

РОЗВ'ЯЗАННЯ ЗАДАЧІ ЕМОЦІЙНОГО АНАЛІЗУ ТЕКСТУ НА ОСНОВІ ЛІНГВІСТИЧНОГО ПАКЕТУ DKPRO CORE

Вінницький національний технічний університет

Анотація

Розроблено засоби автоматизованого емоційного аналізу англomовних текстів на основі вільно доступних бібліотек, алгоритмів та програмних інструментів лінгвістичного пакету DKPro Core. Експериментальне дослідження отриманої програми проведено на основі інавгураційних промов президентів США.

Ключові слова: програмне забезпечення, емоційний аналіз, експеримент, інавгураційна промова, англomовний текст, DKPro Core.

Annotation

The means of automated emotional analysis of English texts on the basis of free available libraries, algorithms and software tools of the linguistic package DKPro Core has been developed. An experimental research of the program was conducted on the basis of inaugural speeches by US presidents.

Key words: Software, emotional analysis, experiment, inaugural speech, English text, DKPro Core.

Вступ

Комп'ютерна лінгвістика – галузь мовознавства, що вивчає мову за допомогою комп'ютера, а також створює лінгвістичне забезпечення для комп'ютерних систем опрацювання інформації [1]. Центральними науковими проблемами комп'ютерної лінгвістики є проблема моделювання процесу розуміння сенсу текстів (переходу від тексту до формалізованого подання його сенсу), а також проблема синтезу мови (переходу від формалізованого подання сенсу до текстів на природній мові).

Означені проблеми виникають при вирішенні ряду прикладних задач, зокрема автоматичного перекладу текстів з одних мов на інші, спілкування з комп'ютером на природній мові, автоматичної класифікації та індексування текстових документів, їх автоматичного реферування, пошуку документів в повнотекстових базах даних [2]. Актуальною задачею, корисною для цілого ряду інших задач комп'ютерної лінгвістики, вважають емоційний аналіз природно-мовних текстів.

Мета роботи полягає в побудові засобів автоматизованого емоційного аналізу англomовних текстів на основі вільно доступних бібліотек, алгоритмів та програмних інструментів лінгвістичного пакету DKPro Core [3].

Результати дослідження

В основу розробки програмного забезпечення було покладено ідеї формалізованого у [4] підходу до визначення сенсу текстової інформації. Для експериментального дослідження такого підходу в умовах задачі автоматизованого емоційного аналізу англomовних текстів було обрано інавгураційні промови усіх президентів США.

Для реалізації програми обробки тексту інавгураційної промови кожного з президентів США на основі інструментальних засобів DKPro Core було розроблено програму, що складається з:

- модуля автоматизованого зчитування інавгураційної промови з веб сайту;
- модуля аналізу промови та підрахунку настрою;
- модуля виведення результатів у файл.

Розроблена програма зчитує промову з веб сторінки та ділить її на речення. Наступним кроком є визначення настрою кожного слова – програма отримує з бібліотеки DKPro Core дані про позитивний та негативний стан слова та присвоює йому той чи інший рівень. Застосовується шкала, що має відповідні поділки від 1 до 5 для позитивних слів та від -1 до -5 для негативних, а основна інформація зчитується з наступних файлів:

- EmotionLookUpTable.txt, в якому міститься словник зі словами та рівнем їх настрою;

- NegatingWordList.txt – в цьому файлі міститься характеристика зниження настрою слова, наприклад, not happy має негативний настрій;
- BoosterWordList.txt вміщує випадки, які підвищують рівень настрою, наприклад, very happy має більш позитивну оцінку, ніж happy.

Далі програмно підраховується загальний рівень настрою по кожному реченню, отримані дані узагальнюються для тексту в цілому і передаються до наступного модуля виведення результатів. Загальна кількість слів, що використовувались для аналізу настрою в проведеному експерименті, складає приблизно 3000. Наприкінці для зручності сприйняття програма переводить значення суми настрою промови у формат від 0 до 1, де 0 – різко негативна, 1 – дуже позитивна.

Проведені дослідження показали що рівень настрою кожної інавгураційної промови помітно різниться, причому суттєво впливають на зниження настрою війни та економічні кризи у відповідні періоди. Проте, в цілому, середньостатистичний рівень настрою очільників США з часом спадає. Результати дослідження відображено на рисунку 1.

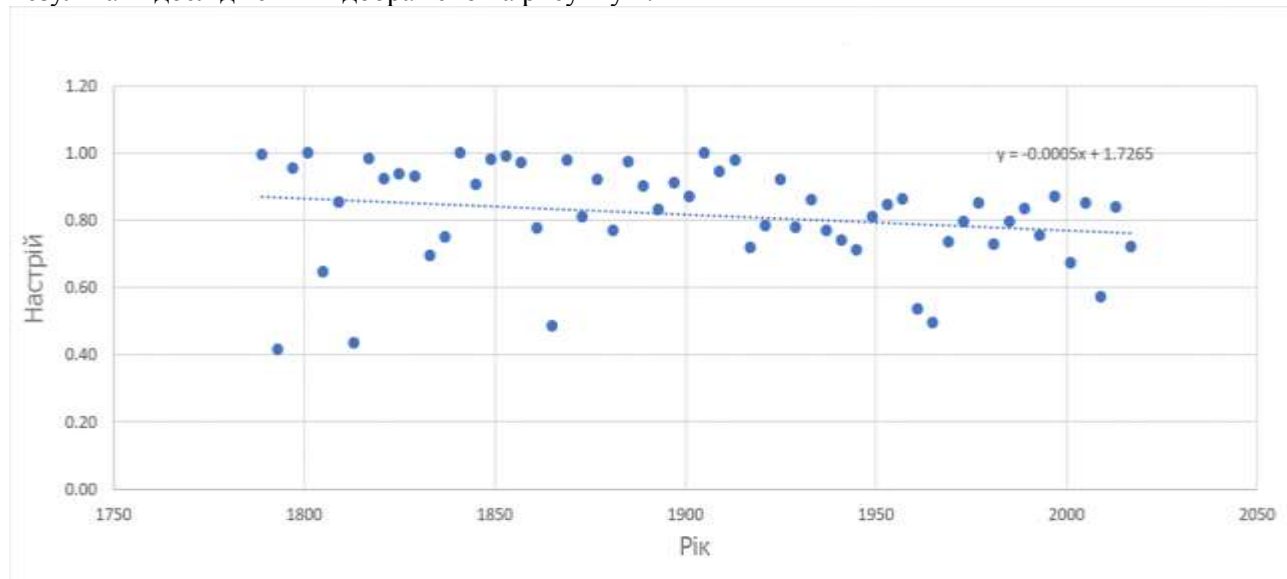


Рис. 1. Рівень настрою інавгураційних промов президентів США

Отримані результати шляхом порівняння з історичними даними свідчать, що розроблене програмне забезпечення успішно виконує поставлену задачу. Важливо, що запропонована програма може бути модернізована для емоційного аналізу тексту будь-якої іншої мови шляхом наповнення відповідних словників.

Висновки

Розроблено та успішно апробовано програмне забезпечення, що базується на інструментальних засобах лінгвістичного пакету DKPro Core та успішно розв'язує задачу автоматизації емоційного аналізу англійського тексту за рахунок своєї гнучкості. Експериментальне дослідження програми емоційного аналізу було проведено на основі текстів інавгураційних промов усіх президентів США, отримані результати корелюють з історичними даними. У майбутньому запропонована програма може бути модернізована і для інших мов шляхом наповнення відповідних файлів-словників.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики. — Донецьк: Юго-Восток, 2003. — 184 с.
2. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник. — К.: Видавничо-поліграфічний центр "Київський університет", 2008. — 351 с.
3. Natural Language Processing: Integration of Automatic and Manual Analysis [Електронний ресурс]. — Режим доступу: <http://tuprints.ulb.tu-darmstadt.de/4151/1/rec-thesis-final.pdf>. — Назва з екрану.
4. Кветний Р.Н. Визначення сенсу текстової інформації на основі моделі розповсюдження обмежень / Р.Н. Кветний, О.В. Бісікало, І.О. Назаров // Вимірювальна та обчислювальна техніка в технологічних процесах. — 2012. — № 1. — С. 93–96.

Котик Ярослав Юрійович — студент групи ЗАКІТ-18м, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, e-mail: kotikyaroslav@gmail.com

Науковий керівник: **Бісікало Олег Володимирович** — д-р техн. наук, професор, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця

Koty Yaroslav Y. — student of group ЗАКІТ-18m, Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, e-mail: kotikyaroslav@gmail.com

Supervisor: **Bisikalo Oleg V.** — Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia