

Рецензия

по процедура за защита на дисертационен труд на тема:

„Автоматично резюмиране на правни текстове“

за придобиване на

образователна и научна степен „доктор“

от

кандидат: **Валентин Венков Змийчаров,**

Област на висше образование: **4. Природни науки, математика и информатика,**

Професионално направление: **4.6. Информатика и компютърни науки,**

Докторска програма: **„Софтуерни технологии“ – Откриване на знания,**

катедра: **„Софтуерни технологии“,**

Факултет по математика и информатика (ФМИ),

Софийски университет „Св. Климент Охридски“ (СУ)

Рецензията е изготвена от: **проф. д-р Мария Михайлова Нишева-Павлова – ФМИ, СУ,** в качеството на член на научното жури, съгласно Заповед № РД-38-278 / 09.06.2025 г. на Ректора на Софийския университет.

1. Обща характеристика на дисертационния труд и представените материали

Представеният дисертационен труд съдържа 114 страници текст, включващ шест глави; списък на използваните литературни източници, съдържащ 58 заглавия; списък на фигурите; списък на таблиците.

Освен дисертационния труд по процедурата са предадени още:

- автореферат на дисертацията на български и английски език;
- лични документи на докторанта, включващи
 - ✓ творческа автобиография;
 - ✓ копие на диплома за завършено висше образование, ОКС „бакалавър“, и съответното приложение;
 - ✓ копие на диплома за завършено висше образование, ОКС „магистър“, и съответното приложение;
 - ✓ заповед за зачисляване в редовна докторантура и заповед за трансформиране на докторантурата в задочна;

- ✓ удостоверение за положените изпити в рамките на докторантурата и натрупаните в резултат ECTS кредити;
- ✓ заповед за отчисляване от докторантура с право на защита;
- ✓ документи от проведено предварително обсъждане на дисертационния труд в катедра „Софтуерни технологии“ на ФМИ;
- ✓ декларация за оригиналност на дисертационния труд и публикациите по него;
- ✓ списък и копия на публикациите по дисертацията;
- ✓ документи по процедурата за проверка на оригиналността на дисертационния труд;
- ✓ справка за изпълнението на минималните национални изисквания по чл. 2б, ал. 2 и 3 от ЗРАСРБ.

2. Данни и лични впечатления за кандидата

Валентин Змийчаров има висше образование – ОКС „бакалавър“, завършена в СУ „Св. Климент Охридски“ – ФМИ, специалност „Софтуерно инженерство“, и ОКС „магистър“, завършена в СУ „Св. Климент Охридски“ – ФМИ, магистърска програма „Извличане на информация и откриване на знания“. Има значителен професионален опит в разработването на проекти в областта на софтуерното инженерство и машинното самообучение. В продължение на повече от пет години е водил упражнения по Изкуствен интелект и WWW технологии на студенти от ФМИ.

Нямам непосредствени лични впечатления от кандидата и неговата работа освен такива от успешното му участие в научни конференции в областта на изкуствения интелект.

3. Съдържателен анализ на научните и научно-приложните постижения на кандидата, съдържащи се в представения дисертационен труд и публикациите към него, включени в материалите по процедурата

Съдържанието на дисертационния труд включва девет обособени части: шест глави, списък на използваните литературни източници, списък на фигурите; списък на таблиците.

Уводната първа глава представя накратко областта на изследванията, обосновава и формулира целите и конкретните изследователски задачи на дисертационния труд.

Очертана е структурата на дисертацията и е резюмирано съдържанието на останалите ѝ глави.

Във втора глава, която има обзореен характер, са въведени основните концепции и алгоритми в областта на обработката на естествен език и автоматичното резюмиране. Характеризирани са основните типове подходи за автоматично резюмиране и потенциалът на предварително обучените езикови модели като BERT, T5 и GPT за решаване на задачата за автоматично резюмиране на правни документи. Анализирани са достъпни набори с правни текстове, като се акцентира върху тяхната приложимост за обучение на съвременни модели за резюмиране.

Трета глава е посветена на анализ на поредица от въпроси, свързани със събирането, почистването и подготовката на данните, използвани в проведените експерименти. Представен е създаденият от дисертанта нов набор от правни текстове, обхващащ текстове на 24 езика, като са описани неговите характеристики, както и процесът на конструирането му и предизвикателствата при многоезиковата обработка.

Четвърта глава дискутира в детайли най-значимите оригинални резултати на дисертанта по избраната тема. Представен е сравнителен анализ на популярни подходи за екстрактивно резюмиране на текстове (базов подход; подход, основан на TF-IDF; TextRank; подходи, основани на BERT/LegalBERT и k -means; PreSumm и др.). Анализирани са предизвикателствата при резюмиране на текст, които поражда работата с различни езици. Въведен е разработеният от дисертанта нов екстрактивен подход StructExtSum за резюмиране на текст, основан на структурата на документа. Представена е разработката на оригинална система за автоматично създаване на резюмета на правни текстове, реализирана в Google Colab.

