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

АНАЛІЗ ЕМОЦІЙНОГО ОКРАСУ ТЕКСТУ В СОЦІАЛЬНИХ МЕРЕЖАХ

Текстова частина до курсової роботи
за спеціальністю „Інженерія програмного забезпечення”

Керівник курсової роботи
к.т.н., доц. Ковалюк Т. В.
_____
(підпис)
“____” _____ 2020 р.

Виконав студент
Баранов К. О.
“____” _____ 2020 р.

Київ 2020

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

ЗАТВЕРДЖУЮ
Зав. кафедри інформатики,
доц., к.ф.-м.н.
_____С. С. Гороховський
(підпис)
„____”_____2020 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ
на курсову роботу

студенту_____ факультету _____ _____курсу
ТЕМА _____
Вихідні дані:

- Текстова частина
- Додаток

Зміст ТЧ до курсової роботи:

Індивідуальне завдання
Вступ
1 Огляд теоретичних відомостей
2 Складові емоційного аналізатора
3 Метрики оцінювання роботи аналізатора
4 Актуальні задачі емоційного аналізу
5 Розробка додатку
Висновки
Список літератури
Додатки

Дата видачі „___” _____ 2020 р. Керівник _____
(підпис)
Завдання отримав _____
(підпис)

Календарний план

Тема: _____

Календарний план виконання роботи:

№ п/п	Назва етапу дипломного проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на дипломну роботу.	12.11.2019	
2.	Огляд технічної літератури за темою роботи.	15.12.2019	
3.	Виконати аналіз сучасних методів емоційного аналізу	20.01.2020	
4.	Програмування додатку	20.03.2020	
5.	Інтегрування додатку з соціальною мережею	05.04.2020	
6.	Створення слайдів для доповіді та написання доповіді.	25.04.2020	
7.	Аналіз отриманих результатів з керівником, написання доповіді та попередній захист магістерської роботи.	02.05.2020	
8.	Корегування роботи за результатами попереднього захисту.	03.05.2020	
9.	Остаточне оформлення пояснювальної роботи та слайдів.	10.05.2020	

Студент _____

Керівник _____

“_____”_____

Зміст

Анотація

Автоматизованому емоційний аналізу починають приділяти все більше уваги, так як обсяг інформації, що може бути оброблена емоційним аналізатором та використана в маркетингових та дослідницьких цілях, безперервно зростає. Джерелами таких даних перш за все являються соціальні мережі, де користувач має змогу опублікувати власну точку зору щодо будь-якого продукту, сервісу, бренду тощо. Емоційний аналіз має на меті зібрати такі дані та вилучити з них необхідну цінну інформацію.

Дана робота містить огляд напрацьованих відомостей про алгоритми емоційного аналізу, методи оцінювання роботи таких алгоритмів та сучасні проблеми емоційного аналізу, що потребують вирішення.

У ході роботи було розроблено додаток, в якому реалізовано роботу з емоційним аналізатором.

Вступ

Поява та швидке розповсюдження соціальних мереж у сучасному світі надало нові інформаційні можливості використання соціальних даних. Кожен користувач такої мережі може вільно поділитися своєю думкою про будь-який товар, послугу, подію, висказати своє ставлення до діяльності тієї чи іншої особи тощо. Переважно, такі дані мають текстовий неструктурований формат. Наприклад, це може бути короткий текстовий запис на сторінці соціальної мережі користувача, чи як відгук на сторінці товару в онлайн-магазині, або ж просто коментар до чужої публікації, чи виважена рецензія до фільму. Соціальні дані містять велику кількість корисної інформації, та надають нові перспективні напрями для удосконалення та автоматизації систем ефективної підтримки прийняття рішень у таких сферах, як маркетинг чи політика. Однак робота з такими даними несе за собою низку проблем, які насамперед тісно пов’язані з особливостями природньої мови.

На сьогодні користувачі соціальних мереж публікують сотні тисяч дописів кожної секунди [1]. Деякі такі статистичні дані наведені в таблиці 1.

Таблиця 1 – Статистичні дані соціальних мереж

Соціальна мережа	Кількість публікацій кожної секунди
Facebook	54 977
Twitter	8 963
Instagram	7 787
Reddit	38

Такий об’єм інформації, що генерується кожної секунди, неможливо і економічно недоцільно обробляти використовуючи лише людський ресурс у

вигляді головного інструменту. Висока потреба у генерації узагальнених звітів, які б відображали ставлення аудиторії до того чи іншого продукту або послуги, потребує нових теоретичних та практичних набутків у галузі лінгвістичної обробки неструктурованого тексту [2]. Серед найбільш перспективних напрямів цієї галузі наразі приділено велику кількість уваги саме розвитку алгоритмів емоційного аналізу тексту, що й є головною темою цієї курсової роботи.

1 ОГЛЯД ТЕОРЕТИЧНИХ ВІДОМОСТЕЙ

Емоційний аналіз поєднує в собі процеси обробки природніх мов та пошуку даних. Також він містить в собі проблему глибокого машинного навчання. Машинне навчання дозволяє комп'ютеру знаходити певні закономірності в наборі інформації, що подають йому на вхід, та інтерпретувати її певним шляхом. Для задачі емоційного аналізу процес машинного навчання можна охарактеризувати, як обробку вхідного тексту та надання йому емоційної оцінки.

Основною задачею емоційного аналізу є перетворення вхідної інформації у вигляді тексту, написаного природньою мовою, у бінарну форму – позитив чи негатив. З розвитком технологій до результату такої обробки тексту також додали третю відповідь, що означала емоційну нейтральність. Оцінювання емоційного сигналу може відбуватися на як на рівні цілого документу, так і на рівні його окремих речень, та навіть окремої характеристики об'єкта про яку йде річ. На сьогодні найбільш перспективним рівнем вважається рівень характеристики об'єкта.

Також з часом мінялося й саме визначення емоційного аналізу [3, 167].

Перше визначення говорило, що результатом процесу емоційного аналізу є двовимірний вектор {e, s}, у якому 'e' – це об'єкт про який власне йде мова у документі, а 's' – емоційна оцінка тексту цього документу.

Таке визначення хоч і надає бажану емоційну оцінку об'єкту, проте не відповідає на запитання, хто був автором тексту, що є важливою складовою для дослідження цільової аудиторії об'єкта обговорення. Також потрібно зауважити на тому, що з часом думка автора може змінитись, щоб врахувати це, нам також бажано зберігати часову мітку цього висловлювання. Таким чином з'явилося нове визначення емоційного аналізу.

Друге визначення описувало результат процесу емоційного аналізу як вектор {g, s, h, t}, де ‘g’ – це згадана у тексті характеристика об’єкта, ‘s’ – емоційне забарвлення тексту, ‘h’ – автор думки, ‘t’ – час публікації.

Оскільки об’єкт може мати декілька характеристик, визначення знову було змінено.

Емоційний аналіз – це процес лінгвістичної обробки тексту написаного природньою мовою, результатом якого є вектор вигляду {e, a, s, h, t}, де ‘e’ – об’єкт, ‘a’ – характеристика об’єкту, ‘e’ – емоційне забарвлення тексту, ‘h’ – автор, ‘t’ – час публікації.

2 СКЛАДОВІ ЕМОЦІЙНОГО АНАЛІЗАТОРА

Побудову емоційного аналізатора можна умовно поділити на три етапи. Перший етап полягає у побудові бази знань, на другому етапі визначається рівень абстрагування тексту, і на третьому етапу обирається метод, який буде використовуватись для пошуку корисної інформації у тексті та оновлення бази знань.

2.1 База знань

База знань об’єктів та емоційних оцінок реального світу використовується для підготовки аналізатора до вилучення емоційної характеристики з реального тексту [4, 299–305]. Джерелами для побудови такої бази знань можуть бути лінгвістичні експерти, лексикони, підручники, журнали та інші експертні дані, що можуть бути використані для навчання системи. Проте експерти з лінгвістики є ключовим джерелом, так як ними надається найбільш якісна інформаційна оцінка. Отримані дані використовуються для побудови моделі аналізатора шляхом машинного навчання.

Лексикони використовуються переважно для наповнення бази знань, оскільки вони містять інформацію про синоніми та семантичну інтерпретацію слова в різних мовах.

Такі джерела, як правило, є трудомісткі та затратні, так як вимагають здійснення їх обробки за допомогою експертів. Тому часто використовуються автоматизовані методи наповнення бази даних. Такі дані поступаються в точності та якості, проте дозволяють обробити більший об’єм за коротший час.

2.2 Рівень абстракції

Умовно аналіз емоційної оцінки тесту можна розглядати на трьох рівнях абстракції: на рівні документа, на рівні речення та на рівні характеристики об’єкта. Рівень документу може передати загальний настрій тексту, тоді як на рівні характеристики емоційний аналіз може працювати з більш детальними і чутливими емоційними віяннями автора.

2.2.1 Рівень документу

Рівень документу розглядається як найбільш верхній рівень абстрагування емоційного аналізу. Перш за все, процес бінарної класифікації тексту починають вже на цьому рівня. Емоційна оцінка на рівні документу передає загальний настрій та тон автора. Можна відразу побачити як автор ставиться до об’єкту висловлювання – негативно чи позитивно. На початкових етапах розвитку емоційного аналізу даний рівень абстракції користувався значним попитом через свою простоту.

Дослідження були головним чином сфокусовані на припущенні Лю Бінга [5, 1037–1043], що протягом усього документу автор оцінював один об’єкт. Проте таке припущення виявилося невірним, оскільки текст міг містити як і декілька авторів, так і певну не одиничну кількість обговорюваних об’єктів.

Спочатку проблему емоційного аналізу було запропоновано вирішувати на основі прикметників [6]. Додатково вони встановили винятки до таких сполучень, як «і», «але», «ні» і їм подібним. База знань для такої моделі містила понад двадцять один мільйон слів англійської мови, і кожне слово в ній було промарковане стосовно емоційної оцінки. Як було сказано раніше, алгоритм потребував, щоб вся навчальна інформація була промаркована, а отже був не

достатньо ефективний, оскільки, коли алгоритм ігнорував не промарковані прикметники, що в свою чергу зменшувало точність результатів.

Згодом було запропоновано проводити навчання алгоритму без промаркованих наперед слів [7, 417–424]. Цей спосіб визначав емоційну складову прикметників та прислівників в три етапи.

Перший етап полягав в виявленні словосполучень, які містять прикметники чи прислівники разом з контекстом вживання, для визначення їх емоційного навантаження.

Другий етап оцінює емоційну забарвленість знайдених фраз за допомогою алгоритму поточкової взаємної інформації. Забарвленість емоції в фразі обчислюється на основі її співставлення з опорним позитивним та негативним словами.

У третьому етапі підраховується кількість позитивних та негативних фраз. Таким чином знаходять загальний емоційний опис документу.

Проте даний алгоритм не набув поширення, так як при його застосуванні до документів написаних на різних мовах, кожна з яких має власні лінгвістичні правила, він не міг видати прийнятний результат.

У 2002 році Бо Панг запропонував використовувати методи машинного навчання для вирішення проблеми емоційного аналізу [8, 79–86]. Алгоритми наївний баєсів класифікатор, максимуму ентропії та метод опорних векторів вже добре зарекомендували себе в вирішенні проблеми визначення теми документу. У ході дослідження було виявлено, що метод опорних векторів має кращі результати ніж міг надати наївний баєсів класифікатор.

Також було запропоновано відносити документ перед обробкою до однієї з п’яти категорій за його якістю:

а) висока якість;

б) середня якість;

в) низька якість;

г) дублікат;

д) спам.

Дана характеристика отримувалася шляхом порівняння обраного документа з певним описом продукту, ставлення до якого досліджується в даний момент.

2.2.2 Рівень речення

Процес емоційної оцінки на рівні документа та речення можна вважати майже однаковими, за винятком того, що на рівні речення головним завданням є виявлення суб'єктивності. Таким чином деталізація результату збільшується на рівні речення, ніж документу. Документ містить суб’єктивні та об’єктивні висловлювання. Емоційне забарвлення залежить від досвіду автора щодо обговорюваного об’єкту, тому важливим завданням на цьому рівні є виключення всіх фактичних об’єктивних висловлювань, та зосередження лише на суб’єктивних [9, 174–181].

Виявлення об’єктивності чи суб’єктивності в реченні відбувається за слідкування за певними прикметниками в реченні. Метод показав гарний результат серед текстів одного домену, проте так як одне слово може мати позитивний характер в одному домені, а в іншому – негативний, то для якісного емоційного аналізу необхідно збирати велику базу знань прикметників для різних доменів, що збільшує вартість розв’язання проблеми емоційного аналізу.

Бо Панг та Ліліан Лі також зробили висновок, що стискання вхідного тексту в загальному не впливає на результат емоційної оцінки [10, 271–278]. Під засобами стиснення мається на увазі відкидання тих частин документу, яка не впливає на загальний рівень емоційної складової документу. Для виявлення корисних, тобто лише суб’єктивних висловлювань у тексті, вони використали

комбінований алгоритм, який складався з поєднання наївного баєсового аналізатору та методу опорних векторів. Проте неоднозначність емоційного прояву слова в різних предметних областях чи мовах породжує низьку якість емоційної оцінки.

Згодом було запропоновано оцінювати емоційну характеристику контексту в якому знаходяться прикметники. Дана процедура виконується в два кроки.

Спочатку проводиться аналіз ключових слів в контексті для відповіді на запитання, чи є вони емоційно нейтральні або ж, навпаки, мають емоційних відтінок. Тут мають враховуватися особливості мови документу, семантика слів у реченні та особливості документа.

Потім знайдені таким чином ключові слова вже зіставляють з емоційною характеристикою. Цього разу використовується десять різних характеристик, таких як особливості слова, особливості емоції, яку вони передають.

Поєднання всіх цих характеристик дає кращий результат.

2.2.3 Рівень характеристики об’єкту

Це найдетальніший рівень аналізу. Тут характеристики можна поділити на явні та неявні, а також на залежні на незалежні від предметної області тексту. Багато досліджень проведено щодо цього рівня аналізу. В загальному випадку було класифіковано чотири способи знаходження сутностей в тексті: на основі лінгвістичних правил, прихована модель маркова, надійний класифікатор мінімізації ризиків та класифікатор максимальної ентропії. У більшості випадках надійний класифікатор мінімізації ризиків працює краще порівняно з іншими. Проте всі чотири способи не можуть надати якісний аналіз в умовах багатомовних текстів.

2.2 Навчання системи

Результат обробки запиту до емоційного аналізатора залежить від обраних моделі та методу навчання системи. На разі не існує жорстких правил для виборі тих чи інших технік.

2.2.1 Вибір моделі

Щоб знайти та використати корисні дані, які мають величезні розміри, виникає потреба в машинному навчанні. Існує декілька моделей навчання, які можна використати у сфері емоційного аналізу. Вибір моделі це складний і необхідний етап. Не існує єдиної правильної моделі. Всі вони мають властивість породжувати хибні дані. На разі їх поділяють на такі типи [11, 1542–1546]:

а) моделі прогнозування, які намагаються інтерпретувати майбутнє значення питання в темі документа. Ефективність даної моделі залежить від точності прогнозування майбутніх значень;

б) Описові моделі, що фактично використовуються для узагальнення результату аналізу. Класифікація та кластеризація належать до такого типу;

в) користувацька модель використовується тоді, коли одне й те саме висловлювання може емоційно по різному трактуватися для різних груп користувачів;

г) авторська модель описує ситуацію, коли думка одного автора може впливати на думки інших користувачів. Це питання авторитету.

2.2.2 Техніки машинного навчання

У даній галузі поширено декілька таких технік [12, 168–171].

Навчання з вчителем, що використовує марковані дані. Дані попередньо оброблюються експертною групою, що робить даний підхід дорогим.

Навчання з вчителем, що використовує марковані та не марковані дані. У 2004 було запропоновано алгоритм часткового маркування для зменшення загального обсягу даних, що мають бути оброблено безпосередньо експертами. Найбільш поширеним методом для анотованих емоційністю слів чи виразів являється метод навчання з частковим маркуванням.

Навчання без вчителя використовує не марковані масиви даних. Правила описані спеціально для обробки даних без ніяких анотацій. Часто використовується з використанням класифікатора на основі активних ознак та їх комбінацій. Також існують класифікатори на основі трансформацій, побудований на основі лексичних правил.

Результатом такого аналізу є власне думка автора про певну обговорюваний об’єкт. Іноді такий аналізатор використовується для розширення вже існуючої бази знань.

3 МЕТРИКИ ОЦІНЮВАННЯ РОБОТИ АНАЛІЗАТОРА

Існує декілька показників для здійснення оцінювання алгоритму емоційного аналізу: точність, повнота, F-міра, експертна оцінка та правильність. Для цього потрібно ввести декілька метрик, які ми будемо використовувати для оцінки алгоритму:

а) істинно позитивна відповідь (TP) – кількість позитивних відповідей до прикладів, які позначені як правильні;

б) хибно позитивна відповідь (FP) – кількість позитивних відповідей до прикладів, які позначені як неправильні;

в) істинно негативна відповідь (TN) – кількість негативних відповідей до прикладів, які позначені як неправильні;

г) хибно негативна відповідь (FN) – кількість негативних відповідей до прикладів, які позначені як правильні;

д) коректний вивід (CO) – кількість відповідей системи, які позначені як правильні експертом.

Отже, повнота – це кількість правильних документів знайдених системою до загальної кількості шуканих документів у колекції:

$$R = TP \,/\, (TP + FN). \tag{4.1}$$

Точність – це кількість правильно знайдених документів відносно до загальної кількості документів, що вернула система на запит

$$P = TP \,/\, (TP + FP) \tag{4.2}$$

Правильність – це відношення кількості правильних відповідей системи до загальної кількості відповідей системи, визначається як

$$A = (TP + TN) / (TP + TN + FP + FN) \quad (4.3)$$

Експертна згода – це відсоток походження між двома чи більше експертами щодо коректної відповіді системи, позначається наступним чином:

$$HA = CO / (TP + TN + FP + FN) \quad (4.4)$$

4 АКТУАЛЬНІ ЗАДАЧІ ЕМОЦІЙНОГО АНАЛІЗУ

Щоб отримати ідеальний результат, алгоритм емоційного аналізу повинен мати інтелект людини. Наближення до високої якості емоційного аналізу можливе, якщо усунути прогалини у кількох суміжних дослідженнях. Зараз ведеться дослідження у різних напрямках та аспектах, що допоможуть усунути ці прогалини та розробити більш якісний алгоритм емоційного аналізу.

4.1 Фільтрація спаму

Зі збільшенням корисної інформації, у світі збільшується і кількість спаму. Якість емоційного аналізу основана правдивості і достовірності думок реальних людей, які реально мають що сказати про певний продукт чи сервіс. Конкуренти можуть генерувати неправдиві відгуки, щоб здобути перевагу у їхній сфері діяльності. Виявлення таких текстів є основною проблемою сьогодення, як фільтрація правдивих думок про продукт [13, 219–230].

Проблеми, які повстають у питанні виявлення спаму:

а) виявлення групи чи окремих спамерів. Результат роботи будь-якої системи підтримки рішень залежить від правильності вхідних даних, тому виявлення фальшивих відгуків є пріоритетною задачею;

б) незалежність від предметної області. Пошук алгоритму виявлення спаму, який би якісно працював незалежно від предметної області об’єкта;

в) ознаки спаму. Зараз вони поділяються на три групи: контент тексту, метадата публікації, така як час, місце, автор, та фактичні дані.

4.2 Врахування часового аспекту

Потрібно враховувати не лише емоційну складову відгуку, але й час коли цей відгук був опублікований, що дозволяє дослідити відношення аудиторії до певного явищу динамічно, у часі. Є необхідність автоматизованого розподілу ваги для більш нових відгуків до відносно більш давніх. Далі наводиться список вразливих аспектів емоційного аналізу до тимчасової природи документів:

а) явні та неявні ключові слова. Більшість досліджень задають ключові слова явно в запиті до системи, або ж динамічно охоплюють його шляхом аналізу документу, що є надійнішим способом, але досить складним завданням;

б) географічний розподіл часу. При обробці емоційного аналізу в часі потрібно враховувати часовий розподіл між авторами текстів;

в) аналіз прогнозу. Час може бути включений для процесу прогнозування емоційної оцінки аудиторії до об’єкта;

г) актуальність відгуку. З плином часом значення емоційної характеристики попередніх оглядів вже не так важливо, варто звертати більшу увагу на актуальні документи.

4.3 Багатомовність

Інтернет використовується у всьому світі. З’являється проблема багатомовності. Далі наведено різні аспекти обробки багатомовних документів:

а) використання перекладу та транслітерації;

б) автоматичне розширення лексикону різними мовами. Для документів написаних різними мовами виникла потреба у підтримці паралельних

лексиконів. Отримати весь лексикон певної мови майже неможливо, адже завжди існують нові слова у такому лексиконі;

в) слова з декількома значенням. Значення слова не є однозначним і залежить від контексту;

г) багатомовність в одному висловлюванні. Слова різних мов можуть бути вживані для побудови одного речення.

4.4 Неструктурованість текстів

На даний момент більшість автоматизованих систем працюють лише з структурованими даними. Такі дані відносяться до інформації з високим рівнем організації, наприклад, реляційні бази даних. З такими даними працювати просто. Відсутність структури значно впливає на методики вирішення проблем, вони призначені здебільшого для людей, які не взаємодіють між собою у строгому форматі [14, 562–570]. Такі неструктуровані дані можуть містити:

а) неформальні форми;

б) сленги;

в) графічні позначення;

г) пропуски букв чи слів;

д) помилки.

Отже, емоційний аналіз потребує якісної обробки природньої мови, потрібні кращі алгоритми для переводу тексту, пошуку жаргонів, трактування графічних символів тощо.

5 РОЗРОБКА ДОДАТКУ

Маємо на меті розробити додаток, який міг би оброблювати запити, які містять в собі певну текстову інформацію, та повертати дані, що описують емоційну оцінку цього тексту.

Серверну частину було вирішено реалізувати на мові програмування Java у поєднанні з фреймворком Spring Boot. Він надаватиме REST API, який ми буде використовувати для здійснення аналізу публікацій соціальних мереж.

Для здійснення власне емоційного аналізу у проект було долучено бібліотеку LingPipe. Ця бібліотека використовує алгоритм з класифікацією висловлювань на об’єктивні та суб’єктивні, та класифікує текст за позитивним чи негативним характером. У ній реалізовані ідеї роботі Ліліана Лі та Бо Панга на тему емоційного аналізу [10].

Використовуючи засоби бібліотеки, спочатку необхідно навчити систему аналізувати дані. Для цього скористаємося демонстраційними даними, які були надані розробниками цієї бібліотеки. Код, який використовувався для навчання системи, наведений у додатку А.

Після проведення тренування системи, отриману модель можна використовувати для проведення емоційного аналізу. На разі вона може розлічити негативну чи позитивну емоцію. Програмний код наведений у додатку Б.

Додаток В містить шаблон для створення запитів до розробленого додатку через IP/TCP протокол. Сервер приймає запит, аналізує переданий у запиті текст на емоційну оцінку та власне вертає цю оцінку у якості відповіді.

Висновки

Отже, у даній роботі було проведено аналіз проблеми емоційного аналізу тексту написаного природньою мовою. У роботі описані основні складові емоційного аналізатора, такі як побудова бази знань, вибір рівня деталізації обробки та вилучення шуканої корисної інформації. Розглянути відомі метрики для оцінювання результату роботи емоційного аналізатора. Також була приділена увага різним сучасним проблемам у даній галузі, що на даний момент не мають універсальних та стабільних вирішень. Більшість таких проблем тісно пов’язані з особливостями природньої мови та відставання суміжних лінгвістичних галузей обробки неструктурованих даних. Також було розроблено додаток, який реалізує роботу з емоційним аналізатором.

Список літератури

1. Chaffey D. Global social media research summary 2020 [Електронний ресурс] / Dave Chaffey // Smart Insights. – 2020. – Режим доступу до ресурсу: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research.
2. Sanh V. The Best and Most Current of Modern Natural Language Processing [Електронний ресурс] / Victor Sanh // Medium. – 2019. – Режим доступу до ресурсу: https://medium.com/huggingface/the-best-and-most-current-of-modern-natural-language-processing-5055f409a1d1.
3. Sentiment Analysis and Opinion Mining [Електронний ресурс] // Morgan & Claypool Publishers. – 2012. – Режим доступу до ресурсу: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf.
4. Hatzivassiloglou V. Effects of Adjective Orientation and Gradability on Sentence Subjectivity [Електронний ресурс] / V. Hatzivassiloglou, J. M. Wiebe // ACL Anthology. – 2000. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/C00-1044.
5. Liu F. A Broad-Coverage Normalization System for Social Media Language [Електронний ресурс] / F. Liu, F. Weng, X. Jiang // Association for Computational Linguistics. – 2012. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/P12-1109.
6. Moghaddam S. Opinion Polarity Identification through Adjectives [Електронний ресурс] / S. Moghaddam, F. Popowich. – 2010. – Режим доступу до ресурсу: https://www.researchgate.net/publication/47820140_Opinion_Polarity_Identification_through_Adjectives.
7. D. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews [Електронний ресурс] / Peter D. Turney // Association for Computational Linguistics. – 2002. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/P02-1053.pdf.

8. Pang B. Thumbs up? Sentiment classification using machine learning techniques [Електронний ресурс] / B. Pang, L. Lee, S. Vaithyanathan // Association for Computational Linguistics. – 2002. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/W02-1011.
9. Hatzivassiloglou V. Predicting the semantic orientation of adjectives [Електронний ресурс] / V. Hatzivassiloglou, K. McKeown // Association for Computational Linguistics. – 1997. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/P97-1023
10. Pang B. A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts [Електронний ресурс] / B. Pang, L. Lee // Association for Computational Linguistics.. – 2004. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/P04-1035
11. Effective and efficient? Bilingual sentiment lexicon extraction using collocation alignment [Електронний ресурс] / [Z. Lin, S. Tan, X. Cheng та ін.]. – 2012. – Режим доступу до ресурсу: https://dl.acm.org/doi/10.1145/2396761.2398469.
12. Named entity recognition through classifier combination [Електронний ресурс] / R.Florian, A. Ittycheriah, H. Jing, T. Zhang. – 2003. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/W03-0425.
13. Jindal N. Opinion Spam and Analysis [Електронний ресурс] / N. Jindal, B. Liu. – 2008. – Режим доступу до ресурсу: https://www.cs.uic.edu/~liub/FBS/opinion-spam-WSDM-08.pdf.

14. Brody S. Coooooooooooooollllllllllllll!!!!!!!!!!!!!! Using Word Lengthening to Detect Sentiment in Microblogs [Електронний ресурс] / S. Brody, N. Diakopoulos // Association for Computational Linguistics. – 2011. – Режим доступу до ресурсу: https://www.aclweb.org/anthology/D11-1052.

Додаток А
(обов'язковий)
Тренування системи емоційного аналізу

```
@Service
public class LingPipeTrainingService implements TrainingService {

   private static final String LING_PIPE_TRAINING_DATA_FOLDER = "data";
   private static final String FILES_ENCODE_FORMAT = "ISO-8859-1";

   private File categoriesFolder;
   private String[] categories;
   private DynamicLMClassifier<NGramProcessLM> classifier;

   public LingPipeTrainingService() {
         categoriesFolder = new File(
               LING_PIPE_TRAINING_DATA_FOLDER, "txt_sentoken");
         categories = categoriesFolder.list();
         classifier = DynamicLMClassifier.createNGramProcess(categories, 8);
   }

   public void train() throws IOException {
         for (String category: categories) {
           Classification classification = new Classification(category);
           File categoryFolder = new File(categoriesFolder, category);
           File[] categoryFiles = categoryFolder.listFiles();
           for (File trainingFile: categoryFiles) {
               if (isTrainingFile(trainingFile)) {
                        String trainingText = Files.readFromFile(trainingFile,
   FILES_ENCODE_FORMAT);
                  Classified<CharSequence> classified = new
Classified<CharSequence>(trainingText, classification);
                  classifier.handle(classified);
                }
           }
       }
   }

   private static boolean isTrainingFile(File file) {
       return file.getName().charAt(2) != '9';
   }

}
```

Додаток Б
(обов’язковий)
Емоційна оцінка тексту

```
@Service
public class LingPipeAnalizationService implements AnalizationService {

   @Autowired
   private TrainingService trainingService;

   public String analize(String text) {
          Classification classification = trainingService.classify(text);
          return classification.bestCategory();
   }

}
```

Додаток В
(обов’язковий)
Формат запиту емоційного аналізу

Method	POST
URI	http://localhost:8080/api/v1/analize
Body	I like this movie. I would like to watch it one more time.
Response	Positive