

УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ

РЕЦЕНЗИЯ

върху дисертационен труд за получаване на научната степен „доктор“ в професионално направление 4.6 „Информатика и компютърни науки“

Автор на дисертационния труд: Стоян Милчев Велев

Тема на дисертационния труд: “Оптимизация на SQL заявки чрез генетичен алгоритъм”

Рецензент: професор, д.т.н. Цветан Атанасов Семерджиев – директор на Институт за научни изследвания и обучение на докторанти в УНИБИТ

1. Актуалност и значимост на разработвания в дисертационния труд проблем

Съвременното технологично развитие се съпровожда от взривообразно генериране на информация и неவிжданo до днес натоварване на системите за нейната обработка. От една страна “чувствителността” на сензорните среди нараства взривообразно, а от друга пропускателната способност на комуникационните канали се увеличава експоненциално, при което на входа на автоматизирани информационни системи (АИС) се образуват значителни по своя обем масиви от данни, чакащи своята обработка. В тези условия съществено значение придобиват характеристиките на базите данни (БД), в които се съхранява все по-голямо количество необработена информация. Времето за достъп, надеждността и защитата от несанкционирани действия, са критични техни параметри. Механизмите за достъп до БД лесно могат да се окажат „тясно“ място и пречка пред превръщането на БД във високоефективни хранилища. Същевременно, тяхното развитие и усъвършенстване е нетривиален и творчески процес на създаване и внедряване на нови методи и алгоритми. Тъй като скоростта на обръщенията към БД определя скоростта на приложенията като цяло, то използването на ефективни алгоритми за изпълнение на заявките придобива особена акту-

алност и значимост. Съответно, оптимизацията на заявките към БД се превръща в актуален оптимизационен проблем.

На работещите с класическите методи, алгоритми и системи за оптимизация на заявки е добре известно, че те са неефективни за заявки с голям брой съединения (над 10-12). За оптимизацията им се използват частни решения с полиномиална сложност, които съществено ограничават вида на заявките, типа на използваната функция за оценка на сложността, конкретния вид на методите за реализация на съединенията и т.н. Общата задача за намиране на оптималния ред за изпълнение на последователност от съединения е NP-пълна. Размерът на пространството на решенията нараства експоненциално при увеличаване на броя на съединените релации и при големи заявки задачата за намиране на оптимално подреждане на съединенията не може да се реши точно. Имайки предвид това обстоятелство, авторът на дисертационния труд е съсредоточил своите усилия върху разработката и изследването на оптимизатор на SQL заявки който да обработва в реално време заявки със стотици съединения. Резултата от намирането на такова решение е повишаване на ефективността на достъпа до данните, в реално време.

2. Обща характеристика и структура на дисертационния труд

Обект на изследване в представения дисертационен труд е управление-то на заявките към релационни БД.

Предмет на изследването е оптимизирането на заявките. Изследователският интерес на автора е фокусиран върху проектирането и реализацията на генетичен алгоритъм, който да е способен да обработва много големи заявки (със стотици съединения), в реално време.

Целта на изследването да се създаде оптимизатор на SQL заявки с голям брой съединения, на базата на разработения алгоритъм.

За постигане на така формулираната цел, авторът на изследването решава следните **задачи**:

- Прави обзор на принципите и методите за обработка на SQL заявки в БД, описва принципната архитектура на натоварените с обработката

на заявките модули на БД и стандартните фази, през които преминава тази обработка.

- Прави обзор и класифицира съществуващите алгоритми за подреждане на съединения, като разкрива техните свойства, ограничения и приложимост.
- Проектира генетичен алгоритъм за оптимизиране на заявки с голям брой съединения преобладаващ част от известните недостатъци на съществуващите класически алгоритми.
- Създава библиотека за реализиране на генетични оптимизатори.
- Проектира и реализира експериментален оптимизатор способен да обработва много големи заявки (със стотици съединения), в реално време.
- Създава система за автоматизиране на експерименти с генетични алгоритми приложени към задачата за оптимизиране на заявки.
- Провежда експерименти за да настрои параметрите и провери работоспособността на реализирания оптимизационен алгоритъм.
- Прави сравнително изследване с аналогични алгоритми описани в литературата

Работната хипотеза на изследването е, че на базата на критичен анализ на принципите и методите за обработка на SQL заявки в БД и на свойствата и приложимостта на съществуващите алгоритми за подреждане на съединения, ще може да се създаде генетичен алгоритъм за оптимизиране на SQL заявки с голям брой съединения, преобладаващ част от известните недостатъци на съществуващите класически алгоритми.

Крайният резултат на изследването е генетичен алгоритъм за оптимизиране на заявки с голям брой съединения и оптимизатор на SQL заявки използващ разработения алгоритъм.

