ТЕХНИЧЕСКИ УНИВЕРСИТЕТ - ВАРНА

ФИТА

МЕСЕЦ НА НАУКАТА 2022

АВТОМАТИЗИРАН АНАЛИЗ НА МНЕНИЯ В ТЕКСТОВЕ НА БЪЛГАРСКИ ЕЗИК

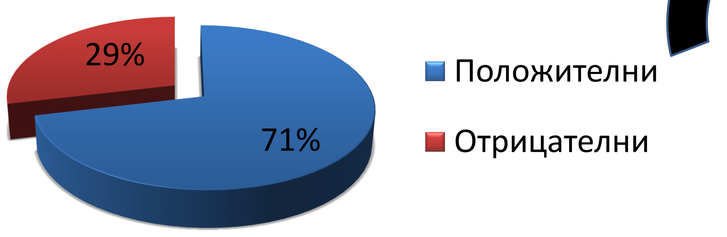
Виолета Божикова, доцент, СИТ
Даниела Петрова, докторант, СИТ

Въведение

Проектът е насочен към провеждане на изследвания в областта на автоматизирани системи за обработка на информация и в частност в извличането на мнения и анализа на настроения в текстове на български език.

Докато за текстове на английски език има изобилие от разработени методологии, алгоритми, методи, приложения, бази данни и готови софтуерни продукти, почти не се откриват такива, приложени върху текстове на български език, което предполага нуждата от провеждането на подробно изследване и търсенето на най-подходящия алгоритъм за извличането на мнения в спецификите и особеностите на българския език.

База данни 1



База данни 2

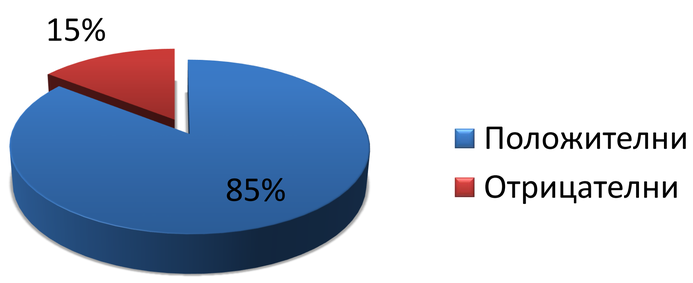


Таблица 1. Бази от данни след предварителната обработка

Коментари	База данни 1	База данни 2
Положителни	63 714	84 489
Отрицателни	25 624	14 357
Общо	**89 341**	**98 846**

Заключение

Във връзка с поставените задачи са направени опити в търсенето на най-подходящия подход за разработка на софтуер за автоматизирането на анализа на настроения в мнения за различни продукти и услуги на български език.

Изпълнените дейности по проекта подпомагат научно-изследователската работа на колектива и на научното звено по продължаване на анализа на методите и подходите за автоматизирани системи за обработка на информация.

Резултати

1. Разработени са две база данни с мнения и коментари на български език: от гости на хотели и от потребители на различни видове услуги и стоки;
2. Разработен е алгоритъм за предварителна обработка на данните, които да се използват в извличането на мнения в спецификите и особеностите на българския език, който дава оптимални резултати;
3. Направен е сравнителен анализ дали обекта и сферата на коментарите влияе върху резултатите от анализа на мнения

Таблица 2. Резултати от приложени методи

Метод	База данни 1	База данни 2
	Без прилагане на stemming	
Naïve Bayes	**0.8683**	0.8530
Логаритмична регресия	0.8503	0.9355
Логаритмична регресия с би-грами	0.8575	0.9338
Метод на опорните вектори	0.8239	0.9375
Метод на опорните вектори с би-грами	0.8413	**0.9451**
	С прилагане на stemming	
Naïve Bayes	0.8639	0.8489
Логаритмична регресия	0.8508	0.9387
Логаритмична регресия с би-грами	**0.8644**	0.9410
Метод на опорните вектори	0.8326	0.9385
Метод на опорните вектори с би-грами	0.8475	**0.9467**

Публикации по проекта

1.D.Petrova, Comparative assay on sentiment analysis on two databases in Bulgarian language, ICMECE 27-28.11.2021 Ankara, Turkey, ISBN:978-625-409-707-2, pp. 43-47. 2.D.Petrova, Automatic Sentiment Analysis on Hotel Reviews in Bulgarian – Basic Approaches and Results, IEMAICLOUD 26-28.04.2021 London, UK, ISBN:978-3-030-92904-6 pp.48-56. 3.D.Petrova, V.Bozhikova, Development of two data bases with comments in Bulgarian language and application of supervised learning approaches on them for comparative sentiment analysis – summary, Е-годишник на ТУ-Варна -под печат

Благодарности

Финансирането е от бюджетната субсидия за наука на Технически университет-Варна за 2021г.