

РЕЦЕНЗИЯ

на докторската дисертация на

Тодор Пламенов Чолаков

на тема “Автоматизиране на анализа и обработката на програмен код”

в професионално направление

4.6 “Информатика и компютърни науки”

по научна специалност 01.01.12 Информатика (Софтуерни технологии)

Рецензент: доц. д-р Минко Марков, ФМИ, СУ

2 септември 2015 г.

1 Анализ на постиженията в дисертацията

Дисертацията е оформена в 11 глави и се простира върху 174 страници плюс 5 приложения, списък на фигурите, списък на таблиците и списък на цитираната литература. Съдържа 80 фигури и 12 таблици. Цитираната литература се състои от 84, 5 от които са на български, а останалите, на английски език. Измежду последните, 8 са уеб сайтове, а 7 от източниците са научни и след 2006.

Дисертацията има научно-приложен характер. Докторантът е извършил изследвания в избраната област с множество публикации и освен това е разработил множество алгоритми, реализирал ги е софтуерно и е тествал реализациите върху реални примери.

1.1 Актуалност на проблема

Създаването на средства за автоматизиране на анализа на програмен код е важна задача с огромно практическо приложение. Докторантът разглежда няколко дейности по автоматичен анализ и трансформация на програмен код: поддръжка на стар (legacy) софтуер, дейности при разработка на нов софтуер, и дейности при анализа на еволюцията на софтуера във времето. Това са задачи от първостепенно значение във време, когато хората все повече и повече зависят от софтуера, който ползват. Един от големите успехи на формалната логика на XX век е

доказателството, че един доста ограничен вариант на тази задача, известен като *The Halting Problem*, е алгоритмично нерешим. Алгоритмичната нерешимост в общия случай съвсем не означава обаче, че няма смисъл да се правят опити за автоматизиране на анализа на код на конкретен софтуер. Напротив, това е бързо развиваща област със забележителен прогрес. Както би могло да се очаква при задача, която дори не може да се формулира с математическа прецизност, прогресът се извършва чрез все по-добри и по-добри евристики, които се сравняват чрез тестване върху реални обекти.

Накратко, смятам, че докторантът е направил много удачен избор на проблематика за изследване.

1.2 Познание на състоянието на изследванията и резултатите в избраната област

Докторантът показва добро познание на областта на изследванията. Глава 3 съдържа обзор на средства за автоматизация на софтуерния анализ, описвайки различни категории от инструменти с по няколко примера от всяка категория. Списъкът на използваната литература съдържа десетки заглавия на трудове, които са цитирани адекватно в дисертацията. След представянето на всеки свой нов алгоритъм (глави 4–10), авторът прави обзор на известни алгоритми за същата или подобна задача на утвърдени автори и сравнителен анализ на своите резултати с техните. Докторантът демонстрира отлично познание на областта и на аналогични, вече извършени изследвания и разработки.

Въпреки това мисля, че обзорът можеше да е дори по-широк и да има повече цитирани работи от последните 5 години.

1.3 Методика за извършените изследвания и разработки

Основните теоретични инструменти, използвани от докторанта, са от теорията на формалните езици, теорията на графите и теорията на алгоритмите. Глава 2 “Основни понятия и термини” съдържа необходимите дефиниции и фундаментални обяснения, които се използват в дисертацията. Дефинициите са пълни и прецизни и обясненията са грамотни, което показва точна и ясна мисъл и много добро разбиране на теоретичния фундамент. Особено е акцентирано върху различните видове използвани графи: граф на извикванията, граф на използванията, ресурсен граф и други, като изложението съдържа многобройни цитати на източници, въвели или използвали тези графи.

Всяко изследване на работата на авторските алгоритми е придружено с евристични критерии за оценка на качеството. Това е напълно разбираемо предвид факта, че разработката е евристична по същността си, тъй като строго формални критерии за целта просто не съществуват.

Като пример и инструмент за изследванията си, докторантът ползва езика Java и системата Stratego/XT. Обект на изследването на еволюцията на софтуера е системата Ant—библиотека на Java и част от свободния софтуер Apache.

2 Принос на дисертацията

Накратко, целта на дисертацията е постигната. Извършени са оригинални, нетривиални изследвания, направено е щателно проучване на съществуващи изследвания в областта, конструирани са работещи алгоритми, те са имплементирани и експерименталните резултати са представени на общността. Приносите може да се резюмират така:

- Осъществено е извличане на информация (знание) от програмен код чрез използване на системи за презаписване на терми.
- Осъществен е анализ на програмен код, основан на графи, които се изграждат от кода.
- Създаден и програмно реализиран е алгоритъм за откриване на компоненти на програмен код, като компонентите не са просто синтактични части, които може лесно да си открият със синтактичен анализ, а са смислови единици, които се откриват чрез анализ на графите, изградени въз основа на кода. Резултатът е докладван (заглавие 52 в библиографията) на конференция в България през 2014 г.
- Създадена е методика за анализ на еволюцията на софтуера.
- Създаден и реализиран е алгоритъм за откриване на повтарящ се код, базиран на конструирани въз основа на кода графи.
- Създаден и реализиран е алгоритъм за откриване на така наречените класове-божества, базиран на конструирани въз основа на кода графи.
- Създаден и реализиран е алгоритъм за откриване на класове подходящи за разделяне на части, базиран на конструирани въз основа на кода графи.
- Тези алгоритми са имплементирани и тествани върху реални примери и резултатите от това са представени на общността.

2.1 Публикуване на получените резултати

Докторантът цитира 6 доклада на конференции в България 2012 г., на които са изнесени основните резултати от работата по дисертацията. Всички доклади са в съавторство с научния ръководител на докторанта, доцент Димитър Биров.

3 Оформяне на дисертацията

Дисертацията е написана прецизно и внимателно. Авторът е положил големи усилия, за да постигне формална прецизност при описанието на компонентите на софтуера, на графите, които описват връзките между компонентите и на алгоритмите, които работят върху тези графи. Дисертацията е написана на добър език и с висок стил.

Предложените алгоритми са евристики и не са податливи на прецизно изследване за коректност, на каквото са податливи алгоритмите, чиято евентуална оптималност може да се изрази с математическа прецизност. Затова отсъствието на анализ на коректността в никакъв случай не е недостатък. Авторът е изследвал обстойно сложността по време на своите алгоритми, изразявайки я в голямо-О нотация.

Авторът е положил големи усилия, за да направи текста по-лесно четим, като го е структурирал внимателно, с обособени дефиниции и множество разбивания на булети и под-булети, а също така ясно обособени подглави и под-подглави.

4 Критика на дисертацията

Дисертацията е написана внимателно и с очевидно внимание към детайлите, но въпреки това има правописни грешки, примерно “баазово” на страница 9 и доста други.

Сериозна забележка е използването на нотацията голямо-О при изследването на сложността по време на алгоритмите. Щателното изследване на сложността по време на създадените алгоритми е похвална, но нотацията голямо-О не е подходяща при сравняване на качествата на два алгоритъма. Голямото-О е асимптотична горна граница. Сравняването на две горни граници не е информативно за обектите, на които те са горни граници. Примерно,

Иван е не по-висок от 10 метра, а Стоян е не по-висок от 20 метра.

може да е истина, но от това не следва, че Стоян е по-висок от Иван. Прецизното, в асимптотичния смисъл, сравняване на сложността по време на алгоритми става чрез голямо-Тета (Θ) нотация.

Дребна забележка по оформянето е, че авторът използва знаците “<” и “>” вместо ъглови скоби “⟨” и “⟩”, когато описва идентификатори в контекста на някакъв формален език (примерно, стр. 20, но и на много други места).

И накрая, истински съществена забележка. Задачата, с която авторът се е захванал, в най-общия си вид изглежда практически нерешима. Засега никой не е постигнал автоматизирано извличане на семантика от синтаксис по начин, по който интелигентен човек би го направил. Прочее, има дълбока мирогледна пропаст между тези, които твърдят, че това е принципно невъзможно, защото човешкият ум е повече от работещ компютър, и тези, които твърдят, че е въпрос на време да бъде постигнато, като и двете страни имат убедителни аргументи. Естествено, никой не очаква от дисертационен труд да постигне голям прогрес по тази мъчително трудна задача. Но аз очаквам от автора да покаже, че е наясно със страховитата трудност на общия вид на задачата и да очертае, макар и разказвателно и неформално, какви особености на синтактичните обекти, които той разглежда, позволяват извличане на смислена семантика.

5 Качества на автореферата

Авторефератът отразява точно и пълно основните резултати, постигнати в дисертацията.

6 Заключение

Дисертацията на Тодор Чолаков отговаря на изискванията на Закона за РАСРБ и на релевантните правилници на МОНМ, СУ и ФМИ. Докторантът е написал сериозен научно-приложен дисертационен труд. Вложил е големи усилия и е постигнал интересни резултати.

Поради всичко това, давам положителна оценка и с убеждение предлагам на уважаемото жури да присъди на Тодор Пламенов Чолаков образователната и научна степен “доктор” в област на висшето образование 4.6 “Информатика и компютърни науки” по научна специалност 01.01.12 Информатика (Софтуерни технологии).

Подпис:
доц. д-р Минко Марков,
катедра ИС, ФМИ, СУ