

РЕЦЕНЗИЯ

от проф. д-р Христо Стефанов Кискинов,

професор във ФМИ при Пловдивския университет „Паисий Хилендарски” (ПУ)

на дисертационен труд за присъждане на образователната и научна степен „доктор“ по: област на висше образование 4. *Природни науки, математика и информатика*; професионално направление 4.6. *Информатика и компютърни науки*; докторска програма *Информатика*.

Автор: Николай Георгиев Ханджийски.

Тема: Итеративен парсиращ алгоритъм с приложение в профилирането на парсери.

Научен ръководител: проф. д-р Елена Петрова Сомова.

1. Общо представяне на процедурата и представените материали

Със заповед № РД-21-654 от 21.03.2024 г. на Ректора на Пловдивския университет „Паисий Хилендарски” (ПУ) съм определен за член на научното жури за осигуряване на процедура за защита на дисертационен труд на тема *„Итеративен парсиращ алгоритъм с приложение в профилирането на парсери“* за придобиване на образователната и научна степен „доктор“ в област на висше образование 4. *Природни науки, математика и информатика*, професионално направление 4.5 *Информатика и компютърни науки*, докторска програма *Информатика*. Автор на дисертационния труд е Николай Георгиев Ханджийски – докторант в редовна форма на обучение към катедра „Компютърна информатика” на Факултета по Математика и информатика (ФМИ) при ПУ „Паисий Хилендарски” с научен ръководител проф. д-р Елена Петрова Сомова. На първото неприсъствено заседание на научното жури, проведено в периода 22.03.2024 г. – 23.03.2024 г. и отразено в Протокол № 1, съм определен да изготвя рецензия.

Представеният от докторанта Николай Георгиев Ханджийски комплект материали на хартиен и електронен носител е в пълно съответствие с Чл.36 (1) от Правилника за развитие на академичния състав на ПУ (ПРАСПУ) и включва следните документи:

- молба до Ректора на ПУ за разкриване на процедура за защита на дисертационен труд;
- автобиография в европейски формат;
- протокол от разширения катедрен съвет, свързан с докладване на готовността за откриване на процедурата и с предварително обсъждане на дисертационния труд;
- дисертационен труд;
- автореферат на български и английски език;
- списък на научните публикации по темата на дисертацията, включващ 6 заглавия;
- копия на научните публикации – 6 бр.;
- декларация за оригиналност и достоверност на приложените документи;
- справки за спазване на минималните национални изисквания.

Документите са много старателно подготвени и нямам никакви забележки по тях.

2. Кратки биографични данни за докторанта

Николай Георгиев Ханджийски е роден на 19.02.1995 г. През 2010 г. придобива образователната степен „бакалавър“ във ФМИ на ПУ „Паисий Хилендарски“ по специалност „Информатика“, като по време на следването си е бил студент по програма Цепус в Университета Йоханес Кеплер в Линц, Австрия и по програма Еразъм в Университета Кобленц-Ландау в Кобленц, Германия. През 2019 г. Придобива образователната степен „магистър” пак в ПУ „Паисий Хилендарски“ по специалност „Софтуерни технологии“. От 2019 г. е редовен докторант към катедра Компютърна информатика във ФМИ при ПУ.

От 2012 г. досега работи като програмист във фирма Софтуерни системи – България (с централа в Пловдив), а от 2018 г. досега и като управител на фирма за софтуерни продукти в Офенбург, Германия, където и живее понастоящем.

3. Актуалност на тематиката и целесъобразност на поставените цели и задачи

Изследванията, свързани с транслаторите (компилатори и интерпретатори) и математическите основи, на които се базират (формални езици, пораждащи граматика, абстрактни машини) е една вечно зелена тема в информатиката. От средата на 30-те (1936 г.), когато се появяват забележителните работи върху изчислимостта на Алън Тюринг, на Емил Пост и на Алонсо Чърч, през военните години 1942-1945, когато е създаден първият алгоритмичен език за програмиране и компилатор за него от Конрад Цузе, през средата на 50-те с теоретичните изследвания върху формалните езици и пораждащите ги граматика на Ноам Чомски, бурното навлизане на електронните изчислителни машини и създаването на първите универсални езици за програмиране от високо ниво FORTRAN и ALGOL с техните компилатори, чак до ден днешен по тази тема непрекъснато се работи. Че темата продължава да е изключително актуална и в днешно време, лесно се установява като се проследи публикационната активност на работещите по тази тематика в специализираните списания.

Фактът, че дисертантът е получил значими (според мен) и оригинални резултати в област, в която се работи интензивно вече близо 90 години, говори за съдържателни и оригинални цели и задачи, които дисертантът е преследвал в дисертацията си.

4. Познание на проблема

Извършеният обзор на научните изследвания по проблема, подбраната литература и най-вече получените резултати говорят недвусмислено за дълбоки познания в изследваната област.

