международната конференция Joint EDBT/ICDT Workshops 2013 и в Arhiv. Класирането на първо и второ място в различни задачи от независимото оценяване на имплементациите за Workshop on Scalable Similarity Search Strings/Join, Genoa, 2013, е недвусмислено свидетелство за качествата на алгоритъма от глава 5 и е потвърждение за практическата му използваемост. В горните съвместни статии приносът на Стефан Герджиков е не по-малък от приноса на другите съавтори.

Резултати от дисертацията са докладвани на 3 реномирани международни конференции и на 2 от Пролетните конференции на ФМИ.

Авторефератът и заключението (резюме) правилно отразяват научните приноси от изследванията на Стефан Герджиков, изложени в представената дисертация.

Избраният стил на изложение е добър, с грижа читателят да бъде информиран какво и как ще бъде направено, какви са затрудненията или пък пречките да се приложи дадена техника. Не си струва да изреждам дори и не дълъг списък от правописни грешки, най-често транспозиции на

букви (например form вместо from), изпуснат знак за коренуване или повторен/излишен определителен член, които не създават затруднения за разбирането. Това се отнася и за началото на глава 2, стр. 17, където скоро се разбира, че за текст ще се използва вече буквата S , а не T . Също не си струва да се впечатляваме, че във формулировката на лема 1.2.5 е изпуснато предположението, че всяко състояние на автомата трябва да е съществено, т.е. да е достижимо от начално и от него да е достижимо някое крайно. Известно недоумение буди наблюдението, че дисертантът не се осмелява да нарече нито едно твърдение от дисертацията си теорема. Не мога обаче да се досетя за причината, породила употребата на израза „се сходя“ вместо „е сходящ/а“ в текста на български език, следствие 2 на стр. 17 от автореферата.

Имайки предвид казаното, убедено смятам, че представената дисертация наред с високите си научни качества по недвусмислен начин показва широката обща математическа култура, богат арсенал от овладяни техники и похвати, способност за задълбочени научни изследвания и творческа изобретателност на Стефан Герджиков. Намирам, че всички изисквания на ЗРАСРБ, Правилника за прилагане на ЗРАСРБ, правилниците за условията и реда за придобиване на научни степени и за заемане на академични длъжности на СУ „Св. Климент Охридски“ и на ФМИ към обсъжданата дисертация и автора ѝ са изпълнени, поради което **убедено предлагам Стефан Владимиров Герджиков да придобие образователната и научна степен „доктор“ в професионално направление 4.5 Математика, научна специалност „Математическа логика“, за дисертационния труд „Ефективен алгоритъм за приближено търсене в регулярни множества“.**

Подпис:

3 януари 2014 г.
София

(проф. д-р Тинко Тинчев)