

РЕЦЕНЗИЯ

от д-р Иван Колев Койчев, професор в СУ "Св. Кл. Охридски"
на докторската дисертация на **Каменка Атанасова Стайкова**
докторантка към **Институт по информационни и комуникационни
технологии - БАН,**
на тема **„Лингвистични и семантични ресурси при компютърно генериране и
анотиране на български текстове”,**
за придобиване на образователната и научна степен „ДОКТОР” в професионално
направление **4.6 „Информатика и компютърни науки”,**

Темата на представената дисертация е в актуалната област: компютърно генериране и аотиране на текстове. Това е една от най-трудните задачи в областта на Изкуствения интелект. Допълнително предимство на дисертацията е, че тя е ориентиран към българския език, който е сравнително беден на компютърни езикови ресурси и инструменти с отворен код.

Дисертацията е добре структурирана. Състои се от увод, три глави и заключение. Приложени са: списък на използваната литература, списък на използваните съкращения и англоезични термини и още няколко приложения представящи, разработени модели, онтологии и други по-важни резултати.

В увода се дава кратка мотивация на темата, ясно се определят целите на дисертацията и произтичащите от тях конкретни задачи. В него накратко се описва и използваната изследователска методология.

В първа глава на дисертацията е представен аналитичен обзор на изследванията, които са свързани с темата на дисертацията, конкретно: компютърното генериране на текст и съответните лингвистични и семантични ресурси, които се използват в процеса на генериране.

Във втора глава са представени създадените ресурс за генериране на български изречения. Представени са теоретични модели за някои граматически явления в българския език и са описани принципите за създаване на приложна българска системично-функционална граматика. Представени са резултати от автоматичното генериране на български изречения със създадената граматика.

В трета глава е представено използването на създадената граматика като лексико-граматически ресурс за генериране на текст на български език в конкретна приложна област. Описан е процесът на планиране на текст. Представени са резултати от автоматично генериране на свързани текстове на български език в различни стилове.

В четвърта глава се разглеждат на семантични технологии, които биха подпомогнали подготвителните работи при компютърно генериране на текст. Използвана е конкретна приложна област за изследване на семантично аотиране с частични граматика на базата на регулярни изрази.

В заключението се обобщава постигнатите резултати и конкретните приноси на докторанта. Предлагат се и насоки за бъдеща работа.

Приложен е и списък с **използвана литература**, съдържащ 83 заглавия, предимно научни статии и книги.

Докторанката демонстрира задълбочени знания в научните области на дисертацията. Направените **обзори** покриват широк спектър от методи и технологии, като **адекватно представят текущото състояние** в съответните области. Като цяло дисертацията прави впечатление с задълбоченото разглеждане на методи за обработка на естествен език, които се използват при автоматичното генерирането на текст. Списъка с използвана литература е представителен за областта и е достатъчен за докторска дисертация. Представянето на разработените подходи е добре структурирано и подкрепено с подходящи илюстриращи примери. Всяка глава завършва с обобщение на представеното в нея и постигнатите резултати. Представената разработка е извървяла целият път, от абстрактните модели до конкретни приложения за решаването на реална практическа задача.

Избраната **методологията на изследване е подходяща** за поставените цели в дисертацията. Тя включва: аналитичен обзор на областта, мотивиран избор на подходящи методи и тяхното творческо доразвиване и прилагане за дадената приложна област. Разработени методи са реализирани като софтуерни прототипи след което експериментално са изследвани и верифицирани с реални данни. Разработените компоненти са внедрени в реални проекти.

Дисертацията е **добре структурирана и естетически оформен**. Издържан е и от гледна точка на правопис, граматика и стилистика. Важните единици в текста са обозначени и почертани с подходящо форматиране, което ги прави открояващи се от останалият текст. Фигурите и таблиците са надлежно номерирани, имат обяснителни надписи и са цитирани в текста.

Също така добро впечатление прави, че като цяло термините са надлежно преведени на български. Към дисертацията е приложен българо-английски речник на термините, и съответните съкращения.

Части от резултатите в дисертацията са представени в 12 статии, като на 9 от тях докторанта е водещ автор. Всичките са в рецензирани издания от които 4 са в списания издавани у нас, една в поредицата LNAI – Springer. Не са забелязани позовавания на публикации по дисертацията. Докторанката е изнесла 7 доклада по темата на дисертацията са международни научни събития. Представените публикации като бройка и качество са достатъчни за дисертационен труд за образователна и научна степен доктор.

Основните приноси на докторанката могат да бъдат обобщени както следва:

- 1) Разработено е **формално описание на базови обекти от българския език** в съответствие с Системично-функционалната теория за естествения език, което позволява компютърно генериране на технически инструкции за създаване на документация на български език.

- 2) Реализация на разработеното описание (1) във вид на приложен компютърен ресурс наречен „системично-функционална граматика на български език“. Реализацията е извършена в средата за многоезиково генериране KPMML, което **позволява пре-използване, разширяване и по-нататъшно развитие.**
- 3) Разработената системично-функционална граматика на български език (2) е реализирана като **модул в многоезиковата система AGILE позволяваща компютърно генериране на технически текстове в различни стилове** (персонален- императивен, безличен- изявителен). Системата е приложена за генериране на свързани текстове на български език при създаване на технически ръководства за CAD/CAM системи.
- 4) **Предложена и реализирана е схема за семантично аотиране на специализирани текстове на български език** в областта на Източно-християнско иконографско изкуство, която използва разработено за целта онтологично описание.

Авторефератът е добре направен и точно отразява дисертационния труд.

Към представената дисертация нямам съществени забележки. На предзащитата направих редица редакционни препоръки. Напълно съм удовлетворен от начина и степента на тяхното отразяване в окончателния вариант.

Можем да обобщим: от представеното дотук е видно, че **представената работа има съществени научно-приложни приноси, напълно достатъчни за присъждане на научно-образователна степен „доктор“.**

В заключение давам **положителна оценка** на представената дисертация и предлагам на уважаемото научно жури, по обявената процедура за защита на докторска дисертация в ИИКТ-БАН да даде образователната и научна степен „доктор“ в професионално направление 4.6. „Информатика и компютърни науки“ на автора на дисертационния труд Каменка Атанасова Стайкова

Рецензент:

Дата 26.06.2015 г.