Нещо повече, аз съм твърдо убеден, че дисертантът познава прекрасно изследваната област. Освен личните впечатления (виж т. 12), има и обективни основания за моето убеждение. То лесно може да бъде аргументирано, ако се вгледаме внимателно в обзорната първа глава на дисертационния труд. В нея за всяко едно дискутирано понятие са указани точно заглавията (а в повечето случаи и точната страница в тях) от литературата, откъдето е взета представената информация. И никъде не се срещат цитирания на повече от две заглавия наведнъж. А са цитирани всички заглавия от библиографията... Това ясно показва, че дисертантът е запознат, и то в детайли, с всеки един от 191-те източника, цитирани в дисертацията.

5. Методика на изследването

Общоприето е, че стандартната методика в изследванията, водеща до постигане на конкретни резултати под формата на работещи софтуерни прототипи е следната: проучване на областта, открояване на проблема, създаване на проектен модел, синтез на подходящо решение, реализация с използване на известни софтуерни инструменти, провеждане на експерименти и анализ на получените резултати. Но все пак ще се осмеля да отбележа, че при изследванията, водещи до значими и оригинални резултати, това не е точно така. Обикновено след първоначалното проучване на областта и избор на проблем, който да се решава, в хода на търсене на неговото решение или на реализацията му, изследователят извършва допълнително проучване областта, което често води до появата на нови идеи, съответно до промяна на проблема за решаване и т.н. Т.е. процесът е така да се каже, итеративен. В подкрепа на тази моя хипотеза е например фактът, че понякога се налага промяна на темата на дисертацията (в случая повече от веднъж). Но в никакъв случай това не бива да се приема за недостатък – точно обратното, това показва искрения стремеж на докторанта да получи значими и оригинални резултати.

6. Характеристика и оценка на дисертационния труд

Дисертационният труд „Итеративен парсиращ алгоритъм с приложение в профилирането на парсери” е написан на 179 страници и се състои от увод, четири глави, заключение, списък на авторските публикации по темата, списък със забелязани цитирания, две приложения от по 1 страница, списък с използваната литература и декларация за оригиналност. Изложението включва още списък на използваните означения и съкращения, 56 фигури и 8 таблици.

Изключително полезни са (виж т. 11 от рецензията) и представените в началото на изложението две таблици на използваните термини от литературата и на въведените от автора термини, с тяхния английски превод и указания за източника или мястото на дефиниране в дисертацията респективно.

В увода, освен формулиране на целите и задачите, се съдържа и кратко описание на структурата на дисертационния труд и на апробацията на резултатите.

Първа глава има обзорен характер и в нея са изложени подбрани от автора основни факти от теорията на формалните езици и обработващите ги абстрактни машини, свързани с методите на транслация. Дискутирани са елементите на транслаторите, извършващи лексикален и синтактичен анализи, както и пораждащите граматика, които се използват от тях. В главата се прави преглед на крайни автомати, стекови автомати, машини на Тюринг, алгоритми на Марков и др. Направеният обзор включва и известни алгоритми за разпознаване и парсиране, работещи на базата на безконтекстни граматика.

Втора глава е посветена на проектираната от автора цялостна концепция за парсиране на данни, наречена от него парсираща машина. В нея освен детайлното и описание, се съдържат и формални дефиниции за въведения от автора вид безконтекстни граматика с допълнителни компоненти, наречени от него напреднали граматика, както и на въведената фразова машина, позволяваща вместо сравнения между низове, да се извършва сравнение между класове от низове, облекчаващо сравнението на лексемите и фразите в хода на парсиране. Определени са и формалните езици, които са дефинирани от новите граматика с допълнителни символи. Показаната парсираща машина съдържа различни видове модули като доставчик, скенер, лексер, парсер, оптимизатор, строител и филтър, всичките описани в детайли, заедно с техните функционалности. Главата завършва с дефинициите на различни видове синтактични дървета и командите за строеж, на базата на които те биват създавани.

В трета глава се съдържа детайлно описание на създадения от автора алгоритъм, наречен Тунелно парсиране, който е вграден в парсер модула на парсиращата машина и се възползва от добавените възможности в нея. Подробно са описани различните обекти, които се създават преди началото на парсирането, на базата на дадена напреднала граматика. Списъкът от тези обекти включва: стек за изпълнение, стек за дълбочина, стек за повторение, архивни стекове, автомати, достижими дървета, конфликти, тунели, маршрутизатори и контролни обекти. Детайлното описание на контролните обекти съдържа стъпките, които парсерът извършва и ефективно представлява псевдо-код на алгоритъма Тунелно парсиране. Главата съдържа пример с дефинираните обекти, които се използват от алгоритъма и завършва с примерно изпълнение на алгоритъма за избран низ от входни данни.

В четвърта глава е представен специално създаден за целта на дисертацията софтуерен инструмент, наречен профилатор, с помощта на който могат да се генерират голям брой безконтекстни граматика и входни данни и да се извършват експерименти с парсиращи машини, които са генерирани от различни парсер генератори на базата на тези граматика. Главата съдържа описанието на специално създаден за целта на дисертацията език на шаблонна граматика за императивно метапрограмиране на граматика. Показани са различните граматика, които дефинират валидните според шаблонния език скриптове. Главата завършва с интерпретацията на резултатите от четири различни експеримента, които са проведени с помощта на профилатора.

В заключението дисертантът е обобщил и систематизирал основните резултати, направил е самооценка на приносите, съдържащи се в дисертационния труд и е описал апробацията на получените резултати. Формулирани са и конкретни перспективите за развитие на дисертационната тематика.

Списъкът на използваната литература включва 191 (сто деветдесет и едно) заглавия, като две са на български език, две са на руски език, а останалите са на английски език.

Дисертацията е написана много грамотно, без правописни или стилистични грешки. Всички използвани или въведени понятия са коректно дефинирани, изложението е строго формализирано, но едновременно с това и разбираемо, а ходът му е логичен.

Не съм констатирал „плагиатство“ в работите на кандидата и в представения дисертационен труд по смисъла на ЗРАС в РБ.

7. Приноси и значимост на разработката за науката и практиката

Подкрепям описаните от докторанта основни приноси в настоящия дисертационен труд. А именно:

Научни приноси:

- Предложен е концептуален модел на парсираща машина, която е подходяща за реализиране на различни стратегии, методи, подходи и алгоритми, предоставяща някои допълнителни възможности спрямо съществуващите;
- Дефинирани са нов вид граматика, наречени от автора напреднали граматика, с допълнителни компоненти (фраза символи, съставени от правила и напреднали символи, според терминологията на дисертанта), с близка структура до безконтекстните граматика в подсилена Бекус-Наур форма;
- Предложен е модел на фразова машина, която предварително категоризира различните фрази в парсер граматиката с цел ускоряване на анализа по време на парсирането.

Научно-приложни приноси:

- Проектирана и описана в детайли е функционалността на въведената парсираща машина;
- Проектирана и описана в детайли е функционалността на нов алгоритъм, наречен Тунелно парсиране, парсиращ на базата на всяка не ляво рекурсивна напреднала граматика, като времето за парсиране е линейно за някои многозначни граматика и за всяка граматика, която може да се изведе от детерминиран автомат с един стек (а ако в изведената граматиката няма рекурсия, парсирането може да се изпълни с константен обем памет);
- Проектиран и създаден е език за метапрограмиране на граматика;
- Проектиран е софтуерен инструмент, наречен профилатор на парсер генератори, позволяващ извършването на тестове (измерване на използваните ресурси по време на разпознаване и парсиране) с парсиращи машини, създадени от различни парсер генератори и компилатори.

Приложни приноси:

- Разработен е работещ прототип на софтуерен инструмент – профилатор на парсер генератори (вкл. с модул за визуализиране на резултатите), позволяващ експериментиране с директно въведена граматика или с програмно изведена от инструмента съвкупност от граматика;
- Извършени и анализирани са експерименти с помощта на профилатора.

Според мен, получените в хода на дисертационното изследване резултати са значими и далеч надхвърлят и като качество и като количество, необходимите за придобиване на образователната и научна степен „доктор“.

8. Преценка на публикациите по дисертационния труд

Част от резултатите от дисертационното изследване са представени в приложените 6 публикации на английски език, като всяка една от тях е индексирана в поне една от

световноизвестните бази от данни с научна информация Web of Science (WoS) и SCOPUS. От тях 5 са в списания, индексирани в SCOPUS с SJR, а 4 са в списания, индексирани във WoS, като 2 от тях имат и Impact Factor (SJR) за съответната година. Те формират общо 168 т., което надхвърля над пет пъти минималните национални критерии по този показател, изискващи 30 т. Всичките 6 представени статии са в съавторство с научния си ръководител. Не е представен разделителен протокол, поради което считам, че участието в тях е равностойно. За мен, личният принос на кандидата в тях е несъмнен, като по-долу (виж т. 9) подробно съм описал мотивите за това мое становище.

Забелязано е и едно цитиране, нещо което не е задължително според минималните национални изисквания, но прави добро впечатление.

9. Лично участие на дисертанта

За мен, личният принос на кандидата в проведеното дисертационно изследване е несъмнен. Без да искам да умаловажа значителния принос на научния ръководител (дано проф. Сомова ме извини за написаното), считам, че водещият автор в проведените изследвания е бил докторантът. Основания за това мое мнение са както перфектното познаване на изследваната област от докторанта (виж т. 4), така и личните ми впечатления от краткото ни общуване (виж т. 12).

