

РЕЦЕНЗИЯ

на дисертационен труд на тема
“Machine Translation Approaches to the Bulgarian Language”
(„Подходи за машинен превод на български език“)
с автор Велислава Райкова Стойкова
за придобиване на научна степен „доктор на науките”
по професионално направление 2.1 „Филология“

Рецензент: проф. д-р Мария Михайлова Нишева-Павлова,
Факултет по математика и информатика на СУ „Св. Климент Охридски“

Със заповед № 3-РК-268/13.07.2022 г. на Ректора на НБУ съм назначена за член на научно жури във връзка с процедура за придобиване на научна степен „доктор на науките“ по професионално направление 2.1 „Филология“ от Велислава Райкова Стойкова с дисертационен труд на тема „Machine Translation Approaches to the Bulgarian Language“ („Подходи за машинен превод на български език“). На първото заседание на научното жури бях определена за рецензент по тази процедура.

1. Обща характеристика на дисертационния труд и представените материали

Дисертационният труд е написан на английски език. Съдържа 110 страници текст и е разделен на две части, включващи общо пет глави, списък на цитираните литературни източници, две приложения, речник и индекс.

Освен дисертационния труд по процедурата са представени още:

- автореферат на български език, съдържащ 43 страници текст;
- автобиография по европейски образец.

Не са представени копия на публикациите във връзка с дисертацията и списък на забелязани цитирания на тези публикации, но такива не се изискват от Наредбата за развитието на академичния състав на НБУ.

2. Данни за кандидата

Д-р Велислава Стойкова има висше образование – ОКС „магистър“ по български език и литература, придобито във Факултета по славянски филологии на Софийския университет „Св. Климент Охридски“. През 2003 г. е защитила дисертация за придобиване на ОНС „доктор“ на тема „Формален морфологичен анализ на категорията определеност/неопределеност в български език, моделирана в DATR – теория“. От 2011 г. заема академична длъжност „главен асистент“ в Института за български език към БАН.

Познавам бегло д-р Стойкова от някои нейни участия в международни научни форуми, на които сме имали възможности за осъществяване на професионални контакти. Имам много добри впечатления от стремежа ѝ към усвояване на нови технологии и осъществяване на задълбочени интердисциплинарни изследвания.

3. Актуалност на областта на изследванията и значимост на изследвания проблем

Дисертационният труд е посветен на изследвания в областта на езиковите технологии. В него са представени резултати на автора от изследването на различни формални езикови теории и тяхната приложимост за целите на автоматичния анализ на българския език и машинния превод на български език. Дискутирани са резултати от прилагането на множество статистически техники за машинен превод на български език.

Областта на изследванията в дисертационния труд е сложна и актуална, като актуалността ѝ се определя от разширяващото се множество реални и потенциални приложения на машинния превод и в частност машинния превод на български език.

Целта на дисертационния труд и задачите за постигането на тези цел, които авторът посочва неявно както в дисертацията, така и в автореферата, са определени на базата на кратък анализ на постиженията и откритите въпроси в избраната област. Поставената цел е значима, а задачите съответстват на целта.

4. Анализ на съдържанието, резултатите и приносите на дисертационния труд

Дисертационният труд включва пет глави, списък на цитираните литературни източници и две приложения.

Уводната първа глава представя областта на изследванията и базови понятия в тази област. Дискутирани са основните подходи за машинен превод, които имат практическо значение – машинен превод, базиран на правила, и статистически машинен превод.

Втора глава е посветена на изследване на приложимостта на езика за представяне на лексикална информация DATR за целите на машинния превод. Лаконично са описани структурата на представянето на граматични знания със средствата на DATR и съществуващите софтуерни инструменти, поддържащи възможности за машинен превод между езици, чиито граматични правила са описани в термините на DATR. Дискутирани са резултатите от осъществения от автора задълбочен анализ на възможностите за използване на DATR за представяне на словоизменението на съществителните имена в българския и руския език за целите на машинния превод. Специално внимание е отделено на използването на DATR като трансформационна граматика за сравняване на представянето на притежателните и възвратното притежателно местоимение в българския език.

В трета глава се разглеждат серия от въпроси, свързани с използването на Универсалния мрежов език UNL за целите на машинния превод. Описани са структурата на UNL и предназначението на основните езикови конструкции. Анализирани са възможностите за използване на UNL за представяне на словоизменението на съществителните имена в българския език и е направено аргументирано сравнение със съответните възможности, предоставяни от DATR. Задълбочено са дискутирани резултатите на автора на дисертационния труд за представяне на притежателните и възвратното притежателно местоимение в българския, английския и руския език със средствата на UNL. Изложението е подкрепено с подходящи примери. Най-съществен приносен елемент на тази глава е представянето на резултати, свързани с използването на UNL за машинен

превод между български и словашки език. Дискутирани са и възможности за използване на UNL за измерване и подобряване на качеството на машинния превод.

Четвърта глава е посветена на изследване на възможностите за машинен превод с помощта на машината за търсене Sketch Engine (SE). Анализирани са приложенията на SE за машинен превод на термини от областта на екологията между български и словашки език. Обърнато е внимание на използване на SE за машинен превод на изрази за време между български и словашки език. Изложението на тази глава акцентира върху описанието на осъществените теоретичен анализ и практически експерименти, насочени към използване на SE за машинен превод между български и словашки език. Демонстрирани са и възможности за използване на SE за машинен превод на математически термини между български и сръбски език.

Заклучителната пета глава съдържа резюмета на представените резултати в предходните три глави и съответните изводи на автора, както и обобщаващи изводи от изследванията на автора по темата на дисертационния труд.

Списъкът на използваните литературни източници включва 105 заглавия. Начинът на цитирането им в текста на дисертацията свидетелства за много добро познаване на състоянието на изследванията и приложените разработки в избраната област.

Приложенията съдържат множество фигури, илюстриращи по подходящ начин изложението на част от секциите от глава 2 и глава 3.

Основните приноси на дисертационния труд може да бъдат формулирани както следва:

- Извършен е сравнителен анализ на съществуващи софтуерни инструменти за машинен превод на български език, основани на използване на формални езикови теории.
- Предложена е оригинална техника за измерване на качеството на резултата от машинен превод на български език и са формулирани възможности за

неговото повишаване чрез подобряване на формалното представяне на езиковите правила.

- Създаден е електронен корпус от специализирани математически текстове, който е използван за статистически машинен превод. Получени са резултати от такъв превод между български и сръбски език.
- Получени са резултати във връзка с установяване на границите на използването на машинен превод на български език като допълващ и подпомагащ превода, извършван от човека.

Дисертационният труд прави отлично впечатление с обхвата и задълбочеността на изложението. Областта на изследванията е актуална и сложна и постигането на съществени резултати в нея изисква сериозна интердисциплинарна подготовка, постоянна и интензивна работа. Постигнатите резултати са оригинални и значими и съответстват напълно на поставената цел.

5. Публикации, които отразяват дисертацията. Отражение на резултатите на дисертацията в трудове на други автори

Резултатите, получени в дисертационния труд, са отразени в общо 12 статии, 11 от които са публикувани както следва:

- пет – в научни издания, реферирани и индексирани в WoS и/или Scopus;
- четири – в сборници с доклади на международни научни форуми;
- две – в авторитетна поредица, реферирани и индексирани в Google Scholar.

Една статия предстои да бъде публикувана в издание на Springer, което се реферира и индексирани в Scopus.

Осем статии от този списък са самостоятелни и четири са в съавторство. Приемам, че всички съавтори имат равностоен принос във всяка от колективните статии.

Не са представени данни за забелязани цитирания на публикации на Велислава Стойкова по резултати от дисертационния ѝ труд в статии на други автори, но при търсене в Scopus бяха открити общо четири, а в Google Scholar – повече от 15 такива цитирания (без автоцитиранията).

По такъв начин са изпълнени и надхвърлени съществено изискванията на чл. 1а от Правилника за прилагане на Закона за развитието на академичния състав в Република България за придобиване на научна степен „доктор на науките“ в

професионално направление 2.1 „Филология“.

6. Автореферат

Авторефератът отговаря на всички изисквания за изготвянето му и представя пълно и точно темата, целта, съдържанието, постигнатите резултати и приносите на дисертационния труд.

7. Критични бележки и препоръки

Като цяло дисертационният труд е оформен грижливо и коректно. Изложението му е компетентно и достатъчно задълбочено. Като по-съществени недостатъци в съдържателно отношение бих посочила липсата на явна формулировка на задачите на представеното изследване и липсата на достатъчна аргументация на избора на използваните езикови и инструментални средства за постигане на целта на изследването. Би било целесъобразно и включването на по-обширен и пълен обзор на състоянието на научните изследвания и практическите резултати в областта на машинния превод, придружен от формулиране и анализ на най-съществените открити въпроси в тази област, на част от които е посветен дисертационният труд.

Смятам обаче, че повечето посочени недостатъци до голяма степен се дължат на обстоятелството, че дисертационният труд има формата на книга, представена за публикуване от реномирано издателство, и авторът е трябвало да се съобрази със съответните изисквания.

8. Заключение

Въз основа на изложеното по-горе смятам, че дисертационният труд напълно удовлетворява и значително надхвърля изискванията на ЗРАСРБ, Правилника за прилагане на ЗРАСРБ и Наредбата за развитието на академичния състав на НБУ. Постигнати са значими изследователски резултати, които представляват оригинален принос в избраната научна област. Оценката ми за дисертационния труд, автореферата и научните приноси на автора им Велислава Стойкова е **положителна**.

Затова предлагам на уважаемото научно жури да присъди на Велислава Райкова Стойкова научната степен „доктор на науките” в професионално направление 2.1 „Филология“.

08.09.2022 г.

Рецензент:

(проф. д-р Мария Нишева-Павлова)