



## ВИРІШЕННЯ ЗАВДАННЯ SENTIMENT ANALYSIS ЗА ДОПОМОГОЮ СЛОВНИКОВОГО ПІДХОДУ (НА ПРИКЛАДІ МАТЕРІАЛУ ФОРУМІВ СМАРТФОНІВ)

**Медведська А.В.**

*Національний технічний університет  
«Харківський політехнічний інститут»  
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,  
e-mail: nyuta3871@inbox.ru*

Вивчення думок і відгуків інших людей стало важливою частиною процесу прийняття рішень. Коли певний вибір пов'язаний з фінансовими витратами, то люди так чи інакше спираються в цьому питанні на досвід інших людей. Сьогодні «social web» дає нові можливості для обговорення та поширення інформації серед людей на форумах, у блогах, соціальних мережах, де люди мають можливість обмінюватися думками щодо різних товарів та послуг. Аналіз думок ще називають тональністю, або opinion mining – це клас методів контент-аналізу в комп'ютерній лінгвістиці, призначений для автоматизованого виявлення в текстах емоційно забарвленої лексики та емоційної оцінки думок по відношенню до об'єктів [1].

В аналізі тональності тексту вважається, що текстова інформація в мережі Інтернет ділиться на два класи: факти і думки. Ключовим поняттям є визначення думок.

Думки діляться на два типи:

1. Проста думка.
2. Порівняння.

Проста думка містить висловлювання автора про один об'єкт. Вона може бути висловлена прямо, або неявно. В обох випадках проста думка зазвичай має емоційне забарвлення – позитивне чи негативне.

Другий тип думок – порівняння – можна розділити на три види [2]:

1. Порівняння аспектів об'єктів на користь одного (Non-equal Gradable).
2. Прирівнювання аспектів різних об'єктів (Equative).
3. Перевага одного об'єкту над іншими (Superlative).

Загальний підхід у вирішенні задачі Opinion Mining виглядає наступним чином: збір колекції документів, попередня обробка текстів, виділення ознак емоційності, визначення полярності тексту.

Визначення полярності тексту зазвичай розглядається на двох рівнях:

1. На рівні документа.
2. На рівні речення.

Завдання на рівні документа полягає у визначенні полярності документа в цілому. Причому, текст документа може одночасно містити речення, як з негативним, так і з позитивним емоційним забарвленням.

Завдання визначення полярності речень вирішується як в якості підзадачі аналізу тональності документа, так і в якості самостійного завдання, наприклад,



при аналізі коротких повідомлень і коментарів у соціальних мережах. Цей метод призначений для більш точного аналізу думки в тексті [3].

При вирішенні задачі аналізу тональності тексту популярними є методи, засновані на знаннях, Machine Learning та прихована Марковська модель.

Етапи вирішення задачі аналізу тональності текстів:

1) Збір текстів для перевірки роботи програми.

Для перевірки роботи програми було зібрано 200 повідомлень позитивної на негативної забарвленості за тематикою смартфони. Для збору повідомлень було обрано англomовний сайт <http://www.pcadvisor.co.uk>. На цьому сайті розміщені огляди різної техніки, у тому числі мобільної, де користувачі мають можливість обмінюватися думками щодо певного товару. Повідомлення мають наступний вигляд: «I love Samsung products! I had an iPhone prior to this GalaxyS5. I would never go back! Samsung products are so user friendly. I have to Samsung tablets that are wonderful. Thus phone has excellent photo quality! To much to name.»

2) Складання словників позитивно на негативно забарвленої лексики.

Були складені словники емоційно забарвленої лексики, що виражають думки користувачів щодо мобільної техніки. Мова повідомлень англійська. Кількість позитивних слів – 103, кількість негативних – 86, позитивних смайликів – 26, негативних – 22. Наприклад, словник позитивно забарвленої лексики містить такі слова: strong, excited, gorgeous, delightful та інші. А негативно забарвленої – такі, як: disgruntled, terrible, dreadful, ugly та інші.

3) Розбір повідомлення на слова. У такому вигляді їх можливо буде використовувати для подальшої обробки.

4) Пошук емоційних елементів за допомогою системи Analysis of emotion. Якщо повідомлення містить емоційні елементи – сильно емоційно забарвлене, якщо не містить – не сильно.

5) Підключення адаптованого словнику емоційно забарвлених слів та перевірка повідомлень на наявність графічних усмішок (смайликів).

6) Визначення емоційної забарвленості повідомлення з використанням математичних методів для підрахунку ваг слів та смайликів. Для подальшого аналізу були обрані методи [4] та [5].

### Список літератури

1. Stefano Baccianella Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining (англ.) // Proceedings of LREC : конференція. – 2010. – С. 2200-2204.
2. Bing Liu. Sentiment Analysis Tutorial // AAAI-2011, San Francisco, USA.
3. Pang B. & Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval, v.2 n.1-2, January, 2008 - pp.1-135.
4. Усталов Д. А. Вилучення термінів з російськомовних текстів за допомогою графових моделей // Теорія графів і додатків. Єкатеринбург: Вид-во Урал. – 2012. – с. 62-69.
5. Клековкіна М.В., Котельников Є.В. Метод автоматичної класифікації текстів за тональністю, заснований на словнику емоційної лексики // XIV Всерос. наук. конф. Переславль-Залеський: Вид-во «Університет міста Переславль». – 2012. – с. 118-123.