

РЕЦЕНЗИЯ

от доц. д-р Красимир Костадинов Марков

доцент в Институт по математика и информатика при БАН

на дисертационен труд за присъждане на образователната и научна степен 'доктор'

в област на висше образование: 4. Природни науки, математика и информатика;
професионално направление: 4.6 Информатика и компютърни науки;
докторска програма: Информатика.

Автор: ДИМИТЪР БЛАГОЕВ БЛАГОЕВ

Тема: СИСТЕМИ ЗА СЕМАНТИЧЕН АНАЛИЗ И ИНФОРМАЦИОННО ТЪРСЕНЕ

Научен ръководител: проф. дмн. Георги Тотков, ПУ „Паисий Хилендарски”.

1. Общо описание на представените материали

Със заповед № Р33-1276/24.4.2012.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 „СИСТЕМИ ЗА СЕМАНТИЧЕН АНАЛИЗ И ИНФОРМАЦИОННО ТЪРСЕНЕ” за придобиване на образователната и научна степен ‘доктор’ в област на висше образование 4. Природни науки, математика и информатика; професионално направление 6. Информатика и компютърни науки; докторска програма Информатика. Автор на дисертационния труд е Димитър Благоев Благоев – докторант в редовна форма на обучение към катедра „Компютърна информатика” на ПУ с научен ръководител проф. дмн. Георги Тотков от ПУ.

Представеният от докторанта Димитър Благоев комплект материали на хартиен носител е в съответствие с Чл.36 (1) от Правилника за развитие на академичния състав на ПУ и включва следните документи:

- молба до Ректора на ПУ за разкриване на процедурата за защита на дисертационен труд;
- автобиография в европейски формат;
- нотариално заверено копие от диплома за висше образование (ОКС ‘магистър’)
- заповеди за записване в докторантура и за продължаване на обучението;
- заповед за провеждане на изпит от индивидуалния план и съответен протокол за издържан изпит по специалността с успех Отличен (6.00);
- протоколи от катедрени съвети, свързани с докладване на готовност за откриване на процедурата и с предварително обсъждане на дисертационния труд;
- дисертационен труд;
- автореферат;
- списък на научните публикации по темата на дисертацията;
- копия на научните публикации;
- списък на цитиранията;
- декларация за оригиналност и достоверност на приложените документи.

2. Кратки биографични данни за докторанта

Докторантът Димитър Благоев Благоев е роден на 05.12.1983 г.

Получава следните степени образование:

- средно образование - физика с английски език в Образцова Математическа Гимназия “Акад. Кирил Попов” (1997 г. – 2002 г.)
- бакалавър по информатика в ПУ “П. Хилендарски” (2002 г. – 2005 г.)
- магистър по информатика в ПУ “П. Хилендарски” (2005 г. – 2006 г.)
- специализирал е в Wolverhampton University (Англия) с отпусната стипендия по програмата Сократ-Еразъм (01.02.2006 – 30.06.2006)

Трудов стаж на докторанта:

- демонстратор (2004-2006), хоноруван асистент (2006-2010) и асистент (2010-до сега) в ПУ „П. Хилендарски”, Факултет по Математика и Информатика. Провеждал е упражнения по дисциплините „Обектно-ориентирано програмиране“, „Операционни системи“, „Информационно моделиране“, „Програмиране на Perl“, „Разпределени приложения“, „Алгоритми и структури от данни“, „Иновационни технологии в обучението“, „Методи за проектиране и съставяне на алгоритми“ и ръководство на дипломанти.
- програмист в Apt Systems ЕООД (2006 – 2007).

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

Разработваният в дисертационния труд проблем е актуален както в научно, така и в научно-приложно отношение. Съвременните технологии за разпространение на информация в световната компютърна среда позволиха ползване на цялата свободно достъпна информация. Обемът ѝ постоянно нараства. Потребителите се нуждаят от семантично филтриране на извичаните документи и по този начин да се улесни анализа им. Това налага средствата, чрез които се намира и извлича необходимата информация, да преминат от количествени методи за търсене към качествени. По същество, както отбелязва докторантът, задачата е документи с неструктурирано съдържание да се трансформират в структури от определен тип (стр.5).

