

## РЕЦЕНЗИЯ

върху дисертационен труд за придобиване на образователна и научна степен “доктор” .

Автор на дисертационния труд: ас. маг. Надежда Пламенова Борисова.

Тема на дисертационния труд : “Семантичен уеб-модел и приложения” по научно направление 4.6 “Информатика и компютърни науки” ,

Рецензент: проф. д-р Нина Василевна Синягина.

Дисертационният труд съдържа 129 стр. основен текст, фигури и таблици. Състои се от увод, 6 раздела, заключение, библиографска справка, справка за приносите, списък на публикации, свързани с дисертацията, и приложения

### 1. Актуалност на проблема

Темата на дисертацията е свързана с изследвания в областта на семантични уеб-технологии, които разширяват съществуващите уеб-услуги, като дават възможност за създаване на хранилища на данни в глобалната мрежа Интернет, изграждане на речници от понятия и задаване на правила за обработка на данни. В тази област се работи много интензивно в целия свят.

Използването на онтологии в семантичната мрежа улеснява споделянето на информация от конкретна предметна област между потребителите и софтуерни агенти. Чрез онтологията се осигурява организацията на знанията от дадена област във вид на обща структура, която може да бъде анализирана, споделяна и използвана многократно.

При развитието на семантичен уеб възникват редица задачи, част от които са езиково независими, а другите са специфични за конкретен език. Един изключително важен фактор за напредъка на изследвания в областта на семантичен уеб е решаване на задача, свързана с *извличане на знания от текст*. Въпреки че са разработени разнообразни подходи за извличане на знания от текст, постигнатите резултати все още не са напълно задоволителни поради семантичната разлика между естествените езици и формалните езици за описание на знанията. Особено това се отнася за българския език, който има редица особености, които затрудняват изследвания, свързани с приложение на съвременните технологии.

В настоящия труд докторантката прави опит да допълни съществуващия инструментариум за изследвания, като предлага подход за онтологично базирано извличане на информация от неструктуриран текст на българския език. Считаю че темата на дисертацията е много интересна и актуална.

### 2. Степен на познаване на състоянието на проблема

Докторантката направи много задълбочен обзор на съвременното състояние на проблематиката, свързана с семантичен уеб. В дисертацията подробно са разгледани теоретичните основи, инструментариум и приложения при семантичен

уеб. Представени са съществуващите стандарти в тази област и са разгледани релациите между семантичната глобална мрежа и онтологията.

Описват се характерните приложения на онтологиите при извличане на информация от текст и управление на знания. Представена е класификация на онтологиите.

Докторантката демонстрира дълбокото си запознаване с основните понятия и проблеми, свързани със семантичен уеб, а също така и с основни положения в онтологиите. Тя показва, че познава много добре изследваната област, като същевременно отразява и критичното си отношение към някои нерешени задачи.

Библиографската справка включва 85 заглавия на литературни източници в периода от 1979г. до 2015г. , което показва, че тя е запозната много добре с последните постижения в избраното научно направление. Направено е подробно разглеждане на резултатите от изследванията , отразени в литературните източници, което показва нейните възможности за творческата интерпретация на литературния материал.

### **3. Цел на дисертацията и избраната методика**

Основната цел на дисертацията е: *“ Изследване на възможностите за онтологично базирано извличане на информация от неструктуриран текст на български език.*

За решаване на поставената цел се формулират следните по-конкретни задачи:

1. Да се изследва текущото развитие на технологиите, свързани със семантичен уеб.
2. Да се анализират характерни конкретни приложения на онтологиите в извличане на информация и управлението на знания.
3. Да се анализират съществуващите платформи с отворен код за анализ на текст, реализиращи дейности по извличане на информация, като се анализират и съществуващите в тях средства за обработка на български текст.
4. Да се реализира инструмент за обработка на български текст в платформа с отворен код с цел изпълнение на дейности по извличане на информация и да се изследва ефективността на предложения инструмент за обработка на български текст.
5. Да се създаде модел за извличане на онтологично базирани понятия от неструктуриран текст на български език.

За решаване на поставените задачи се използват методи и подходи на компютърното моделиране, семантичен уеб и онтологиите. За оценка на получените резултати се използва честотното разпределение на разпознатите понятия в зависимост от използваните алгоритми.

Считам, че избраните методи и методики за изследване съответстват с поставената цел и конкретните задачи, свързани с целта.

#### 4. Аналитичната характеристика на дисертационния труд

В **първи** раздел на дисертацията е изложено направеното проучване на съвременното състояние на проблема, свързан с темата на изследването. Разгледани са основни понятия от онтология в информатиката, а също така технологии и стандарти при семантичен уеб. Описани са редактори на онтологии и характерни приложения на онтологиите в извличане на информация и управление на знания. Извършена е класификация на онтологиите и среди за проектиране. Показани са най-широко използваните езици за представяне на знания в семантичен уеб. Особено внимание е отделено на използване на семантични уеб технологии за подпомагане при оценяване на отворени въпроси.

**Вторият** раздел е посветен на задачи, свързани с извличане на информация при семантичен уеб. Разгледани са основните задачи, изпълнявани от системите за IE ( Information Extraction ), степента им на зависимост от тематичната област и ефективността на постигнатите резултати за всяка от тях. Представена е обща схема за онтологично базирано извличане на информация от неструктуриран текст. Отбелязани са и предизвикателствата за онтологично базирано извличане на информация.

**Третият** раздел е посветен на лингвистични технологии , свързани със семантичен уеб. Описват се редица известни платформи за анализ на текст , които биха могли да се използват при определяне на граматична категория на думите в текст на български език. Разгледани и са анализирани следните платформи :

- GATE;
- UIMA;
- Open NLP

Направено е заключение, че за анализ на текст на български език и извличане на съдържание е подходяща платформата GATE. Отбелязани и причините за това.

**В четвъртия раздел** е представен подход за автоматично маркиране на неправилно съгласуване на прилагателно име със съществително име в български текст. За реализация на предложения подход се използва възможността за прилагане на регулярни изрази върху анотации в GATE с цел откриване на неправилно съгласуване на прости фрази, образувани от две думи-атрибутивно прилагателно и съществително, като атрибутивното прилагателно предхожда съществително име. Направен е извод ,че многоезичната поддръжка и езиково независими компоненти в платформата осигуряват възможността за разработване на приложения за автоматично откриване и коригиране на широк спектър от типове граматични грешки в текстове на български език.

В **петия** раздел е описана реализация на инструмент за лематизация на български текст в GATE.

Речниково базиран лематизатор позволява извличане на лексикални ресурси / списъци с основни форми на думи, категоризирани по съответни типове / от BG Office и Bulgarian-language Dictionary. Разработена е софтуерна библиотека за автоматично генериране на словоформи по зададени основна форма и тип.

В **шестия** раздел е показано извличане на онтологично базирани понятия от неструктуриран текст на български език. Предложен е подход за OBIЕ (Ontology based Information Extraction ) от неструктуриран текст, съобразен със спецификата на български език. За реализацията на подхода се използва приложение Corpus Pipeline в платформата GATE ,което изпълнява набор от средства за обработка на текст в определена посредователност. Предложен е модел за онтологично базирано извличане на понятия от неструктуриран текст на български език.

В **заключението** е направено резюме на основните резултати от проведеното изследване.

В **Приложение 1** е представена таблица , включваща честотното разпределение на изследвания корпус по части на речта.

В **Приложение 2** е показан примерен текст с Lookup анотации.

## **5. Приноси на дисертационния труд**

Приносите в дисертационния труд могат да бъдат обобщени по следния начин:

### ***Научно-приложни:***

- 1. Предложен е модел за онтологично базирано извличане на понятия от неструктуриран текст на български език, който може да се използва като основа за автоматично извличане на семантична анотация.*
- 2. Разработен е речниково базиран лематизатор за текст на български език, който се разпространява със свободен достъп под GPL v3 license.*

Тези приноси си отнасят към група “**Създаване на нови методи, инструменти и класификации.**”

### ***Приложни:***

*- Извършена е реализация на предложението речниково базиран лематизатор като инструмент с отворен код за платформата GATE , който е публично достъпен в GitHub..*

*- Изследвана е ефективността на предложението инструмент за лематизация на текст на български език.*

Тези приноси си отнасят към група “ **Получаване на потвърдителни факти**”

### **7. Публикации, свързани с дисертацията**

Във връзка с дисертацията са публикувани 5 труда. От тях една публикация е в списание “Information Technologies and Control (ITC)“, останалите са доклади на международни конференции в страната. Статията в списанието е самостоятелна ( под печат , има бележка, потвърждаваща , че е приета за публикуване ). Останалите трудове са в съавторство с нейния ръководител и други специалисти.

Забелязано е едно цитиране на публикация [1] от други автори.

Публикацията е цитирана в монографията на Ирена Атанасова *QLIFEX-базирана на правила експертна система за оценка на качеството на живот* , ISBN 978-954-680-943-8, 2014.

**8. Представеният автореферат** отразява същността на направените изследвания, описани в дисертационния труд и получените резултати.

### **9. Личното участие**

Анализ на дисертационния труд и представеният списък на публикации, свързани с дисертационния труд ми дава основание да направя заключение, че дисертантката познава много добре състоянието на проблема и нямам никакво основание да се съмнявам в личното ѝ участие в проведените изследвания и в написване на дисертацията.

### **10. Забележки, мнения и препоръки**

Имах възможността да прегледам дисертационния труд преди обсъждането му в катедрата, в резултат на което направих редица принципни и редакционни забележки. Тези забележки и откритите грешки и пропуски са отстранени. По настоящия вариант на дисертацията могат да се направят следните забележки и препоръки :

- В дисертацията много подробно се описват известните неща / раздел 1 /, което увеличава обема на представения труд за сметка на собствените изследвания.
- Общоприето е съдържанието на дисертацията да се представя във вид на глави , а не раздели, както е в настоящия случай;
- Дисертационният труд не е балансиран много добре. Например раздел 1 заема 58 стр. , а раздел 2 само 7 стр.
- Таблиците 5.5, 5.6 и 5.7 не са обяснени в текста на дисертацията.
- Фигурите 1-9 и 6.6 са с прекалено дребен шрифт.

- Имам и следната **препоръка**: Хубаво би било направените изследвания да бъдат отразени в някакво учебно пособие.

### **11. Заключение**

Независимо от направените забележки общата оценка на дисертационния труд е положителна. Като имам предвид съдържащите се в дисертацията безспорни научно-приложни и приложни приноси, актуалността и значимостта на изследваната проблематика, разгласата на постиженията на докторантката в медиите и задълбочените ѝ познания в изследваната област, считам, че дисертационният труд отговаря на изискванията на **Закона за развитие на академичния състав в република България** и на **Правилника за условията и реда за придобиване на научните степени в ЮЗУ-Благоевград**, поради което предлагам на маг. Надежда Пламенова Борисова да бъде присъдена образователната и научна степен “доктор” по научно направление **4.6 Информатика и компютърни науки** „.

19.09.2015г.

Благоевград

РЕЦЕНЗЕНТ:  
/проф. д-р Нина В. Синягина/