

СТАНОВИЩЕ

за дисертацията на Каменка Атанасова Стайкова
на тема „Лингвистични и семантични ресурси при компютърно
генериране и аотиране на български текстове”
представена за присъждане на образователната и научна степен
„доктор”
по научна специалност 01.01.12 „Информатика“
в професионално направление 4.6 „Информатика и компютърни
науки“

от проф. д-р Радослав Димов Павлов,
Институт по математика и информатика – БАН

Настоящото становище е написано и представено съгласно заповед № 70/22.05.2015 на Директора на ИИКТ – БАН. То е изработено въз основа на ЗРАСРБ, правилника за приложението му и специфичния правилник на ИИКТ-БАН.

Дисертационният труд на Каменка Стайкова е посветен на модели и средства за създаване на компютърни лингвистични и семантични ресурси и засяга дълговременно актуални направления на изследванията по обработка на естествен език и изграждане на семантични системи. Обект на научно и научно-приложно изследване са етапите, формалните модели и средствата за създаване на ресурси за компютърно генериране на български текстове, както и аспекти на прилагането на семантични технологии при компютърна обработка на естествен език. Изследванията адресират широк клас възможни приложения, свързани със

автоматично търсене, резюмиране и подходящо представяне за различни потребители на огромните масиви от неструктурирани текстове и частично структурирани данни в световната информационна мрежа.

Дисертационният труд на Каменка Стайкова е в обем от 180 стр. и се състои от увод, 4 глави, заключение, списък на използваните съкращения и означения, библиография от 83 заглавия и 5 приложения.

В Увода е разгледана актуалността на проблематиката и са представени целите и задачите на дисертацията.

В обзорната Глава 1 е направен преглед на подходите и съвременните инструменти за компютърно генериране на текстове. Обсъден е избора на системично-функционалната граматика като подходящ формален модел за създаване на компютърни лингвистични ресурси за генериране на текст. Накратко са разгледани и семантичните технологии, използвани при компютърна обработка на естествен език.

В Глава 2 е представен процеса на създаване от дисертантката на приложна българска системично-функционална граматика посредством моделиране на основни езикови явления в българския език чрез неголям специализиран корпус от процедурни текстове (технически инструкции) на английски, български, руски и чешки. Използван е сравнителен анализ със системично-функционалните структури на английската приложна граматика NIGEL и специфичните структури за другите разглеждани славянски езици.

В Глава 3 е описано автоматичното генериране на български процедурни текстове в системата AGILE, разработена при международен проект. Приложната българска системично-функционална граматика е използвана като български лексико-граматически ресурс. Системата генерира технически инструкции (фрагменти от CAD/CAM

ръководства) в различни стилове (персонален императивен или безличен изявителен; с кратки команди или с обяснения и агрегация) паралелно на български, руски и чешки от еднаква входна информация.

В Глава 4 е представено използването на семантични технологии при извличане на онтологични понятия от текст. Този процес може да се приложи както при семантично търсене на обекти в големи архиви, така и при подготовка за компютърно генериране. Описано е семантично аотиране на български текстове с използване на два типа ресурси: а/ предметна онтология за български иконографски обекти ОБИО (съвместима със стандартния концептуален модел CIDOC – CRM за представяне на обекти на културното наследство) - разширен вариант на онтология, разработена по-рано в рамките на международния проект LOGOS, в който дисертантката участваше; б/ аотационни граматки на базата на регулярни изрази, разработени в системата ClaRK. Ресурсите са подготвени и експериментирани в рамките на националния проект СИНУС върху част от описанията на мултимедийни обекти от цифровата библиотека „Виртуална енциклопедия на българската иконография“.

Приемам напълно научните и научно-приложни приноси на дисертантката, формулирани в заключението на дисертационния труд. Представеният автореферат правилно отразява съдържанието и приносите в дисертационния труд.

Приложеният списък от дванадесет публикации (1 на български и 11 на английски език) потвърждава сериозната и задълбочена работа на дисертантката и показва нейното съществено участие в разработката на дисертационния труд (в 8 публикации тя е единствен или първи автор).

Нямам сериозни критични забележки.

Представеният дисертационен труд съдържа значими резултати и решения на сериозни изследователски и приложни проблеми при обработката на естествен език и създаване на семантични системи. Авторката демонстрира много добро владение на професионалния инструментариум на редица информационни технологии при разработените от нея методи, модели и приложни решения.

ЗАКЛЮЧЕНИЕ

Всичко казано по-горе ми дава 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“ в професионално направление 4.6 „Информатика и компютърни науки“ за дисертационния ѝ труд „Лингвистични и семантични ресурси при компютърно генериране и аотиране на български текстове”.

София, 25.06.2015 г.

Член на научното жури: