

Wschodnioeuropejskie Czasopismo Naukowe
(Warszawa, Polska)

Czasopismo jest zarejestrowane i publikowane w Polsce. W czasopiśmie publikowane są artykuły ze wszystkich dziedzin naukowych. Czasopismo publikowane jest w języku polskim, angielskim, niemieckim i rosyjskim.

Artykuły przyjmowane są do dnia 30 każdego miesiąca.

Częstotliwość: 12 wydań rocznie.

Format - A4, kolorowy druk

Wszystkie artykuły są recenzowane

Każdy autor otrzymuje jeden bezpłatny egzemplarz czasopisma.

Bezpłatny dostęp do wersji elektronicznej czasopisma.

East European Scientific Journal

(Warsaw, Poland)

The journal is registered and published in Poland.

Articles in all spheres of sciences are published in the journal. Journal is published in English, German, Polish and Russian.

Articles are accepted till the 30th day of each month.

Periodicity: 12 issues per year.

Format - A4, color printing

All articles are reviewed

Each author receives one free printed copy of the journal

Free access to the electronic version of journal

Zespół redakcyjny

Redaktor naczelny - Adam Barczuk

Mikołaj Wiśniewski

Szymon Andrzejewski

Dominik Makowski

Paweł Lewandowski

Rada naukowa

Adam Nowicki (Uniwersytet Warszawski)

Michał Adamczyk (Instytut Stosunków Międzynarodowych)

Peter Cohan (Princeton University)

Mateusz Jabłoński (Politechnika Krakowska im. Tadeusza Kościuszki)

Piotr Michalak (Uniwersytet Warszawski)

Jerzy Czarnecki (Uniwersytet Jagielloński)

Kolub Frennen (University of Tübingen)

Bartosz Wysocki (Instytut Stosunków Międzynarodowych)

Patrick O'Connell (Paris IV Sorbonne)

Maciej Kaczmarczyk (Uniwersytet Warszawski)

Dawid Kowalik (Politechnika Krakowska im. Tadeusza Kościuszki)

Peter Clarkwood (University College London)

Igor Dziedzic (Polska Akademia Nauk)

Alexander Klimek (Polska Akademia Nauk)

Alexander Rogowski (Uniwersytet Jagielloński)

Kehan Schreiner(Hebrew University)

Bartosz Mazurkiewicz (Politechnika Krakowska im. Tadeusza Kościuszki)

Anthony Maverick(Bar-Ilan University)

Mikołaj Żukowski (Uniwersytet Warszawski)

Mateusz Marszałek (Uniwersytet Jagielloński)

Szymon Matysiak (Polska Akademia Nauk)

Michał Niewiadomski (Instytut Stosunków Międzynarodowych)

Redaktor naczelny - Adam Barczuk

1000 kopii.

Wydrukowano w «Aleje Jerozolimskie 85/21, 02-001 Warszawa, Polska»

