

РЕЦЕНЗИЯ

от доц. д-р Юлиана Пенева Пашкова

преподавател в Нов български университет, деп. “Информатика“

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: Мария Минева Жекова

Тема: „Естественоезиков интерфейс за информационна система от тип ‘електронен университет’

Научен ръководител: проф. д.м.н. Георги Атанасов Тотков

1. Общо описание на представените материали

Със заповед № РД-81-616 от 25.03.2022 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 Естественоезиков интерфейс за информационна система от тип ‘електронен университет’ 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 Мария Минева Жекова – докторант в редовна форма на обучение към катедра „Компютърна информатика“ с научен ръководител проф. д.м.н. Георги Атанасов Тотков.

Представеният комплект материали на хартиен носител (общо 9 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 откриване на процедура за защита на дисертационен труд;
- автобиография в европейски формат;
- протокол от катедрен съвет, свързан предварително обсъждане на дисертационния труд;
- автореферат;
- декларация за оригиналност и достоверност на приложените документи;
- справка за спазване на специфичните изисквания на ФМИ;

- списък на научните публикации по темата на дисертацията (6 заглавия на доклади от конференции и статии в списания);
- дисертационен труд;
- копия на научните публикации.

Докторантът е приложил 6 публикации.

2. Кратки биографични данни за докторанта

Докторант Мария Жекова е завършила три магистърски специалности „Математика и информатика – педагогически профил“ и „Математика“ (със специализация по Информатика) в ПУ „Паисий Хилендарски“ както и магистърска специалност по „Финанси“ в Технически университет - гр. Варна. От 2016 г. до 2019 г. тя е хоноруван преподавател в катедра „Компютърна информатика+ във Факултета по математика и информатика на ПУ, а от 2020 г. е асистент в катедра „Компютърни системи и технологии“ в Технически факултет на Университет по хранителни технологии - гр. Пловдив. Мария Жекова е работила като администратор на бази от данни и програмист и има опит при разработване, внедряване и поддържане на софтуерни продукти като част от екипа на „Информационно обслужване“ АД - клон Пловдив.

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

Разработваният в дисертационния труд проблем е свързан със задачата за създаване на интерфейси на естествен език, чрез които потребителите по удобен за тях начин да формулират заявки към база от данни - основен елемент на всяка информационна система. Научните резултати по тематиката на рецензираната разработка са от несъмнен практически интерес за бизнеса поради тоталната му електронизация. Достъпът до данни, съхранени в база е дейност, която трябва да бъде изпълнявана от персонала на всички нива на мениджмънт в организацията, а не само от специално обучен персонал. Това подчертава актуалността на усилията за повишаване на ефективността на управленските дейности чрез улесняване на крайните потребители при достъп до данни.

Както повечето научни разработки в това направление, в дисертацията се създава естествено-езиков интерфейс върху ограничена предметна област, но са търсени обобщения (основно в методически план) за приложимост към произволна информационна система. Предизвикателство е и генерирането на заявки от текст на български език.

Целта на дисертационното изследване е да се предложат, изследват и апробират средства, подпомагащи реализирането на интерфейс за формулиране на заявки на български

език към дадена информационна система. Поставените цел и произтичащите от нея задачи са в съответствие със съвременността на проблема.

4. Познаване на проблема

Прегледът на цитираната литература (82 заглавия, от които 8 са на кирилица, 74 – на латиница) позволява да се твърди, че докторантът е навлязъл добре в проблематиката. Списъкът е представителен по брой и по разпределение на авторите, включени са публикации от последните 5 години (19), което е особено важно в областта “информатика”. Съотношението между чуждестранни и български източници намирам за напълно естествено. Обзорът на източниците оставя впечатление за добри познания по разглежданата проблематика.

Изготвеният от докторанта обзор систематизира базирани на изкуствен интелект подходи и технологии за обработка и систематичен анализ на естествен език. Разгледани са начините за трансформиране на естественоезикови въпроси в специфични за база от данни заявки. Проучването на разработените софтуерни системи с интерфейс на естествен език показва, че липсват методики за създаване на подобен такъв, който ефективно да се справя със задачата по извличане на данни от бази, свързани с различни информационни системи, но в една и съща предметна област.

Въз основа на това докторантът избира хибриден подход, който включва методика за моделиране на предметната област и софтуерен инструмент за транслиране на естествено-езиковата заявка.

5. Методика на изследването

Дисертацията следва традиционната за научно-приложна разработка схема: преглед на източниците и анализ на проблематиката; формулировка на целите и задачите; проектни и методически решения; софтуерно приложение; експеримент; анализ на резултатите.

След проведеното аналитично проучване в областта (вкл. използван инструментариум и налични реализации) и извличайки от това полезни и нетривиални задачи, дисертантът разработва подходящо обоснован модел на интерфейс на естествен език към базата от данни на информационната система. Предложена е методика за информационно моделиране на предметната област и архитектура на софтуерно средство за естествено езикова обработка на текст. Реализирани са съответните експериментални интерфейси за различни информационни системи в област „висше образование“ Проведени са експерименти с формулирани на български език задания към релационни бази данни за информационните системи, обслужващи висшето образование.

