

багато інших переваг оновленої інфраструктури.

Розглянуті переваги та впровадження хмарних технологій Microsoft Office 365 в ІТ-інфраструктуру вищого навчального закладу дозволить зробити висновки:

– головним критерієм визначення хмарної технології є можливість роботи з її ресурсами, незважаючи на апаратно-програмне забезпечення користувача, а також його географічного розташування;

– впровадження хмарних технологій в ІТ-інфраструктуру вищого навчального закладу значно спрощує життя студентів та викладачів.

ЛІТЕРАТУРА

1. Биков В.Ю. Хмарна комп'ютерно-технологічна платформа відкритої освіти та відповідний розвиток організаційно-технологічної будови іт-підрозділів навчальних закладів / В.Ю. Биков // Теорія і практика управління соціальними системами. – 2013. – № 1. – с. 81-98.
2. Олексюк В. П. Досвід інтеграції хмарних сервісів GoogleApps у інформаційно-освітній простір вищого навчального закладу [Електронний ресурс] / В. П. Олексюк // Інформаційні технології і засоби навчання. – 2013. – № 3.
3. Співаковський О. В. Побудова ІКТ інфраструктури ВНЗ: проблеми та шляхи вирішення [Електронний ресурс] / О. В. Співаковський, М. О. Вінник, Ю. Г. Тарасіч // Інформаційні технології і засоби навчання. – 2014. – № 1 (39). – С. 99–116.
4. Шишкіна М.П. Хмаро орієнтоване освітнє середовище навчального закладу: сучасний стан і перспективи розвитку досліджень М. П. Шишкіна, М.В. Попель // Інформаційні технології і засоби навчання. – 2013. – №5 (37).
5. Gillam, Lee. CloudComputing: Principles, Systemsand Applications / NickAntonopoulos, LeeGillam. — L.: Springer, 2010. — 379p. — (ComputerCommunicationsandNetworks). —ISBN 9781849962407.

Хом'як М.

Науковий керівник – доц. Карабін О.Й.

РАНЖУВАННЯ ТЕКСТУ ЗА ТОНАЛЬНІСТЮ

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 [4, с.13].

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 природномовні тексти. Обробка природної мови є важливою темою, про яку багато розмовляють не лише у наукових колах. Такий напрям 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».

Мета статті. Розглянути методи ранжування та аналізу тональності природномовних текстів і розробити етапи написання модифікованого алгоритму нечіткої класифікації текстових даних.

Виклад основного матеріалу. 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.

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 [4, с.87].

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.

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.

Вирішення такої проблеми можливе через розширення спектру тональностей, які здатна розрізняти програма. Використання такого рішення може активно застосовуватись у багатьох сферах: маркетинг, психологія, соціальні мережі тощо.

Розглянемо основні поняття, які використовуються у ранжуванні тексту за тональністю. Природномовними текстовими даними (текстом) є сукупність речень будь-якою природною мовою [4, с. 86].

Емоційний зміст (тональність) – це певна емоційна забарвленість тексту, яка формується тональністю його емоційно забарвлених складових одиниць та правил їх поєднання, що визначає

належність тексту до однієї з 4-х категорій, наприклад: радість, страх, задоволеність, агресія [2, с. 118].

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 [1].

Зазначені зв'язки носять синтаксично-семантичний характер і є невід'ємним елементом будь-якого речення природної мовою. Наявність 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.

Дані методи характеризуються 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 правильної роботи системи потрібно 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 рамках цих методів 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.

Такі методи дозволяють використання великого обсягу даних та надають істотну допомогу в навчанні та задіянні нейронних мереж, а 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: вона буде деревом з оцінкою тональності кожного слова, кожної фрази і всього тексту цілком. Програма враховує, що зміна порядку слів змінює тональність тексту. Такий підхід 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 розробки програмного продукту має такі етапи:

Крок 1. Попередній аналіз тексту

Перед будь-якою обробкою даних необхідна попередня обробка. На даному етапі видаляються всі HTML теги, пунктуації та символи. Дана операція здійснюється за допомогою бібліотеки мови програмування Python – «Beautiful Soup». Також всі числа і посилання в тексті замінюються на теги [3].

У тексті присутні так звані «стоп слова» – це часті слова в мові, які в основному не мають ніякого смислового навантаження (напр., «і, та, або ...»). Стоп слова видаляються за допомогою пакета Python Natural Language Toolkit (NLTK).

Після попередньої обробки вихідного тексту одержуємо:

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

Крок 2. Подання у вигляді вектора

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

Наприклад, нехай існують такі набори слів:

[біографія, частина, майбутнє];

[фільм, пам'ять, вулиця];

[бачити, театр, оригінал].

Об'єднання всіх слів із списку в одне, створює відсортований словник (базис вектор): *[біографія, частина, майбутнє, фільм, пам'ять, вулиця, бачити, театр, оригінал].*

Зміна попередніх векторів на індекс слова в словнику покаже наступне:

[1, 0, 1, 0, 0, 0, 1, 0, 0];

[0, 0, 0, 1, 1, 0, 0, 1, 0];

[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-Words» [3]. Нижче подано фрагмент коду, який

реалізовує такий підхід.

```
from sklearn.feature_extraction.text import CountVectorizer
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.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 даної тематики вважається 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 в колективі.

ЛІТЕРАТУРА

1. Бегунов А. Б. Алгоритм автоматизованого визначення емоційного забарвлення текстових даних / А. Б. Бегунов. // Інженерія програмного забезпечення. – 2011. – №2(6).
2. Гаспаров Б. М. Язык, память, образ. Лингвистика языкового существования / Б. М. Гаспаров. – М.: Новое Литературное Обозрение, 1996. – 352 с.
3. Зармагамбетов А. І. Автоматичне визначення тональності тексту [Електронний ресурс] / А. І. Зармагамбетов. – 2015. – Режим доступу до ресурсу: <http://it-ua.info/news/2015/07/21/avtomatichne-viznachennya-tonalnost-tekstu-sentiment-analysis.html>.
4. Ланде Д. В. Интернетика. Навигация в сложных сетях: модели и алгоритмы / Д. В. Ланде, А. А. Снарский, И. В. Безсуднов. – М.: Либроком, 2009. – 264 с.

Шимків Н.

Науковий керівник – доц. Олексюк В.П.

ТЕХНОЛОГІЧНІ АСПЕКТИ ПРОЕКТУВАННЯ СИСТЕМИ ЄДИНОЇ АВТЕНТИФІКАЦІЇ НА ОСНОВІ ТЕХНОЛОГІЇ SHIBBOLETH

Протягом останніх двох десятиліть процес розвитку інформаційно-комунікаційних технологій має значний вплив на розвиток освіти. Велику популярність, на сьогоднішній день, мають системи єдиного входу і саме вони спрощують користування багатьма сервісами.

Метою дослідження є система єдиної автентифікації на основі технології Shibboleth для надання доступу до освітніх порталів.

Зазначена система може бути реалізована з використанням однієї із концепцій:

- 1) синхронізація даних з обліковими записами користувачів усіх сервісів;
- 2) використання спільної бази даних облікових записів користувачів, проте для доступу до кожного сервісу необхідна окрема автентифікація;
- 3) механізм, за допомогою якого єдина дія щодо автентифікації і авторизації користувача надає йому доступ до всіх сервісів системи [4].

Нами було використано саме третій підхід, як найбільш функціональний та безпечний.

Технологія єдиного входу (англ. Single Sign-On) – дозволяє користувачам використовувати різні сервіси без повторної автентифікації.

Основними перевагами зазначеної технології є :

- зменшення кількості паролів, необхідних для доступу до різних програмних продуктів;
- зменшення часу, необхідного для повторного введення паролів;
- зменшення навантаження на мережу, пов'язане з багаторазовими процедурами автентифікації;
- зменшення вартості інформаційної системи за рахунок зниження кількості інцидентів ідентифікаційної бази, пов'язаних з обліковими даними користувачів.

Відкрита для доступу технологія Shibboleth, розроблена корпорацією US Internet2, надає відкритий стандарт управління автентифікацією. Shibboleth та архітектура, яку вона підтримує, робить можливим