

ЗБІРНИК НАУКОВИХ ПРАЦЬ

З МАТЕРІАЛАМИ V МІЖНАРОДНОЇ НАУКОВОЇ КОНФЕРЕНЦІЇ

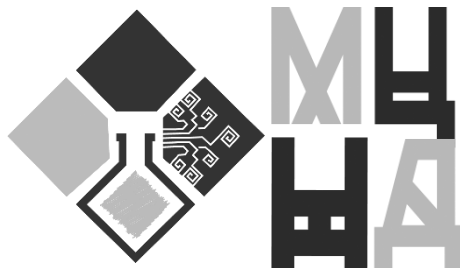
8 ЛИСТОПАДА 2024 РІК

М. ПОЛТАВА, УКРАЇНА

**«НАУКОВІ ВІДКРИТТЯ ТА ФУНДАМЕНТАЛЬНІ
НАУКОВІ ДОСЛІДЖЕННЯ: СВІТОВИЙ ДОСВІД»**



ЗБІРНИК НАУКОВИХ
ПРАЦЬ З МАТЕРІАЛАМИ
У МІЖНАРОДНОЇ
НАУКОВОЇ КОНФЕРЕНЦІЇ



НАУКОВІ ВІДКРИТТЯ ТА ФУНДАМЕНТАЛЬНІ НАУКОВІ ДОСЛІДЖЕННЯ: СВІТОВИЙ ДОСВІД

| 8 листопада 2024 рік
м. Полтава, Україна

Вінниця, Україна
«UKRLOGOS Group»
2024

УДК 082:001
Н 34



Організація, від імені якої випущено видання:

ГО «Міжнародний центр наукових досліджень»

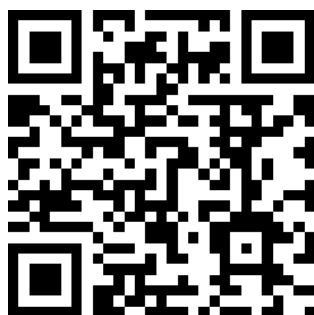
Номер запису організації в Єдиному реєстрі громадських об'єднань: 1499141.

Голова оргкомітету: Сотник С.Г.

Верстка: Білоус Т.В.

Дизайн: Бондаренко І.В.

Рекомендовано до видання Вченою Радою Інституту науково-технічної інтеграції та співпраці. Протокол № 61 від 07.11.2024 року.



Конференцію зареєстровано Державною науковою установою у сфері управління Міністерства освіти і науки «Український інститут науково-технічної експертизи та інформації» в базі даних науково-технічних заходів України на поточний рік та бюлетені «План проведення наукових, науково-технічних заходів в Україні» (**Посвідчення № 355 від 12.06.2024**).

Збірник наукових праць з матеріалами конференції видано офіційно суб'єктом видавничої справи зі **Свідоцтвом ДК № 7860 від 22.06.2023**.

Матеріали конференції знаходяться у відкритому доступі на умовах ліцензії Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Н 34 **Наукові відкриття та фундаментальні наукові дослідження: світовий досвід:** збірник наукових праць з матеріалами V Міжнародної наукової конференції, м. Полтава, 8 листопада, 2024 р. / Міжнародний центр наукових досліджень. — Вінниця: ТОВ «УКРЛОГОС Груп, 2024. — 510 с.

ISBN 978-617-8440-20-6

DOI 10.62731/mcnd-08.11.2024

Викладено матеріали учасників V Міжнародної наукової конференції «Наукові відкриття та фундаментальні наукові дослідження: світовий досвід», яка відбулася 8 листопада 2024 року у місті Полтава.

УДК 082:001

© Колектив учасників конференції, 2024

© ГО «Міжнародний центр наукових досліджень», 2024

ISBN 978-617-8440-20-6

© ТОВ «УКРЛОГОС Груп», 2024

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗРОБКИ ВЕБ-САЙТУ ОНЛАЙН ШКОЛИ АНГЛІЙСЬКОЇ МОВИ Невідома Д.А.	304
МЕТОДИ OSINT ДЛЯ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В КІБЕРПРОСТОРІ Зубко М.О., Делембовський М.М.	307
СИСТЕМИ КІБЕРБЕЗПЕКИ ДЛЯ ІОТ-ПРИСТРОЇВ В ДОМАШНІХ УМОВАХ Павлюк І.А., Делембовський М.М.	310
ТЕХНОЛОГІЇ OSINT ДЛЯ АНАЛІЗУ ТА УПРАВЛІННЯ КІБЕРРИЗИКАМИ Лопатюк І.В., Делембовський М.М.	313
ХАРАКТЕРИСТИКИ СИСТЕМИ ТИПУ «РОЗУМНИЙ БУДИНОК» ТА МЕТОДИ ВИЯВЛЕННЯ АТАК Деркач Я.О., Агеєв Д.В.	316
СЕКЦІЯ XVII. ФІЛОЛОГІЯ ТА ЖУРНАЛІСТИКА	
THE CULTURAL CONTEXT OF TRANSLATION Zaremska I.M.	320
ЗАМІНА ЯК ЗАСІБ ГРАМАТИЧНОЇ ТРАНСФОРМАЦІЇ В ХУДОЖНЬОМУ ПЕРЕКЛАДІ Давиденко Д.С.	323
ЛЕКСИЧНІ ОДИНИЦІ ЯК ІНСТРУМЕНТ АПЕЛЯЦІЇ ДО ПОТРЕБ ЛЮДИНИ У ПРОМОВАХ КАНДИДАТІВ У ПРЕЗИДЕНТИ США 2024 Зажитько В.О.	327
ЛЕКСИЧНІ ОСОБЛИВОСТІ КИТАЙСЬКИХ ГАЗЕТНО-ПУБЛІЦИСТИЧНИХ ТЕКСТІВ Симоненко Д.Ю.	336
НЕПЕРЕСІЧНА ПОСТАТЬ МАРІЇ БІЛОУС-ГАРАСЕВИЧ В ІСТОРІЇ УКРАЇНСЬКОГО ЛІТЕРАТУРОЗНАВСТВА ТА КРИТИКИ Щербань А.М.	338
ОСНОВНІ ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТУ Черняк О.П., Круглій О.Р.	341
СТРАХ В ОПОВІДАННІ ОЛЕКСИ СЛІСАРЕНКА «ГОРБАТЕ ЖИТТЯ» Золотарьов А.С.	348

ОСНОВНІ ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТУ

Черняк Оксана Павлівна

ORCID ID: 0000-0002-3848-0690

кандидат філологічних наук, доцент,
доцент кафедри іноземних мов та перекладу

Волинський національний університет імені Лесі Українки, Україна

Круглій Олена Ростиславівна

ORCID ID: 0000-0001-6346-8821

кандидат філологічних наук, доцент,
доцент кафедри іноземних мов та перекладу

Волинський національний університет імені Лесі Українки, Україна

Кількісне вираження якісних ознак текстового матеріалу отримало широке поширення, і в середині 20 ст. для цього була розроблена спеціальна технологія, відома як «контент-аналіз» – перетворення масивів текстової (або цифрової) інформації в кількісні показники з подальшою статистичною обробкою. Сьогодні контент-аналіз як ніколи важливий для забезпечення достовірності результатів досліджень. Слід зауважити, що контент-аналіз – це техніка збору інформації, виробленого на основі систематичного виявлення відповідних цілям і завданням дослідження характеристик текстів (понять, дієслів, словосполучень та ін.) Його основні операції були розроблені американськими аналітиками Х. Лассуеллом і Б. Берельсоном. Даною проблематикою займались й українські вчені Н. Танатар, А.Федорчук та ін.

Розвиток комп'ютерних технологій та, зокрема, мережі Інтернет як практично невичерпного і такого, що постійно оновлюється, джерела текстів для аналізу не лише відкриває нові можливості для дослідників комунікації, а й спричиняє появу нових методів опрацювання інформації.

Мета статті – дослідити особливості та проаналізувати основні технології інтелектуального аналізу тексту.

У соціальних науках останнім часом досить популярний *Text Mining*. Його можна розглядати, з одного боку, як набір технік для виявлення в текстових даних нетривіальних тенденцій, які можуть зацікавити дослідника (в т.ч. контент-аналітика), а з іншого – як міждисциплінарний напрямок досліджень, що охоплює методики обробки інформації,

машинного навчання, статистичної класифікації, роботи з базами даних та ін., які можуть бути застосовані до тексту [1, с. 350].

Text Mining (інтелектуальний аналіз тексту) – це процес перетворення неструктурованого тексту в структурований формат для виявлення значущих закономірностей і нових ідей. За допомогою текстового аналізу можна аналізувати величезні обсяги текстових матеріалів, щоб виявити ключові концепції, тенденції та приховані взаємозв'язки.

Штучний інтелект тепер здатен автоматично класифікувати тексти за настроєм, темою чи наміром. Наприклад, алгоритм текстового аналізу може переглядати відгуки про товари, щоб визначити, чи є вони переважно позитивними, нейтральними чи негативними. Також можна визначити ключові слова, які найчастіше використовуються.

Таким чином, компанії можуть аналізувати великі та складні масиви даних у простий, швидкий та ефективний спосіб. Інтелектуальний аналіз тексту також скорочує час, що витрачається на ручні та повторювані завдання.

В основі текстового аналізу лежить *Machine Learning* – підкатегорія штучного інтелекту, яка охоплює багато методів та інструментів, що дозволяють комп'ютерам навчитися виконувати завдання автономно.

Моделі машинного навчання навчаються на даних, щоб мати можливість робити точні прогнози. *Text Mining* – це автоматизація аналізу тексту за допомогою машинного навчання. Першим кроком є збір даних. Ці дані можуть надходити з внутрішніх джерел, таких як взаємодія в чаті, електронні листи, опитування або бази даних компанії. Вони також можуть надходити із зовнішніх джерел, таких як соціальні мережі, сайти з відгуками або новинні статті. Потім дані мають бути підготовлені за допомогою різних методів обробки природної мови. Ця «попередня обробка даних» має на меті очистити і перетворити дані у придатний для використання формат.

Це важливий аспект обробки природної мови, що передбачає використання різних методів, таких як ідентифікація мови, токенізація, маркування частин мови, розбиття на фрагменти та синтаксичний аналіз. Метою цих різних методів є форматування даних для аналізу.

Після завершення цієї «попередньої обробки» тексту настає час для аналізу даних. Для вилучення інформації з даних використовуються різні алгоритми текстового пошуку.

Інтелектуальний аналіз даних передбачає інтеграцію понять з інформатики, математики та статистики. Він спрямований на вилучення корисної інформації та виявлення цікавих кореляцій і закономірностей з будь-якої форми даних, особливо числових. Інтелектуальний аналіз даних найбільше асоціюється з ширшим процесом виявлення знань у базах даних (*Knowledge Discovery in Databases, KDD*), «нетривіальним процесом виявлення достовірних, нових, потенційно корисних і, зрештою, зрозумілих закономірностей у даних» [2]. За аналогією, ця стаття розглядає «інтелектуальний аналіз тексту» як процес, що використовує великі колекції текстів для отримання достовірних, потенційно корисних і, зрештою, зрозумілих знань.

Основними технологіями інтелектуального аналізу тексту є:

Пошук інформації

Найбільш відомі інформаційно-пошукові (ІП) системи – це пошукові системи Google, які розпізнають ті документи на World Wide Web, з якими пов'язаний набір заданих слів. Він є розширенням до пошуку документів, де документи обробляються з метою вилучення корисної інформації, яка є важливою для користувача [5]. Таким чином, за пошуком і вилученням інформації слідує етап узагальнення, орієнтований на запит, заданий користувачем, або етап вилучення інформації. Пошук і вилучення інформації у більш широкому сенсі стосується всього спектру обробки інформації, від пошуку інформації до пошуку знань [4]. Це порівняно давня область досліджень, оскільки перші спроби автоматичної індексації були зроблені ще у 1975 р. Увага до неї зросла зі зростанням World Wide Web і в зв'язку з необхідністю в досконалих пошукових системах.

Вилучення інформації

Метою вилучення інформації (ВІ) є вилучення корисної інформації з тексту. Вона включає вилучення суб'єктів, подій і відносин з напівструктурованого або неструктурованого тексту. Найбільш корисна інформація, така як назва особи, місцезнаходження та організація, вилучається без належного розуміння тексту [7]. ВІ стосується вилучення семантичної інформації з тексту. ВІ може бути описана як побудова структурованого зображення вибраної відповідної частини інформації, отриманої з текстів.

ВІ може відбуватися у двох формах: одна з них полягає у визначенні конкретного поля вилученого об'єкта, такого як ім'я, дата або адреса, а

інша – у визначенні частин мови з текстового масиву за допомогою технології обробки природної мови – *natural language processing (NLP)*

Попередня обробка даних

Попередня обробка даних, або очищення даних, - це алгоритм, який виявляє і видаляє помилки або невідповідності з даних, а також консолідує схожі дані з метою підвищення якості подальшого аналізу. Ці очищені дані потім подаються для процесу аналізу. Для очищення даних можна використовувати кілька методів. Три методи, які використовуються в VantagePoint для очищення списків, – це алгоритм стемінгу, елементарна нечітка логіка для консолідації схожих термінів і тезауруси. Для швидкого наближення до кореня слова можна використовувати алгоритм стемінгу або усічення. Слово, для якого потрібно знайти точний або близький відповідник, можна записати як основу або корінь слова, а пошукову систему попросити знайти слова, які відповідають кореню. Методи нечіткого зіставлення можуть бути використані для ідентифікації, асоціювання та відповідного скорочення даних. Наприклад, це стосується помилок у написанні слів, альтернативних переносів і написання з великої літери. Тезаурус визначається як групування термінів у певні поняття. Його можна використовувати для спеціалізованого скорочення даних.

Категоризація

Категоризація є методом навчання під наглядом, оскільки базується на прикладах вихідних даних для класифікації нових документів. Її ключовим проектом було індексування наукової літератури за допомогою контрольованих слів. Ще у 1990-х роках ця сфера розвивалася повною мірою у зв'язку з наявністю постійно зростаючої кількості текстових документів у цифровій формі та вимогою їх організувати для більш зручного використання [3]. Категоризація – це призначення звичайних мовних документів на заданий набір тем відповідно до їх змісту. Це набір текстових документів, процес пошуку точної теми або тем для кожного документа. Метою категоризації є підготовка класифікатора на основі відомих прикладів, а потім невідомі приклади класифікуються автоматично. В даний час автоматизована категоризація тексту застосовується в різних контекстах від класичної автоматичної або напівавтоматичної індексації текстів до доставки персоналізованої реклами, фільтрації спаму та категоризації веб-сторінки за ієрархічними каталогами, автоматичного створення

метаданих та виявлення текстового жанру, відстеження тем тощо [6]. Вивчення автоматизованої класифікації тексту починається з 1960-х років. І до сьогодні це актуальна тема в галузі машинного дослідження тексту.

Кластеризація

Інший метод, який використовується для аналізу текстових даних, - це кластерний аналіз. Кластерний аналіз, який також називають сегментацією даних, - це підхід, який групує або сегментує набір елементів даних (наприклад, документів або термінів) на підмножини або «кластери» таким чином, що об'єкти всередині кожного кластера тісніше пов'язані один з одним, ніж об'єкти, віднесені до різних кластерів. Методи кластерного аналізу можна розділити на дві категорії: ієрархічні та часткові.

Кластеризація є однією з найбільш цікавих і важливих тем у *text mining*. Її мета полягає у пошуку власних структур в інформації, а також впорядкування їх у значні підгрупи для подальшого вивчення та аналізу. Це процес без нагляду, за допомогою якого об'єкти класифікуються на групи, які називаються кластерами. Проблема полягає в групуванні даної непозначеної колекції у значущі кластери без будь-якої попередньої інформації. Будь-які мітки, пов'язані з об'єктами, отримують виключно з даних. Наприклад, кластеризація документів допомагає при пошуку шляхом створення зв'язків між відповідними документами, що, в свою чергу, дозволяє отримати відповідні документи після того, як один з документів знайдено відповідно до запиту [4].

Кластеризація використовується у багатьох галузях, таких як біологія, інтелектуальний аналіз даних, розпізнавання образів, пошук документів, сегментація зображень, класифікація шаблонів, безпека, бізнес-аналітика та веб-пошук. Кластерний аналіз може використовуватися як автономний інструмент *Text Mining* для досягнення розподілу даних, або як етап попередньої обробки для інших алгоритмів видобування тексту, що працюють на виявлених кластерах.

Підведення підсумків

Підсумовування тексту є старим методом у *Text Mining*, але потребує уваги дослідників в галузі обчислювального інтелекту, машинних знань і обробки природних мов. Підсумовування тексту - це процес автоматичного створення стислої версії даного тексту, що надає корисну

інформацію для користувача. У великій організації або компанії дослідник не встигає прочитати всі документи, тому вони підсумовують документ і висвітлюють резюме з основними пунктами [6]. Резюме – це текст, який складається з одного або більше текстів, які містять значну частину інформації, стислої форми, але в якому збережений загальний зміст, як це є у вихідних текстах. Підсумовування тексту включає в себе різні методи, що використовують категоризацію тексту, такі як нейронні мережі, дерева рішень, семантичні графіки, регресійні моделі, нечітка логіка і ройовий інтелект. Однак, всі ці методи мають загальну проблему, а саме – якість розробки класифікаторів є змінною і сильно залежить від типу тексту, який потрібно підсумувати.

Отже, *Text Mining* використовує різну кількість методів, які відіграють важливу роль. Методи відрізняються один від одного. Інформаційно-пошукову технологію використовують для неструктурованого тексту, де можна отримати цінну інформацію, тоді як вилучення інформації відбувається зі структурованої бази даних. Метод узагальнення використовується для узагальнення документа у стислій формі, зберігаючи при цьому його первинне значення.

Категоризація – це контрольований процес, який використовує попередньо визначені документи відповідно до їх змісту. Відповідальність і гнучкість посткоординаційної системи ефективно забороняють встановлення значущих відносин, оскільки категорія створюється індивідом, а не системою. При кластеризації використовують для пошуку внутрішні структури в інформації і організують їх у відповідні підгрупи для подальшого вивчення та аналізу. Це процес без нагляду, при якому об'єкти об'єднують у групи, які називаються кластерами. Кластеризація має справу з великими за розмірами даними. Іншою особливістю є те, що це група подібного типу даних і існує взаємозв'язок між ними.

Можна без сумніву сказати, що текстовий аналіз має багато переваг, адже компанії та приватні особи щодня генерують величезні обсяги даних. Дійсно, майже 80% текстових даних є неструктурованими. Тому їх неможливо проаналізувати без використання текстового аналізу. Текстова аналітика дозволяє аналізувати великі обсяги даних всього за кілька секунд, тим самим підвищуючи продуктивність. Цей аналіз можна виконувати в режимі реального часу, а отже, можна негайно втрутитися, якщо виявлено проблему.

Список використаних джерел:

1. Черняк О.П. Інтелектуальний аналіз тексту та сфери його застосування. *The XXXIV International Scientific and Practical Conference «Problems of the development of modern science», August 30 – September 02, 2022, Madrid, Spain*, p. 350-352
2. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining to Knowledge Discovery in Databases. *AI magazine*. 17(3), 1996. pp. 37-54
3. Hearst, M. A. (1997) Text data mining: Issues, techniques, and the relationship to information access. *Presentation notes for UW/MS workshop on data mining*, July 1997.
4. Mr. Rahul Patel, Mr. Gaurav Sharma, A survey on text mining techniques. *International Journal Of Engineering And Computer Science* ISSN:2319- 7242, Vol 3 Issue 5, May 2014, pp.5621-5625
5. Rashmi Agrawal, Mridula Batra, A Detailed Study on Text Mining Techniques", *IJSCE*, ISSN: 2231-2307, Vol. 2, Issue-6, January 2013.
6. Sagayam R., Srinivasan S., Roshini S., A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques. *Internaltional Journal of Computational Engineering Research (ijceronline.com)* Vol.2 Issue 5. 2012.
7. Vishal Gupta and Guruprit Lehal, "A Survey of Text Mining Techniques and Applications. *Journal Of Emerging Technologies In Web Intelligence*, Vol. 1, No. 1, August 2009.