

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА
АКАДЕМІЯ»
Кафедра інформатики

**Аналіз повідомлень в рядках новин сайтів методом
сентимент аналізу
Текстова частина до курсової роботи
за спеціальністю «Інженерія Програмного забезпечення»**

Керівник курсової роботи:
доцент, кандидат технічних наук
Ковалюк Т.В.

(підпис)

“ ____ ” _____ 2020 р.

Виконала студентка

Якимчук Соломія Олегівна

“ ____ ” _____ 2020 р.

Київ 2020

КАЛЕНДАРНИЙ ПЛАН ВИКОНАННЯ КУРСОВОЇ РОБОТИ

Тема: Аналіз повідомлень в рядках новин сайтів методом сентимент аналізу

Календарний план виконання роботи:

№ п/п	Назва етапу курсової роботи	Термін виконання етапу	Пр имітка
1.	Отримання завдання на курсову роботу.	15.11.2019	
2.	Огляд технічних статей за темою роботи.	15.01.2020	
3.	Ознайомлення з можливостями сентимент аналізу	10.02.2020	
4.	Збір інформації	10.03.2020	
5.	Розгляд і аналіз існуючих застосунків	20.03.2020	
6.	Переваги та недоліки застосунків	20.03.2020	
7.	Розробка застосунку	30.03.2020	
8.	Тестування застосунку	05.04.2020	
9.	Аналіз отриманих даних	08.04.2020	
10.	Створення слайдів для доповіді та написання доповіді.	17.04.2020	
11.	Аналіз отриманих результатів з керівником, написання доповіді та попередній захист курсової роботи.	20.04.2020	
12.	Корегування роботи	21.04.2020	
13.	Остаточне оформлення пояснювальної роботи та слайдів.	23.04.2020	
14.	Захист курсової роботи (проекту)	25.04.2020	

Студент: Якимчук Соломія Олегівна

Керівник: Ковалюк Тетяна Володимирівна

“ ____ ” _____ 2020 р.

Зміст

АНОТАЦІЯ.....	4
ВСТУП.....	6
РОЗДІЛ 1. ХАРАКТЕРИСТИКА СЕНТИМЕНТ АНАЛІЗУ	10
Розділ 1.1 Класифікація тональностей	10
Розділ 1.2. Задачі аналізу тональності	15
Завдання аналізу тональності	20
Розділ 2 Аналіз аналогів	25
Розділ 2.1 Переваги та недоліки SentiStrength	25
Розділ 2.2 «Brand Analytics»	28
Розділ 2.3 «YouScan»	32
Розділ 3 Опис Розробки програмного продукту.....	36
Розділ 3.1 Вимоги до програмного продукту	36
Функціональні вимоги.....	36
Нефункціональні вимоги.....	36
Розділ 3.2 Архітектура та структура програмного продукту	36
Розділ 3.3 Засоби та технології розробки	38
Висновок до розділу	38
Загальний ВИСНОВОК.....	39
Список використаної літератури.....	41

АНОТАЦІЯ

Анотація: приймаючи рішення, ми звикли спиратися на інформацію ззовні, будь то відгуки, соціальні опитування або чії-небудь судження про цікаві нам предмет, а тому зовсім не дивно, що така область лінгвістики, як сентимент-аналіз, міцно увійшла в наше життя. І тим очевидніше необхідність досліджень з розроблення методики, яка зробить можливим визначення тональності тексту як в цілому, так і його окремих частин. Дана робота присвячена аналізу повідомлень в рядках новин сайтів методом сентимент аналізу.

Ключові слова: тональність тексту, сентимент-аналіз.

ANALYSIS OF POSTS IN NEWS LINES OF SITES BY THE METHOD OF SENTIMENT ANALYSIS

Yakymchuck S.

Abstract: *making decisions, we are accustomed to rely on information from outside, whether it is feedback, social surveys or someone's opinions about the subject of interest to us, and therefore it is not surprising that such a field of linguistics as sentiment-analysis has firmly entered our life. And the more obvious is the need for research on the development of a methodology that will make it possible to determine the tonality of the text both in its entirety and its individual parts. This work is devoted to the analysis of messages in news lines of sites by the method of sentiment analysis.*

Keywords: *tonality of the text, sentiment-analysis.*

ВСТУП

Сентимент - аналіз інформаційних потоків, який має великий потенціал застосування для моніторингових, аналітичних і сигнальних систем, для систем документообігу і рекламних платформ, націлених по тематиці веб-сторінок.

Сентимент аналіз – це контекстний розбір тексту, для того щоб визначити тональність повідомлення, яке аналізується. Тональність може бути позитивною, негативною, або нейтральною. [20]

Текст на природній мові, крім інформації, може висловлювати емоційну оцінку того, про що повідомляється.

Виражена в тексті емоційна оцінка називається тональністю або сентиментом (від англ. Sentiment - почуття; думка, настрої) тексту.

Ми зазвичай оцінюємо все за різними критеріями: красивий, бідний, розумний, сміливий, слабкий, тощо. Проте, всі вони мають дві сторони позитивну і негативну(красивий, багатий, розумний, сміливий, сильний – позитивні, страшний, бідний, дурний, боязкий, слабкий - негативні). Тому для простішої обробки запитів ми будемо брати лише характеристику позитивна, або негативна.

Історично склалося так, що традиційний підхід до сентимент аналізу це так звана задача, яка поділяє текст, чи частину тексту на дві категорії: позитивний і негативний, або якщо бути детальним то можна розділити на три і додати ще нейтральну категорію. Саме з такого завдання почав свій розвиток аналіз тональності: оцінити сентимент оціночних відгуків з якої-небудь тематики (кіно, ресторани, електроніка та ін.).

Тим не менш, це не єдиний і не визначальний тип завдання, яке має вирішувати сентимент аналіз тексту. В даний час читачів цікавить не загальна емоційна оцінка тексту, а відношення сентимент до конкретного об'єкта, про що йдеться в тексті, або відношення суб'єкта висловлювання до обговорюваного об'єкту.

Те, що ми оцінюємо ми зазвичай називаємо об'єктом тональності.

В кожного такого об'єкту, який аналізується є людина, яка його написала, або процитувала її можна назвати автором об'єкту, якщо це оригінальний текст, то автором є людина, яка написала його. Якщо текст процитований, то автором буде людина, яка оригінально його висказала.

Таким чином, тональність висловлювання визначається трьома компонентами: суб'єктом тональності (хто висловив оцінку), об'єктом тональності (про кого або про що висловлена оцінка) і власне тональної оцінкою (як оцінили).

Ще одним напрямком сентимент аналізу є виявлення негативності / позитивності атрибутів об'єкта тональності (feature-based / aspect-based sentiment analysis).

Крім самої тональності, текст можна оцінювати по суб'єктивності / об'єктивності судження (Opinion Mining). Якщо ця думка автора висловлювання, що містить суб'єктивну оцінку описуваного, то текст вважається суб'єктивним. І навпаки, якщо це повідомлення ЗМІ або думка, за замовчуванням розділяється учасниками діалогу, то воно вважається об'єктивним.

До суб'єктивної інформації будуть ставитися пряма і непряма мова в тексті, а також цитування. У таких випадках автоматичне визначення суб'єктивності / об'єктивності висловлювання реалізувати технічно набагато простіше, ніж в загальному випадку.

Існує два основні методи вирішення цього завдання автоматичного визначення тональності:

- Статистичний метод. Для нього потрібні заздалегідь розмічені по тональності колекції (корпусу) текстів, на яких відбувається навчання моделі, за допомогою якої і відбувається визначення тональності тексту або фрази.
- Метод, заснований на словниках і правилах. Для цього заздалегідь складаються словники позитивних і негативних слів і виразів. Цей метод може

використовувати як списки шаблонів, так і правила з'єднання тональної лексики всередині пропозиції, засновані на граматичному і синтаксичному розборі.

- Крім того, іноді використовують змішаний метод (комбінацію першого і другого підходів).

При статистичному підході для вирішення завдання загальної класифікації текстів на класи тональності широко використовують метод опорних векторів (SVM), Байєсови моделі, різного роду регресії [Chetviorkin & Loukachevitch - опис змагання ROMIP-2011 по сентимент аналізу даних, практично всі учасники використовували SVM або Байєс] .

Якщо ж метою є визначення тональності у певного, заздалегідь заданого об'єкта (декількох об'єктів), то застосовують більш складні статистичні алгоритми, такі як CRF [Антонова і Соловйов], алгоритми семантичної близькості (наприклад, латентно-семантичний аналіз - LSA, латентне розміщення Дирихле - LDA) та ін., а також методи, засновані на правилах [Пазельская і Соловйов]. Для визначення атрибутивної тональності використовують мовні моделі [García-Moya & all], нейромережі [Tarasov], або тематичні тезауруси.

Уміння розпізнавати тональність текстів широко застосовується в різних видах діяльності людини, таких як вивчення переваг споживача; системах рекомендацій, покликаніх допомогти споживачеві при виборі товарів або послуг; аналізі новинних ресурсів на предмет тональності повідомлень щодо різних персон або подій; підтримки пошукових систем і систем вилучення інформації; психологічних дослідженнях тощо.[22]

Одним з найбільш актуальних завдань в галузі аналізу текстових повідомлень в сайтах з елементами соціальних мереж є задача розпізнавання емоційного забарвлення тексту, яка дозволяє витягти з текстової інформації полярність думки людини про об'єкт висловлювання.

Для досягнення поставленої мети необхідно виконати наступні завдання:

1. Аналіз предметної галузі, виявлення переваг і недоліків існуючих рішень.
2. Аналіз, порівняння і вибір підходу до автоматичного визначення тональності тексту.
3. Реалізація обраного методу визначення тональності, вибір параметрів.
4. Оцінка якості роботи.

РОЗДІЛ 1. ХАРАКТЕРИСТИКА СЕНТИМЕНТ АНАЛІЗУ

Розділ 1.1 Класифікація тональностей

Тональністю(думкою) називають судження про будь-яку сутність, чи її аспекти, висловлене деякими суб'єктом і забарвлене позитивно, негативно, або нейтрально.

Тобто можна тональність представити кортежем з 5 елементів

< holder, aspect, sentiment, entity, time > ,

де:

- (holder) - виразник думки (суб'єкт);
- (aspect) - властивість об'єкта, по відношенню до якого висловлюється думка;
- (sentiment) - тональність думки по відношенню до і-му аспекту сутності
- (entity) - об'єкт, про аспект якого було висказано думку автором в момент часу. Об'єктом може бути товар, людина, подія, організація або тема обговорення;
- (time) - час вираження думки.[16,17]

В аналізі тональності тексту вважається, що текстова інформація ділиться на два типи: факти і думки. Ключовим поняттям є визначення думки.

Тональність поділяються на 2 типи:

- Думка проста;
- порівняння.

Думка проста має в собі слова автора про деякий об'єкт . Ця думка може бути висловлена наступним чином: «Восьма частина Гаррі Поттера не забуття», або не прямо: «Оцінивши всі частини даної фантастики можна прийти до висновку, що йде покращення».

Думку просту можна оформити формально використовуючи наш кортеж. І описати як те, що автор сказав думку про деякий аспект в деякий період часу.

Зазвичай виділяють два види емоцій (sentiment value): позитивна та негативна, тобто класифікація виконується за двома класами. Іноді додається третій - нейтральна думка.[19]

Часто об'єкт може бути поданий у вигляді ієрархічної структури.

З кожним компонентом можна пов'язати певну кількість атрибутів. Вище сказані слова щодо думки аспектом називають і атрибути, і компоненти. Є один виняток, це коли об'єкт і є аспектом.

Наступний вид думок, а саме порівняння можна розділити на 3 види

- порівняння аспектів об'єктів на користь одного
- прирівнювання аспектів різних об'єктів
- перевага одного об'єкта над іншими

В першому варіанті порівняння буде вигляду «об'єкт один є кращим за об'єкт два за деяким аспектом», наприклад: «Комп'ютер «Х» кращий за комп'ютер «У» по розмірах»

В другому варіанті у нас є подібність деяких аспектів деяких об'єктів «Розміри в комп'ютерів «Х» та «У» майже такі ж».

Прикладом третього типу може слугувати речення «В конкурсі на кращий комп'ютер року виграє «Х» у «У»»

Думка другого типу визначається як кортеж

<O1, O2, A, holder, time>

В даному кортежі O1 и O2 – об'єкти, яких порівнюють за аспектом A, де є певний автор і було написано в певний час. Якщо порівнювати кортежі першого і другого варіантів, то другий не має прямої оцінки емоцій автора.

Якщо ми говоримо про аналіз тональності, то часто впливатиме такі визначення, як суб'єктивність і об'єктивність. Давайме розберемось з цими двома термінами:

- об'єктивне – це про факт, а суб'єктивне це про емоції, які виражає якась людина
- об'єктивні речення зазвичай не мають забарвлення, тому аналіз тексту на існування суб'єктивної, чи об'єктивної інформації часто є під завданням визначення полярності тексту.

Прийнято розрізняти такі підходи до розпізнавання тональності тексту:

1. Підхід, заснований на правилах, полягає в використанні наборів правил, виявлених експертами на основі аналізу предметної області, на підставі яких з тексту витягуються n-грами (n-компонентні ланцюжка).
2. Метод, заснований на застосуванні тональних словників. Кожне слово тексту зіставляється зі словником, підсумкова тональність тексту визначається як середнє арифметичне всіх тональностей. Залежно від словника можуть бути застосовуються різні шкали оцінок. Наприклад, в даній роботі використовується NRC Word-Emotion Association Lexicon - словник, що приписує словам різні емоції.
3. Навчання з учителем. Алгоритм тренується на деякому заздалегідь розміченому корпусі текстів. Кожен розмічений текст представлений у вигляді пари - вектора ознак тексту (набір слів і словосполучень з відповідними їм вагами) і приписаної йому тональності. На підставі такої вибірки будується потім статистичний класифікатор, який використовується для визначення тональності нової колекції документів.
4. Навчання без вчителя. Для тренування алгоритму використовується навчальна вибірка нерозмічену заздалегідь текстів. При такому підході найбільшу вагу отримують найбільш часто зустрічаються в тексті терміни, але які при цьому присутні тільки в обмеженій кількості текстів всієї множини.

Так, аналізуючи різні тексти, можна наочно побачити, як в цілому забарвлений текст: позитивно чи негативно.

Окрім такої класифікації, звичайні думки також розділяються на явні (explicit) та неявні (implicit) думки (що мають на увазі).

Отже думка звичайна думка дуже часто називають думкою простою в літературі, та має два основні під види.

Думка явна – це прямо виражені думки щодо певного аспекту деякого об'єкту. Наприклад: «Книжка просто неймовірна».

Думка неявна – це думка, що виражається неявно щодо аспекту або сутності, на основі ефекту, що є у цієї сутності на інші сутності. Такий підтип досить часто виникає в медичній області. Наприклад, речення «Після проведеної процедури моє самопочуття погіршилось» неявна негативну думку висловлюється щодо процедури, адже ця сутність «процедура» має негативний ефект на іншу сутність «стан».

Більшість актуальних досліджень є на основі думок явних, адже їх легше аналізувати. Думки неявні в то й ж час не просто аналізувати. Наприклад, в сфері обслуговування потрібно розуміти, чи процедури надані ефективно. Прикладом може бути речення: «Оскільки у мене був поганий манікюр я скористалась послугами деякого салону краси». Але ж поганий манікюр не відноситься до цього салону краси.

Порівняльна думка висловлює відношення подібності або відмінності між однією чи двома сутностями, та визначає, яку саме сутність врешті було обрано власником думки. Наприклад, речення «Кола на смак краще, ніж Пепсі» та «Кола найкраща» висловлюють дві порівняльні думки. Порівняльні думки, як правило, висловлюються в порівняльній формі прикметника або прислівника, хоча і не завжди (наприклад, «я надаю перевагу»).[12]

Явна думка. Явна думка – це суб'єктивне твердження, що надає звичайну чи порівняльну думку, наприклад:

«Яблучний сік смачний» або «Яблучний сік смачніший за томатний»

Неявна думка. Неявна думка – це об'єктивне твердження, що має в собі

звичайну чи порівняльну думку. Таке твердження, як правило, виражає бажаний чи небажаний факт, наприклад:

«Я купив цей матрац два тижні тому, і він деформувався», та «Заряд батареї телефонів Nokia кращий, ніж в телефонів Samsung».

Явні думки простіше виявити та класифікувати, ніж неявні. Більшість сучасних досліджень фокусується саме на явних думках. Досить мало досліджень було проведено щодо визначення та класифікації неявних думок.

Існують дві концепції, що є дуже тісно пов'язаними з класифікацією емоційного забарвлення думок – суб'єктивність та емоції.[11]

Об'єктивне речення визначає певну фактичну інформацію щодо навколишнього світу, а от суб'єктивне твердження надає інформацію про деякі емоції певної особи.

Прикладом об'єктивного твердження є наступне речення: «iPhone є продуктом компанії Apple». Прикладом суб'єктивного твердження є наступне речення: «Мені подобається iPhone». Суб'єктивні твердження можуть бути у різних інтерпретаціях це може бути і примха, і бажання, і деяка думка.

Є досить таки серйозна проблема зрозуміти де у нас суб'єктивність, а де об'єктивність, особливо у літературі.[10]

Деяке виказування думки говорить про те, що документ чи окреме речення виказує щодо певної сутності деяке забарвлення, чи негативне, чи позитивне. Ці дві концепції не є еквівалентними, не дивлячись на те, що вони досить тісно перетинаються. Задача, що займається визначенням того, чи має документ суб'єктивне чи об'єктивне твердження носить назву «Класифікація суб'єктивності». Ми повинні мати на увазі наступне:

- суб'єктивне речення може не висловлювати ніякої думки. Наприклад, речення «Я думаю, що він пішов додому» є повністю суб'єктивним, але не висловлює нічого. Речення в попередньому прикладі також є повністю суб'єктивним та не висловлює позитивну, негативну чи нейтральну думку щодо камери або одного з її аспектів;

- об'єктивні речення можуть висловлювати думки на основі ствердження бажаних чи небажаних фактів. Наприклад, наступні два речення, що просто висловлюють факти об'єктивно також і висловлюють негативну думку (неявну негативну думку) щодо певних продуктів: «Цей комп'ютер перестав працювати на третій день після покупки» або «Браслет порвався після першого ж одягання».[7]

Окрім явних думок, які несуть в собі суб'єктивні висловлювання, дуже багато інших типів суб'єктивності вивчалися аналітиками, але не так інтенсивне. Дуже багато таких типів можуть також висловлювати ту чи іншу думку. Емоції досить тісно пов'язані з висловленням думки. Вони були класифіковані та виділені в певні категорії – любов, здивування, радість, сум, страх. Сила емоційного забарвлення, що присутня в думці, як правило пов'язана з силою певної емоції, тому можна деколи зустріти не бінарну класифікацію на гарні/погані емоції а класифікацію за типами емоцій, що містить висловлювання. Найчастіше така класифікація відбувається за допомогою словникових методів, адже дуже важко отримати достатньо велику вибірку для того щоб гарно розрізняти декілька видів емоцій у висловлюваннях.

Розділ 1.2. Задачі аналізу тональності

Основна задача сентимент аналізу тексту, що містить висловлення думок, може бути сформульована наступним чином: якщо надано текст, що є суб'єктивним висловлюванням, то за припущення, що висловлювання має єдиний об'єкт, виявити емоційний відтінок тексту як одну з двох полярностей: позитивну чи негативну (sentiment polarity).

Визначення, чи має наданий текст в цілому позитивне чи негативне забарвлення називається “класифікацією полярності настроїв” (sentiment polarity classification) або “класифікацією полярності” (polarity classification). Одним з мінусів даного підходу є те, що емоційну складову документа не

завжди можна однозначно визначити, тобто документ може містити як ознаки позитивної оцінки, так і ознаки негативної.[5]

У комп'ютерній лінгвістиці текст природною мовою вважається неструктурованою інформацією. У завданнях, об'єднаних терміном «аналіз емоційного забарвлення тексту», визначається те, яким чином з тексту природною мовою витягується, аналізується і структурується інформація.

Отже, аналіз тональності тексту зазвичай включає в себе наступні основні завдання:

- - визначення наявності емоційного забарвлення;
- - визначення полярності тексту;
- - вилучення аспектів з емоційного забарвленого тексту.

Задача визначення полярності тексту формуються наступним чином:

• «визначити, яке емоційне забарвлення тексту, позитивне чи негативне?» Визначення полярності тексту зазвичай розглядається на декількох рівнях:

- на рівні документу. Основною задачею на цьому рівні є класифікація, чи повністю весь документ є відображенням позитивної чи негативної думки. Наприклад, для певної рецензії на товар, система автоматичного аналізу тональності тексту визначає, чи висловлює ця рецензія позитивний настрій в цілому. Це завдання має назву «класифікація полярності настроїв на рівні документу». Така деталізація має на увазі, що кожен документ виказує думку, яка є сутністю єдиною. Отже його не можна використовувати для документів, об'єктами яких є декілька сутностей. Однак в рамках даної задачі, зазвичай документи мають одне чітке емоційне забарвлення, бо коментарі до новин зазвичай є досить короткими емоційними текстами;

- на рівні речення. На цьому рівні об'єктом дослідження є окреме речення. Оцінюється чи має сутність взагалі якесь забарвлення негативне, чи позитивне. Це близько зв'язано з «класифікацією суб'єктивності, яка розрізняє речення (так звані об'єктивні речення) що висловлюють факти, від тих які

надають емоції людини.

- на рівні сутності та аспекту. Обидва попередніх рівня не включають аналіз того, що саме сподобалося чи не сподобалося власнику думки. Аспектний рівень дозволяє виконати більш детальний аналіз. Замість того, щоб аналізувати лінгвістичні конструкції, аспектний рівень аналізує саму думку. Зазвичай, аналіз думки без аналізу її об'єкту має обмежене використання. Окрім того, визнання важливості аналізу об'єкту думки допомагає глибше зрозуміти проблему аналізу тональності. Наприклад, речення «Не зважаючи на поганий сервіс, мені все одно сподобався цей ресторан» має позитивний сентимент, проте важко казати, що все є позитивним. Якщо ми розглядаємо лише щодо об'єкту «ресторан», то ми можемо все вважати позитивним. В іншому випадку, коли об'єктом є «сервіс», то думка є повністю негативною. В багатьох дослідженнях об'єкти думки описуються сутностями та їх аспектами.

Тобто, метою аналізу такого рівня є виявлення сентименту об'єкту та його властивостей. На такому рівні аналізу, можна отримати структурований підсумок думок щодо не тільки самого об'єкту, але і його властивостей, що перетворить неструктурований текст в структурований масив даних що може бути використаний для будь-яких типів аналізу. Взагалі, аналіз на рівні документу та на рівні речення є досить складними, а аналіз на рівні аспектів є неймовірно важкою задачею.

Найбільш важливими індикаторами сентименту є «слова емоційного забарвлення» (sentiment words). Ми використовуємо їх для виказання думки негативної, чи позитивної. Наприклад, «добре», «чудово», «неймовірно» – ми використовуємо дані слова для позитивного забарвлення, а от «погано», «жахливо», «сумно» – це уже слова для відображення негативного забарвлення. З зрозумілих причин інструментами для аналізу думки є вирази і слова. Якщо ми отримаємо набір з наших інструментів, то отримаємо «лексичний словник».[9]

Незважаючи на те, що слова та вирази для вираження емоційного забарвлення ж дуже важливими при аналізі тональності, просто використання їх не є досить ефективним. Проблема є комплексною та набагато складнішою. Нижче показані основні проблеми, з якими можна стикнутися при виконанні аналізу тональності за словником:

- позитивне чи негативне слово може приймати протилежний відтінок при використанні в іншій предметній області. Так, наприклад, «У цього фільму передбачуваний сюжет» є негативною характеристикою, а «У цього коду передбачувана поведінка» є позитивною;
- наявність неологізмів та помилок у словах, а це досить поширене у інтернет культурі явище, у багатьох випадках робить використання словникових методів недоцільним;
- речення, що містить в собі слово емоційного забарвлення, може мати нейтральний сентимент. Цей феномен виникає зазвичай в декількох типах речень. Питальні речення та умовні речення це найважливіші 2 види, наприклад «Чи не могли б ви дати пораду, яка з фотокамер Sony є найкращою?», та «Якщо буде камера, яка буде мені до вподоби в цьому магазині, я її обов'язково куплю.» В обох цих реченнях є слова, що виражають позитивне забарвлення («до вподоби», «найкраща»), але ці речення не висловлюють забарвлення щодо потрібних товарів.
- речення, що містять сарказм з вмістом, чи навпаки без вмісту слів з певною тональністю є дуже складними для аналізу, наприклад «Яка прекрасна парасоля. Зламалась після другого використання». Сарказм досить часто з'являється в дебатах чи дискусіях, особливо політичних тому аналізувати політику, зазвичай, важко. Це є досить актуальна проблема і вона широко обговорювана та навіть є деякі результати обговорень.

Якщо проаналізувати вище сказані слова, то стає зрозуміло, що аналіз на основі лексичного словника є досить важким. . Тому, методи, що

використовують лексичний словник, як правило, застосовуються рідше, ніж альтернативні методи.

Уміння розпізнавати тональність текстів широко застосовується в різних видах діяльності людини, таких як вивчення переваг споживача; системах рекомендацій покликаних допомогти споживачеві при виборі товарів або послуг; аналізі новинних ресурсів на предмет тональності повідомлень щодо різних персон або подій; підтримки пошукових систем і систем вилучення інформації; психологічних дослідженнях тощо.

Проте, існує безліч проблем, з якими стикаються дослідники при спробі визначення тональності тексту. Ось деякі з них:

1. Тональність може залежати від предметної області тексту. Так, наприклад, слово «величезний» застосовне до опису телевізора, має позитивну тональність, але як характеристика телефону, це слово приймає вже кілька негативний відтінок.

2. Використання заперечення здатне змінити тональність висловлювання на протилежну. Наприклад, в висловлюванні «Раніше мені подобалося ходити в кафе поряд з будинком. Там була приголомшлива кухня. На жаль, тепер там занадто голосно грає музика ». Перше і друге речення позитивно забарвлені, але через використання заперечення в останньому реченні загальна тональність об'єкта «кафе» змінюється на негативну.

3. Сентимент-аналіз погано справляється з розпізнаванням саркастичний висловлювань. Такі вирази можуть мати загальну тональність, зворотний тональності окремих слів: «Відмінна машина для любителів ходити пішки».

4. Значення тональності залежить від того, хто проводить аналіз. Наприклад, фраза «У KFC відмінно йдуть справи» має позитивний забарвлення для компанії KFC і негативний для McDonald's.

Завдання аналізу тональності

Розглянемо модель сутності. У нас є об'єкт c_i , нехай вона має скінченну кількість характеристик $F_i = \{f_{i1}, f_{i2}, \dots, f_{in}\}$. C_i може складатись і по іншому, використовуючи різні вирази $\{cc_{i1}, cci_2, \dots, cc_{in}\}$. Кожен аспект сутності може бути представлений як набір виразів $\{fc_{i1}, fc_{i2}, \dots, fc_{in}\}$. [6]

Тепер у нас є текст t , який має певну тональність і має думку про наші об'єкти $\{c_1, c_2, \dots, c_n\}$ та їх характеристики від деякого автора в деякий час

Оскільки, ми описали основні елементи, які нам потрібні, тому можна перейти до задач

Завдання 1. Вилучення об'єктів та їх розподіл. Вилучити всі вирази об'єктів у деякому документі та виконати розподіл та розбиття їх на класи об'єктів. Кожен клас характеризує окремий об'єкт.

Завдання 2. Вилучення характеристик об'єктів та їх розподіл. Вилучити всі вирази об'єктів в емоційно забарвленому документі та виконати категоризацію та розбиття їх на класи сутностей. Кожен клас характеризує окрему сутність.

Завдання 3. Вилучення авторів думки та їх класифікація. Вилучити авторів думки та класифікувати їх. Задач 1 та 2 схожі на цю.

Завдання 4. Вилучення часу та стандартизація. Необхідно вилучити всі часові періоди, коли думки були написані їх власниками та стандартизувати різні формати представлення часу.

Завдання 5. Розподіл емоційного забарвлення характеристик. Необхідно визначити, чи є думка що до характеристики f_{ij} нейтральною, негативною чи позитивною, або надати числове значення що визначає думку. Кортеж, про який ми говорили раніше містить в собі багато інформації щ допомогою якої можна робити висновки, та базу для створення підсумків. Даний тип підсумків, який використовує характеристик і має назву підсумки на основі характеристик

Наприклад зараз ми представимо як ж може виглядати підсумок на основі інформації, яка була вище.

В даному прикладі проводиться підсумок коментарів щодо Дональда Трампа. Цей підсумок є структурованим, на відміну від звичайного текстового підсумку. В даному прикладі, зверху ми бачимо висновок відносно всього об'єкта загалом. 23 відгуки мали позитивну думку щодо об'єкту і 134 - негативну.

Об'єкт «Трамп» містить в собі дві найпопулярніші характеристики бізнес і президенство. Так, щодо президенства 10 та 93 позитивних негативних відгуків відповідно. Кількість коментарів також може бути посиленням на власне самі коментарі. Маючи підсумок такого типу, можна швидко та зручно провести аналіз того, як користувачі висловлюються щодо всієї сутності в цілому, так її до окремих його компонентів. Якщо людина, яка аналізує зацікавлена в деякому компоненті об'єкту або деяких деталях, він може скористатися покликанням та подивитись самі думки чи повні документи.[6]

Висновок до розділу 1

Метод сентимент аналізу містить в собі багато класифікацій і задач, якщо виокремити основне поняття сентимент аналізу, то це про кортеж з 5 елементів, так звану тональність (думку): < holder, aspect, sentiment, entity, time >, яке містить судження про будь-яку сутність, чи її аспекти, висловлене деякими суб'єктом і забарвлене позитивно, негативно, або нейтрально. Сам метод аналізує даний кортеж та робить висновки, щодо забарвлення, яке містить кортеж.

Також варто звернути увагу на види самої думки – це проста і порівняльна. В кожного з виду є свої особливості із-за яких стає важче аналізувати кортеж.

Щодо самого аналізу, то виокремлюють об'єктивне і суб'єктивне твердження, як правило об'єктивне не беруть аналізувати, адже це факт, який не має забарвлення, а більшість аналізу припадає якраз на суб'єктивну думку.

В розділі ми виділили основні підходи до розпізнавання тональностей, а саме:

1. Підхід, заснований на правилах
2. Метод, заснований на застосуванні тональних словників
3. Навчання з учителем
4. Навчання без вчителя

Також варто виділити класифікацію думок на явні і неявні, адже неявні думки не беруться для аналізу, адже містять інформацію, яка важко запрограмувати на аналіз.

Розділ містив інформацію ще про задачі сентимент аналізу, можна виділити основні:

- - визначення полярності тексту;
- визначення наявності емоційного забарвлення;
- - вилучення аспектів з емоційного забарвленого тексту.

Якщо розглянути задачу визначення полярності тексту, то вона зазвичай розглядається на декількох рівнях:

- на рівні документу.
- на рівні речення.
- на рівні сутності та аспекту.

Окрім основних задач, було виділено і основні проблеми, з якими можна стикнутися при виконанні аналізу тональності за словником

1. позитивне чи негативне слово може приймати протилежний відтінок при використанні в іншій предметній області
2. наявність неологізмів та помилок у словах
3. речення, що містить в собі слово емоційного забарвлення, може мати нейтральний сентимент
4. речення, що містять сарказм з вмістом, чи навпаки без вмісту слів з певною тональністю є дуже складними для аналізу

Проблеми виникають і при самому визначенню тональності, а саме:

1. Тональність може залежати від предметної області тексту.
2. Використання заперечення здатне змінити тональність висловлювання на протилежну.
3. Сентимент-аналіз погано справляється з розпізнаванням саркастичний висловлювань
4. Значення тональності залежить від того, хто проводить аналіз.

Окрім основних задач ми також виділили і під задачі:

1. Вилучення об'єктів та їх розподіл
2. Вилучення характеристик об'єктів та їх розподіл
3. Вилучення авторів думки та їх класифікація
4. Вилучення часу та стандартизація
5. Розподіл емоційного забарвлення характеристик

Якщо підсумувати, то сентимент аналіз має достатню кількість, як проблем так і задач, багато з яких досі важко вирішити, проте основну частину можна зробити, якщо використовувати деякі правила.

РОЗДІЛ 2 АНАЛІЗ АНАЛОГІВ

Розділ 2.1 Переваги та недоліки SentiStrength

SentiStrength — система, розроблена M. Thelwall, K. Buckley, G. Paltoglou та D. Cai. Перш за все, дана система була розроблена для аналізу коротких структурованих неформальних текстів англійською мовою. Однак, вона може бути налаштована для роботи з текстами на ряді інших мов. Результат видається у вигляді двох оцінок – оцінка позитивної та негативної складової тексту (за шкалами від +1 до +5 та від -1 до -5 відповідно). [2]

Крім того, існує можливість надання оцінок у бінарній чи тернарній шкалах. Алгоритм шукає максимальне значення тональності в тексті для кожної шкали (як позитивної, так і негативної). При роботі алгоритм враховує принаймні найпростішу взаємодію слів (наприклад, слова-підсилювачі посилюють значення тональності для слова, на яке вони діють) і ідіоматичні вирази. Недоліки системи: хоча система може бути налаштована для різних мов, але алгоритм не враховує специфіку кожної мови. Крім того, враховується лише загальна тональність тексту, без виділення аспектів. [2]

SentiStrength – це автоматичний аналіз настроїв до 16000 текстів соціальних веб-сайтів за секунду з точністю до рівня людини для англійської - інші мови доступні або легко додаються. SentiStrength оцінює силу позитивних та негативних настроїв у коротких текстах, навіть для неофіційної мови. Він має людський рівень точності для коротких текстів соціальних веб-сайтів англійською мовою, крім політичних. SentiStrength повідомляє про два сильні сторони настрою:

Від -1 (не негативно) до -5 (вкрай негативно)

Від 1 (не позитивно) до 5 (надзвичайно позитивно)

Оригінально даний застосунок був створений для англійської мови та пізніше оптимізований для інших мов, такі як: Finnish, German, Dutch Spanish, Russian, Portuguese, French, Arabic, Polish, Persian, Swedish, Greek, Welsh, Italian, Turkish.

Даний застосунок можна тестувати на веб сайті, та скачати. Для того, щоб завантажити застосунок потрібно вести свої дані: ім'я, email та організацію, в якій працюєте. Тоді ви отримуєте інструкцію з завантаження, потрібно завантажити два файли, один zip, його розархівувати та додати на диск c, інший застосунок, який має запуститись, якщо у вас все правильно виконано. Даний застосунок зроблений для Windows, проте може працювати і на Mac, Linux, Android через Crossover. Застосунок є безплатним для академічних цілей, щоб придбати і використовувати для комерційних цілей потрібно заплатити £1000.

При завантаженні на Windows даний застосунок має вигляд:

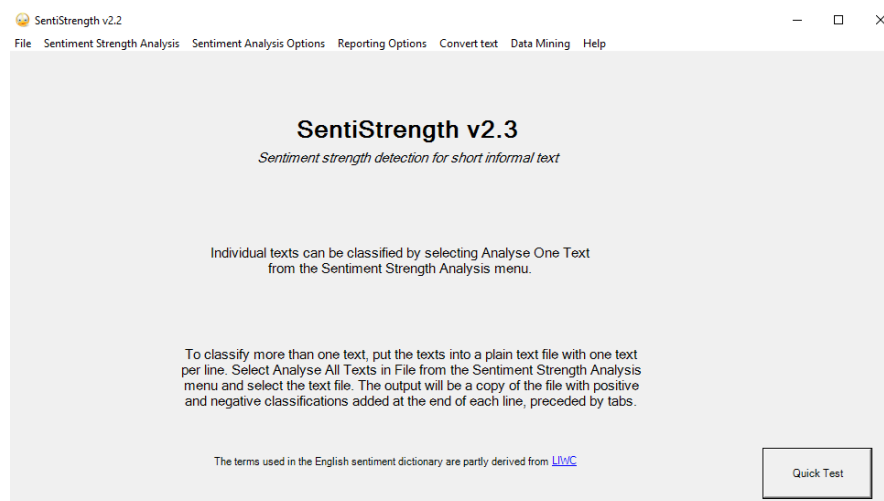


Рисунок 2.1 – Головне вікно програми SentiStrength

Можна проаналізувати короткий текст натиснувши кнопку Quick Test, або великі тексти вибравши в меню “Sentiment Strength Analysis”. Зробимо тести. Введемо в швидкий текст фразу «My name is Solomia. I am chess champion of Ukraine. Chess is so fantastic». Результат:

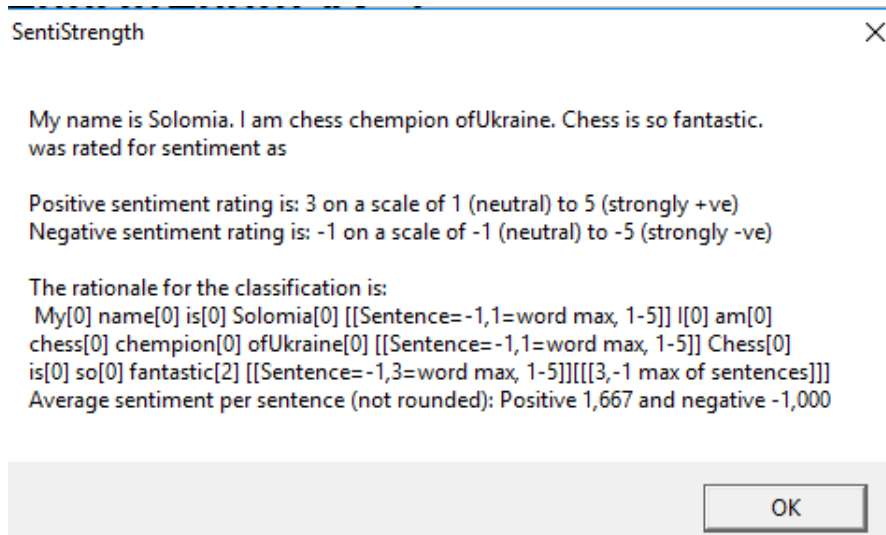


Рисунок 2.2 – Результат аналізу тексту в програмі SentiStrength

Всі слова, окрім “fantastic” отримали оцінку 0 (1 позитивний, -1 негативний, тобто по їх системі оцінки ці слова були ні не позитивні, ні не негативні, по-іншому можна сказати, що вони були нейтральні), а от слово “fantastic” отримало оцінку 2, тобто трохи більше аніж не позитивне.

Давайте поспробуємо другий метод оцінки, не через фразу, а через текст. В текстовому документі ввели наступний текст

«I love you.

I hate you.

Can you help me?

I am scary.

London is the capital of Great Britian»

Кожен рядок вважається окремим текстом і аналізується окремо. Ось результат:

Translation	Positive	Negative	EmotionRationale
I love you.	3	-1	I[0] love[2] you[0] [[Sentence=-1,3=word max, 1-5]] [[[3,-1 max of sentences]]]

I hate you.	1	-4	I[0] hate[-3] you[0] [[Sentence=-4,1=word max, 1-5]] [[[1,-4 max of sentences]]]
Can you help me?	1	-1	Can[0] you[0] help[0] me[0] [[Sentence=-1,1=word max, 1-5]] [[[1,-1 max of sentences]]]
I am scary.	1	-4	I[0] am[0] scary[-3] [[Sentence=-4,1=word max, 1-5]] [[[1,-4 max of sentences]]]
London is the capital of Great Britain	3	-1	London[0] is[0] the[0] capital[0] of[0] Great[2] Britain[0] [[Sentence=-1,3=word max, 1-5]] [[[3,-1 max of sentences]]]

Якщо розглянути детальніше, то можна побачити, що слова “love” і “hate” оцінюють в різну кількість позитивних\негативних балів. Якщо слово “love” має оцінку позитивності 2, то слово “hate” має оцінку негативності -3. Також цікавим є момент, що власну назву «Great Britain» визначило як окремі два слова і дало оцінку 2 слову “Great”, що не є правильним, адже фраза «London is the capital of Great Britain» є фактом, отже немає мати ні позитивних, ні негативних оцінок.

Отже, даний застосунок має багато переваг, такі як веб версія і додаток на комп'ютер\телефон. Багато мов, цікаву методологію оцінки тональності, проте є і мінуси, такі як нерівномірна оцінка відповідних слів позитивно і негативно забарвлених, також не враховуються власні назви та немає української мови.

Розділ 2.2 «Brand Analytics»

Brand Analytics –це система моніторингу та аналізу згадок в соціальних медіа (блогах, мікроблогах, соціальних мережах, форумах та ін.) і онлайн ЗМІ.

Моніторинг відбувається по всім типам повідомлень і даних: текстам повідомлень, розпізнаваними текстами на картинка, розшифровкам відео, чекіни, stories. [3]

Система має такі функції:

- відстежує обговорення в інтернеті компанії, її продукції, послуг, персоналу і конкурентів;
- аналізує автоматично всі згадки, які були знайдені
- генерує звіти про теперішню ситуацію

Brand Analytics призначена для маркетологів, PR–менеджерів, фахівців служби підтримки користувачів, менеджерів з продажу та всім, кому важлива думка користувачів про компанію та її послуги. В системі Brand Analytics використовується модуль визначення тональності Eureka Engine, який може класифікувати російськомовні тексти за трьома видами – позитивні, негативні та нейтральні.[3]

Модуль заснований на алгоритмі випадкових марковських полів з використанням тональних словників. Це дозволило досягти достатньої якості (середня точність складає близько 87%) і високої швидкості обробки текстів. [3]

На відміну від більшості інших реалізацій, поданий модуль дозволяє оцінити силу тональності. Таким чином, користувачу надається можливість не тільки отримати якісну емотивну оцінку документа в цілому щодо об'єкта, що цікавить тональності, а й кількісне співвідношення сили негативного та позитивного відношення до нього.

Модуль може працювати текстами новинного потоку, які написані відносно літературною мовою, так і «некласичною» мовою повідомлень у соціальних медіа.

Зібрані дані автоматично насичуються метаданими:

1. Мова повідомлення
2. Тональність

3. Розмір потенційної аудиторії повідомлення
4. Географія: країна, область, місто, координата
5. Вік або стать автора
6. Зацікавленість: кількість коментарів, лайків і репостів повідомлення
7. Наявність і ступінь агресії.

Також автоматична аналітика класифікує повідомлення по темам, а серед авторів виділяються акаунти ЗМІ і повідомлення спам-ботів. Для оцінки ефективності маркетингу і PR система вираховує відповідність бенчмарки.[3]

Система автоматично будує наочні звіти з графіками і таблицями і надає аналітичні інструментарії:

1. Стрічка повідомлень по об'єкту моніторингу
2. Детальний звіт з графіками і табличними даними: по джерелам, авторам, географії, геолокації, тегам, популярним словам, персонам, компаніям, найбільш цитованим публікаціям.
3. Фільтрація повідомлення за будь яким набором параметрів з підрахунком інтегральних показників, статистикою, можливістю формувати автоматичний звіт
4. Вибірка повідомлень за будь який період часу з початку моніторингу
5. Сортування повідомлень – за залученням аудиторії, розміром аудиторії, кількістю коментарів, лайків, репостів, переглядів, дублювання.
6. Повнотекстовий пошук за зібраним повідомленням
7. Порівняння об'єктів моніторингу за ключовими метриками, в демографічному і географічному розрізах, а також за будь якими зрізами медіа поля.

Pricing				
	"STARTUP"	"STARTUP +"	BASIC	EXPANDED
	25 000 RUB/month <small>pre-paid per 1 month</small>	37 000 RUB/month <small>pre-paid per 1 month</small>	55 000 RUB/month <small>pre-paid per 1 month</small>	105 000 RUB/month <small>pre-paid per 1 month</small>
Collection of new messages per month, including deleted topics	5 000	20 000	100 000	350 000
Number of topics	3	5	10	15
Automated marking of topics (tags per topic)	20	30	40	50
Monitoring of social media and online mass-media	✓	✓	✓	✓
Monitoring by message text,				

Рисунок 2.3 – Ціни на застосунок

Тобто ціни варіюються від 25 000 до 105 000 за місяць, все залежить від розміру колекції повідомлень, кількості тем, кількість тегів, тощо.

Також є можливість отримати демо, проте щоб його отримати потрібно ввести наступні дані:

**Try
Brand Analytics
at work!**

Fill in the form, and our consultant analysts will help you set the system up, taking into account your specific business tasks.

**Free-of-charge
demonstration access**

Specify the brand, company to monitor during your demo access

☐ I agree to the [terms and conditions](#) and the [Confidentiality Policy](#)

Рисунок 2.4 – Вікно для введення вхідних даних

Якщо ви власник компанії, то для вас підберуть демо версію і налаштують спеціально для вас, якщо ви викладач і хочете отримати для наукових цілей, вам також дадуть доступ, але тільки якщо ви пройдете безплатне навчання, проте якщо ви студент, то доступу вам не дадуть.

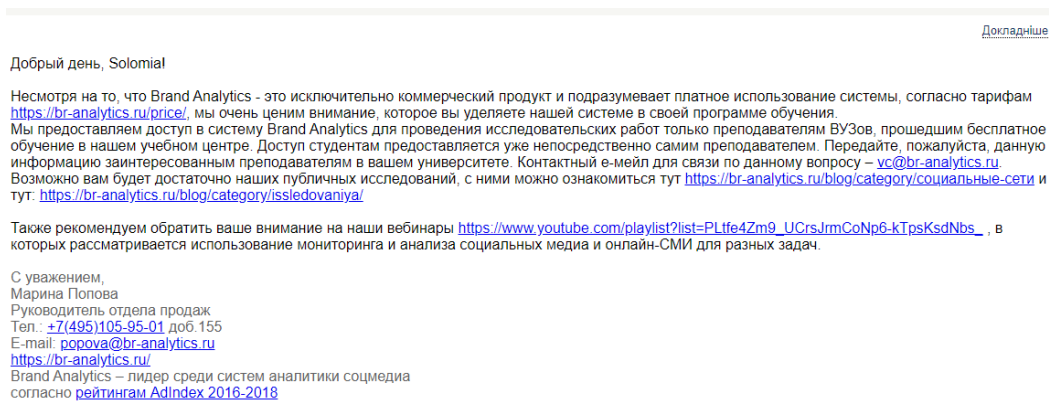


Рисунок 2.4 – Результат комунікації зі студентом

Отже даний застосунок важко протестувати і дізнатись, як він працює, адже доступу до нього у мене немає, проте з функціоналу видно, що застосунок є для ширшого аналізу, окрім тональності є багато інших факторів, яким з опису вони надають більший аналіз. Також цінова політика є досить таки немалою і якщо вам потрібне тільки аналіз тональності, то немає сенсу переплачувати.

Розділ 2.3 «YouScan»

YouScan – система для моніторингу та аналітики соціальних медіа. Система моніторингу YouScan дає змогу без зайвих зусиль стежити за згадуваннями брендів в соціальних мережах і онлайн-ЗМІ, знаходити цінні споживчі інсайти, реагувати на проблеми користувачів та аналізувати сприйняття продуктів і послуг споживачами. [4]

Перейдемо до функцій, які надає YouScan:

1. YouScan моніторить соцмережі, блоги, форуми, сайти відгуків, месенджери, а також онлайн-ЗМІ. Результати моніторингу так автоматичні звіти доступні в онлайн-інтерфейсі в режимі реального часу
2. Звіти за тональністю, географічним розподілом згадувань, демографією авторів, джерелами, а також хмара слів і візуальні інсайти допоможуть максимально зручно проаналізувати те, що відбувається у вашій темі моніторингу
3. YouScan автоматично проаналізує всі сплески обговорень, допоможе

моментально знайти найважливіші згадування й не пропустити нові тренди в дискусіях

4. Завдяки функції «Правила» можна максимально автоматизувати роботу по розмітці згадувань і відправку автоматичних повідомлень. За допомогою правил можна автоматично тегувати згадування, відправляти повідомлення, корегувати тональність, видаляти або позначати як «оброблені»[3,21]

Ще деякі плюси YouScan:

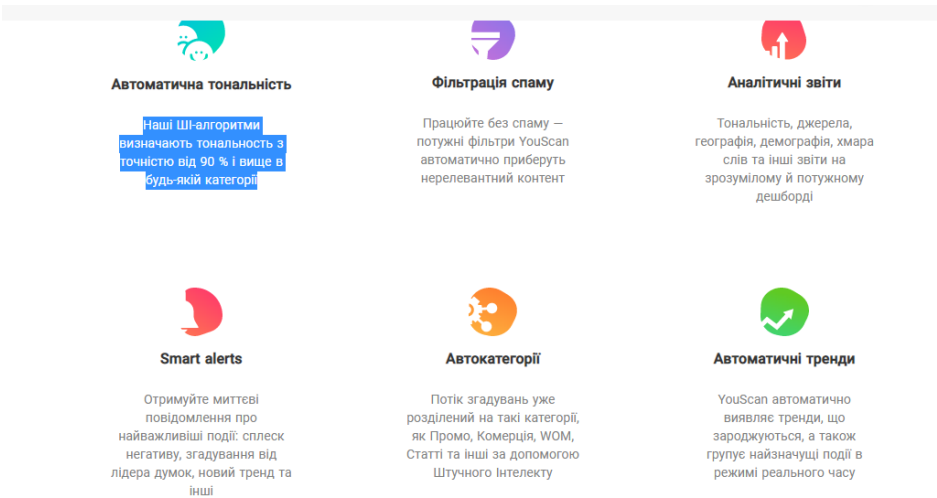


Рисунок 2.5 – Можливості програми YouScan

Що важливе для нас, це те що ШІ-алгоритми YouScan визначають тональність з точністю від 90 % і вище в будь-якій категорії [3]

Цінова політика:

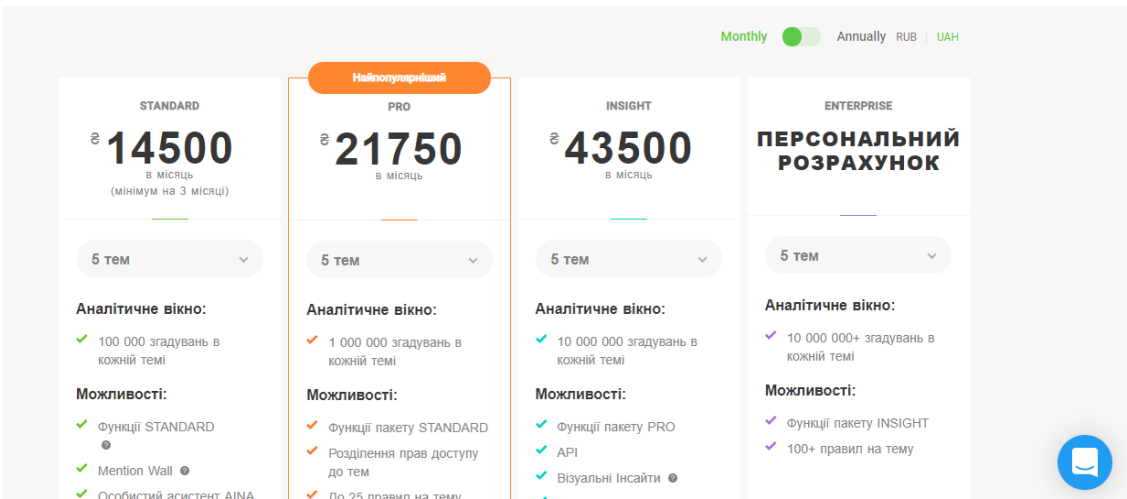


Рисунок 2.6 – Цінова політика YouScan

Ціни в діапазоні від 14500 до 43500 грн за місяць, або через персональний розрахунок.

Ціни залежать від:

1. Кількість тем(5-25). Наприклад , якщо ми візьмемо в пакеті стандарт 25 тем ціна збільшиться до 72500
2. Від кількості згадувань, які будуть обробляться
3. Від підтримки різних функцій

Щоб отримати демо потрібно ввести наступну інформацію:

YOUSCAN Продукт Дослідження Вартість Клієнти Про нас

YouScan в дії

Зареєструйтеся на безкоштовну демонстрацію системи. Заповніть цю форму, щоб побачити, як дані з соціальних медіа допоможуть вашій компанії змінитися на краще, слухаючи своїх споживачів.

Чого мені чекати?

- Ми допоможемо налаштувати для моніторингу цікаві для вас теми.
- Розповімо, як працювати з системою.
- Підкажемо, як отримати максимум від роботи з даними з соціальних медіа для вирішення ваших завдань.

Нам довіряють

Avito Ipsos PEPSICO DDB® BURGER KING Мегалон BBDO

ІМ'Я

ПРІЗВИЩЕ

НАЗВА КОМПАНІЇ

РОБОЧИЙ E-MAIL

НОМЕР ТЕЛЕФОНУ

UA +380

☐ Під час реєстрації ви приймаєте [Ліцензійну угоду](#) та даєте дозвіл на обробку персональних даних згідно з [Політикою конфіденційності YouScan](#).

НАДІСЛАТИ

Даний застосунок пропонує версію для компаній, щоб можна було налаштувати під їх потреби. При отриманні демонстрації ви можете очікувати:

1. Допоможуть налаштувати для моніторингу цікаві для вас теми
2. Розкажуть, як працювати з системою
3. Підкажуть, як отримати максимум від роботи з даними з соціальних медіа для вирішення ваших завдань. [4]

Отже, як і в попередньому варіанті, демо версію ми протестувати не змогли, адже не отримали її. Проте з опису функціоналу даний ресурс звертає більшу увагу на тональність та обіцяє 90% правильність своїх результатів.

Висновок до розділу 2

Даний розділ був для аналізу аналогів у сфері сентимент аналізу, для аналізу було взято три аналоги: «SentiStrength», «YouScan», «Brand Analytics».

Було оцінено ціну, можливості, які надає кожен із аналогів тощо. Лише одного із аналогів, а саме YouScan можна було протестувати у роботі, решта не мали вільного доступу, лише з додатковими умовами.

Щодо ціни, то цінова політика досить різна і залежить від того, які додаткові функції ви хочете. В середньому можна взяти певні можливості приблизно за однакових цін.

Щодо можливостей, то SentiStrength створений лише для аналізу тональності і запропоновані досить великі відсотки вірності, щодо інших аналогів, то аналіз тональності є одним із атрибутів для аналізу, який проводить дані продукти.

Якщо підсумувати, то аналоги мають свої плюси і мінуси, немає дуже хорошого застосунку, всюди є свої нюанси, є куди розвиватись.

РОЗДІЛ 3 ОПИС РОЗРОБКИ ПРОГРАМНОГО ПРОДУКТУ

Розділ 3.1 Вимоги до програмного продукту

Функціональні вимоги

Розробка застосунку на мові Java, використовуючи метод tf-idf для побудови даних стоп слів. Повинні бути дані для аналізу методом tf-idf та дані саме для аналізу тональності.

Нефункціональні вимоги

Застосунок має правильно оцінювати тональність відповідно до наявних даних.

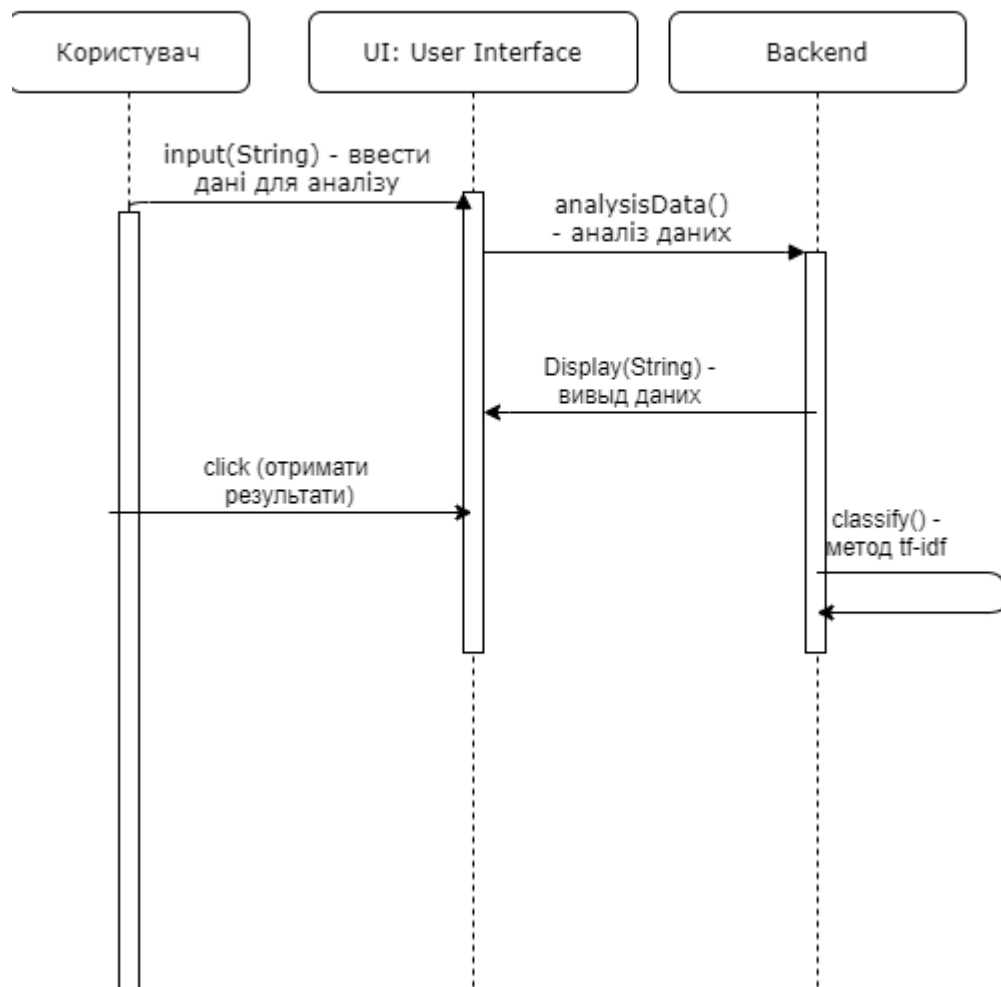
Інтерфейс має містити поле для вводу та поле для виводу даних.

Розділ 3.2 Архітектура та структура програмного продукту

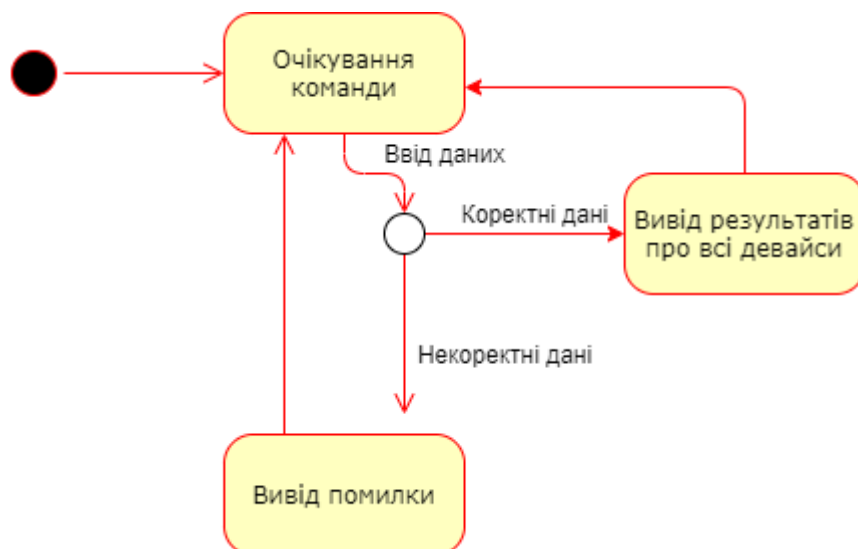
UML Activity



UML Sequence



UML State Chart



Розділ 3.3 Засоби та технології розробки

Для розробки було використано мову програмування JAVA та WindowBuilder та IDE eclipse.

Висновок до розділу

Даний застосунок показує роботу методу сентимент аналізу, містить потрібну інформацію для аналізу та отримання результатів. Проте, занадто мало реалізовано, щоб можна було якісно і в повній мірі оцінити тональність застосунку. Оцінка проводилась лише для англійської мови та для прямої, суб'єктивної думки.

ЗАГАЛЬНИЙ ВИСНОВОК

У курсовій роботі був проведений огляд методу сентимент аналізу для аналізу тональності (думки) та було проведений огляд аналогів розробки у даній сфері, а саме трьох різних застосунків: SentiStrength, YouScan, Brand Analytics.

Для аналізу тональності нам потрібно вибрати задачу, яку буде виконувати даний аналіз, адже задач і під задач у аналізу тональності є багато. Наразі є декілька застосунків які допомагають у цьому, всі вони створенні в більшості для маркетингу. При виборі застосунку потрібно звернути увагу на наступні критерії, як того що ви будете аналізувати, так і сферу використання. По-перше треба визначитись який тип у нас думки, ми розглянули, що може бути простою, чи порівняльною; об'єктивною, чи суб'єктивною; явною, чи неявною. По-друге треба визначитись із задачею, яку буде виконувати метод сентимент аналізу, наведемо декілька варіантів:

- визначення наявності емоційного забарвлення;
- визначення полярності тексту;
- вилучення аспектів з емоційного забарвленого тексту.

Це основні задачі, проте є ще і під задачі, як наприклад видалення якихось частин думки, тощо.

На основі цієї інформації можна вибрати чим саме користуватись, адже є досить серйозні проблеми з оцінкою об'єктивних, чи неявних думок. Не всі застосунки забезпечують основні задачі, що уже говорити про під задачі.

В роботі ми розглянули тестування можливостей деяких застосунків, які мали вільний доступ. Такий застосунок, на жаль, був лише один і його результати були не 100% і він виконував лише основну функцію визначення полярності тексту.

В розробленому нами застосунку, ми продемонстрували роботу методу сентимент аналізу для основної задачі це визначення полярності тексту.

Отже, у курсовій роботі було проведено аналіз застосунків, визначено основні критерії вибору застосунку для аналізу, визначені основні задачі та під задачі методу сентимент аналізу, визначено основні проблеми з якими стикаються при розробці застосунку на основі методу сентимент аналізу, проаналізовані плюси і мінуси уже існуючих застосунків. Результати роботи допоможуть на далі в розробці якісного застосунку на основі методу сентимент аналізу, який буде вирішувати основні проблеми та виконувати різного роду задачі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
2. SentiStrength Platform/ [Електронний ресурс]: [Веб-сайт]. – Електронні дані. — Режим доступу: <http://sentistrength.wlv.ac.uk/>
3. Система мониторинга и анализа социальных медиа и СМИ /[Електронний ресурс]: [Веб-сайт]. – Електронні дані.— Режим доступу: <https://br-analytics.ru/>
4. YouScan — платформа для аналітики соцмедіа #1 в Східній Європі. [Електронний ресурс]: [Веб-сайт]. – Електронні дані.— Режим доступу: <https://youscan.io/ua>
5. Котельников Е.В. Распознавание эмоциональной составляющей в текстах: проблемы и подходы / Е.В. Котельников, М.В. Клековкина, Т.А. Пескишева, О.А. Пестов; под ред. С.М. Окулова. - Киров: Изд-во ВятГГУ, 2012. 103 с.
6. Лукашевич Н.В. Автоматический анализ тональности текстов по отношению к заданному объекту и его характеристикам // Электронные библиотеки, 2015. Т. 18. № 3-4. С. 88-119.
7. Opinion Mining on the Web by Extracting Subject-Aspect-Evaluation Relations [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Nozomi Kobayashi, Ryu Iida, Kentaro Inui, Yuji Matsumoto – 2006 – Режим доступу: <https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-018.pdf>
8. The Importance of Neutral Examples for Learning Sentiment [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Koppel, Moshe; Schler, Jonathan – 2005 – Режим доступу: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.9735&rep=rep1&type=pdf>

9. Bing Liu. Sentiment Analysis and Subjectivity [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Bing Liu – 2010 – Режим доступу: <https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>
10. Sentiment strength detection in short informal text / [Mike Thelwall, Kevan Buckley, Georgios Paltoglou et al]. // Journal of the American Society for Information Science and Technology.-2010.- No61.- P. 2544–2558 Making objective decisions from subjective data: detecting irony in customer reviews. [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Reyes Antonio, Rosso Paolo – 2012 – Режим доступу: https://www.researchgate.net/publication/257015931_Making_objective_decisions_from_subjective_data_Detecting_irony_in_customer_reviews
11. Feature Selection and Weighting in Sentiment Analysis [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – O’Keefe, Tim; Koprinska, Irena – 2006 – Режим доступу: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.709.1463&rep=rep1&type=pdf> 76
12. Thumbs up or thumbs down. Semantic orientation applied to unsupervised classification of reviews [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – P. Turney. – 2002 – <http://www.aclweb.org/anthology/P02-1053.pdf>
13. Sentiment analysis algorithms and applications: A survey [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Medhata, Walaa; Hassan, Ahmed – 2014 – Режим доступу: <https://www.sciencedirect.com/science/article/pii/S2090447914000550>
14. Sentiment Analysis on Movie Reviews [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Режим доступу: <https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/data>
15. A Comparison of the Accuracy of Support Vector Machine and Naive Bayes Algorithms In Spam Classification [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Rita McCue, Jürgen Schmidhuber – 29.11.2009 – Режим доступу:

<https://pdfs.semanticscholar.org/55c5/9874114617b57eedd8636c5d0e7785fb885f.pdf>

16. Opinion lexicon [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Режим доступу: <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>
17. Self-training from labeled features for sentiment analysis. Inf Process Manage [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – He Yulan, Zhou Deyu – 2014 – Режим доступу: <https://www.sciencedirect.com/science/article/abs/pii/S0306457310000932>
18. TIOBE Index for March 2018 [Електронний ресурс]: [Веб-сайт]. – 2018 – Електронні дані. Режим доступу: <https://www.tiobe.com/tiobe-index/>
19. Text Classification using Naive Bayes [Електронний ресурс] – 2015. – Режим доступу: <http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learnnote07-2up.pdf>
20. The importance of neutral examples for learning sentiment [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Rita McCue, Jonathan Schler – 21.10.2005 – <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.9735.77>
21. Too much information [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Режим доступу: <https://www.economist.com/node/18895468>
22. How Americans get their news [Електронний ресурс]: [Веб-сайт]. – 2014 – Електронні дані. – Режим доступу: <https://www.americanpressinstitute.org/publications/reports/survey-research/howamericans-get-news/>