Може да се твърди, че докторантът Мария Жекова е изпълнила методически коректно научно изследване и показва умения за систематични научни изследвания.

6. Характеристика и оценка на дисертационния труд

Дисертационният труд е с общ обем от 144 страници, от които 136 страници представляват основен текст вкл. 2 приложения, 6 страници – библиография и 2-страници – допълнителна информация. Състои се от Списък на използваните съкращения, Списък на таблиците, Списък на фигурите, Увод, 4 (четири) глави, Приложения, Заключение, Списък на публикациите, Списък на използвана литература и Декларация за оригиналност и достоверност. Библиографията включва 82 източника, основно на английски език, цитирани коректно в текста на дисертацията. Приложенията съдържат разработените от докторанта лингвистичен модел за експеримент и резултати от проведени тестове. Работата е добре илюстрирана с подходящо подбрани фигури (52) и таблици с резултати (32), и като обем е напълно в нормите. Всяка глава завършва със заключение, в което са представени получените резултати.

В увода се обосновава актуалността и значимостта на избраната тема. Представени са целта и задачите на научното изследване, описва се структурата на дисертацията. Формулираните задачи произтичат от целта на дисертационното изследване: да се предложат, изследват и апробират средства, подходящи за автоматизация на процеса по създаване на интерфейси на естествен език към информационна система в област“висше образование“.

Глава 1 съдържа обзор на литературата по дисертационния проблем. Систематизирани са основните понятия и определения, разгледани са различни подходи за обработка и анализ на свободен текст на естествен език. Особено внимание е отделено върху съществуващи инструменти и системи с естественоезиков интерфейс както и върху методите за автоматизиране на подобни обработки. Проучени са начините за проектиране на интерфейси на естествен език, в които определени лексикални елементи се използват като входни параметри в заявки към базата с данни. Направен е обоснован избор на подход (хибриден) за създаване на естественоезиков интерфейс към информационна система. Счита поставената цел и произтичащите от нея задачи, формулирани след анализ на изследванията в областта, за целесъобразни.

В Глава 2 е представен общ модел на процес за създаване интерфейс на естествен език за дадена информационна система. Процесът включва следните стъпки: проучване на потребителските изисквания и анализ на базата от данни и услугите, предоставяни от

информационната система; моделиране на предметната област; обработване на потребителската заявка, формулирана като текст на български език; трансформиране на текста в заявка към базата от данни. Предложена е методика за моделиране на предметната област на информационната система, с помощта на която се генерира списък от понятия, описващи базата от данни; изгражда се лингвистичен модел въз основа на получения списък и се формират езикови шаблони за отговор на заявки на естествен език. По този начин се осигурява функционирането на естественоезиковия интерфейс и възможността за извличане на данни от базата. Методиката е инвариантна по отношение на предметната област и връзката ѝ с естествения език. Избрана е архитектура (трислоен клиент-сървър) на софтуерно средство с интерфейс на естествен език.

Глава 3 разглежда реализацията на софтуерния инструмент за интерфейс на естествен език, който може да се приложи към различни информационни системи. Задачата му е да преобразува потребителската заявка от български или английски език в израз на езика за работа с релационни бази от данни SQL. За изпълнение на поставената задача са предвидени два основни модула. Първият от тях осъществява анализ и обработване на въведените лингвистични данни, като използва метаданни от библиотеки, които позволяват извличане неявна информация за понятията от предметната област. Този модул е основен за намирането на граматични модели, конкретизиращи даден контекст и осигуряване на комуникацията с клиента. По-точно, първият модул извършва сегментация на изречението, проверка за принадлежност на отделните езикови единици (токени) към модела на предметната област и заменя избраните токени с основни форми от собствената езикова база. Вторият модул аотира входните токени като допълнително етикетира участващи в него единици от потребителското задание до параметри, изграждащи заявка към базата от данни. Провежда се търсене на отношения и асоциации между получените токени. В краен резултат, в зависимост от структурата и елементите на потребителския въпрос, като се прилагат разработените структурни шаблони, се формира SQL заявка към базата.

В Глава 4 са дискутирани резултати от два експеримента, проведени за целите на дисертационното проучване с различни информационни системи в област „висше образование“. За първия от тях е създаден интерфейс на естествен език по предложената от докторанта методика към информационна система от тип „електронен университет“. Проведени са три теста с достатъчно на брой въпроси, позволяващи настройване на някои характеристики (напр. добавяне на нови модели на въпроси) в зависимост от спецификата на информационната система. Вторият експеримент обхваща разработването на естественоезиков интерфейс към база от данни на НАОА, свързана с оценяване на

качеството на висшето образование в България. За целта създаденият в първия експеримент модел на предметната област е допълнен с още езикови единици, синонимни представяния и специфични шаблони на въпроси. Проведените експерименти доказват приложимостта на предложения подход за разработването на интерфейс на естествен език към база от данни.

