

УДК 303.316:004.912

DOI: <https://doi.org/10.31651/2076-5843-2022-3-4-48-57>

АХМЕДОВ Ренат Рамазанович¹

викладач

ORCID ID: <https://orcid.org/0000-0002-3084-4672>

akhmedov.kneu@gmail.com

ДЕРБЕНЦЕВ Василь Джоржович¹

кандидат економічних наук, доцент

ORCID ID: <https://orcid.org/0000-0002-8988-2526>

derbv@kneu.edu.ua

БЕЗКОРОВАЙНИЙ Віталій Сергійович¹

кандидат економічних наук

ORCID ID: <https://orcid.org/0000-0002-4998-8385>

retal.vs@kneu.edu.ua

¹ Державний вищий навчальний заклад «Київський національний економічний університет імені Вадима Гетьмана», м. Київ, Україна

КОНТЕНТ-АНАЛІЗ СОЦІАЛЬНИХ МЕДІА МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Робота присвячена проблемам проведення контент-аналізу електронних соціальних медіа. Проаналізовано особливості соціальних медіа та їх відмінності від традиційних засобів масової інформації. Визначено основні завдання, що дозволяє вирішувати застосування контент-аналізу соціальних медіа для бізнесу. Обґрунтовано використання технології машинного та глибокого навчання для створення систем контент-аналізу. Запропоновано застосування технології Transfer Learning для перенесення знань із переднавчених мовних моделей на інший домен або іншу мову, зокрема, українську з інших слов'янських мов.

Ключові слова: контент-аналіз, соціальні медіа, машинне навчання, глибоке навчання.

Постановка проблеми. Протягом двох останніх десятиріч бурхливий розвиток соціальних медіа (інтернет-ресурсів, які наповнюють контентом зареєстровані користувачі, вони також споживають інформацію інших учасників мережі) спричинив революцію в засобах комунікації сучасного суспільства. Різноманітні соціальні медіа (соціальні мережі, блоги, форуми, дошки оголошень, подкасти, контент-співтовариства, фото- та відеохостінги тощо) створили принципово нову екосистему людського спілкування та стали невід'ємною частиною практично усіх сфер людської діяльності.

Через постійне збільшення кількості інформації звичні раніше комунікативні технології втрачають свою ефективність. За сучасних умов ведення бізнесу здатність швидко відстежувати та контролювати громадську думку стає одним із ключових факторів досягнення успіху. Тому значна кількість провідних світових компаній почала перебудовувати свої бізнес-моделі із використанням можливостей сучасних засобів комунікації через соціальні мережі та інші платформи, для чого використовують технології контент-аналізу. Контент-аналіз широко використовується у маркетингових дослідженнях (зокрема, для моніторингу лояльності аудиторії до різних брендів, товарів, послуг; аналізі діяльності конкурентів; продуктивній аналітиці та підтримці клієнтів; пошуку інсайтів та трендів в індустріях; моніторингу думок співробітників тощо), в соціології (оцінці полярності думок стосовно суспільно важливих тем), політології (наприклад, для оцінки ставлення виборців до кандидатів) тощо.

Величезні обсяги інформації на різноманітних електронних платформах вимагають адекватних засобів їх моніторингу та обробки, аналізу контенту тощо. За даними компанії Finance Yahoo, ринок програмного забезпечення для аналізу контенту оцінюється в 4,7 мільярда доларів і, як очікується, буде зростати приблизно на 17% щорічно [1]. Тому розроблення автоматизованих систем контент-аналізу соціальних медіа, зокрема, із використанням технологій штучного інтелекту (машинного та глибокого навчання), набуває особливої актуальності.

Аналіз останніх досліджень і публікацій. Оскільки розвиток та вдосконалення технології контент аналізу має майже сторічну історію, методологічним аспектам проведення контент-

аналізу присвячено багато наукових праць закордонних та вітчизняних вчених, зокрема, К. Кніпендорфа [2], А. Таршиса [3], А. Берко [4], В. Іванова, Н. Костенко [5-6]. Їх дослідження були присвячені теоретичним та прикладним аспектам контент-аналізу, спрямованого переважно на традиційні засоби масової інформації (ЗМІ).

Проте протягом останніх двох десятиріч виник принципово новий феномен – соціальні мережі та соціальні медіа, що вимагає адаптації існуючих та розроблення нових методів та підходів. Останнім часом у вітчизняній науковій літературі було опубліковано низку робіт, присвячених методології та практичній реалізації контент аналізу в соціальних медіа [7-10].

В роботі [7] досліджено соціальні мережі як джерела великих даних для отримання та аналізу даних про різноманітні події. Проведено критичний аналіз підходів та методів дослідження соціальних мереж, розглянуто основні метрики, які використовуються в якості вхідних даних. В роботі [8] здійснено аналіз основних проблем синтаксичного та семантичного аналізу та функціональних сервісів управління контентом, вдосконалено формальний опис цього процесу та сформовано множину параметрів його супроводу, що дало змогу розробити вимоги та рекомендації для проектування систем електронної контент-комерції. В статті [9] проаналізовано методи і засоби проектування системи контент-аналізу сайту, що дозволило авторам розробити систему кількісного контент-аналізу текстової інформації інтернет-газети (аналіз статистики, рейтингове оцінювання тощо). В роботі [10] було проведено систематизацію сучасних поглядів на використання контент-аналізу та запропоновано методологію аналізу контенту електронних ЗМІ для таргетування цільової аудиторії, визначення уподобань щодо змісту публікацій, що складається з таких етапів: веб-скрапінг, парсінг, якісний аналіз, кількісний аналіз, та аналіз тональності текстів.

Проведення контент-аналізу сучасних соціальних медіа вимагає поєднання традиційних підходів із сучасними методами аналізу надвеликих даних, методів автоматизованої обробки природної мови, технологій штучного інтелекту, зокрема, машинного та глибокого навчання.

Метою статті є розроблення методологічних засад проведення контент-аналізу електронних ресурсів (соціальних медіа) із використанням технологій штучного інтелекту, зокрема, машинного навчання.

Викладення основного матеріалу дослідження. Поява соціальних медіа дозволила користувачам самостійно створювати контент і ділитися ним з іншими користувачами. Сьогодні кожен окремий користувач – цінне джерело інформації, що транслює оточуючим свої думки, погляди, уподобання. За останні роки соціальні мережі зазнали стрімкого розповсюдження, особливо в країнах, що розвиваються. Зокрема, мобільні та стрімінгові платформи, онлайніві форуми, сайти та служби обміну миттєвими повідомленнями демонструють експоненційне зростання, що призвело до широкого використання соціальних мереж.

За даними агенції We Are Social, яка нещодавно опублікувала щорічний звіт із дослідження стану й перспектив розвитку світової сфери діджиталізації (Digital 2022 Global Overview Report, [11]), частка користувачів інтернету в світі на початок 2022 р. склала близько 5,3 млрд. осіб або 60% населення (біля 88% – у розвинених країнах), а кількість активних користувачів різноманітних соціальних мереж за останні 10 років зросла майже втричі – з 1,48 млрд. у 2012 р. до 4,62 млрд. у 2022 р. (рис. 1).

Необхідно зауважити, що соціальні медіа і соціальні мережі не є тотожними поняттями. Головні відмінності соціальних медіа від соціальних мереж – функціональність та масовість. Так, наприклад, Telegram – це месенджер, яким користується 500 млн. користувачів [12], але при цьому він не є соціальною мережею. Теж саме стосується і Slack – чат-платформи для роботи, денна аудиторія якої перевищує 10 млн. осіб. Все це – соціальні медіа.

Основними інструментами соціальних медіа є можливість створення, дистрибуції та редагування контенту з вибудовуванням комунікації. Основні функції соціальних медіа:

- ✓ обмін контентом;
- ✓ створення комунікаційних зв'язків між користувачами;
- ✓ вирішення певних завдань в Інтернеті;
- ✓ дозволя та розважальні послуги тощо.

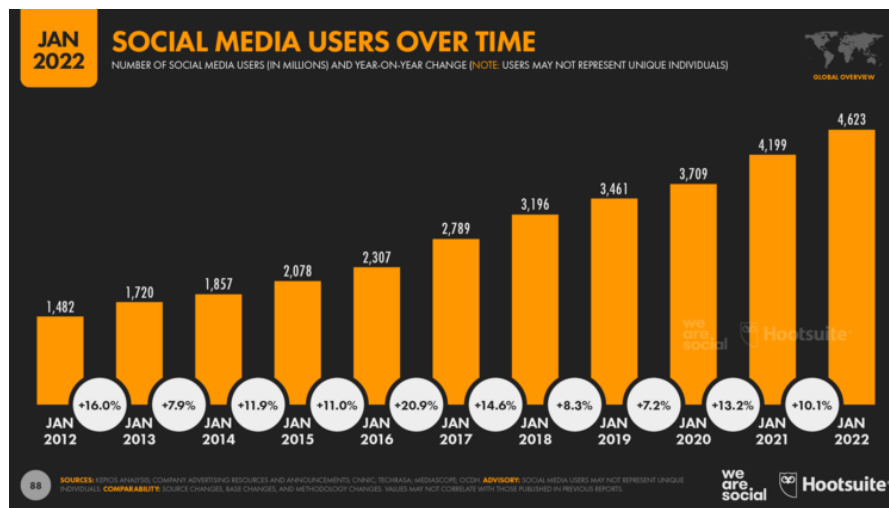


Рис. 1 – Динаміка активних користувачів соціальних мереж протягом 2012-2022 рр. (Джерело інформації [11])

При цьому, за даними агенції Statista [12], в топ 10 за популярністю мереж як у світі, так і в Україні увійшли Facebook, YouTube, WhatsApp, Instagram, Tik Tok (рис. 2).

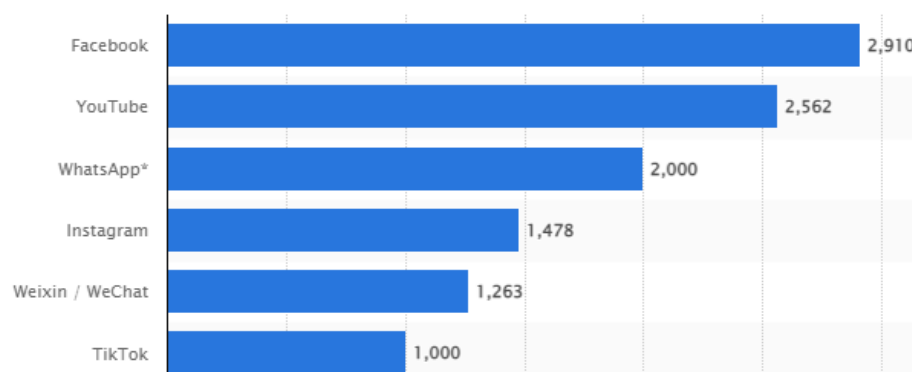


Рис. 2 – Рейтинг популярності соціальних мереж: кількість активних користувачів на місяць, млн. осіб (джерело інформації [12])

Хоча зараз спостерігається тенденція до універсалізації специфіки соціальних медіа, можна виділити такі їх види: комунікаційні; медіа, контент яких створюється користувачами; обмінні тощо. Комунікаційні включають соціальні медіа, що об'єднують людей для спілкування: чати, мікроблоги, сайти з відгуками, форуми, наприклад: Tinder, Clubhouse, IRecommend тощо. Щодо соціальних медіа, в яких можна додавати, змінювати та редагувати інформацію, то найбільш відомим представником цього типу є Вікіпедія. В обмінних медіа відбувається обмін будь-яким типом інформації, зокрема, фото, відео, текстами, музикою тощо (наприклад, Youtube, Twitch, Tik Tok, Instagram).

Таким чином, із точки зору технології та інструментарію для здійснення аналітики, зокрема, контент-аналізу, необхідно враховувати спеціалізацію соціальних медіа за типом основного контенту: текст, зображення, відео. Так, основу Twitter складають мікроблоги, тобто текстові повідомлення невеликої довжини; в Instagram користувачі переважно діляться зображеннями зі своїми передплатниками; Youtube позиціонує себе як сервіс відеохостингу тощо.

Необхідно зауважити, що використання соціальних медіа виявилось корисним як компаніям, так і споживачам. Так, компанії зацікавлені у взаємодії зі своїми споживачами через соціальні медіа, адже це суттєво підвищує впізнаваність бренду і лояльність аудиторії, можуть здійснювати підтримку та зворотній зв'язок із наявними й потенційними клієнтами, контрагентами, партнерами тощо.

З іншого боку, споживачі дедалі частіше використовують соціальні медіа для ухвалення рішень про покупку. Вони шукають відгуки, аналізують думки відомих блогерів, нерідко приймають рішення на підставі того, як часто вони контактують із брендом у соціальних мережах. У публікаціях користувачі залишають відгуки про товари та послуги, формують тренди, діляться своїми уподобаннями та інтересами. З цієї причини великі компанії мають все більш зростаючу зацікавленість у аналізі надвеликих обсягів даних соціальних медіа.

Контент-аналіз (від англ., Content Analysis) є формалізованим методом вивчення текстової та графічної інформації, що полягає в систематичній та надійній фіксації певних елементів змісту деякої сукупності документів з подальшою квантифікацією (кількісною обробкою) отриманих даних [3-5].

Таким чином, контент-аналіз – це метод кількісного вивчення великих обсягів інформації (текстових документів, відео-, аудіоматеріалів, зображень). Цей метод є провідним у дослідженні змісту повідомлень, записів на сайтах соціальних мереж, відповідей на відкриті питання опитувань та інших масивах електронних документів [6].

В найбільш простому випадку контент-аналіз передбачає виявлення частоти використання в тексті певних характеристик (наприклад, термінів, біграм, триграм, емоційно забарвленої лексики тощо), що дозволяє робити висновки щодо намірів авторів тексту або можливих реакцій споживачів цього контенту.

При цьому можна виділити три напрями застосування контент-аналізу текстової інформації:

- ✓ виявлення того, що *“існувало до тексту”* і в певний спосіб отримало в ньому відображення (текст як індикатор певних властивостей досліджуваного об'єкта, навколишньої дійсності, автора тексту або адресата, споживача);
- ✓ визначення того, що *“існує лише у тексті”* (різні характеристики та форми подання інформації; стиль, структура і жанр; тональність та емоційна спрямованість);
- ✓ виявлення того, що *“буде (або може) існувати після тексту”*, тобто після сприйняття споживачем контенту (оцінка різних ефектів впливу).

До основних завдань для бізнесу, що дозволяє вирішувати застосування контент-аналізу, можна віднести такі.

Медіа-моніторинг, аналіз медіа-іміджу. Напіваавтоматизоване та автоматизоване відстеження в інформаційному потоці ЗМІ контексту згадування брендів, назв компаній, партій, громадських або політичних діячів, бізнесменів або інших суб'єктів.

Аналіз успішності медіа-кампанії. Відстеження розповсюдження інформаційних повідомлень у ЗМІ та соціальних мережах. Вивчення динаміки уваги, емоційної полярності (позитивної, негативної, нейтральної) відгуків та коментарів, виявлення ключових інформаційних стейкхолдерів, які сприяють або заважають очікуваному поширенню інформації.

Аналіз соціальних мереж. Відстеження розповсюдження інформаційних повідомлень у соціальних мережах. Вивчення онлайн поведінки певної категорії користувачів: наприклад, молоді, прихильників певної політичної партії, споживачів певних типів продуктів або товарів тощо. Аналіз соціально-демографічного складу та онлайн поведінки учасників виділених груп у соціальних мережах.

Конкурентний аналіз веб-сайтів. Визначення глобального та національного рейтингу сайту, рейтингу у певній тематичній ніші (наприклад, серед сайтів новин), головних конкурентів, джерел трафіку, контексту згадок, популярності окремих розділів, найперспективніших майданчиків для просування сайту.

Одним із корисних інструментів контент-аналізу є створення аналітичних звітів – медіа досліджень певного ринку (галузі, бренду, теми тощо) з наочним представленням результатів у вигляді графіків, рейтингів, діаграм тощо. Окрім кількісних даних можна також вилучати якісні характеристики: тональність контенту (сентимент-аналіз), інформаційні приводи тощо.

Отже, контент аналіз соціальних медіа дозволяє виявляти тренди, позицію учасників ринку з того чи іншого приводу. Такі звіти дозволяють:

- ✓ планувати інформаційну стратегію розвитку компанії;

- ✓ своєчасно корегувати свої дії на ринку;
- ✓ відстежувати динаміку згадувань о компанії чи її брендах, о певних подіях, чи персонах по ключових словах;
- ✓ виявляти ключові інформаційні приводи опублікованих матеріалів;
- ✓ визначати емоційну спрямованість як публікацій в цілому (негативна, нейтральна, позитивна тощо), так і по певних аспектах (аспектний сентимент-аналіз).

Базовою операцією в контент-аналізі можна вважати кодування або присвоєння коду (тегу) фрагменту тексту або зображення. Таким фрагментом, як правило, є речення, абзац тексту або область зображення.

Присвоєння кодів фрагментам тексту дозволяє перетворювати якісні дані у цифровий формат. Так, наприклад, текстовий документ можна подати числовою матрицею, рядки якої відповідають реченням, а стовпці – числовим кодам, що відповідають певним поняттям, (образам, шаблонам, патернам), які ми шукаємо та виділяємо у тексті. Якщо у розглянутому реченні розпізнаний той чи інший образ, то у відповідна комірка матриці містить “1”, в протилежному випадку “0”.

Таким чином, в результаті контент-аналізу відбувається формалізація, що забезпечує відволікання від значення понять та змісту виразів в тексті з метою дослідження їх логічних особливостей, дедуктивних та виразних можливостей.

Наступні кроки в обробці даних, які спочатку були якісними, приймають форму операцій над векторами (рядками та стовпцями матриць), що здійснюються з використанням лінійної алгебри та методів і алгоритмів машинного навчання.

Якщо контент-аналіз здійснюється з метою оцінки ефективності присутності бренду в соціальних медіа, то вона зазвичай визначається набором статичних даних. Для цього у соціальних медіа проводиться кількісний аналіз, який враховує такі метрики: 1) кількість згадок про бренд; 2) охоплення контенту у співвідношенні з діями користувачів; 3) зростання аудиторії та її конверсія у покупки тощо. При цьому за допомогою кількісного (статистичного) підходу можна здійснити лише поверхневий аналіз, зокрема вилучення певних фактів та оцінки їхньої частоти (або інших статистичних характеристик).

Але більшість актуальних завдань контент-аналізу є набагато складнішими за частотний аналіз – це вилучення знань, і тут потрібні алгоритми машинного навчання. Залежно від типу конкретного завдання створюється та налаштовується своя окрема модель.

Створення автоматизованих (напівавтоматизованих) систем для контент-аналізу текстових даних ґрунтується на використанні сучасних методів та моделей обробки природної мови (від англ., Natural Language Processing, NLP). NLP – це міждисциплінарна галузь на перетині інформатики, штучного інтелекту та прикладної лінгвістики, метою якої є створення методів, моделей та алгоритмів для автоматизованих комп’ютерних систем аналізу природної мови [13].

Переважає більшість прикладних задач NLP включає виконання таких етапів [14] (рис. 3).

Препроцесінг тексту полягає в поданні тексту, поданого природною мовою, у форматі, зручному для подальшої роботи. Цей етап, залежно від мети дослідження, може включати: видалення знаків пунктуації; цифр та пробілів; видалення тегів та інших елементів розмітки, які переважно не містять корисної інформації та лише додають шум до даних; переведення всіх слів до нижнього регістру тощо.

Сегментація або токенізація полягає у поділу тексту на речення, а речення – на окремі слова, словоформи або морфеми (токени).

Визначення контекстно-незалежних ознак, що характеризують кожен з токенів, тобто не залежать від елементів, що стоять поруч.

Уточнення значущості та застосування фільтрів до так званих “стоп-слів”. Як для англійської, так і для української (та, зокрема, переважної більшості європейських мов) притаманна наявність багатьох допоміжних частин мовлення (прийменники, сполучники, займенники, артиклі тощо), які спотворюють статистичну значущість інших слів у досліджуваних текстах через їх досить велику як абсолютну, так і відносну частоту появи в текстах, незалежно від прикладної області.

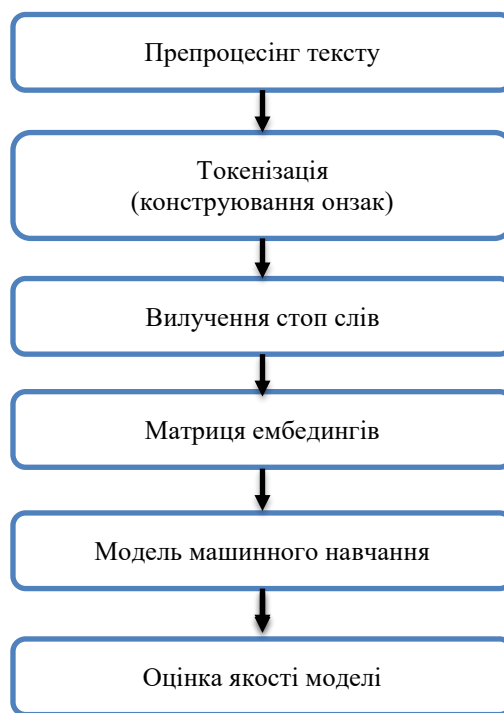


Рис. 3 – Етапи оброблення природної мови при застосуванні методів машинного навчання
(джерело інформації: складено авторами на основі [13-14])

Парсинг залежностей, або семантичний (змістовний) аналіз тексту – виділення семантичних відношень. В загальному випадку семантичне представлення є графом, мережею, що відбиває бінарні відношення між двома вузлами – змістовними одиницями тексту.

Векторне представлення токенів, або створення ембедінгів (від англ., *embedding*), що дозволяє виділити слова, які застосовуються у схожому контексті.

Формування моделі машинного навчання, що відповідає заданій меті (наприклад, класифікації, оцінці тональності, вилученню іменованих сутностей тощо).

Серед прикладних задач NLP, які безпосередньо або дотично пов'язані з проблематикою контент-аналізу, є *класифікація; виділення сутностей та фактів; POS-тегування; семантичний аналіз*.

Найбільш поширеною є задача класифікації, яка передбачає:

- ✓ диференціацію за бінарним та мультикласовим критеріями;
- ✓ відображення змістовної складової та оцінка тональності;
- ✓ визначення релевантності написаного заданим параметрам;
- ✓ поділ на тематичні класи та категорії;
- ✓ підготовка стислого переказу або анотації;
- ✓ виявлення думок та властивостей, що залежать від поставленого завдання.

Виділення сутностей та фактів полягає у ідентифікації посилань на іменовані об'єкти, що належать до однієї із заданих категорій, а також встановлення семантичного взаємозв'язку між ними. До найбільш важливих іменованих сутностей належать:

- ✓ назви компаній та брендів;
- ✓ імена політиків, бізнесменів, громадських діячів або інших персон залежно від мети дослідження;
- ✓ географічні назви, адреси компаній;
- ✓ теги, що пов'язані із геолокацією, датою та часом;
- ✓ числові дані (зокрема, грошові суми) тощо.

POS-тегування (від англ., *Part-of-Speech Tagging*) використовується в NLP для визначення частини мови та граматичних характеристик слів у тексті з присвоєнням їм відповідних тегів. Це

особливо актуально для випадку, коли зміст слова залежить від контексту. POS-тегування дозволяє зіставити словам в тексті спеціальні теги на основі їх змісту та контексту.

Семантичний (змістовний) аналіз тексту – виділення семантичних відношень, формування семантичного уявлення. В загальному випадку семантичне уявлення є графом, семантичної мережею, що відбиває бінарні відношення між двома вузлами – змістовними одиницями тексту.

Глибина семантичного аналізу може бути різною, а в реальних системах найчастіше здійснюється лише синтаксично-семантичне подання тексту або окремих речень. Семантичний аналіз застосовується, зокрема, у задачах аналізу тональності тексту (англ., Sentiment Analysis), наприклад, для автоматизованого визначення позитивності чи негативності відгуків.

Необхідно зауважити, що сьогодні на ринку присутня достатня кількість як умовно безкоштовних, так і платних систем для здійснення контент-аналізу, зокрема, Google Analytics, Word Stat, Rapid Miner Text Extension, Meaning Cloud, Atlas.ti, MAXQDA, Text2data, SAS Text Miner тощо [15-17].

Переважає більшість наявних програмних сервісів та систем контент-аналізу дозволяють виконувати кількісний аналіз (та певні елементи якісного аналізу) змісту як традиційних, так і соціальних медіа: досліджувати частоту згадок, визначати тональність, здійснювати розподіл аналізованих тем тощо.

Серед найбільш поширених завдань, які дозволяють вирішувати ці системи, можна вказати такі:

- ✓ автоматичне виявлення ключових слів у документах;
- ✓ парсінг веб-сторінок;
- ✓ автоматичне відстеження повідомлень у соціальних мережах та веб-ЗМІ на задану тематику;
- ✓ виявлення іменованих сутностей (персон, назв брендів та компаній, географічних тегів);
- ✓ здійснення аналізу тональності документів (сентимент-аналіз);
- ✓ автоматичне транскрибування аудіо-, відеоматеріалів з високою точністю розпізнавання тощо.

Але, на нашу думку, незважаючи на різноманіття готових програмних рішень, розроблення користувацьких автоматизованих систем із використанням технологій машинного навчання має свої переваги. По-перше, універсальні системи мають меншу ефективність обробки контенту, зокрема, текстового, ніж системи, розроблені під певний домен або певні прикладні задачі. Це пов'язано із специфікою лексики та стилістики певних предметних галузей та платформ: сайтів політичних, фінансових, спортивних новин; аналізу певних галузевих ринків у традиційних та соціальних медіа; аналізу постів та твітів у соціальних мережах тощо.

По-друге, на ринку є достатньо обмежена лінійка інструментів для роботи із українською мовою.

Альтернативою до наявних сервісів контент-аналізу є побудова користувацьких систем на основі таких потужних переднавчених мовних моделей, як BERT від компанії Google [18], або GPT-3 від Open AI [19], які було реалізовано на архітектурі глибоких нейронних мереж типу трансформер [20].

Застосування архітектури трансформер, що ґрунтується на незалежному використанні кодувальника та декодера (які, в свою чергу, є також глибокими нейронними мережами), мало величезний вплив для вирішення багатьох задач автоматичної обробки природної мови. Застосування цих компонентів передбачає попереднього навчання моделі на великому обсязі нерозмічених даних та донавчання під конкретне завдання обробки текстів.

У загальному випадку, під переносом знань (англ., Transfer Learning) мається на увазі процес навчання нової моделі на новому завданні з використанням моделі, навченої раніше на схожій задачі [14]. Наприклад, «знання», отримане під час навчання передбачення наступного (або прихованого) слова багатомовної моделі, може бути використане для навчання одномовної моделі (для мови, на якій багатомовна модель була вже навчена раніше), або, в більш загальному випадку, перенесення «знань» з однієї групи мов на іншу.

Таке перенесення знань на практиці реалізується таким чином: як початкова ініціалізація ваг нової моделі (глибокої нейронної мережі) беруться ваги моделі, навченої раніше на схожій задачі, та проводиться навчання на новому завданні.

Тому перспективним підходом, з нашої точки зору, є застосування технології Transfer Learning для перенесення знань із переднавчених мовних моделей на інший домен або іншу мову, зокрема, українську з інших слов'янських мов.

Висновки. Роль соціальних медіа в сучасному світі величезна – вони замінюють або зливаються з класичними медіа: телебачення поступово переходить у Youtube, радіо перетворюється на стрімінгові майданчики з подкастами, новини подають у телеграм-каналах та агрегаторах, а речі купують у маркетплейсах. І на всіх цих платформах ми можемо спілкуватися чи впливати на контент: ставити лайки, писати коментарі чи додавати свої публікації.

Накопичення достатнього обсягу навчальних даних, розробка багатоядерних процесорів та графічних карт, а також формування потужних преднавчених мовних моделей та розробка ефективних алгоритмів обробки надвеликих обсягів інформації – фактори, що зумовлюють популярність застосування технології машинного та глибокого навчання для завдань контент-аналізу в останні роки.

Але необхідно зауважити, що застосування комп'ютерних систем обробки природної мови не гарантує досягнення бажаного результату за замовчанням. Якість обробки обумовлюється багатьма чинниками – починаючи з приналежності текстів до певної лінгвістичної групи, і до особливостей вимови, якщо йдеться про голосовий формат.

