

СТАНОВИЩЕ

за присъждане на образователната и научна степен "доктор"
по научна специалност 01.01.12 Информатика
професионално направление 4.6 Информатика и компютърни науки
(Информатика - Изкуствен интелект)

на Тодор Васков Цонков

задочен докторант към катедра "Софтуерни технологии" на ФМИ,
СУ "Климент Охридски"

на тема: "Изследване на мнения и чувства от текстове от социалните мрежи"

научен ръководител: проф. д-р Иван Койчев

от доц. д-р Красимира Иванова,

Институт по математика и информатика, БАН

Това становище е написано и представено на основание на заповед
No. РД 38-99/17.02.2016 на Ректора на СУ "Св.Кл. Охридски".

1 Актуалност на тематиката

С развитието на информационните и комуникационни технологии за определянето на общия облик на една компания голямо значение започна да има нейното представяне във виртуалното пространство. Особено значение придобива влиянието, което оказват мненията на потребителите върху другите потенциални потребители – не толкова в сайтовете на самите компании в частта "мнения на клиенти" (това няма особена тежест доколкото тези текстове могат да бъдат редактирани от администратора на сайта), колкото в общите портали и свободните социални мрежи.

Чрез възможностите, които интернет предоставя, за лесно събиране на първичен материал за анализиране, задачите, които си поставя математическата лингвистика, стават все по-амбициозни. В последните години все по-силно започва да се развива областта по извличане на мнения от социалните мрежи с отчитане на "сложната кривина" на функцията "харесвам-не харесвам" предвид спецификата на употребата на езика за изразяване на едно отношение с противоположни изразни средства. Задачата още повече се усложнява от богатството на изказа, който естественият език предоставя чрез използването на изрази с двойно значение.

В случая Тодор Цонков се е заел точно с тази нелесна задача – определяне на наличието на двойно значение в текстове от социалните мрежи, които ежедневната употреба на езика е формирала като практики.

2 Общо описание на представените материали и документи

За разглеждане бяха представени:

- дисертационен труд в обем от 101 страници;
- автореферат на български език (32 стр.);
- 4 публикации, в които са изложени части от дисертационния труд, от които една е в сборник на международна конференция (Balkan Conference on Informatics 2015), две в национални конференции (НК "Образованието и изследванията в ИО – 2013 и 2014 г) и една в списание ТУ-София, филиал Пловдив;
- резюме на доклада, изнесен на конференцията "125 години математика и природни науки в СУ Св. Климент Охридски" (5-7.12.2014).

Дисертацията се състои от: въвеждаща част, в която ясно са очертани мотивацията и поставените цели; следващите три глави са посветени на описание на предлаганите подходи и проведените експерименти за доказване на качеството на хипотезите за откриване в текстове от социалните мрежи на двузначно значение (ирония, сарказъм и сатира), определяне на преувеличения и невярна информация, класифициране на мнения за наличие на определени отношения към обсъжданите обекти и събития. Пета глава е посветена на базиран на правила подход за автоматично откриване на наличието на невярна информация, както и на алгоритъм за подобряване на списъка от думи и изрази, носещи определено отношение. Следващата глава описва принципната схема и конкретната реализация на система за автоматично извличане и анализ на текстове от социалната мрежа с вграден модул за оценяване на мненията на потребителите. Заключението обобщава постигнатите резултати и конкретните приноси според автора и съдържа плановете му за бъдещо развитие на работата.

Авторефератът коректно отразява основните моменти на дисертационния труд и съответства по форма и по съдържание на изискванията.

3 Оценка на основните приноси на дисертационния труд

Авторът показва добро познаване на методите и средствата за разрешаване на проблемите по откриване на двойно значение в текстове от социалните мрежи. Той обръща внимание както на дефинирането на определени правила за търсенето на наличие на изследваните нюанси на двойно значение, така и на другата съществена част от процеса – изграждането на пространство на характеристиките, върху които да работят съответните алгоритми.

Експериментите са добре планирани и проведени в достатъчен обем. Резултатите от експериментите доказват повишаване на качеството на разпознаване при използването на предложените правила и извлечени характеристики.

За самостоятелността на извършването на работата личи и по описанието на изградената система за извличане на мнения от текстове от социалната мрежа, като авторът предлага и принципни изисквания, на които следва да отговаря такава система.

Разнообразието от теми – търсене на двойно значение, преувеличение, лъжа, оценка на наличието на слухове, всъщност са подчинени на общия проблем по

идентифициране на разминаване на прякото значение на думите със смисъла на преносното съобщение.

Добро впечатление прави оценката на реалната задача – в частта за оценка на качеството на разпознаване се отчита невъзможността за получаване на единно мнение от самите съдии за наличието или неналичието на двойственото значение, което само по себе си води до невъзможност за получаване на 100% точност на разпознаването, както и необходимостта от въвеждане на знания за конкретната област за постигане на по-добро разпознаване.

4 Бележки и препоръки

Като цяло текстът е добре структуриран. Следването на схемата – предложен алгоритъм, реализация, постигнати резултати, изводи – спомага за стройното представяне на предложените идеи, свършената работа и аргументиране на постигнатите резултати.

Забелязани граматически грешки (неправилно членуване, липсващи или излишни предлози, объркан словоред и др.) съм изпратила на автора за поправка.

Предложените алгоритми и правилата за извод по откриване на различните ситуации имат доста общи черти. Работата дотук предлага едно богато поле за действие по-нататък. В бъдеще би могло да се продължи с изследванията като се анализира взаимовръзката между отделните явления и обобщаване на алгоритъма с параметризиране на съответните случаи.

5 Заключение

От моя гледна точка представеният труд напълно съответства по форма и съдържание на изискванията за дисертация за получаване на образователната и научна степен "доктор" по научна специалност 01.01.12 Информатика (Изкуствен интелект); професионално направление 4.6 Информатика и компютърни науки, затова препоръчвам на членовете на уважаемото жури да гласуват за присъждане на тази степен на Тодор Васков Цонков.

25.03.2016

Подпис: