

РЕЦЕНЗИЯ

от д-р Елена Петрова Сомова
доцент в Пловдивския университет „П. Хилендарски“

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. д-мн Георги Атанасов Тотков – Пловдивски университет „П. Хилендарски“

1. Общо описание на представените материали

Със заповед № Р33-1276 от 24.4.2012 г. на ректора на Пловдивския университет „Паисий Хилендарски“ (ПУ) съм определена за член на научно 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 (ОНС) ‘доктор’ в област на висше образование 4. *Природни науки, математика и информатика*, професионално направление 4.6 *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“ с научен ръководител д-мн Георги Атанасов Тотков, професор в Пловдивския Университет „Паисий Хилендарски“.

Представеният комплект материали на хартиен носител (общо 20 документа) е в съответствие с Чл.36 (1) от Правилника за развитие на академичния състав на ПУ, и съдържа:

- 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’), което в комплекта на рецензента не е нотариално заверено;
- заповеди за записване в докторантура, отчисляване от докторантура с право на защита и назначаване на комисии за провеждане на изпити от индивидуалния учебен план;

- протоколи за издържани изпити по специалността с успех Отличен (6.00) и по английски език 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 (9 заглавия на доклади от конференции и статии в списания);
- списък на всички научни публикации;
- копия на научните публикации;
- списък на забелязани цитирания;
- справка за спазване на специфичните изисквания на ФМИ;
- декларация за оригиналност и достоверност на приложените документи
- заповеди за определяне на тема на дисертацията и удължаване срока на докторантурата.

Докторантът е приложил 1 монография и 8 публикации.

2. Кратки биографични данни за докторанта

Докторантът Д. Благоев е завършил Образцовата математическа гимназия „Акад. Кирил Попов” в гр. Пловдив. Завършва бакалавърска специалност „Информатика“ в ПУ „Паисий Хилендарски“ за 3 години и магистърска специалност „Софтуерни технологии” в ПУ, като дипломните му работи за придобиване на бакалавърска и магистърска степен са свързани със създаването на софтуерна система в компютърната лингвистика за визуално редактиране на граматика с вграден генератор на парсери (с научен ръководител проф. дмн Г. Тотков). Още като студент, а по-късно и като докторант, Д. Благоев участва в национални проекти, финансирани от Фонд „Научни изследвания“, свързани с тематиката на дисертационното изследване (с ръководител проф. Г. Тотков). Докторантът има преподавателски опит като хоноруван асистент и асистент във Факултета по математика и информатика на ПУ. Д. Благоев също така има опит в разработването и на фирмени софтуерни системи, т. к. е работил като програмист в софтуерна компания за период от една година.

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

Тематиката на дисертационното изследване е особено актуална като се има предвид огромното количество информация, което се разпространява чрез нейния основен източник – Интернет. Въпреки, че търсенето на информация е относително разработена технология, то

извличането на информация, използвайки семантичен анализ, все още е голямо предизвикателство в научните сфери.

Основната цел на докторанта е да моделира и създаде система за бързо проектиране на процеси в областта на информационното търсене и семантичния анализ. В дисертационното изследване поставената цел е конкретизирана в четири подцели: а) изследване на съществуващи методи и системи за информационно търсене и семантичен анализ; б) предлагане на концептуален модел за описание и управление на работни процеси за информационно търсене и семантичен анализ; в) проектиране и изграждане на прототип на система, базирана на предложения модел; г) създаване на приложения в различни области и експериментиране с конкретни описания на процеси, създадени с реализираната система.

За постигане на поставените цели са формулирани 8 (осем) задачи – предмет на решаване в хода на проведеното изследване.

В резултат е създаден прототип на система за описание и управление на работни потоци, наречен ILNET. Определено достоинство на системата е възможността за моделиране и управление на процеси в различни области, доказано от създадените приложения. Прототипът е приложен в системи за извличане на контактна информация от документи на български език; автоматизирано генериране на метаданни за учебни обекти; моделиране и управление на виртуални адаптивни курсове.

Дисертационното изследване изисква познания в различни области на информатиката – проектиране и създаване на софтуерни системи, описание и управление на процеси, компютърна лингвистика, системи за е-обучение, стандарти за описание на учебни обекти и др.

