

РЕЦЕНЗИЯ

на дисертационен труд

за придобиване на образователната и научна степен „Доктор”

Област на висше образование: 5. Технически науки

Професионално направление: 5.3. Комуникационна и компютърна техника

Научна специалност: Автоматизирани системи за обработка на
информацията и управление

Автор: Илия Димитров Тронков,
докторант към катедра „Компютърни системи и технологии”
на ТУ-Габрово

Тема: „Проектиране и разработка на външни индексирани структури
за работа в реално време”

Рецензент: доц. д-р Валентин Пенев Бакоев,
ФМИ на ВТУ „Св. св. Кирил и Методий”

Настоящата рецензия е изготвена на основание заповедите на Ректора на ТУ-Габрово № 3-01-309/24.06.2016 г. и № 3-01-339/06.07.2016 г., както и решенията от първото заседание на научното жури, проведено на 29.06.2016 г.

1. Тема и актуалност на дисертационния труд

Изборът на темата на дисертационния труд и нейната актуалност са достатъчно добре формулирани и обосновани в началото на увода и на първа глава от дисертацията. Както посочва авторът, над 90% от всички съществуващи данни са генерирани през последните две години. Източниците на тези данни непрекъснато се увеличават, както и данните, които те произвеждат. Посочени са примери за огромните количества данни (от порядъка на стотици терабайта дневно), получавани, съхранявани, обработвани, анализирани и т.н. в Европейска организация за ядрени изследвания (CERN), във Facebook, при Интернет на нещата (Internet of Things), в New York Stock Exchange. Те достатъчно убедително демонстрират необходимостта от ефективни решения при работата с огромни по обем данни, т.е. в областта, наречена BigData.

В увода на дисертацията са представени:

- *обектът* на изследване – външните индексни структури от данни във връзка с търсенето на по-ефективни решения при BigData;
- *предметът* на изследване – обръщенията за произволен достъп до външните запаметяващи устройства, извършвани от външни индексни структури от данни;
- *целта* на дисертацията – проектиране и реализиране на нова външна индексна структура от данни (ИСД), която да подобри ефективността на работа в сравнение с най-известните съществуващи ИСД.

Представени са произтичащите от целта *задачи, методът* на изследване и *мястото*, където са проведени изследванията (фирма СТС Софт АД, гр. Габрово). Дадени са резюмета на четирите глави от дисертацията.

2. Обзор на цитираната литература

Дисертационният труд е в обем от 106 стр. и е структуриран както следва: Съдържание, Списък на фигурите, Списък на таблиците, Увод, четири глави, Заключение, Приноси, Литература и Авторска справка. Литературата включва 115 заглавия, подредени по реда на цитирането им. От тях 4 са на български език, а останалите са на английски. Почти половината от източниците са в Интернет, като за всеки е посочено кога за последно е достъпван – предимно през последната година. Останалите са книги (сред тях, като изключения, има по-стари, но класически издания), учебници, доклади и статии – най-вече издания и публикации от последните години. Добро впечатление прави познаването на по-важните източници по темата и умелото им използване.

3. Методика на изследване

Методиката на изследване произтича от формулираните в увода задачи. В резюме тя включва:

- унифициран подход за сравнение на различните фамилии от ИСД и сравнителен анализ;
- проектиране и разработване на нова дървовидна ИСД, наречена от автора WaterfallTree;
- формулиране на концепции и изисквания, проектиране и разработване на нова недървовидна ИСД, наречена NoTree;
- двете ИСД WaterfallTree и NoTree да бъдат реализирани като Storage Engines на СУБД STSdb.
- създаване на приложение Database Benchmark за сравнителни тестове на различни индексни структури от данни, както и на сайт за

сравнителен анализ на резултатите, получени с помощта на Database Benchmark приложението.

Изследванията в дисертацията са структурирани в 4 глави. В Глава 1 са разгледани основните характеристики на СУБД, основната роля на ИСД при тях и различните видове ИСД при най-известните БД (Таблица 1). Като се проследява развитието на ИСД се извеждат три концепции, които да залегнат в основата на една нова ИСД. Представен е йерархичният модел на паметта във връзка с дизайна на външните ИСД. В следващите части (раздели 5, 6 и 7) е представена нова дървовидна ИСД, наречена WaterfallTree, разработената и патентована¹ от дисертанта. Тя решава един от фундаменталните проблеми на ИСД – индексването на записи със случайни ключове без влошаване на скоростта на индексване. Структурата WaterfallTree включва в себе си много идеи от областта на индексването и сортирането. Изложението ѝ изяснява различни аспекти на техниките за индексване на данни като изграждане и балансиране на индексно дърво, използване на съобщения и операции за отложено изпълнение, търсене на запис по зададен ключ и др. В изводите в края на главата е посочено мястото на новата ИСД сред съществуващите подобни образци, обуславя се нуждата от изграждането на нов модел ИСД от по-високо ниво.

В Глава 2 са представени теоретичните постановки на разработения от автора интервален модел на ИСД. В Раздел 2.1, на базата на сравнителен анализ на съществуващите ИСД, са изведени техните най-базови компоненти: интервал, ниво от интервали, асоциираната с тях операция за размяна на интервали между нивата от интервали и сливане на интервали. При интервалния модел се акцентира не върху броя на сравненията, които се извършват между ключовете, а върху броя на входно-изходните операции. Направена е оценка на силните и слабите страни на различните индексирани технологии, без да се навлиза в детайлите на техните реализации. Като се разкриват общите фундаментални принципи между различните класове ИСД се достига до идеята за ново решение за индексване на данни, наречено от автора NoTree и патентовано като нова ИСД. В Раздел 2.2 са представени принципите ѝ на работа, основани на интервалния модел. Централно място в NoTree ИСД заема

¹ По-точно подадени са патентни заявки, които са приети и регистрирани по установения ред и чакат изминаването на установения срок, за да бъдат официално признати за патенти.

k -разбиването на множество от интервали (дефинирано в Раздел 2.2.2) и трите свойства на k -разбиванията (доказани в Раздел 2.2.3) – необходими, но недостатъчни условия за минималност на k -разбиване. Представен е алгоритъм за намиране на k -разбиване, което да удовлетворява тези условия. След това е описан алгоритъм за едновременно сливане на интервали. В него е добавена функционалност за прескачане на интервали, която допълнително намалява общия брой на входно-изходните операции, необходими за сливане на интервалите. Дадена е и паралелна версия на алгоритъма – каскадно паралелно сортиране. Нуждата от него произтича от това, че тясното място от входно-изходни операции, премахнато с k -разбиването, се премества на ниво сравнение на ключове при сливане. В останалата част на втора глава са представени други елементи на изграждането на NoTree индекса: цикъл за индексирание, обработка на съобщения, изпълнения на заявки за четене, каскадна паралелизация на сливането на интервали.

В Глава 3 е представена реализацията на NoTree ИСД, реализирана като StorageEngine в базата от данни STSdb 5.0. Тя е проектирана и разработена под ръководството на автора, като първата ѝ версия е анонсирана през 2006 година. В разработката на STSdb са участвали повече от 20 души и са написани десетки хиляди редове програмен код. Клиенти на STSdb са редица световно известни компании и организации като DELL, Omron, NASA и др. За реализацията на NoTree е използвана .NET платформата на Microsoft и програмният език C# (реализациите на авторските алгоритми от втора глава са дадени на електронен носител, приложен към дисертацията). STSdb има три базови слоя – *виртуална файлова система* (представена в Раздел 3.1.1), *StorageEngine* (Раздел 3.1.2) и *приложно-програмен интерфейс* (Раздел 3.1.3). Виртуалната файлова система е отговорна за изпълнението на всички заявки към външните запамятаващи устройства. StorageEngine е ядрото на базата от данни. STSdb поддържа два StorageEngines, основани на WaterfallTree и на NoTree ИСД. Приложно-програмен интерфейс е необходим за достъпване на функционалността на базата от данни STSdb.

В Глава 4 са представени резултатите от сравнителни тестове между различните индексирани технологии: NoTree, B-tree, LSM-tree и Fractal Tree, използвани в някои от най-известните СУБД. За целта е използвано приложението с отворен код Database Benchmark, разработено под ръководството на

автора и достъпно на Sourceforge и Github, за да няма съмнения в резултатите от тестовете. В допълнение е разработен и сайт за сравнителен анализ на резултатите. Освен средата за тестване, в тази част на дисертацията са описани видовете тестове (със случайни и с последователни ключове), методиката на тестване, тестовите данни (от вида ключ, запис и избраните размери), мотивиран е изборът те да бъдат генерирани в реално време (вместо да бъдат записани във файл), както и за броя на записите в тестовите примери: 10^8 и 10^9 . Представени са измерваните параметри и данните за компютърните конфигурации, върху които са проведени тестовете. В Раздел 4.7 са изложени резултатите от тестовете, илюстрирани с поредица от таблици и диаграми. В следващия раздел е направен анализ на получените резултати, който показва превъзходството на ИСД NoTree и STSdb над останалите индексирани технологии и бази от данни.

4. Приноси на дисертационния труд

Приносните моменти в дисертацията са систематизирани от автора както следва:

1) Научни приноси:

- Създадени са нови ИСД: дървовидна WaterfallTree и недървовидна NoTree.
- Въведен е интервален модел, който представлява стилизирано представяне на различните класове от ИСД.
- Въведено е k -разбиване и минимално k -разбиване. Доказани са три необходими, но не и достатъчни условия за дадено k -разбиване да бъде минимално.

2) Научно-приложни приноси:

- Разработен е алгоритъм за намиране на k -разбиване.
- Разработен е алгоритъм за едновременно сливане на k нива от интервали с прескачане и каскадна паралелизация.
- Разработен е цикъл за индексирание на NoTree.

3) Приложни приноси:

- Разработените ИСД WaterfallTree и NoTree са реализирани като StorageEngine за STSdb. Подадени са патентни заявки за WaterfallTree и за NoTree.

- Разработено е приложението Database Benchmark за тест и оценка на производителността на ИСД. Разработен е сайт за сравнителен анализ на резултатите получени от Database Benchmark.
- Разработена е Key/Value база от данни STSdb, която има множество корпоративни клиенти, както е видно от представените документи.

Считам, че приносните моменти в дисертацията са представени аргументирано и коректно. Приемам ги за състоятелни по всички точки. Основания за това ми дават самата дисертация, публикациите, двете заявки за международни патенти, сертификатът Seal of EXCELLENCE, получен за проект по европейската програма Horizon 2020, многобройните корпоративни клиенти на STSdb, както и фактът, че към днешна дата Database Benchmark има хиляди сваляния и излиза на първо място при търсене в Google.

5. Публикации и цитирания на публикации по дисертационния труд

Публикациите по темата на дисертацията са 5 на брой, всички са самостоятелни. Четири от тях са доклади на международни конференции и са публикувани на английски език. Според номерацията в авторската справка, първият доклад е на реномираната конференция „Automation, Quality and Testing, Robotics“ (AQTR), 2014 IEEE International Conference, Cluj-Napoca в Румъния. На конференцията UNITECH'2014, организирана от ТУ-Габрово, Илия Тронков е участвал с 2 доклада. Четвъртият доклад е на ICEST 2016, 28-30 юни 2016, Охрид, Македония. Петата публикация е статия на български език в Journal of the Technical University of Gabrovo през 2016 г. Всички те представят най-съществените моменти и резултати от изследванията в дисертацията.

Както споменах, Илия Тронков (в съавторство) има подадени две заявки за международни патенти в Patent Cooperation Treaty (PCT). Те са упоменати в авторската справка, представена е и пълната документация по тях. Пак там са посочени и софтуерните продукти, разработени под ръководството на Илия Тронков: STSdb, Database Benchmark и сайта за сравнителен анализ на резултатите, получени от Database Benchmark.

Авторът не е посочил данни за цитирания, аз също не успях да намеря такива.

6. Авторство на получените резултати

В дисертацията и в автореферата авторът нееднократно споменава, че базата данни STSdb и сайтът Database Benchmark са дело на колектив от около 20 души. Продуктите са разработвани в продължение на повече от 10 години във фирма СТС Софт АД, в която Илия Тронков е бил технически директор. Като такъв той има водеща роля в идеологията, проектирането и разработването на продуктите. В този смисъл Илия Тронков има най-съществен принос за получаване на резултатите от дисертацията. Самата дисертация е негов научен труд, в който авторството и приносите му са безспорни. Същото важи и за публикациите му по темата на дисертацията. Убедеността ми в това се основава на самите трудове (дисертация, публикации), на представените документи (служебни бележки, сертификата Seal of EXCELLENCE, заявки за патенти), на неговите представяния (на семинари и на предварителната защита), а също и на личните ми впечатления от комуникацията помежду ни при обсъжданията на дисертацията в предварителните ѝ версии.

7. Автореферат и авторска справка

Авторефератът е в обем от 31 страници и съдържа: Обща характеристика на дисертационния труд, Структура и съдържание на дисертационния труд, Авторска справка (със списък на публикации, патенти и продукти), Кратка автобиография и Благодарности. Авторефератът достатъчно добре отразява най-съществените моменти от изследванията, получените резултати и приносите в дисертацията. Те са подходящо илюстрирани с няколко подобрени фигури и таблици от дисертацията.

8. Забележки по дисертационния труд

Имах възможността да се запозная с три предварителни версии на дисертацията. По всяка от тях направих редица критични бележки от различно естество, които бяха отчетени. Към настоящата (окончателна) версия нямам забележки по същество. Като изключения са останали съвсем малко правописни и пунктуационни грешки.

9. Заключение

Настоящата дисертация е един много хубав и рядко срещан пример за съчетаване на теория, практика и бизнес. Тя показва как за актуални и реални задачи от бизнеса могат да бъдат търсени и открити нови иновативни решения

и по този начин практиката да стимулира развитието на науката и обратно. В основата на постигнатите в дисертацията резултати стоят задълбочени теоретични познания и практически умения на докторанта, не само в областта на базите данни, но и в редица свързани области: програмиране, софтуерно инженерство, IT-мениджмънт и маркетинг, работа в екип и т.н. Според мен включените в дисертацията изследвания, предложените решения и резултати съдържат потенциал за поне две подобни дисертации, както и за продължаване на изследванията. Например ще ми бъде интересно да видя резултати и анализ, подобни на тези от Глава 4, като вместо NoTree се тества WaterfallTree.

На основание казаното дотук считам, че представеният дисертационен труд **отговаря** на изискванията на Закона за развитие на академичния състав в Република България. Постигнатите резултати ми дават основание да **предложа** да бъде придобита образователната и научна степен „Доктор” от Илия Димитров Тронков в:

област на висше образование – 5. Технически науки,
професионално направление – 5.3. Комуникационна и компютърна,
техника;
специалност – Автоматизирани системи за обработка на информацията
и управление.

12.09.2016 г.

Подпис:

Заличено обстоятелство,
на основание чл.2 от ЗЗЛД

(доц. д-р Валентин Бакоев)