Ето защо, поставената в труда основна цел на дисертационното изследване – да се моделира и създаде система за бързо проектиране на процеси в областта на информационното търсене и семантичният анализ, е актуална и изисква творчески подход за достигането ѝ.

Правилно докторантът е декомпозирал основната цел на четири основни компонента:

1. Изследване на съществуващи методи и системи за информационно търсене и семантичен анализ;
2. Предлагане на общ модел за изграждане на системи за моделиране и процеси решаващи разглежданите задачи;
3. Изграждане на конкретна система, базирана на предложеният модел;
4. Създаване и експериментиране с конкретни описания на процеси, използвайки реализираната система.

Основните задачи на изследването са:

1. Да се проучи общата теория за информационно търсене и семантичен анализ, да се класифицират съществуващите системи и методи;
2. Да се създаде концептуален модел за описание на процеси за информационно търсене и семантичен анализ;
3. Да се проектира система за описание и управление на работни процеси и да се разработи прототип, базиран на съвременни технологии;
 - 3.1. Да се проектира система за управление на работни процеси, базирана на описания модел от точка 2;
 - 3.2. Да се разработи прототип на СУРП (система за управление на работни процеси);

3.3. Да се опишат инструментите, които СУРП предоставя

4. Да се създадат приложения чрез разработения прототип на система за управление на работни процеси, които да докажат приложимостта му за решаване на конкретни практически проблеми;

4.1. Създаване на система за извличане на контактна информация от документи на български език;

4.2. Създаване на методи за автоматизирано генериране на метаданни за учебни обекти;

4.3. Създаване на модул за моделиране и управление на виртуални адаптивни курсове, базирани на moodle.

Както поставената цел с нейните четири компонента, така конкретните задачи, разработени в дисертацията са актуални и на високо ниво.

4. Познаване на проблема

От изложението личи задълбочено познаване състоянието на проблема и критична творческа преценка на литературния материал. Подбраните заглавия в библиографията показват добро ниво на филтрация на литературните източници. От анализа на областта на дисертационното изследване, проведен в първа глава, може да се съди, че докторантът детайлно е проучил и изследвал разглеждания проблем.

5. Методика на изследването

Методиката за провеждане на дисертационното изследване следва класическа схема: изграждане на теоретична основа – създаване на информационен модел – построяване на съответна компютърна система – приложения и оценка на нейната полезност. Избраната методика на изследване позволява постигане на поставената цел и получаване на адекватни решения на задачите, поставени в дисертационния труд.

6. Характеристика и оценка на дисертационния труд

Дисертацията има обем от 134 страници, включвайки 31 фигури и 4 таблици и е разпределена в увод, 4 глави, заключение и библиография от 127 заглавия – книги, статии и доклади на други автори, от които 10 на кирилица, 105 на латиница и 12 интернет-източници, 9 собствени статии и доклади.

В дисертационния труд е направен преглед на типовете системи за семантичен анализ и информационно търсене. Разгледани са съвременните направления в разработката на системи за семантичен анализ и информационно търсене. Изводите от направения обзор са използвани за създаване на концепция на модел за описание на процеси за информационно търсене и семантичен анализ. Изведена е архитектура, удовлетворяваща поставените изисквания. Направена е реализация на архитектурата. Моделът е приложен успешно за разработка на система за извличане на контактна информация от документи на български език, методи за автоматизирано генериране на метаданни за учебни обекти и модул за моделиране и управление на виртуални адаптивни курсове, базирани на moodle.

В първа глава е направен обзор в областта на информационното търсене и извличането на информация, който представя основните модели, задачи, методи и средства за тяхното прилагане и решаване, формулира най-важните изисквания към съвременните информационно-търсещи системи (ИТС), дискутира нерешени проблеми и систематизира съответната специализирана терминология. Специално внимание е обърнато на лингвистичните и статистически методи за анализ и оценяване на съдържанието на текстови документи. Във връзка

с моделите за информационно търсене са разгледани основните методи за тематичен анализ и обработка на текстова информация.

Във втора глава е представен концептуален модел, който може да бъде използван за описание и бързо прототипизиране на лингвистични процеси.

В трета глава е проектирана система за описание и управление на работни процеси използваща предложеният модел. Извършен е анализ на функционалните и нефункционалните изисквания. Описани са основните архитектурни компоненти и реализацията на системата.