Wschodnioeuropejskie Czasopismo Naukowe

Aleje Jerozolimskie 85/21, 02-001 Warszawa, Polska

E-mail: info@eesa-journal.com , <http://eesa-journal.com/>

SPIS TREŚCI

NAUKI INŻYNIERYJNE I TECHNICZNE | ТЕХНИЧЕСКИЕ НАУКИ

Белошицкий Н. В., Белошицкая Н. И. РЕЦИКЛИНГ – ПУТЬ ПОЛУЧЕНИЯ НОВЫХ МАТЕРИАЛОВ И ТЕХНОЛОГИЙ В ПОРОШКОВОЙ МЕТАЛЛУРГИИ.....	5
OSADCZY W., ANISIMOW K., BONDARENKO A SPRECYZOWANIE MODELU ODDZIAŁYWAŃ SEJSMICZNYCH PRZY OBECNOŚCI AKCELEROGRAMÓW TRZĘSIENIA ZIEMI	10
Васильченко Г.М. ПРОГНОЗИРОВАНИЕ СОСТОЯНИЯ НАРУЖНЫХ ОГРАЖДАЮЩИХ КОНСТРУКЦИЙ НА ОСНОВЕ ИНСТРУМЕНТАЛЬНОГО ЭКСПРЕСС-АНАЛИЗА ТЕМПЕРАТУРНЫХ ПОЛЕЙ НА ВНУТРЕННИХ ПОВЕРХНОСТЯХ ОГРАЖДЕНИЙ ПРИ ТЕРМОМОДЕРНИЗАЦИИ ЗДАНИЯ ВЕНТИЛИРУЕМОЙ ФАСАДНОЙ СИСТЕМОЙ	14
Войтенко И. В. ОЦЕНКА ВЛИЯНИЯ АНИЗОТРОПИИ ПРОЧНОСТИ В ЗАДАЧАХ ВЗАИМОДЕЙСТВИЯ ПОДПОРНЫХ СТЕН С НЕОДНОРОДНЫМИ ОСНОВАНИЯМИ.....	21
Georgieva A. V. DEVELOPMENT OF TECHNOLOGY FOR RECEIPT OF FUNCTIONAL PRODUCTS OF CEREAL-BASED ENRICHED WITH PUREE MIXTURE OF PUMPKIN AND BEET.....	26
Джафаров Н. Д. ПОСТРОЕНИЕ АЛГОРИТМА МНОГОКРАТНОГО УПЛОТНЕНИЯ ПОКРЫТИЯ ИЗ ПОРИСТОГО МАТЕРИАЛА.....	32
Затучный Д. А. ОБЕСПЕЧЕНИЕ РАДИОСВЯЗИ В ГРАЖДАНСКОЙ АВИАЦИИ В КОРОТКОВОЛНОВОМ ДИАПАЗОНЕ.....	38
Зинченко З.А., Самихов Ш. Р. О КИНЕТИКЕ АЗОТНОКИСЛОТНОГО ВЫЩЕЛАЧИВАНИЯ КОНЦЕНТРАТА МЕСТОРОЖДЕНИЯ ТАРОП.....	41
Иванова С. С. КОМПЛЕКСНАЯ ОЦЕНКА КАЧЕСТВА ПОДЗЕМНЫХ ГАЗОВЫХ СЕТЕЙ.....	45
Иовдальский В. А. НОВАЯ ПАРАДИГМА РАЗВИТИЯ ТЕХНИКИ ГИС СВЧ – ДИАПАЗОНА.....	48
Aidana Irmanova, Nazira Ospanova GIS AND KNOWLEDGE SHARING: DESIGNING GIS FOR CULTURAL HERITAGE IN PAVLODAR REGION, KAZAKHSTAN.....	54
Казимиренко Ю.А. ВЛИЯНИЕ ЭНЕРГЕТИЧЕСКИХ ПАРАМЕТРОВ РАДИОАКТИВНЫХ ГРУЗОВ НА ПРОЦЕССЫ ТЕПЛОПЕРЕНОСА И ПРОЧНОСТЬ КОМПОЗИТНЫХ КОНСТРУКЦИЙ ПЛАВУЧИХ СООРУЖЕНИЙ.....	58
Kamildzhanov B.I. OPTIMIZATION OF FREIGHT TRAFFIC FLOW ON THE AUTOMOBILE TRANSPORT.....	63
Карабін Д.Д., Вербицький В. Г. ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ ПЛІВОК АЛЮМІНІЮ ТА ЇХ НАПИЛЕННЯ В ТЕХНОЛОГІЇ ІНТЕГРАЛЬНИХ МІКРОСХЕМ.....	65
Кравченко Л. В. Черноволов В. А. Протасов Д. Н. МЕТОДИКА МОДЕЛИРОВАНИЯ РАСПРЕДЕЛЕНИЯ ВОДЫ МАШИНАМИ КРУГОВОГО ДВИЖЕНИЯ.....	69
Кучеренко Д. Е МЕТОДЫ И СРЕДСТВА КОНТРОЛЯ СОСТОЯНИЯ КАБЕЛЬНЫХ ЛИНИЙ.....	74
Кучеренко Д. Е., Гранкина Н. А. МЕТОД АНАЛГЕЗИЙНОЙ ЭЛЕКТРОФИЗИОТЕРАПИИ КРС.....	78
A. Zhulynskyj , L. Ladieva SIMPLIFIED MATHEMATICAL MODEL OF THE CONTACT MEMBRANE DISTILLATION PROCESS.....	83
Meshcheryakov Y.V., Griguletsky V.G., Ivakin R.A. IMPROVEMENT OF DESIGNS OF DEEP SEA WELLS, TAKING INTO ACCOUNT RESULTS OF THE ACCIDENT ON MS-252 WELL IN THE GULF OF MEXICO.....	87

Larin A.N., Chernobay G.A., Nazarenko S.Y., Lipovoy V.A. DEFINITION OF ELASTIC PROPERTIES OF FIRE HOSES OF TYPE "T" WITH A DIAMETER OF 51 MM UNDER TORSION.	90
Мазакoвa Б.М., Нурбековa А.Х., Найзагареевa А.А. РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ АЛГОРИТМОВ ПО ОБРАБОТКЕ БИОПОСЛЕДОВАТЕЛЬНОСТЕЙ.	95
Невзорова Н. А. УПРАВЛЕНИЕ ФАЛЬСИФИКАЦИЕЙ ПРИ СБОРЕ СТАТИСТИЧЕСКИХ ДАННЫХ.	97
Пизинцали Л. В. УКРАИНА – ПРОБЛЕМЫ УТИЛИЗАЦИИ МОРСКИХ СУДОВ.	100
Прибыльский В., Олейник С., Нгуен Фионг Донг ПОТРЕБИТЕЛЬСКИЕ СВОЙСТВА ФЕРМЕНТИРОВАННОГО НАПИТКА НА ОСНОВЕ РИСА.	104
Радин Ю.А. Свицерский А.Г., Лыско В.В. СОЗДАНИЕ ЦЕНТРОВ ТЕХНИЧЕСКОЙ ПОДДЕРЖКИ (ЦТП) ЭКСПЛУАТАЦИИ ПАРОГАЗОВЫХ УСТАНОВОК С КОТЛАМИ – УТИЛИЗАТОРАМИ.	108
Saltykov V.M. Saltykov A.V. DEFINITION OF THE FIELD OF ELECTROMAGNETIC COMPATIBILITY OF ARC FURNACES AND ELECTRIC SUPPLY SYSTEM.	112
Saltykov A.V. Saltykov V.M. CONDITIONS ELECTROMAGNETIC COMPATIBILITY BY THE MAGNETIC FIELD OF INDUSTRIAL FREQUENCY IN THE SURROUNDING AREA FROM CABLE SYSTEMS POWER SUPPLY.	117
Сергеева О. В., Пивоваров А.А. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОЦЕССА ПЛАЗМОХИМИЧЕСКОЙ ОБРАБОТКИ ВОДНЫХ РАСТВОРОВ ИГРАЮЩИХ РОЛЬ ЖИДКОГО КАТОДА.	121
Скрябин И. О, Аракелян С. М., Евстунин Г. А., Абрахин С. И., Новикова О. А. ЛАЗЕРНОЕ МНОГОЛУЧЕВОЕ ТЕРМОУПРОЧНЕНИЕ ОТВЕТСТВЕННЫХ ДЕТАЛЕЙ МАШИНОСТРОЕНИЯ.	127
Tarasenko V. N., Degtev I. A., Chernysh N. D. ACOUSTIC COMFORT OF A MULTIPURPOSE HALL PALACE OF CULTURE FOR UNIVERSITY STUDENTS OF BSTU NAMED AFTER V. G. SHUKHOV.	132
Тульский Г.Г. Диаб Хассан Сенкевич И.В. Тульская А.Г. ОКСИДНО КОБАЛЬТОВЫЙ ТИТАНОВЫЙ ЭЛЕКТРОД ДЛЯ ЭЛЕКТРОХИМИЧЕСКОГО СИНТЕЗА ГИПОХЛОРИТА НАТРИЯ.	135
Khakimov Z.T. DEVELOPMENT OF METHODS FOR THE ANALYSIS OF THE SPECTRAL CHARACTERISTICS OF OPTICAL FIBER COMMUNICATION SYSTEMS.	140
Хом'як М.І., Карабін О.Й. РАНЖУВАННЯ ТЕКСТУ ЗА ТОНАЛЬНІСТЮ.	141
Цеханович О.М., Макальский Л. М., Кухно А. В. ОЧИСТКА МИНЕРАЛИЗОВАННЫХ ВОД ПОДЗЕМНОЙ ОТКАЧКИ.	144
Скрыпников А. В., Чернышова Е. В., Поротиков А. В. ПОСТРОЕНИЕ МАТРИЦЫ СВЯЗНОСТИ СЛОЕВ И ДИСКРЕТНОЙ МАКРО-МОДЕЛИ СЕТИ СОЦИАЛЬНЫХ ЗАКЛАДОК.	148
Шевцов О. В. МОДЕЛЬ АТАКИ ПІДРОБКИ ПІДПИСУ NTRUSIGN З ПОСИЛЕНИМИ ПАРАМЕТРАМИ.	152
Элеманова Р.Ш., Мусульманова М. М. К ВОПРОСУ ПОВЫШЕНИЯ ФУНКЦИОНАЛЬНЫХ СВОЙСТВ ФЕРМЕНТИРОВАННОГО ЗЕРНОВОГО НАПИТКА.	156

of the amplifier is further amplified in the voltage amplifier 11 with built-in filter that cuts the high-frequency interference and network (over 30 Hz).

Information is entered into the computer of the spectrograph through COM port (RS-232). On the monitor screen 12a shown the recorded spectra.

Also, spectra can be directly displayed on the XY recorder 12b, and recorded on graph paper. Turning the laser in a stand made with the toggle switch 17, the inclusion of AOPF performed using a toggle switch 18.

In this scientific and measuring complex set of spectral characteristics of the research was conducted, including spectrogram removed without the use of laser radiation and using AOPF.

The optical signal generated by the semiconductor laser, consisting of several transverse modes, through the microlens is entered into single-mode fiber.

As is known, in systems with wavelength division multiplexing applied several (the number of channels) single-frequency laser diodes emitting equidistant along the length of the radiation, with a difference of a few nanometers in the spectrum.

Thus, the apparent refractive index of the resonator has a group refractive index. Depending on the values $\lambda = 1,52 \text{ m}$, n and $L = 200 - 500 \text{ microns}$, the interval between the modes is $\Delta\lambda \approx 1,5 - 3 \text{ nm}$. This elementary calculation explains the regular line structure of the spectrum of the semiconductor laser (Figure 2). [2-3].

LITERATURE.

1. Radjabov TD, AM Nazarov, Davranbek DA, AA Simonov, Khakimov ZT, Pichko SV The device for the diagnosis and optimization of the spectral characteristics of optical fiber communication systems number IAP20090142 // State Patent Office of the Republic of Uzbekistan "OFFICIAL BULLETIN" - Tashkent. 2010. №4 (108), pp 19-20.

2. T.D.Radzhabov, O.A.Kobildzhanov, V.V.Kurepin, S.V.Pichko. A method of forming an optical signal for fiber optic testing devices with ASU Uzbekistan Patent IDP 04 788 B, 2000.

3. A.Kozanne, J. Fleury, Mr. Maitre, M. Rousseau, optics and communication // Moscow, Because of the "Mir", 1984, 326-327, 502 p.

РАНЖУВАННЯ ТЕКСТУ ЗА ТОНАЛЬНІСТЮ

Хом'як Мар'яна Іванівна

студентка, фізико-математичний факультет,
Тернопільський національний педагогічний університет
імені Володимира Гнатюка

Карабін Оксана Йосифівна

канд. пед. наук, доцент
кафедри інформатики і методики

її викладання Тернопільського педагогічного університету імені Володимира Гнатюка

RANKING TEXT TONE

Homiak M.I. student, department of physics and mathematic, Ternopil Volodymyr Hnatiuk National Pedagogical University

Karabin O.Y. PhD at Ternopil Volodymyr Hnatiuk National Pedagogical University,

Department of Computer Science and Methods of Its Teaching

АННОТАЦІЯ

Розглянуто питання аналізу тональності тексту. З'ясовано зміст понять «природномовні тексти» і «емоційний зміст». Також у статті висвітленню питання щодо розробки алгоритму програми для визначення емоційного змісту текстів. У роботі розглянуто різні методи аналізу тональності природно мовних текстів. При тому за основу дослідження було використано комбінований підхід аналізу текстових даних і синтаксичних зв'язків у тексті. При використанні даного підходу враховується не лише емоційний зміст окремих слів, але і словосполучень. Це дозволяє ранжувати текстові дані за емоційним забарвленням. На основі цього підходу реалізовано алгоритм, який дозволяє визначити емоційне забарвлення природномовних текстів.

ABSTRACT

This article describes problem of analyzing text tone. Found out the meaning of "natural language texts" and "emotional content". Also, the article describes an algorithm that allows determining the emotional content of the lyrics. This article examines the various methods of analysis tone of natural language texts. As the main method of text data and syntactic links in the text used a combined analysis. This approach is used not only emotional meaning of each word, but also meaning of phrases of words. This allows you to rank text data for emotional content. Based on this was implemented algorithm, which determines the emotional content of natural language texts.

Ключові слова: природномовні тексти, тональність тексту, емоційний зміст, ранжування, сентиментаналіз текстових даних, семантика тексту.

Keywords: natural language texts, the tone of the text, emotional content, ranking, sentiment analysis, semantics of the text.

Постановка проблеми. Однією з важливих переваг використання інформаційних технологій є можливість автоматизованого аналізу великих масивів даних. Спектр

завдань, які відносяться до інформатизації різних сфер життя, є надзвичайно широким і включає, зокрема, отримання інформації з метою прийняття рішень, навчання,

розв'язання наукових та організаційних задач тощо [1].

Нині інформаційні технології застосовують практично у кожній сфері і тому збільшується необхідність у розробці спеціалізованого програмного забезпечення, що забезпечить автоматизовану обробку даних різного типу. Значна частина даних, що опрацьовуються, припадає саме на природно мовні тексти. Обробка природної мови є важливою темою, про яку багато розмовляють не лише у наукових колах. Дана концепція вважається основою для майбутнього розвитку штучного інтелекту.

Серед найбільш цікавих і популярних методів цього широкого наукового напрямку є один, який називається *sentiment analysis*, що в перекладі на означає «аналіз тональності текстів». Одним із основних аспектів дослідження текстових даних є оцінювання тональності, або емоційного змісту тексту. Аналіз тональності тексту – набір методів контент-аналізу в комп'ютерній лінгвістиці, який призначений для автоматизованого виявлення в текстах певного емоційного змісту тексту та його емоційно забарвленої лексики.

Тональність – це емоційне ставлення автора висловлювання до деякого об'єкту, інформація про який виражена в тексті. Емоційна складова, виражена на рівні лексеми або певного комунікативного фрагмента, називається лексичною тональністю (або лексичним сентиментом). Тональність цілого тексту, що розглядається, можна визначити як функцію, яка є сумою лексичних тональностей складових (слів і речень) і правил за якими вони поєднані у тексті.

Існує і інше визначення тональності тексту – це позитивне, негативне або нейтральне забарвлення як цілого текстового документу, так і його окремих частин, які мають відношення до певних понять, таких як персони, організації, бренди тощо [1].

Аналогічні програмні рішення аналізують текст лише за його позитивним вмістом і результати аналізу є не точними, тому такі програмні продукти використовують лише у деяких сферах. Проте, таку проблему можна вирішити, якщо розширити спектр тональностей, які може розрізняти програма. Використання такого рішення може активно застосовуватись у багатьох сферах: маркетинг, психологія, соціальні мережі тощо.

Огляд останніх досліджень та публікацій.

Проблема аналізу природномовних текстів описується дослідниками та науковцями. Сарбасова А.Н., розглядала різні методи аналізу тональності текстів, описала існуючі програмні додатки, які сьогодні використовують для сентимент-аналізу, проаналізувала результати дослідження і методи покращення результатів [5].

Науковці Олександр Прохоров, Олександр Керимов розкрили питання використання автоматизованих систем сентимент-аналізу та їх популярність на ринку, здійснили детальний опис та сфери застосування, провели дослідження щодо популярності кожного додатку та його вартості [6].

Олексій Соловійов описав алгоритм визначення емоційного змісту речень та ранжування тексту за тональністю. 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 тональною оцінкою (як оцінили). Автор показав важливість використання модуля SentiFinder у програмі, який визначає три види тональності текстів (позитивну, негативну і нейтральну) щодо заданого об'єкта тональності [7].

Мета статті: розглянути методи аналізу тональності тексту, дослідити особливості існуючих методів сентимент-аналізу природномовних текстів, проаналізувати методи для вдосконалення існуючих програмних рішень, розглянути алгоритм розробки програми автоматизованого аналізу емоційного змісту тексту, зробити аналіз отриманих результатів.

Виклад основного матеріалу. Важливим завданням є розгляд та визначення сутності базових понять у тональності тексту.

Природномовними текстовими даними (текстом) є сукупність речень будь-якою природною мовою [1].

Емоційний зміст (тональність) – це певна емоційна забарвленість тексту, яка формується тональністю його емоційно забарвлених складових одиниць та правил їх поєднання [2], що визначає належність тексту до однієї з 4-х категорій, наприклад:

- радість;
- страх;
- задоволеність;
- агресія.

Кожна з категорій має перелік ознак, за якими вона може бути ідентифікована. Емоційна забарвленість тексту визначається такими параметрами:

– терми (емоційно забарвлені слова), які належать до конкретної емоційної категорії;

– відношення (зв'язки) між цими словами (термами) у тексті та правила за якими вони поєднані між собою [4].

Такі зв'язки носять синтаксично-семантичний характер і є невід'ємним елементом будь-якого речення природною мовою. Наявність таких відношень відображається на емоційній забарвленості всього тексту в цілому.

Аналіз тональності тексту передбачає використання одного або декількох методів, наприклад:

I. Методи, засновані на правилах і словниках.

Даний підхід характеризується тим, що текст аналізується на основі заздалегідь складених тональних словників. Однак процес їх створення дуже трудомісткий; основною проблемою є той факт, що одне і те ж слово в різних контекстах може мати різну тональність. Це означає, що для адекватної роботи системи потрібна скласти велику кількість правил – тому найчастіше системи аналізу тональності тексту створюються з прив'язкою до певної предметної області.

II. Методи, засновані на теоретико-графових моделях.

У таких методах текст зображується у вигляді графа на підставі того припущення, що деякі слова мають більшу вагу і, отже, сильніше впливають на тональність всього тексту. Після ранжування вершин графа слова класифікуються відповідно зі словником тональності, де кожному слову присвоюється певна оцінка («позитивне», «негативне» або «нейтральне»). Результат обчислюється як співвідношення кількості слів з позитивною оцінкою до кількості слів з негативною оцінкою.

III. Методи, засновані на машинному навчанні – з учителем і без.

Великі дані можуть надати істотну допомогу в навчанні нейронних мереж, які також використовуються в аналізі тональності тексту. Причому, точність оцінки тональності таким способом зростає до 85% – принаймні, такої цифри вдалося досягти вченим зі Стенфорду. Принцип роботи програми простий: вона будує дерево з оцінкою тональності кожного слова, кожної фрази і всього тексту цілком. Найцікавіше: програма розуміє, що зміна порядку слів змінює тональність тексту. Можна припустити, що саме цей факт і забезпечує таку високу точність оцінки тексту і дозволяє вважати нейронні мережі перспективним інструментом такого аналізу.

Поставлена задача є задачею нечіткої класифікації (ранжування): потрібно знайти ступінь належності вхідних даних до кожної із заздалегідь заданих категорій.

Алгоритм роботи створеного програмного продукту має такі етапи:

Крок 1. Попередній аналіз тексту

Перед будь-якою обробкою даних необхідна попередня обробка. У цій стадії видаляються всі html теги, пунктуація, символи. Дана операція здійснюється за допомогою бібліотеки python - «Beautiful Soup». Також всі числа і посилення в тексті замінюються на теги [3].

У тексті присутні так звані «стоп слова» – це часті слова в мові, які в основному не несуть ніякої смислової навантаженості (напр., в англійській мові це такі слова як «the, at, about ...»). Стоп слова видаляються за допомогою пакета Python Natural Language Toolkit (NLTK).

Після попередньої обробки вихідного тексту отримано наступне:

[Біографія, частина, майбутнє, фільм, пам'ять, вулиця, бачити, театр, оригінал] – тобто набір слів.

Крок 2. Подання у вигляді вектора

Уведений текст необхідно представити у вигляді вектора з чисел, оскільки комп'ютер краще опрацьовує числові дані. Для цього необхідно скласти словник з усіма словами або ж використати готові словники, і замінити слова з тексту індексом в словнику.

Наприклад, нехай існують такі набори слів:

1. [біографія, частина, майбутнє]
2. [фільм, пам'ять, вулиця]
3. [бачити, театр, оригінал]

Об'єднуючи все слова зі списку в один, створюється відсортований словник (базис вектор):

[біографія, частина, майбутнє, фільм, пам'ять, вулиця, бачити, театр, оригінал]

Замінюючи попередні вектори на індекс слова в словнику буде наступне:

1. [1, 0, 1, 0, 0, 0, 1, 0, 0]
2. [0, 0, 0, 1, 1, 0, 0, 1, 0]
3. [0, 1, 0, 0, 1, 0, 0, 0, 1]

Такі вектори називаються «векторами властивостей» або ж «features vector».

У результаті кожне текстове повідомлення має свій вектор, який можна порівнювати з іншими за допомогою стандартних вимірювань, таких як Евклідова відстань, косинусна відстань тощо. Даний підхід називається «мішок

слів» або ж «Bag-Of-Word». Нижче фрагмент коду, який реалізовує даний підхід.

```
from sklearn.feature_extraction.text import
CountVectorizer
# в пакеті sklearn вже вбудований даний метод
vectorizer = CountVectorizer(analyzer = "word",
\ tokenizer = None,
\ preprocessor = None,
\ stop_words = None,
\ max_features = 5000)
```

```
train_data_features = vectorizer.fit_transform(clean_train_
reviews) train_data_features = train_data_features.toarray()
```

Крок 3. Класифікація текстів

Алгоритм класифікації Random Forest використовується для класифікації документів в цьому експерименті. Алгоритм вже реалізований в пакеті scikit-learn, тому потрібно лише перевірити дані і вказати кількість дерев. Далі алгоритм виконує завдання використовуючи навчальну вибірку та зберігає всі необхідні дані.

Висновки. На основі класичного підходу до ранжування текстових даних розроблено модифікований алгоритм нечіткої класифікації природномовних текстових даних, який дозволяє більш точно визначати емоційне забарвлення останніх, ніж існуючі алгоритми, а також робить механізм для управління категоріями емоцій (їх додаванням та видаленням) більш гнучким. Подальше вивчення питання видається автору перспективним, оскільки розроблений алгоритм може бути застосований як основа системи автоматизованого визначення емоційного змісту природномовних текстових даних, яка може бути використана при вирішенні широкого спектру задач, зокрема, для комп'ютеризованого аналізу впливу інформації із ЗМІ на людей, аналізу психоемоційного стану колективу у великих корпораціях тощо.

Використана література

1. Ландэ, Д.В. Интернетика. Навигация в сложных сетях: модели и алгоритмы [Текст] / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. — М.: Либроком, 2009. — 264с.
2. Гаспаров, Б. М. Язык, память, образ. Лингвистика языкового существования [Текст] / Б. М. Гаспаров. — М.: Новое Литературное Обозрение, 1996. — 352 с.
3. Archive.IS [Електронний ресурс]. — Режим доступу: <http://archive.is/>
4. ВААЛ [Електронний ресурс]. — Режим доступу : <http://vaal.ru/>
5. Сарбасова А. Н. Исследование методов sentiment-анализа русскоязычных текстов // Молодой ученый. — 2015. — №8. — С. 143-146.
6. Керимов А. Сентимент-анализ и продвижение в социальных медиа [Електронний ресурс] / А. Керимов, А. Прохоров. — 2012. — Режим доступу до ресурсу: <http://www.moluch.ru/archive/88/17413/>.
7. Соловьёв А. Сентимент анализ текста [Електронний ресурс] / Алексей Соловьёв. — 2015. — Режим доступу до ресурсу: <http://eurekaengine.ru/blog/2015/07/10/sentiment-analysis/>.

Wschodnioeuropejskie Czasopismo Naukowe
(Warszawa, Polska)

Czasopismo jest zarejestrowane i publikowane w Polsce. W czasopiśmie publikowane są artykuły ze wszystkich dziedzin naukowych. Czasopismo publikowane jest w języku polskim, angielskim, niemieckim i rosyjskim.

Artykuły przyjmowane są do dnia 30 każdego miesiąca.

Częstotliwość: 12 wydań rocznie.

Format - A4, kolorowy druk

Wszystkie artykuły są recenzowane

Każdy autor otrzymuje jeden bezpłatny egzemplarz czasopisma.

Bezpłatny dostęp do wersji elektronicznej czasopisma.

East European Scientific Journal

(Warsaw, Poland)

The journal is registered and published in Poland.

Articles in all spheres of sciences are published in the journal. Journal is published in English, German, Polish and Russian.

Articles are accepted till the 30th day of each month.

Periodicity: 12 issues per year.

Format - A4, color printing

All articles are reviewed

Each author receives one free printed copy of the journal

Free access to the electronic version of journal

Zespół redakcyjny

Redaktor naczelny - Adam Barczuk

Mikołaj Wiśniewski

Szymon Andrzejewski

Dominik Makowski

Paweł Lewandowski

Rada naukowa

Adam Nowicki (Uniwersytet Warszawski)

Michał Adamczyk (Instytut Stosunków Międzynarodowych)

Peter Cohan (Princeton University)

Mateusz Jabłoński (Politechnika Krakowska im. Tadeusza Kościuszki)

Piotr Michalak (Uniwersytet Warszawski)

Jerzy Czarnecki (Uniwersytet Jagielloński)

Kolub Frennen (University of Tübingen)

Bartosz Wysocki (Instytut Stosunków Międzynarodowych)

Patrick O'Connell (Paris IV Sorbonne)

Maciej Kaczmarczyk (Uniwersytet Warszawski)

Dawid Kowalik (Politechnika Krakowska im. Tadeusza Kościuszki)

Peter Clarkwood (University College London)

Igor Dziedzic (Polska Akademia Nauk)

Alexander Klimek (Polska Akademia Nauk)

Alexander Rogowski (Uniwersytet Jagielloński)

Kehan Schreiner (Hebrew University)

Bartosz Mazurkiewicz (Politechnika Krakowska im. Tadeusza Kościuszki)

Anthony Maverick (Bar-Ilan University)

Mikołaj Żukowski (Uniwersytet Warszawski)

Mateusz Marszałek (Uniwersytet Jagielloński)

Szymon Matysiak (Polska Akademia Nauk)

Michał Niewiadomski (Instytut Stosunków Międzynarodowych)

Redaktor naczelny - Adam Barczuk

Wschodnioeuropejskie Czasopismo Naukowe — 162 st.

Aleje Jerozolimskie 85/21, 02-001 Warszawa, Polska

E-mail: info@eesa-journal.com , <http://eesa-journal.com/>