



РЕЦЕНЗИЯ

върху дисертационен труд за придобиване на
образователна и научна степен „доктор”

Автор на дисертационния труд: маг. инж. Камен Николаев Стойков

Тема на дисертационния труд: **Обработка, анализ и извличане на
знания от големи масиви данни на естествен език чрез машинно
обучение**

Научен ръководител: 1: проф. д-р Милена Кирилова Лазарова - Мицева

2: проф. д-р Огнян Наков Наков

Рецензент: доц. д-р Мария Ангелова Ангелова – Димитрова

Рецензията е възложена с решение на Научно жури, назначено със
Заповед № ОЖ-5.3-30 от 22.07.2021 г на Ректора на ТУ-София.

*1. Актуалност на разработвания в дисертационния труд
проблем в научно и научно-приложно отношение.*

Дисертационният труд е посветен на методите за обработка на големи масиви от неструктурирани данни и извличане на знания от тях с използване на подходящи структури от данни, методи, алгоритми и софтуерни архитектури за постигане на оптимални резултати.

Обект на изследването в дисертационния труд са големи масиви от неструктурирани данни като специално внимание се обръща на обработването на текстови документи на естествен език с правен характер.

Основната поставена цел е да се намерят методи и алгоритми за обработка на текстови документи на естествен език чрез машинно обучение, за да се изгради модел за обработка на съдържанието на правни документи от вида „Декларации за поверителност“. Вниманието е насочено към употребата на невронно-мрежови модели за класификация на документи на естествен език.

2. Степен на познаване състоянието на проблема и творческа интерпретация на литературния материал.

Авторът добре познава състоянието на проблемите в областта на методите и алгоритмите за обработка на текстови документи на естествен език. Това ясно личи от подробния литературен обзор, направен въз основа на голям брой литературни източници. Посочени са общо 139 литературни източници, от които: 115 на латиница и 24 интернет адреса.

3. Съответствие на избраната методика на изследване с поставената цел и задачи на дисертационния труд.

Избраната методика на изследване включва употребата на невронно-мрежови модели за класификация на документи на естествен език.

Тази методика съответства напълно на поставените цели и задачи на дисертационния труд.

4. Кратка аналитична характеристика на естеството и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд.

Дисертационният труд на маг. инж. Камен Николаев Стойков е посветен на методите за обработка на големи масиви от неструктурирани данни и извличане на знания от тях.

Трудът съдържа 156 страници и е разделен на четири глави. В текста са включени 139 литературни източника, 39 фигури, 5 графики и 6 таблици.

Първа глава на дисертационния труд разглежда данните като обект за анализ и обобщава методите за тяхната обработка.

Втора глава очертава методология за обработване на неструктурирани данни на естествен език.

Трета глава разработва методика за имплементиране на невронно-мрежов модел за обработване и анализ на декларации за поверителност.

Четвърта глава на дисертационния труд представя планирането и провеждането на експерименти за оценка на приложението на предложената методология за обработване и анализ на декларации за поверителност.

След всяка глава са посочени изводи от постигнатото в съответния раздел и се поставят задачи за последващо изпълнение. Това в голяма

степен демонстрира аналитичното и последователно мислене на докторанта.

В края на труда са представени списък с приноси по дисертационния труд и списък на изготвените публикации. Посочени са и списъци с включените в текста на дисертацията фигури, графики и таблици, а също и списък с използваните съкращения.

Текстът на дисертацията е добре оформен и форматиран. Фигурите са с добро качество.

5. Научни и/или научно-приложни приноси на дисертационния труд.

В списък с приноси по дисертационния труд са посочени 2 научни, 2 научно-приложни и 3 приложни приноса.

Основният постигнат научен резултат е методология за обработване на неструктурирани данни на естествен език с цел анализ на съдържанието и структурата на правни документи от вида декларации за поверителност с използване на машинно обучение. В допълнение е предложен подход за формиране на обучаващо множество данни чрез автоматично аотиране на декларации за поверителност с използване на невронни мрежи и клъстеризация.

Разработени са методи и алгоритми за всеки от етапите на подготовка, аотиране и класификация на декларации за поверителност, които позволяват създаване на модел за структурата на правните документи и извличане на знания от тях. Изведена е методика за проектиране, конфигуриране и имплементиране на невронно-мрежов модел за обработка на данни от декларации за поверителност.

Към приложните приноси спада разработването на софтуерен инструмент, прилагащ всички разработени методи и алгоритми и провеждането на серия от експерименти с него. Всички получени резултати са обстойно анализирани и показват практическата полезност и приложимост на научните и научно-приложните приноси на кандидата.

Признавам получените приноси. Те са добре формулирани и правилно разпределени в съответните подвидове.

6. Оценка за степента на личното участие на дисертанта в приносите.

Направените три самостоятелни публикации по темата на дисертацията, останалите публикации в съавторство с научните

ръководители и други автори, както и приложените материали, показващи доброто познаване на разглежданата тематика, ми дават основание да считам, че получените резултати са предимно лично дело на дисертанта.

7. Преценка на публикациите по дисертационния труд.

По дисертационния труд са направени седем публикации, от които три самостоятелни и четири в съавторство с научните ръководители на дисертанта и други изследователи. Пет от публикациите са представени на научни конференции. Две публикации са отпечатани в научни списания. Публикациите отразяват основните приноси по темата на дисертационния труд.

Цитирания не са известни.

8. Използване на резултатите от дисертационния труд в научната и социалната практика.

Във връзка с научните изследвания по дисертационния труд докторантът е участвал в два научно-изследователски проекта, данни за които са посочени в труда. Първият е финансиран от Научноизследователски сектор при Технически университет - София, а вторият – от Министерство на образованието и науката.

9. Оценка на съответствието на автореферата с изискванията за изготвянето му, както и на адекватността на отразяване на основните положения и приносите на дисертационния труд.

Авторефератът е в обем от 32 страници и отразява съдържателно съществени моменти от дисертационния труд. Спазени са изискванията в съответствие с образеца за изготвяне на авторефератите по дисертационните трудове, посочен в сайта "Развитие на АС - Работни документи".

10. Мнения, препоръки и бележки.

Образователните цели на дисертацията са изпълнени.

Приносите, постигнати от дисертанта отразяват правилно, постигнатите резултати и могат да се използват за придобиване на образователна и научна степен "доктор".

Препоръчвам на дисертанта да продължи да работи в избраната научна област и да продължи задълбочената си експериментална работа с цел допълнително подобряване на вече представените резултати.

11. Заключение с ясна положителна или отрицателна оценка на дисертационния труд.

От представения дисертационен труд се вижда, че дисертантът е добре запознат с изследвания проблем. Темата на дисертацията е актуална и в резултат на направените разработки има постигнати научни, научно-приложни и приложни приноси. С това авторът показва, че може самостоятелно да решава научни проблеми.

Докторантът е положил успешно изпити, изготвил е необходимия брой научни публикации и е повишил компетенциите и квалификацията си. С това е изпълнена образователната част на исканата степен.

Като имам предвид постигнатите приноси, оценката ми за дисертационния труд и материалите към него е изцяло положителна.

Считам, че изискванията на ЗРАСРБ и Правилника за условията и реда за придобиване на научни степени в ТУ – София са изпълнени и препоръчвам на членовете на Почитаемото Научно жури да вземе решение за присъждане на научната и образователна степен „доктор“ на маг. инж. Камен Николаев Стойков.

Дата: 15.10.2021

РЕЦЕНЗЕНТ:
(доц. д-р Мария Ангелова – Димитрова)