В четвърта глава са разгледани приложения на прототипа на система за управление на работни процеси. Обърнато е особено внимание на широката приложимост на разработеният модел. Описани са реализации на система за извличане на контактна информация от документи на български език, подходи за автоматизирано генериране на метаданни за учебни обекти и модул за моделиране и управление на виртуални адаптивни курсове, базирани на moodle.

В заключението е направено обобщение на архитектурата и реализираната система. Отбелязани са основните резултати и са очертани възможностите за бъдещо развитие.

7. Приноси и значимост на разработката за науката и практиката

Основните приноси на дисертационния труд са:

1. Предложен е модел за описание на процеси за информационно търсене и семантичен анализ, сравним, в частност с мощността на мрежите на Петри;
2. Проектирана е архитектура на система за управление на работни процеси, базирана на описаният модел;
3. Реализиран е прототип на системата;
4. Системата е апробирана, като са описани различни процеси от областта на информационното търсене и семантичният анализ, и е използвана за извличане на контактна информация от документи на български език, автоматизирано генериране на метаданни за учебни обекти, моделиране и управление на виртуални адаптивни курсове и извличане на структурирани данни от уеб документи.

Приноси 1, 2 и 4 имат научно-приложен характер, а принос 3 – приложен.

8. Преценка на публикациите по дисертационния труд

Резултати на дисертационния труд са използвани в национални и европейски проекти в областта на е-обучението – МИ-203 (Моделиране на учебни процеси и управление на проекти за е-обучение) и Д002-308 (Автоматизирано извличане на метаданни от е-документи) към Националния фонд за научни изследвания, BG-19 (изследователска програма между България и Германия), и др.

Създадените методики и софтуерни средства са използвани при провеждане на 3 учебни курса с повече от 60 университетски преподаватели от страната, и в редица университетски дисциплини.

В публикациите са отразени основните резултати, получени в дисертационния труд и може да се счита, че резултатите са апробирани в достатъчна степен пред специализирана научна аудитория. Налице е много добро разгласяване на резултатите от изследванията в дисертацията. Авторът е дал списък с 9 публикации, от които 4 лично е докладвал на конференции. Публикациите са: части от книги: 2 бр.; статии в списания: 1 бр.; статии в сборници от конференции: 7 бр. Докладите са: на конференции в чужбина: 1 бр.; на конференции у нас: 3 бр. Списъкът с докладвания на различни форуми се припокрива с основния списък от публикации, поради което той не се рецензира, а се взема под внимание само информативно.

Освен това авторът е дал списък от изследователски проекти (2 броя) по които е работил и в които са използвани резултатите от дисертацията.

Има 6 цитирания на публикации по дисертационния труд.

От деветте труда към дисертацията, 1 е самостоятелен, а 8 са в съавторство.

9. Лично участие на докторанта

Нямам съмнение в самостоятелния принос на докторанта за получаване на цитираните резултати.

10. Автореферат

Авторефератът е направен според изискванията и адекватно отразява съдържанието, основните резултати и приноси на дисертационния труд.

11. Критични забележки и препоръки

Към Рецензията труд могат да се отправят и критики и забележки, като само една има съществен характер - *отсъствие на ясно определени изводи от обзора, обосноваващи целта и задачите на дисертационния труд*. Тези изводи съществуват имплицитно в текста, но липсва заключителен обобщаващ анализ и обосновка.

Останалите забележки са несъществени и очевидно се дължат на претовареността на докторанта и недостига на време. Това са често срещани в дисертациите типове грешки (числата в скоби указват страницата, на която са срещнати):