4. Познание на проблема

От обзора и анализа на проблема в дисертационния труд (общо 44 страници), може да се съди, че докторантът обстойно е проучил и изследвал състоянието на изследванията в областта. Докторантът е направил обзор на изследванията в областта в периода от 1957 г. до 2011 г. Списъкът на използваната литература съдържа 127 заглавия, от които 10 са на кирилица, 117 – на латиница и 14 са интернет източници. 17 от цитираните източници са от последните 5 години (от 2007 г.).

5. Методика на изследването

Методиката за провеждане на дисертационното изследване следва класическата схема: изграждане на теоретична основа – моделиране на предметната област – построяване на съответен компютърен модел – реализация на софтуерна система – провеждане на експерименти и оценяване на резултатите. Самото изследване, създаването на съответния

прототип ILNET и приложените софтуерни експерименти са поддържани от няколко проекта към Националния фонд „Научни изследвания“.

Избраната методика позволява постигането на основната цел на дисертационното изследване и изпълнение на поставените задачи.

6. Характеристика и оценка на дисертационния труд

Текстът на дисертационния труд (общо 138 нестандартни страници, с около 30 реда по 90 знака) включва основен текст (121 страници), който илюстрира основното изложение с допълнителни схеми, фигури и представяния на създадените софтуерни средства и системи и библиография (12 страници).

Текстът на дисертационния труд е организиран в увод, 4 (четири) глави и заключение.

В Глава 1. (Обзор на съвременни методи и системи) е направен обзор в областта на информационното търсене и извличане на информация. Разгледани са подробно видовете и характеристиките на съвременните системи за информационно търсене.

Глава 2. (Метамодел за описание на процеси ILNET) представя концептуален модел, който може да бъде използван за описание на процеси за информационно търсене и семантичен анализ, а Глава 3. (Архитектура и реализация на система за управление на работни процеси) показва проектирането на прототип на система за описание и управление на работни процеси, използваща предложения концептуален модел. Описани са основните компоненти на архитектурата и реализацията на системата.

В глава 4. (Приложения на разработения модел) са разгледани три приложения на прототипа на системата за управление на работни процеси, а именно: система за извличане на контактна информация от документи на български език, автоматизирано генериране на метаданни за учебни обекти и модул за моделиране и управление на виртуални адаптивни курсове, базирани на Moodle.

В заключението са представени четирите приноса на дисертационния труд и насоките за бъдещо развитие.

7. Приноси и значимост на разработката за науката и практиката

Основният *научен принос* на дисертационното изследване е предлагания формален модел ILNET на работен поток за моделиране на процеси в различни области, съвместим с Мрежи на Петри.

Като *научно-приложни приноси* на изследването могат да бъдат определени създадените:

– архитектура на система за моделиране и управление на процеси;

- графичен редактор за визуално моделиране на процеси;
- среда за изпълнение на моделирани процеси;
- приложна система за извличане на контактна информация от документи на български език;
- приложна система и методи за автоматизирано генериране на метаданни за учебни обекти;
- приложна система за моделиране и управление на виртуални адаптивни курсове, базирани на среда за е-обучение Moodle;
- приложна система СЛОГ за трансформиране на български текст в реч.

Системата СЛОГ е отличена на университетско изложение ЕКСПО'2003.

Като *приложни приноси* на дисертационното изследване отбелязвам направения анализ на концепциите и методите на информационно търсене и семантичен анализ и съвременните системи за информационно търсене, както и проектирането и реализацията на прототипа на система за управление на работни потоци ILNET.

Резултати на дисертационния труд са използвани в два национални проекта в областта на е-обучението и автоматизираното генериране на метаданни за е-документи.

8. Преценка на публикациите по дисертационния труд

В публикациите на докторанта са отразени основните резултати, получени в дисертационния труд и може да се счита, че те са апробирани в достатъчна степен пред специализирана научна аудитория. Резултати на дисертационното изследване са представени в 10 публикации, от които 2 (две) – в списание (Information Technologies and Control и Information Science & Computing), 1 (една) – като част от два раздела в монография, 2 (две) – в трудове на международни научни конференции, проведени в България (Recent Advances in Natural Language Processing и Computer Systems and Technologies) и 4 (четири) – в трудове на национални научни конференции ("Образованието в информационното общество", „10 години катедра Компютърни системи към ТУ – София, филиал Пловдив“, конференция за студенти, докторанти и млади научни работници и конференция на НБУ). 4 (четири) от публикациите са на английски език, а останалите 5 (пет) - на български. 2 (две) от публикациите по дисертационното изследване са цитирани общо 6 (пъти).