Пета глава съдържа задълбочен анализ на ефективността на дискутираните в дисертацията методи за резюмиране на текст. Представени са резултатите, постигнати върху правни текстове на различни езици и научни статии на английски език, като се акцентира на приложимостта на подходите за специфични езици и типове документи. Показано е преимуществото на предложения от дисертанта модел StructExtSum при решаването на задачата за резюмиране на правни текстове на български и английски език. Включена е и секция, която съдържа анализ на резултатите, постигнати с помощта на големи езикови модели при резюмиране на правни текстове на 24 езика.

Заключителната шеста глава съдържа формулировка на приносите на дисертацията и насоки за бъдещи изследвания по темата. Дискутирани са и етичните аспекти на автоматичното резюмиране на правни документи.

Основните научни и научно-приложни приноси на дисертационния труд могат да бъдат обобщени както следва:

- Разработен е нов метод за екстрактивно резюмиране. Предложеният метод StructExtSum е тестван върху многоезикови данни и е показал конкурентни резултати в сравнение със съществуващите подходи. Той може да бъде внедрен в платформи за управление на документи, правни информационни системи и др. за реализация на семантично търсене и извличане на съществени части от текстове и по този начин да предостави на различни типове потребители инструменти за оптимизация на достъпа до необходимата правна информация;
- Създаден е нов многоезиков набор от данни за резюмиране на правни документи, който включва документи на 24 езика;
- Извършен е сравнителен анализ на множество предварително обучени езикови модели и са показани силните страни и ограниченията на тези модели при обработката на правни текстове. Резултатите от този анализ може да бъдат използвани при избора на подходящ модел в зависимост от контекста.

Научните и научно-приложните постижения на кандидата, представени в дисертационния труд, са оригинални и значими и предоставят редица възможности за по-нататъшни изследвания в областта на извличането на информация и обработката на естествен език. Възможни са също така реални практически приложения на предложените нови теоретични решения.

4. Аprobация на резултатите

Резултатите, получени в дисертационния труд, са отразени в четири статии, публикувани в специализирани научни издания с SJR както следва:

- една – в сборник с трудове на International Conference Computational Linguistics in Bulgaria (SJR 0,220 за годината на публикуването);
- две – в CEUR Workshop Proceedings (SJR съответно 0,202 и 0,228 за годината на публикуването);
- една – в сборник с трудове на Международната научна конференция RANLP (SJR 0,316 за годината на публикуването).

Всички статии по резултати от дисертацията са в съавторство, като са представени декларации на съответните авторски колективи, в които е посочен явно равностойният личен принос на отделните съавтори. По такъв начин са изпълнени и многократно

надхвърлени минималните национални изисквания по чл. 2б, ал. 2 и 3 на ЗРАСРБ и допълнителните изисквания на СУ „Св. Климент Охридски“ за придобиване на ОНС „доктор“ в ПН 4.6. Информатика и компютърни науки.

Представените от кандидата резултати в дисертационния труд и научните трудове по резултати от него не повтарят такива от предишни процедури за придобиване на научна степен и заемане на академична длъжност.

Кандидатът има и още една научна публикация в издание, индексирано в Scopus. Към датата на финализиране на настоящата рецензия в Scopus има данни за общо седем цитирания на трудове на кандидата от други автори, три от които са на статии по резултати от дисертационния труд.

Няма доказано по законоустановения ред плагиатство в представения дисертационен труд и научните трудове по тази процедура.

5. Качества на автореферата

Авторефератът, представен на български и английски език, отговаря на всички изисквания за изготвянето му и представя пълно и точно темата, целите, съдържанието, постигнатите резултати и приносите на дисертационния труд.

6. Критични бележки и препоръки

Нямам съществени критични бележки към формата и съдържанието на представената за официалната процедура версия дисертационния труд, както и към формулировката на основните резултати и приносите, посочени в дисертационния труд и придружаващите го научни трудове. Повечето от забележките ми към версията на дисертацията, представена на предварителното обсъждане в катедра „Софтуерни технологии“ на ФМИ, са взети предвид от дисертанта. Останали са неотстранени малък брой граматически грешки, които не нарушават отличното общо впечатление от дисертацията и значимите резултати, представени в нея. Препоръчвам на дисертанта да продължи активната си изследователска и експериментална работа в избраната област.

7. Заключение

След като се запознах с представените в процедурата дисертационен труд и придружаващите го научни трудове и въз основа на направения анализ на тяхната значимост и съдържащите се в тях научни и научноприложни приноси, **потвърждавам**, че представеният дисертационен труд и научните публикации към него, както и

качеството и оригиналността на представените в тях резултати и постижения, отговарят на изискванията на ЗРАСРБ, Правилника за приложението му и съответния Правилник на СУ „Св. Климент Охридски“ за придобиване от кандидата на образователната и научна степен „доктор“ в професионално направление 4.6. Информатика и компютърни науки. В частност, кандидатът удовлетворява минималните национални изисквания в професионалното направление и не е установено плагиатство в представените по процедурата научни трудове.

Въз основа на гореизложеното, **препоръчвам** на научното жури да присъди на Валентин Венков Змийчаров образователна и научна степен „доктор“ в професионално направление 4.6. Информатика и компютърни науки.

25.07.2025 г.

Изготвил рецензията:

(проф. д-р Мария Нишева-Павлова)