- **правописни грешки, пропуснати думи, недовършени фрази:** реализираната системи (3); k-мерен вектора (14); за конкретния документа (16); експерименти; в Рунете (19); и. ; потребителя] (24); възникват изисквания (30); висока динамики (34); тях ме умеят (36); висока квалифицирани (39); компресията може да бъде прилагана. (41); отрицателно качеството (45); системата може да се окаже претоварени (45); заявка]. (47); изпращат входни параметри (51); Семантиката на изпълнението на модела е дефинирана както следва: ???? (56); на синтаксисът (94); в основният (94); процеса на анализът. (94); извличане на текстът. (94); намалявайки както размерът на (95); правят анализаторът (96); на анализаторът (98); WfMC1 (стр.109); материали,, дейности (115)
- **общии фрази и пасажии без указване на източника, противоречиви фрази:** стотици системи; хиляди функции; съществува мнение (1) и (19); общата теория за информационно търсене (2); в някои дескрипторни системи (11); древна задача (93); Обикновено се използват два основни подхода ... Общото и в трите подхода е, ... (102);
- **терминологична неизясненост** – объркване на понятията „класификация”, „клъстеризация”, „филтриране” (6) , но вярно представени на стр.7; ILNET е метамодел, система или наредената n-торка ? (глава 2)
- **погрешна номерация на литературни източници:** указване на различни заглавия с едни и същи номера - в точки 4.2 и 4.3 номерацията на литературните източници започва от 1 и несъответства на тази в списъка от библиографията, а на статията, където е публикуван текста; при което без цитиране остават 43 заглавия с номера 53, 55, 57, 59, 62, 63, 64, 65, 66, 67, 68, 69, 70, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 113, 114, 126, 127; неправилно е цитирано заглавие 74: (this volume);
- **объркана номерация при цитирането на фигурите, както и липсваща дума „Фиг.”:** (49); (54); (60); (64); (66); (82);

- **указани несъществуващи елементи на дисертацията:** детерминирани математически модели и модул за получаване на неатрактивни мрежи (3); Приложение 1. (71); (Ландауер и Литман) (28); В настоящата статия (93)
- **указани са само две използвани съкращения** - СУРП (система за управление на работни процеси) и СУБД (система за управление на бази от данни) – (4), но има поне 24 други съкращения: (IR, SMART, TF, IDF, LSI, ИТС, LSA, БД, HITS, SIGIR, NIST, TREC, ACM, ИИБ, КП, ХОРП, ИРП, ОРП, МС, LO, LOM, BPM, LMS; WMS и др.)
- **някои интернет връзки не работят или сочат към източник различен от указания:** заглавия 3; 23; 26; 33; 84.

Тези забележки отразяват до голяма степен и личните виждания и вкусове на рецензента и като цяло не влошават общото много добро впечатление, което прави рецензираният труд.

Представеният труд по тематика и съдържание изцяло се отнася към научната специалност **01.01.12 – Информатика**.

По обем и значимост на научните приноси, оригиналност на творческите постижения и по степен на разгласяване на научните резултати в различни издания и форуми, трудът има качествата и отговаря на изискванията за дисертация за получаване на образователната и научната степен "доктор".

12. Лични впечатления

Познавам докторанта от времето когато беше демонстрант в ПУ. Сега, като рецензент, с удовлетворение отбелязвам академичното израстване на Димитър Благоев като изследовател и преподавател (в ПУ), и успешното достигане до нов качествен етап на неговото научно развитие.

13. Препоръки за бъдещо използване на дисертационните приноси и резултати

В критичен план ще отбележа, че обемът на текста на дисертационния труд би следвало да бъде редуциран, със спестяване на някои подробности (спец. при представянето на съдържанието на конференции, стандарти и пр.).

Качеството и количеството на получените дисертационни резултати дава основания и за препоръка – публикуване в периодични научни издания.

Предложените в дисертацията реализации на модела ILNET са предпоставка за по-широко внедряване в практиката. В тази посока, особено внимание заслужават указанията от докторанта:

- а) провеждане на следващи експерименти в различни предметни области – за автоматизация на процесите по създаване и калибриране на бази от тестови единици и въпроси,
- б) популяризиране (вкл. свободно разпространение и споделяне) на създадените тестови модули – специално в рамките на Moodle-общността.

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“**. Представените материали и дисертационни резултати **напълно** съответстват на специфичните изисквания на

Факултета по математика и информатика, приети във връзка с Правилника на ПУ за приложение на ЗРАСРБ.

Дисертационният труд показва, че докторантът Димитър Благоев Благоев притежава задълбочени теоретични знания и професионални умения по научна специалност 01.01.12 Информатика, като демонстрира качества и умения за самостоятелно провеждане на научно изследване.

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 Димитър Благоев Благоев. в област на висше образование: 4. Природни науки, математика и информатика; професионално направление 4.6 Информатика и компютърни науки; докторска програма Информатика; научна специалност 01.01.12. Информатика

05.05.2012 г.

Рецензент:

(доц. д-р Красимир Марков)