

## СТАНОВИЩЕ

от д-р **Иван Колев Койчев, професор** в ФМИ на СУ“Св. Климент Охридски“

за дисертационния труд на тема

**„Изследване на мнения и чувства от текстове от социалните мрежи“,**

с автор **Тодор Васков Цонков**

за придобиване на образователната и научна **степен „доктор“** по:

област на висше образование 4. Природни науки, математика и информатика;

професионално направление **4.6. Информатика и компютърни науки;**

докторска програма Информатика“ (Изкуствен интелект)

В последното десетилетие, разработването на методи и алгоритми за анализ на мнения в текстове от интернет се радваше на особена популярност. Това направление бързо намери и своите приложения, поради важноста за бизнеса, какво е мнението на клиентите за продуктите и услугите им. Основен проблем, който пречи на системи да достигнат висока точност са мненията с двойно значение (ирония, сарказъм и др.). За трудността на задачата говори и факта, че и ние хората не винаги разбираме такова значение. Представената дисертация предлага разработки за откриване на двойно значение и преувеличения в текст. Следователно може да заключим че темата на представената дисертация е в **много актуалната област**.

Дисертацията е **добре структурирана**. Състои се от увод (глава 1), пет глави, заключение (глава 7) и списък на използваната литература.

В **глава 1 (увод)** се дава кратка мотивация на темата, ясно се определят целите на дисертацията и произтичащите от тях конкретни задачи. Представя се съдържанието на дисертацията. Накрая е даден и речник с преводите на български на най-често използваните термини в дисертацията.

В **глава 2** се прави обзор на областта на изследванията, като се обосновава защо е избрана предметната област за автоматично засичане на ирония, сарказъм, хипербола и неверни твърдения в текст.

В **глава 3** се описани разработен метод за откриване на двойно значение в текстове от социалните мрежи. Предложени са емпирични правила за откриване на двойно значение. Представят се резултати от експерименти с реални данни за оценка предсказващата точност на предложените правила и на различни класификатора. Направена е количествена оценка на предсказващата сила на описващи характеристики. Разглеждан е и случая с прилагането на тези правила и алгоритми за откриване на ирония, сарказъм и сатира върху текстове на български език.

В **глава 4** се разглеждат разработените алгоритми за определяне на невярна информация и преувеличение в текст. Представят се резултати от проведени експерименти с данни от социалните мрежи, като разработените алгоритми са сравнени с подобни съществуващите алгоритми.

В **глава 5** е представен подход за класификация на мнения от текстове на български език, основан на правила за автоматично засичане на слухове от текстове от социалните мрежи на български език и алгоритъм за подобряване на съответния списък от думи за предметна област.

В глава 6 е представена система за извличане на потребителски мнения от социални мрежи, която прилага разработените подходи. Направени са различни статистики за да се покаже ползата за потребителя на приложението. Предлага се търсене по ключова дума и се определя чувството в текста по скалата (-5;+5).

Глава 7 е заключението, което обобщава постигнатите резултати, систематизира приносите на докторанта и предлага препоръки за бъдещи изследвания по темата.

Приложен е списък с използвана литература, който съдържа 85 заглавия, предимно научни статии и книги.

Дисертацията е добре форматирана и естетически оформена. Важните единици в текста са обозначени и почертани с подходящо форматиране. Съдържа 19 таблици и 26 фигури, които са надлежно номерирани, имат обяснителни надписи и са коментирани в текста. Представянето на разработените алгоритми е добре структурирано и е подкрепено с подходящи илюстриращи примери. Всяка глава завършва с обобщение на представеното в нея и постигнатите резултати. Дисертацията и автореферата, са относително добре от гледна точка на правопис и граматика, макар че в тази насока имам какво да се подобри. Термините са надлежно преведени на български език, като е даден и речник с преводите на български език на най-често използваните термини.

От представената дисертация е видно че докторанта има задълбочени знания в научните области на дисертацията. Направените обзори покриват широк спектър от методи, технологии и софтуерни инструменти, като адекватно представят текущото състояние в съответните области. Списъка с използвана литература е представителен за областта и е достатъчен за докторска дисертация.

