

СТАНОВИЩЕ

Относно: дисертационен труд за придобиване на образователната и научна степен „доктор“ по специалност 01.01.12 „Информатика“, професионално направление 4.6 „Информатика и компютърни науки“

Автор на дисертационния труд: Каменка Атанасова Стайкова

Тема на дисертационния труд: „Лингвистични и семантични ресурси при компютърно генериране и аотиране на български текстове“

Член на научното жури: доц. д-р Данаил Антонов Дочев - научен ръководител (заповед № 70/22.05.2015 на Директора на ИИКТ - БАН)

1. Актуалност на проблема

Представеният дисертационен труд е посветен на безспорно актуална изследователска тема – изследване и прилагане на модели и софтуерни инструменти за създаване на лингвистични и семантични компютърни ресурси, предназначени за прилагане при добре дефинирани подобласти на компютърните обработки на естествен език - компютърно генериране и аотиране на български текстове. Реалната практическа значимост на тази област е пряко свързана с нуждата от автоматично търсене, резюмиране, обработка и подходящо представяне за най-разнообразни цели и групи потребители на огромните масиви от неструктурирани текстове, налични в глобалното информационно пространство.

2. Обща характеристика на представения труд

Трудът е написан в обем от 180 стр., от които 51 стр. в 5 приложения, 27 фигури, 12 таблици и 5 стр. литература, включваща 83 заглавия. 35 от посочените публикации са издадени след 2004 г., а 21 от тях - след 2009 г. Разгледаната литература демонстрира актуалността и перспективността на тематичната област на дисертацията, както и литературната осведоменост и компетентност на докторантката.

Глава 1 представлява обзорна част на дисертационния труд и представя етапите на компютърното генериране на текстове според широко признати в литературата схеми. Направен е кратък преглед на съвременни модели и инструменти за формализиране и информационни обработки на необходимите лингвистични знания, както и на семантични инструменти за привличане на семантични знания за конкретни предметни области при автоматично генериране на достатъчно сложни, смислово свързани и оценявани от потребителите като достатъчно естествени текстове. Обзорът подкрепя направения подходящ подбор на изследователска методология и естествено обосновава целите и задачите на дисертацията.

На основата на направения анализ аргументирано е избрана Системично-функционалната граматика като формален модел при създаване на български лексико-граматически ресурси за компютърно генериране. Глава 2 представя извършено от авторката моделиране на основни езикови явления в българския език чрез избрания формален модел върху материала на неголям специализиран корпус от текстове (технически инструкции) на английски и три славянски езика (български, руски и чешки). Резултатите от моделирането водят до създаване на приложна българска системично-функционална граматика чрез сравнителен анализ с английската приложна системично-функционална граматика NIGEL и системично-функционалните структури, подходящи за другите разглеждани славянски езици.

Глава 3 описва прилагането на разработената българска системично-функционална граматика като информационен ресурс в средата за многоезиково генериране KPMIL. Средата използва опростен бърз алгоритъм (без механизъм за възврат) за обработка на информационните структури в големи по обем системично-функционални лексико-граматически ресурси.

Приведени са примери на генерирани български текстове – CAD/CAM технически инструкции в различни стилове (персонален императивен или безличен изявителен; с кратки команди или с обяснения и агрегация) от еднаква входна информация.

Глава 4 е посветена на използване на семантични технологии при извличане на онтологични понятия от текст, насочено към автоматизирано/автоматично семантично аотиране на български текстове. Резултатите от този тип информационни обработки могат да се приложат при семантично търсене на обекти в големи архиви; те могат принципно да подпомогнат и процесите на компютърно генериране. Разгледано е формализиране на предметни знания в онтологичен модел - в приложение е показана предметна онтология за български иконографски обекти, съвместима със стандартния концептуален модел CIDOC – CRM за представяне на обекти на културното наследство. За целите на семантично аотиране на специализирани текстове онтологията е използвана съвместно с аотационни граматика на базата на регулярни изрази, разработени в системата ClaRK.

3. Оценка на приносите и личното участие на кандидата

Приемам и оценявам положително научните и научно-приложни приноси на дисертантката, формулирани в заключението на дисертационния труд.

Нямам сериозни критични забележки по дисертационния труд.

По дисертацията са представени дванадесет публикации (1 на български и 11 на английски език), от които 3 самостоятелни, 5 с двама автори, 2 с трима автори и по една с четири и десет автори (в последната всички съавтори са от чужбина). 6 публикации са в трудове на международни конференции, 3 - в списанието CIT (издавано у нас с международна редколегия), 2 – в други български списания и 1 – в издавана поредица от работни статии.

Цялостната разработка на дисертационния труд е лично дело на докторантката, което се потвърждава и от 8-те публикации, в която тя е единствен или първи автор.

Авторката е придобила ценен опит за съвместна работа при решаване на сложни задачи с изследователски характер, включително в многонационален колектив при работата по международен проект. Личните ми впечатления от дълговременната съвместна работа потвърждават нейната мотивираност и упоритост при атакуване на изследователски проблеми.

Авторефератът отразява основните положения и приноси в дисертационната работа.

4. Заключение

При разработването на дисертационния труд дисертантката е показала добро вникване в изследователските проблеми и задълбоченост при тяхното поставяне и решаване. Тя демонстрира широки познания по съвременни методи и инструменти на информационните технологии за обработка на лингвистични и предметни семантични знания, както и добри професионални умения при адаптиране и използване на софтуерни приложения с изследователски характер.

Дисертационният труд напълно отговаря на изискванията на ЗРАСРБ и правилника за приложението му, както и на специфичния правилник на ИИКТ-БАН.

Получените резултати ми дават основание убедено да препоръчам на почитаемото Научно жури да присъди на Каменка Атанасова Стайкова образователна и научна степен “Доктор” по научната специалност 01.01.12 „Информатика“.

София, 24.06.2015 г.