2. Степен на проникване в проблема и оценка за състоянието на решаването му към настоящия момент

Авторът на дисертационния труд е навлязъл в научния проблем до степен, която му е позволила да проведе необходимите изследвания и му е дала възможност да получи важни, практически полезни научни резултати. Използваните и обработени от него литературни източници включват значими публикации по темата и осигуряват реална оценка на постигнатото в научната област, до момента. Те са цитирани от автора коректно и позволяват да се определи мястото и значението на получените резултати в общата конструкция от знания по разглеждания проблем.

3. Относно избраната методика на изследванията

Методиката на изследването се базира на изучаването на натрупания до момента опит и известните добри практики в областта на оптимизацията на SQL заявките в базите данни. Тя включва анализ и оценка на наличните публикации по темата, критична оценка на получените до сега резултати и последователно преминаване от теоретични знания към проектиране и експериментално изследване на конкретен алгоритъм. Избраната методика осигурява ефективно получаване на нови знания и гарантира обективността на получените резултати.

4. Кратка аналитическа характеристика на естеството и достоверността на материала върху който се градят приносите на дисертационния труд

Материалът, върху който се градят приносите на дисертационния труд, е представен под формата на завършено, самостоятелно научно изследване, в конкретна, тясна област на човешката практика, а именно – оптимизация на SQL заявките в базите данни. Той е с общ обем от 173 страници и е структуриран в увод, осем глави и три приложения. Всяка глава е обособена част от работата. Постигнатите резултати, приносните моменти и авторските публикации по темата са описани в последната осма глава. Връзката между главите е осигурена от логиката на изложението и дава цялостна представа за научното изследване. Съдържанието на изследването е подредено както следва:

В първа глава се дефинира изследователската задача за намиране на оптималната подредба на съединенията и се разглежда нейната сложност.

Във втора глава се представят резултатите от направения преглед на основните стъпки при обработката и оптимизацията на SQL заявки в СУБД. Описана е архитектурата на оптимизаторите на заявки, както и функционалността и взаимодействието на техните модули. Представени са методите за достъп до релации и реализация на съединения, профилните формули за основните операции на релационната алгебра, ценовите модели на заявки към БД, както и основните техники за определяне на размера на релациите и разпределението на стойностите на атрибутите. Дефинирано е пространството на решенията и неговото най-важно подпространство – това на ляворекурсивните решения.

В трета глава е направен обзор и класификация на известните стратегии и алгоритми за подреждане на съединения. Разглеждат се детерминистични алгоритми, недетерминистични – рандомизирани и генетични, както и хибриди алгоритми, притежаващи характеристиките на повече от един от тези класове. Главата завършва със сравнителен анализ на разгледаните алгоритми според тяхната приложимост, свойства, ограничения, предимствата и недостатъци.

В четвърта глава е направен кратък исторически преглед на системите за оптимизиране на заявки. Проследено е развитието им от създаването на езика SQL в края на 70-те години на миналия век, до днес. Очертани са посоките на бъдещото развитие на техниките оптимизиране на заявки към БД.

В пета глава е дефиниран оригинален генетичен алгоритъм за оптимизиране на заявки. Разглежда се избраното представяне на решенията и реализацията на операторите за мутация, кръстосване и селекция. Особено внимание се отделя на мотивацията и реализацията на предложения алгоритъм за селекция с адаптивен, динамично изменящ се размер на популацията. Доказано е, че при избор на три генетични оператора, оптимизационният алгоритъм е глобално сходящ.

В шеста глава е описана реализацията на генетичен оптимизатор на SQL заявки. Описана е архитектурата на системата, реализацията на отделните модули и входно-изходния ѝ интерфейс. Дефинират се класовете и

методите на библиотеката за конструиране на генетични оптимизатори и е описан програмният интерфейс. Описани са поддържаните методи за достъп до релация и методи за съединение, ценовият модел и поддържаното подмножество на езика SQL. Специфицира се и форматът на резултата, т.е. синтаксисът на плана за изпълнение на заявката.

В седма глава са дефинирани методите за оценяване на алгоритмите за оптимизация и основните трудности при подобни изследвания. Дадено е описание на използваните тестови модули, включително инструментите за генериране на случайни системни каталози и заявки над тях, които се използват като входни данни за оптимизатора. Приведени са резултатите от сравнителното представяне на алгоритмите и от тестовете за локалната ефективност на операторите за мутация и кръстосване.

В осма глава са обобщени изводите от предходните глави и са описани получените научни, научно - приложни и приложни приноси. Представен е списъкът с публикациите на автора по дисертационния труд. Главата завършва с някои насоки за бъдещо развитие на изследването.

В края на изложението е даден пълен списък на използваната библиография. Като приложения са включени: пълен изходен код на библиотеката за изграждане на генетични оптимизатори, примерен системен каталог и примерен ценови модел.

Достоверността на получените резултати се потвърждава от направените библиографски проучвания и проведените експериментални изследвания. Детайлният анализ на дисертационния труд позволява да се проследят логиката и причинно-следствените връзки на анализа и синтеза, при което не остават съмнения относно оригиналността и достоверността на представения научен материал.

5. Основни научни и научно - приложни приноси в дисертационния труд (същност и характер)

1. На базата на аналитичен обзор на методите за обработка и оптимизация на заявките към БД е предложена изчерпателна класификация на

съществуващите стратегии и алгоритми за подреждане на съединения и са описани техните свойства, ограничения и приложимост.

2. Разработен е оригинален адаптивен генетичен алгоритъм за оптимизиране на заявки с голям брой съединения за полиномиално време, без налагане на ограничения върху формата на ценовата функция. Алгоритмът показва резултати, превъзхождащи по скорост и качество предложените в литературата подобни алгоритми и класическите рандомизирани алгоритми. Нямам информация за подобен *адаптивен* генетичен алгоритъм за намиране на оптимална подредба на съединения, което ми дава основание да го считам за първи от този вид.
3. Получени са експериментални резултати от сравнението на класическите и разработения алгоритми. Сравнението им е по скорост и качество на намираните решения, при различни параметри. В тези опити са конкретизирани параметрите на алгоритъма от гледна точка на качеството на намираните решения за типични заявки.
4. Конкретизирани са кодирането на решенията и генетичните оператори за мутация и кръстосване. При този избор на операторите и използвайки оригиналния оператор за селекция, генетичният алгоритъм е доказано сходящ
5. Проектирана е библиотека за разработване на генетични оптимизатори която позволява бъдещо разширение и лесно интегриране в цялостни системи.
6. Проектирана е разширяема система за оптимизиране на заявки, използваща предложения оригинален алгоритъм
7. Създадена е библиотека на ANSI C++ за разработване на генетични оптимизатори, с която могат да се реализират оптимизационни модули на СУБД с различна архитектура. Постигната е независимост на оптимизиращите модули от конкретната дефиниция и представяне на оптимизационната задача. Осигурена е платформена независимост като библиотеката е компилирана и тествана на различни компютърни архитектури и под различни операционни системи

8. Създаден е оптимизатор на заявки, който обработва заявки с много голям брой съединения в реално време
9. Създадена е система за експериментиране с генетични и рандомизирани оптимизационни алгоритми, която включва инструменти за генериране на случайни речници на данни, случайни релации с много голям брой съединения и случайни заявки над тези релации

6. Оценка на авторското участие в получаването на приносите

Личният принос на докторанта в получаването на резултатите в дисертационния труд е неоспоримо. Авторството му е без съмнение. Дисертацията и публикациите по нея са написани в специфичен стил. Изброените погоре научни приноси са оригинални и са изцяло негово лично дело.

7. Преценка на публикациите по дисертационния труд

Получените в дисертационния труд резултати са публикувани в четири публикации (една статия и три доклада на научни конференции с международно участие). Те позволяват да се получи цялостна представа за получените резултати и осигуряват необходимата публичност на получените приноси и авторските претенции.

8. Използване на получените в дисертационния труд резултати и препоръки за бъдещото им внедряване

Приложимостта на получените резултати е безспорна. Те могат да се използват в процеса на изграждане на бази данни. Дисертационното изследване ще е от полза и за всички дипломанти, специализанти, докторанти и изследователи, работещи в тази сфера.

9. Относно автореферата към дисертационния труд

Авторефератът към дисертационния труд изпълнява функциите си съгласно изискванията на ЗРАСРБ и ППЗРАСРБ. Той вярно и точно отразява дисертационния труд, а именно: заглавието, целта, задачите, получените резултати, изводите, препоръките, възможностите за използване на резултатите, приносите и списъка на публикациите.

10. Критични бележки

В процеса на рецензирането на представения дисертационен труд бяха забелязани някои неточности и неизчерпателно разработени въпроси. Те не намаляват актуалността и значимостта на получените в дисертационния труд приноси, поради което не се обсъждат в настоящата рецензия.

11. Други въпроси

Няма други въпроси, по които да се вземе отношение в настоящата рецензия.

12. Заключение

Цялостната оценка на дисертационния труд позволява да се стигне до заключението, че той има качествата на дисертация за присъждане на образователната и научна степен "доктор". Той е актуално и значимо, самостоятелно научно изследване, в което са получени научни резултати, представляващи интерес за работещите в областта на релационните бази данни. Работата е доказателство за израстването на нейния автор като учен-изследовател, който е постигнал ново квалификационно ниво в своето развитие и е получил актуални и значими приноси.

Така изброените качества на представения дисертационен труд, а също така получените в него приноси, ми дават основание да му дам положителна оценка и да гласувам положително за присъждането на образователната и научна степен "доктор " на Стоян Милчев Велев.

РЕЦЕНЗЕНТ: _____

/проф. Ц. Семерджиев/

17.02.2012 г. / София