До проблемних моментів, що зустрічаються на практиці, відносяться:

- ✓ розкриття анафор, тобто змістовної інтерпретації займенників, що використовуються;
- ✓ присутність у формулюваннях сленгу або неологізму, що не є друкарською помилкою, та має для автора та споживача контенту цілком конкретний зміст;
- ✓ використання вільного порядку розташування слів, що впливає на тлумачення фрази в цілому в українській та інших мов слов'янської групи, що частково компенсується морфологією та синтаксисом, які стають додатковим навантаженням для системи;
- ✓ інтерпретація омонімів, що належать до фонетичної категорії.

Підсумовуючи, зауважимо, що розроблення комп'ютерних систем контент-аналізу соціальних медіа, зокрема, із використанням сучасних технологій штучного інтелекту (машинного та глибокого навчання), не втрачає актуальності та потребує подальших досліджень.

Список використаних джерел

1. Yahoo Finance. URL: <https://finance.yahoo.com> (дата звернення 15.09.2022).
2. Krippendorff K. Content Analysis: An introduction to its methodology. London: Sage, 1980.
3. Костенко Н., Іванов В. Досвід контент-аналізу: Моделі та практики. Київ: Центр вільної преси, 2003. 200 с.
4. Таршис А. Е. Контент-анализ: Принципы методологии. (Построение теоретической базы. Онтология, аналитика и феноменология текста. Программа исследования). м.: Книжный дом ЛИБРОКОМ, 2014. 182 с.
5. Іванов В.Ф., Костенко Н.В. Контент-аналіз. Велика українська енциклопедія. URL: <https://vue.gov.ua/Контент-аналіз> (дата звернення: 15.09.2022).
6. Берко А. Ю. Системи електронної контент-комерції: монографія / А. Ю. Берко, В. А. Висоцька, В. В. Пасічник. Львів: Вид-во Нац. ун-ту "Львівська політехніка", 2009. 612 с.
7. Войтович О. П., Буда А. Г., Головенько В. О. Дослідження методів аналізу соціальних мереж як середовища інформаційних війн. URL: <https://epsi.vntu.edu.ua/uploads/2017/76-86yucc0hnc6o8o3xgkr97hrynqd5m0obr.pdf> (дата звернення 15.09.2022).
8. Кісь Я. П., Висоцька В. А., Чирун Л. Б. Застосування контент-аналізу для опрацювання текстових масивів даних. *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. 2015. Вип. 814. С. 282-292.
9. Фольтович В., Коробчинський М., Чирун Л., Висоцька В. Метод контент-аналізу текстової інформації Інтернет-газети. *Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології*. 2017. Вип. 864. С. 7-19.

10. Ахмедов Р.Р., Безкоровайний В.С., Данильченко Т.В. Методологія аналізу контенту електронних засобів масової інформації. Економічний простір. 2021. Вип.176. С. 141-145. DOI: <https://doi.org/10.32782/2224-6282/176-25>
11. Digital 2022 Global Overview Report. URL: <https://datareportal.com/reports/digital-2022-global-overview-report> (дата звернення 15.09.2022).
12. Statista. URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (дата звернення 15.09.2022).
13. Hobson, L., Cole, H., Hannes, H. Natural Language Processing in Action Understanding, analyzing, and generating text with Python. Manning Publications (P), 2019.
14. Kamath, U., Liu, J., Whitaker, J. Deep Learning for NLP and Speech Recognition. Springer Nature Switzerland AG, 2019. DOI: <https://doi.org/10.1007/978-3-030-14596-5>.
15. Predictive Analytics. Today. URL: <https://www.predictiveanalyticstoday.com/top-qualitative-data-analysis-software/> (дата звернення 15.11.2022).
16. Intellspot. URL: <https://www.intellspot.com/content-analysis-software/> (дата звернення 15.11.2022).
17. Bizzzdev. URL: <https://bizzzdev.com/top-10-content-analysis-tools-in-2022/> (дата звернення 15.11.2022).
18. Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT, P. 4171-4186, 2019.
19. Brown T. et al. Language models are few-shot learners. arXiv:2005.14165. 2020.
20. Vaswani, A., Shazeer, N. et al. Attention is all you need. In Proc. of the 1st Conference on Neural Information Processing Systems (NIPS 2017), 2017, pp. 6000-6010.

References

1. Yahoo Finance (2022). Retrieved from: <https://finance.yahoo.com>.
2. Krippendorff, K. (1980). *Content Analysis: An introduction to its methodology*. London: Sage.
3. Kostenko, N., Ivanov, V. (2003). *Content analysis experience: Models and practices*. Kyiv: Free Press Center (In Ukr.)
4. Tarshis, A. (2014). *Methodology principles. (Building a theoretical base. Ontology, analytics and phenomenology of the text. Research program)*. moscow, Book house LIBROKOM. (In rus.)
5. Ivanov, V., Kostenko, N. (2017). *Content analysis*. Great Ukrainian encyclopedia. Retrieved from: <https://vue.gov.ua/Контент-аналіз> (In Ukr.)
6. Berko, A. et al. (2009). *Systems of electronic content commerce: a monograph*. Lviv: Publishing House of the National University "Lviv Polytechnic" (In Ukr.)
7. Voytovich, O. et al. (2018). *Study of methods of analysis of social networks as an environment of information wars*. Retrieved from: <https://epsi.vntu.edu.ua/uploads/2017/76-86ycc0hnc6o8o3xgkr97hrynqd5m0obr.pdf> (In Ukr.)
8. Kys, Ya. et al. (2015). Application of content analysis for the processing of text corpus of data. *Bulletin of the Lviv Polytechnic National University. Information systems and networks*, 814, 282-292. Retrieved from: http://nbuv.gov.ua/UJRN/VNULPICM_2015_814_28 (In Ukr.)
9. Foltovych, V, et al. (2017). Method of content analysis of textual information of Internet newspapers. *Bulletin of the Lviv Polytechnic National University. Computer science and information technology*, 864, 7-19. Retrieved from: http://nbuv.gov.ua/UJRN/VNULPKNIT_2017_864_4 (In Ukr.)
10. Akhmedov, R. at al. (2021). Methodology of content analysis of electronic mass media. *Economic space*, 176, 141-145. DOI: <https://doi.org/10.32782/2224-6282/176-25>
11. Digital 2022 Global Overview Report. Retrieved from: <https://datareportal.com/reports/digital-2022-global-overview-report>
12. Statista. Retrieved from: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users>
13. Hobson, L., Cole, H., Hannes, H. (2019). *Natural Language Processing in Action Understanding, analyzing, and generating text with Python*. Manning Publications (P)
14. Kamath, U., Liu, J., Whitaker, J. (2019). *Deep Learning for NLP and Speech Recognition*. Springer Nature Switzerland AG 2019. <https://doi.org/10.1007/978-3-030-14596-5>
15. Predictive Analytics. Today. Retrieved from: <https://www.predictiveanalyticstoday.com/top-qualitative-data-analysis-software>
16. Intellspot. Retrieved from: <https://www.intellspot.com/content-analysis-software>
17. Bizzzdev. Retrieved from: <https://bizzzdev.com/top-10-content-analysis-tools-in-2022>

18. Devlin, J., Chang, M-W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2
19. Brown, T., W. et al. (2020). Language models are few-shot learners. arXiv:2005.14165
20. Vaswani, A., Shazeer, N. et al. (2017). Attention is all you need. *In Proc. of the 1st Conference on Neural Information Processing Systems (NIPS 2017)*

AKHMEDOV Renat¹

asistent professor

DERBENTSEV Vasy¹

PhD (Economics), Associate Professor

BEZKOROVAINYI Vitalii¹

PhD (Economics)

¹ Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine

CONTENT ANALYSIS OF SOCIAL MEDIA USING MACHINE LEARNING

Introduction. During the last two decades, the rapid development of social media has caused a revolution in means of communication in modern society. Therefore, a significant number of the world's leading companies began to rebuild their business models using the capabilities of modern means of communication through social networks and other platforms, for which content analysis technologies are successfully used.

Purpose. The purpose of the article is to develop methodological principles for conducting content analysis of electronic resources (social media) based on using Artificial Intelligence technologies, in particular, Machine and Deep Learning.

Results. The paper analyzes the phenomenon of social media and identifies the key factors that determine the effectiveness of their use for both business and consumers. Based on this, the paper explores the features of the content analysis of social media, which take into account their mass character, as well as the presence of large arrays of unstructured information. The large amount of information on various electronic platforms requires adequate means for their monitoring and processing, analysis of content as well. To solve these problems, the paper substantiates the use of modern Natural Language Processing technologies based on Machine and Deep Learning approaches. An alternative to the existing services for content analysis is developing systems based on such tight forwarding motion models, like BERT (provided by Google), or GPT-3 (provided by Open AI), which was implemented on the Transformer Deep Neural Networks architecture. The article also proposes the use of Transfer Learning technology to transfer knowledge from pre-learned language models to another domain or another language, in particular, Ukrainian from other Slavic languages.

Originality. The main findings of this paper are the follows: (i) the main advantages of using social media for businesses and consumers are substantiated; (ii) the characteristic features of conducting content analysis in social media are determined; (iii) the advantages and disadvantages of using Natural Language Processing methods for solving problems of content analysis in social media are shown; (iv) Transfer Learning approach to transfer knowledge from pre-learned language models to another domain or another language, in particular, Ukrainian from other Slavic languages has been proposed for solving content analysis tasks.

Conclusion. The accumulation of a sufficient amount of training data, the development of multi-core CPU and graphics processors, as well as the formation of powerful pre-trained language models and the development of effective algorithms for processing extremely large amounts of information are factors that determine the efficiency of the use of Machine and Deep Learning technology for content analysis tasks in recent years. Therefore, the development of computer systems for content analysis of social media, in particular, using modern technologies of Artificial Intelligence (Machine and Deep Learning), does not lose its relevance and requires further research.

Keywords: content analysis, social media, Machine Learning, Deep Learning.

Одержано редакцією: 20.09.2022
Прийнято до публікації: 24.11.2022