9. Лично участие на докторанта

Д. Благоев е единствен автор на 1 (една) от 9-те публикации по дисертационната тема, в 2 (две) има един съавтор, в 4 (четири) е с още двама съавтори, и в 2 (две) – с трима и повече съавтори. В 3 (три) от съвместните работи докторантът е посочен като първи съавтор.

Нямам съмнение в личния 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

На стр. 49 от дисертацията, както и в публикация No 2 от списъка с публикациите по дисертационния труд, се споменава за четвърто приложение на създадения модел – система СЛОГ за озвучаване на български текст, която не е представена подробно от докторанта. Поради значимостта на такава система дори и не само от 'информатична' гледна точка, препоръчвам включването на това приложение към основния текст на дисертационния труд.

Разработените приложни софтуерни системи са тествани с контролни и реални данни, както е отбелязано на стр. 117, но не е предоставена информация в дисертационната работа за тези експерименти или получените резултати са представени съвсем накратко (стр. 98 и стр. 102).

Следващите забележки са несъществени по отношение на качеството на научното изследване и касаят по-скоро текстовото оформление на работата.

Не са цитирани литературните източници на някои от представените факти в обзора, като напр.: моделите, основани на класификатори (стр. 12); лингвистичните методи за българския език (стр. 23); тестовете на Дируестър и Харигман (стр. 27); извличане на информация от колекция от текстови документи на Ландауер и Литман (стр. 28); библиографско търсене (стр. 33) и тематично търсене (стр. 34).

Списъкът със съкращения е почти празен, защото повечето съкращения не са изнесени в него. Съществуват сгрешени препратки към фигури (напр. препратката на стр. 60 към Фиг. 2 би трябвало да е към Фиг. 4). Таблиците в Глава 3. не са номерирани. Всички цитирания на литературни източници в Глава 4. са сгрешени.

Библиографията не е записана по стандартните изисквания и подредена по азбучен ред.

Всички тези забележки затрудняват четенето на дисертацията, но по никакъв начин не намаляват нейната значимост.

12. Лични впечатления

Познавам Д. Благоев още като студент в 1-ви курс на спец. Информатика към ФМИ, когато се включва в изследванията, провеждани в областта на компютърната лингвистика в кат. „Компютърна информатика“. От този период датират и първите му публикации, участия в научни конференции и национални проекти. Тематиката на отлично защитените от него дипломни работи (за бакалавърска и магистърска степен) е свързана със създаването на

софтуерна система в компютърната лингвистика за визуално редактиране на граматика с вграден генератор на парсери.

13. Препоръки за бъдещо използване на дисертационните приноси и резултати

Развитието на функционалността на прототипа ILNET и използването му в други области на приложение са сред най-важните перспективи за развитие на изследването. Дисертационното изследване дава големи възможности за изследвания в областта на моделирането и управлението на бизнес процеси като напр. моделиране на компонентите на известните стандарти за описание и управление на процеси (BPMN, UML, EPC, BPEL, XPDL).

ЗАКЛЮЧЕНИЕ

Независимо от критичните бележки, определено считам, че рецензията дисертационен труд съдържа *научни, научно-приложни и приложни резултати, които могат да бъдат разглеждани като приноси в науката* и отговарят на изискванията за присъждане на образователната и научна степен 'доктор' в Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за прилагане на ЗРАСРБ и съответния Правилник на ПУ.

Проведеното дисертационно изследване показва, че докторантът Димитър Благоев Благоев притежава основни теоретични знания и професионални умения в област на висше образование 4. *Природни науки, математика и информатика*, професионално направление 4.6 *Информатика и компютърни науки*, докторска програма *Информатика*, като демонстрира качества и умения за самостоятелно провеждане на научни изследвания.

Поради гореизложеното, убедено давам своята *положителна оценка* за проведеното изследване, представено от рецензираните дисертационен труд, автореферат, постигнати резултати и приноси и *предлагам на почитаемото научно жури да присъди на Димитър Благоев Благоев 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*.

11.05. 2012 г.

Рецензент:

(доц. д-р Елена Сомова)