В заключението са представени направените от дисертационното изследване изводи и са систематизирани приносите на докторанта. Предложени са перспективи за развитие на тематиката.

7. Приноси и значимост на разработката за науката и практиката

Приносите в дисертационното изследване обхващат създаване на модел на процеса за изграждане интерфейс на естествен език за дадена информационна система и установяване на съответна методика за неговото осъществяване. Предложена е архитектура на софтуерно средство за достъп до база от данни чрез заявки на естествен език и е реализиран негов прототип. Представени са резултати от проведени еталонни тествания с различни информационни системи в област „висше образование“.

Приемам заявените от докторанта приноси и ги характеризирам като научно-приложни.

Основен извод от извършените проучвания е, че разработените модели и средства са достатъчно универсални, което ги прави приложими за автоматизиране на процеса по изграждане на естественоезикови интерфейси към информационни системи с различни предметни области. Адекватността им е демонстрирана с различни бази от област „висше образование“. За мен няма никакво съмнение относно използваемостта на приносите в практиката, но препоръчвам в бъдеще така описаните експерименти да се разширят и задълбочат.

Перспективите за развитие на дисертационното изследване могат да се обобщят както следва:

1. Добавяне на нови функционалности към разработеното софтуерно средство за достъп до база от данни чрез заявки на естествен език.
2. Повишаване дружелюбността на естественоезиковия интерфейс чрез диалог и визуализация на отговора.
3. Развитие и обогатяване на модела на предметната област както и подобряване на лингвистичния модел.
4. Прилагане на разработените средства в други предметни области.

8. Преценка на публикациите по дисертационния труд

Докторантът е представил списък от шест публикации по дисертационното изследване. В тях са отразени основните резултати, получени в дисертационния труд. Докладвани са на катедрени семинари, на национални и международни научни форуми. Резултатите са представяни и обсъждани пред специализирана научна аудитория. Като брой и представителност публикациите са достатъчни. Изпълнени са специфичните изисквания на ФМИ.

Четири от публикациите са в съавторство и две са самостоятелни. На английски език са представени три от публикациите. Три от публикациите са в специализирани научни списания (*International Journal on Information Technologies and Security*, *Knowledge for research and practice International Journal Scientific Papers*, *Научни трудове на Съюза на учените в България – Пловдив*) и три – в трудове на конференции (две национални - *Годишник на Националната научна конференция за студенти и докторанти, Пловдив 2019*, *Сборник доклади на Национален младежки форум пролет „Наука, технологии, иновации, бизнес”, Пловдив 2019*) и една международна - *IOP Conference Series: Materials Science and Engineering*). Две от публикациите са индексирани в международни бази от данни Scopus и Web of Science, и двете са публикувани в издания с SJR.

Авторът е не представил списък с цитирания.

9. Лично участие на докторанта

Четири от публикациите (от общо 6), представени по дисертационния труд са в съавторство и две – самостоятелни, като във всички от тях Мария Жекова е водещ автор. В представената документация липсва разделителен протокол относно авторското участие. Приемам, че авторството на докторанта е преобладаващо.

10. Автореферат

Намирам, че авторефератът в обем 36 страници е съставен съобразно изискванията и представлява състоятелно кратко изложение на дисертационния труд, в което основните положения и приносите са представени идентично, като са обобщени постигнатите резултати.

11. Критични забележки и препоръки

По текста могат да се направят някои терминологични забележки, но предвид разликата в превода от английски на отделните понятия, не считам, че това е силно наложително. В тази връзка намирам за удачно посочването на съответния термин на английски език, за да се

избягнат неправилни интерпретации. За съжаление, в работата има някои технически грешки, списъкът със съкращения не е пълен. Неудачно е също така използването на източник като Wikipedia в дисертационно изследване. Препоръчително е също така спазването на изискванията на БДС за библиографско цитиране – електронните източници са без година и дата на последно влизане на потребителя.

12. Лични впечатления

Не познавам докторант Мария Жекова и не мога да споделя лични впечатления.

13. Препоръки за бъдещо използване на дисертационните приноси и резултати

Дисертационните приноси и резултати могат да се използват при изграждането на естественоезикови интерфейси към различни информационни системи. Предложената методика е инвариантна по отношение на предметната област и връзката ѝ с естествения език. Интересна перспектива за бъдещо развитие на постигнатите резултати представлява повишаване дружелюбността на естественоезиковия интерфейс чрез диалог и визуализация на отговора. Развитието и обогатяването на модела на предметната област както и подобряване на лингвистичния модел ще позволят подобряване на връзката „понятие-обект в базата от данни“. Нямам никакво съмнение относно използваемостта на приносите в практиката.

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 Мария Минева Жекова **притежава** задълбочени теоретични знания и професионални умения по научна специалност 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 Мария Минева Жекова в област на висше образование: 4. Природни науки, математика и информатика, професионално направление 4.6 Информатика и компютърни науки, докторска програма „Информатика“.

26.04. 2022 г.

Изготвил рецензията:

доц. д-р Юлиана Пенева