10. Автореферат

Авторефератът, написан на български и на английски език, е на 32 страници, съответства на изискванията на ПРАСПУ и съдържа основните резултати, получени в дисертационния труд.

11. Критични бележки и препоръки

Нямам съществени критични бележки. Все пак ще изброя няколко не само несъществени, но и субективни.

– Някои английски термини не са преведени добре (според мен) на български. Например *powerset* би трябвало да се преведе като *множество от всички подмножества*, а не като *степенно множество*. Това е може би поради малкото литература на български, използвана от дисертанта. Но наличието на таблица с използваните в дисертацията термини, английския им първообраз и литературния източник, почти напълно отстраняват проблема.

– Някои от въведените в дисертацията термини на мен не ми звучат/изглеждат добре (най-вече на български). Например напреднала граматика, фраза символ, напреднал символ и т.н. Може би *разширена граматика* (*extended* вместо *advanced grammar*), *фраза-символ* или *фразов символ* вместо *фраза символ* и т.н. биха звучали/изглеждали по-добре. Но тук трябва веднага да отбележа, че първо – всички те са строго дефинирани в дисертацията и второ – че не е лесно да бъдат променени, понеже са използвани така в публикуваните вече статии. И тук наличието на таблица с въведените в дисертацията термини, английския им превод и мястото на дефиниране в дисертацията, правят тази критична бележка несъществена.

– Аз бих структурирал обзорната глава малко по-различно, но пък настоящата структура си има своите предимства (виж т. 4).

12. Лични впечатления

Познавам докторанта Николай Ханджийски още от студентските му години, когато съм му преподавал дискретна математика (преди 16 години). За жалост, тогава той не проявяваше особен интерес към тази област. Актуални лични впечатления за него придобих едва около предварителната защита на неговия проект за дисертационен труд, понеже бях включен в разширения катедрен съвет като дългогодишен лектор по дисциплината Дискретна математика. Позволих си да му дам някои препоръки относно дисертационния труд, но единствено и само по оформлението и представянето, които той обмисли, прие и последва, независимо от краткия срок за корекции. От своя страна, по време на няколкото наши кратки дискусии успях лично да се убедя в неговите енциклопедични познания в

изследваната област. Разбрах, че по време на това дългогодишно изследване докторантът е написал огромно количество програмен код и че познава областта еднакво добре както в теоретичен, така и в практически аспект. Впечатлен бях и от неговата логична и критична мисъл, но най-вече от неговия ентузиазъм и жажда за творчески изследвания.

13. Препоръки за бъдещо използване на дисертационните приноси и резултати

Силно впечатление прави фактът, че дисертантът правилно е описал шест конкретни възможности за продължаване на изследванията, описани в дисертационния труд. Със сигурност има и други. Надявам се скоро да видим и съответните публикации, описващи техните реализации. Бих препоръчал също за в бъдеще местата за публикуване да стават все по-престижни, което незабавно ще се отрази и на цитирането на съответните статии.

Предвид сериозните постигнати резултати, както и голямото количество теоретични знания и практически умения, придобити от дисертанта по време на неговите изследвания, предлагам на дисертанта да помисли за издаване на монография на базата на настоящия дисертационен труд, която да бъде публикувана от сериозно издателство.

Пожелавам на докторанта да продължи да работи в тази интересна област и то със същия ентузиазъм.

ЗАКЛЮЧЕНИЕ

Дисертационният труд *съдържа научни, научно-приложни и приложни резултати, които представляват оригинален принос в науката* и отговарят на всички изисквания на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за прилагане на ЗРАСРБ и съответния Правилник на ПУ „Паисий Хилендарски“. Не констатирам плагиатство. Представените материали и дисертационни резултати надхвърлят минималните национални изисквания, приети с Правилника за приложение на ЗРАСРБ.

Дисертационният труд показва, че докторантът Николай Георгиев Ханджийски **притежава** задълбочени теоретични знания и професионални умения по научна специалност Информатика, като **демонстрира** качества и умения за провеждане на изследвания с получаване на оригинални и значими научни приноси.

Поради гореизложеното, убедено давам своята *положителна оценка* за проведеното изследване, представено от рецензираните по-горе дисертационен труд, автореферат, постигнати резултати и приноси, и *предлагам на почитаемото научно жури да присъди образователната и научна степен „доктор“* на Николай Георгиев Ханджийски в област на висше образование 4. *Природни науки, математика и информатика*; професионално направление 4.5. *Информатика и компютърни науки*, докторска програма *Информатика*.

14.05. 2024 г.

Изготвил рецензията:

Гр. Пловдив

Проф. д-р Христо Стефанов Кискинов