

СТАНОВИЩЕ

от проф. д.н. Иван Асенов Мирчев

за дисертационен труд на тема:

”Семантичен уеб - модели и приложения“,

на Надежда Пламенова Борисова

за присъждане на образователната и научна степен “доктор“

в област 4. Природни науки, математика и информатика

Професионално направление 4.6 „Информатика и компютърни науки“

1. ОБЩИ ПОЛОЖЕНИЯ

Дисертационният труд се състои от увод, шест глави, заключение, списък на авторските приноси, две приложения, списъци на използвана литература, фигури, таблици и съкращения, в общ обем от 129 страници.

Цитираната литература (85 заглавия) е подредена според азбучен класификатор и включва източници на английски и български езици.

2. АКТУАЛНОСТ НА ПРОБЛЕМА

Научните изследвания и резултати на докторанта се отнасят до актуални въпроси, свързани със семантичните уеб технологии. Развитието на съвременните методи, свързани с изграждане на машинно обработваем слой от метаданни над съществуващата мрежа от хипертекстови документи, с което се предоставя възможност за търсене и споделяне на ресурси по смисловото им съдържание, има голямо значение в компютърните науки.

Независимо от множеството изследвания и разработки в тази област, проблемите, свързани със семантичен уеб пораждат нови подходи за създаване на ефективни модели и алгоритми, голяма част от които все още не са решени.

Вниманието на докторанта е насочено към изследване, свързано с решаване на задача за *извличане на знания от неструктуриран текст, събразена със спецификата на българския език.*

Считам, че темата е актуална и съдържа нови моменти, които могат да бъдат използвани при бъдещето развитие на системите за семантичен уеб и по-специално за български език.

3. АНАЛИЗ НА ДИСЕРТАЦИОННИЯ ТРУД

В дисертационния труд авторът си поставя следната основна цел:

Изследване на възможностите за онтологично базирано извличане на информация от неструктуриран текст на български език.

За постигане на поставената цел са формулирани пет конкретни задачи.

В първата глава е изложено направеното проучване за съвременното състояние на проблема, като са представени класификации на различните видове онтологии, както и основни дефиниции и твърдения, използвани в следващите глави. Описани са конкретни приложения на онтологиите в извличането на информация и управлението на знания.

Във втора глава са представени класификации на задачите, свързани с извличане на информация. Разгледани са и са анализирани отделните фази на процеса на IE (Information Extraction).

В трета глава са проучени и анализирани платформи с отворен код за анализ на текст. Направен е извод, че за анализ на текст на български език и извличане на съдържание е най-подходяща платформата GATE.

В четвърта глава се предлага подход за автоматично маркиране на неправилно съгласуване на прилагателно име със съществително име в български текст. Показано е как при този процес могат да се използват функционалните възможности на платформата GATE. Прилагането на регулярни изрази върху анотации в GATE предоставя широк набор от техники за решаване на реални проблеми, свързани с компютърната обработка на естествен език.

В пета глава се описва реализацията на разработения от автора инструмент за лематизация на български текст в GATE, наречен BGLemmatizer. За лематизатора са използвани лексикални ресурси със свободен достъп BG Office и Bulgarian-language Wiktionary.

В шеста глава се описва предложеният от автора подход за онтологично базирано извличане на понятия от неструктуриран текст на български език, който може да се използва като основа за автоматично извличане на метаданни.

В заключението накратко е обобщено проведеното изследване и са направени съответните изводи.

4. ЛИЧНОТО УЧАСТИЕ

Анализът на дисертацията и представеният списък на публикациите, свързани с дисертационния труд, ми позволяват да направя заключение, че авторът познава много добре състоянието на проблема и лично е извършил изследванията и разработил дисертацията си.

5. АВТОРЕФЕРАТ

Авторефератът съдържа 50 страници и е съставен съгласно изискванията. Съдържанието му правилно отразява основните моменти на дисертационния труд, актуалността на темата и проведените изследвания, като са отразени и получените резултати.

6. ПУБЛИКАЦИИ, СВЪРЗАНИ С ДИСЕРТАЦИЯТА

Представеният списък на публикациите, свързани с дисертацията включва 5 наименования. От тях една е самостоятелна и е приета за публикуване в списание “Information Technologies and Control“, а останалите са отпечатани в сборници от международни конференции.

7. НАУЧНИ И НАУЧНО-ПРИЛОЖНИ ПРИНОСИ

Докторантът е формулирал четири основни приноса. Приемам тези приноси и считам, че те биха могли да бъдат обобщени по следния начин:

Научно-приложни:

Към тях спадат приносите [1] и [4].

Приложни:

Към тях се отнасят приносите [2] и [3].

9. ЗАБЕЛЕЖКИ И ПРЕПОРЪКИ

Нямам принципни забележки към дисертационния труд. Редица грешки и недостатъци, забелязани от мен, бяха отстранени преди и след предварителното му обсъждане в катедра „Информатика“. По представения вариант на дисертацията могат да се отбележат някои неточности, стилови и редакционни пропуски:

- Част от представените фигури са на английски език, а другата част са на български език. Хубаво би било да бъде възприет еднакъв начин на оформянето им.
- Не навсякъде в дисертацията е отбелязан авторският принос.
- Заключение е кратко и не отразява всички моменти на проведеното изследване.

Имам следната препоръка:

- Резултатите от проведеното изследване да бъдат отразени в учебно пособие.
- Да се засили публикационната дейност както у нас, така и в международни реномирани издания.

10. ЗАКЛЮЧЕНИЕ

Представеният дисертационен труд съдържа оригинални идеи, които могат да бъдат използвани при проектиране на нови системи за *Семантичен уеб*. Докторантът показва задълбочени познания в изследваната област. Като цяло дисертационният труд е изпълнен на добро професионално ниво и отговаря на изискванията на закона за присъждане на образователната и научна степен „**доктор**”.

Считам, че представената дисертация като обем, ниво и приноси отговаря на изискванията на ЗРАСРБ, ППЗРАСРБ и ВПРАС в ЮЗУ „Неофит Рилски“ за придобиване на ОНС „доктор“ по ПН 4.6 „*Информатика и компютърни науки*“.

18.09.2015 год.

Подпис:

Проф. д.н. Иван Асенов Мирчев