Избраната методологията на изследване е подходяща за поставените цели в дисертацията. Тя включва: аналитичен обзор на областта, мотивиран избор на подходящи методи, разработване на нови методи, тяхната експериментална проверка и тяхното прилагане за решаването на конкретни приложни задачи.

Видно е че представената разработка е извървяла целият път, от абстрактните модели, пред експериментална проверка на разработените алгоритми до прилагането им за решаването на реални практически задачи. Разработените алгоритми са реализирани със съвременни средства за разработка на софтуерна, като е разгледана и възможността за разпаралеляване на алгоритмите с цел използването им за обработка на големи масиви от данни.

Основните научни и научно-приложни приноси на докторанта са свързани с разработки на подходи за автоматично разпознаване на ирония, сарказъм, сатира, невярна информация и преувеличаване в текст (хипербола и литота). Това са слабо изследвани области поради трудността на задачата. Изведени са емпирични правила, чиято предсказваща точност е проверена експериментално върху реални текстове от социалните мрежи на английски и български език. Събрани са корпуси от социалните мрежи за ирония, сарказъм, сатира, хипербола и литота. Експериментално са установени предсказващи характеристики за двойно значение на текст. Разработен е алгоритъм за подобряване на списъка от положителни и отрицателни термини за определяне на мнение в текст. Експериментално е проверена на точността на

предсказване на класификатори, научено от различни алгоритми, за определяне на двойно значение на мненията от социалните мрежи. Разработен и експериментално е проверена точността на предсказване на метод за засичане на слухове от социалните мрежи. Изградена е система за автоматично извличане на мнения от основните социални мрежи (Twitter, Facebook, Google +, LinkedIn, Yahoo), в която се прилагат и разработените методи. Основната функционалност на система дава възможност по дадена единица (например: продукт, личност, марка и др.) да се извлекат съответните мненията от социалните мрежи, на които автоматично се оценява отношението на потребителите относно единицата. Съгласен съм с представения в дисертацията по-подобен списък с приноси и смятам че те са **напълно достатъчни за докторска дисертация**.

Части от резултатите в дисертацията са **публикувани** в 5 статии, като на 4 от тях докторанта е водещ автор. Всичките са в рецензирани издания. Една е с сборник на международна конференция. Три от тях са в сборници на рецензирани национални конференции (с международно участие), а една е в списание на Техническият Университет. Всичките статии ма конференции са докладвани от докторанта, като в допълнение има два изнесени доклада по темата на дисертацията на международни научни работни срещи и един на докторантско училище в България. Като негов научен ръководител съм съавтор на 4 от публикациите и декларирам, че в тях, докторанта има изразена водеща роля от идеята до реализацията. Представените публикации, като бройка и качество, **изпълняват изискванията на закона и съответните правилници** за дисертационен труд за образователна и научна степен „доктор“.

**Авторефератът е добре направен** и точно отразява дисертационния труд.

Към представената дисертация **нямат съществени забележки**. Определено има място за подобрене в текста на дисертацията, но да ги коментирам в това становище ще е излишно издребняване.

От представеното до тук е видно, че разглежданата **дисертация напълно отговаря на изискванията** на ЗРАСРБ, Правилника за прилагане на ЗРАСРБ, Правилника за развитие на академичния състав на СУ ”Св. Климент Охридски” и специфичните изисквания на ФМИ за **присъждане на образователната и научна степен „доктор”**.

Личните ми впечатления от докторанта са, че той е високо квалифициран специалист с богат професионален опит и изследователски дух. Активно се включи в обучението на студенти, като хоноруван асистент.

В заключение, давам **положителна** оценка на представената дисертация и предлагам на уважаемото научно жури, по обявената процедура за защита на докторска дисертация в ФМИ на СУ ”Св. Кл. Охридски”, да присъди образователната и научна степен „доктор” в професионално направление 4.6. „Информатика и компютърни науки“ на автора на дисертационния труд Тодор Васков Цонков.

03.05.2016 г.  
гр. София

Подпис:

/проф. д-р Иван Койчев/