

## МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 303.316:004.912

DOI: <https://doi.org/10.32782/2224-6282/176-25>**Ахмедов Р. Р.**

асистент,

Київський національний економічний університет імені Вадима Гетьмана

ORCID: <https://orcid.org/0000-0002-3084-4672>**Безкоровайний В. С.**

кандидат економічних наук,

Київський національний економічний університет імені Вадима Гетьмана

ORCID: <https://orcid.org/0000-0002-4998-8385>**Данильченко Т. В.**

кандидат технічних наук, доцент,

Київський національний економічний університет імені Вадима Гетьмана

ORCID: <https://orcid.org/0000-0001-6174-946X>**Akhmedov Renat, Bezkorovainyi Vitalii, Danylchenko Tetiana**

Kyiv National Economic University named after Vadym Hetman

### МЕТОДОЛОГІЯ АНАЛІЗУ КОНТЕНТУ ЕЛЕКТРОННИХ ЗАСОБІВ МАСОВОЇ ІНФОРМАЦІЇ

*В сучасному інформаційному суспільстві особливого значення набуває аналіз потоку інформації ЗМІ, тому особливої актуальності набувають питання проведення контент-аналізу, який ґрунтується на переведенні якісних даних у цифровий формат задля одержання можливостей оброблення інформації з використанням технологій біг-дата-аналітики та штучного інтелекту. У рамках дослідження узагальнено погляди на використання контент-аналізу та запропоновано методологію аналізу контенту електронних засобів масової інформації для таргетування цільової аудиторії, визначення вподобань щодо змісту публікацій, що складається з таких етапів, як вебскрапінг, паркінг, якісний аналіз, кількісний аналіз, аналіз тональності текстів. Запропонована в дослідженні методологія є основою для реалізації сучасних засобів контент-аналізу з використанням сучасних інформаційних технологій та методів і моделей штучного інтелекту.*

**Ключові слова:** контент-аналіз, засоби масової інформації, великі дані, парсінг, аналіз тональності тексту.

### METHODOLOGY OF ANALYSIS OF THE CONTENT OF ELECTRONIC MEDIA

*In the conditions of modern information society, the analysis of the flow of media information is of particular importance. Therefore, the issues of content analysis, which is based on the conversion of qualitative data into digital format in order to obtain information processing capabilities using technologies of big-data analytics and artificial intelligence approaches, are especially relevant. This study summarized the views on the use of content analysis and proposed a methodology for analyzing the content of electronic media for targeting and determining preferences for the content of publications. For this purpose, were proposed were proposed a conceptual scheme for conducting content analysis, which consists of 5 stages. The first stage consists of creating and updating a full-text database on a daily basis by automated data collection and retrieval of content from media websites (web scraping). The second stage involves using of modules for determining useful content (parsing) and its subsequent storage in the database. The quantitative analysis stage is especially important when generalization, aggregation of results, or to calculate the frequency of words (frequency analysis), categories, tags and expressions, as well as their joint use in the text. The purpose of qualitative analysis is to reduce the material so that the main content remains. Similarly, qualitative analysis can provide additional information about individual components of the text (terms, definitions) in order to improve understanding, explanation, and interpretation of a particular passage. Finally, the last stage is to analyze the tone of the texts (sentiment analysis), i.e. to determine the emotional (positive or negative) attitude of the authors of the text to the described event, person or phenomenon. Authors (or users who leave comments) actively use various means in order to create the desired picture of an event in the reader's imagination, and to evoke the appropriate emotions. The proposed methodology is the basis for the implementation of modern tools of content analysis using information technologies and methods artificial intelligence approaches.*

**Keywords:** content analysis, mass-media, big data, parsing, sentiment analysis.

**JEL classification:** C80, L82

**Постановка проблеми.** Внаслідок бурхливого впровадження цифровізації традиційні засоби масової інформації повинні адаптуватись до нових реалій цифрового суспільства. При цьому медіабізнес має трансформуватись до інтегрованої моделі, яка здатна задовольнити потреби нових споживчих звичок усе більш вимогливих користувачів.

Медіаіндустрія та ЗМІ розвиваються з безпрецедентною швидкістю, керуючись подвійними потребами щодо зниження операційних витрат і одночасного отримання більшого доходу від дедалі зростаючої конкуренції з боку соціальних мереж та інших сучасних цифрових засобів поширення інформації, а також невизначеністю ринків. Медіакомпанії багато в чому є одними з перших, хто взяв на озброєння технології великих даних, оскільки це дає їм змогу керувати цифровою трансформацією, більш повно використовувати не лише дані, які вже були доступні, але й нові джерела даних як всередині організації, так і за її межами.

Використовуючи великі дані, розширену аналітику та штучний інтелект, медіакомпанії можуть скористатися потенціалом новітніх технологій, щоби покращити знання про вподобання споживачів та оптимізувати свої стратегії продажів, реклами та контенту. Вони також можуть розробляти нові послуги, цифрові медіа та програми для створення нових можливостей для бізнесу та способів комунікації зі своєю аудиторією.

Процес створення великих даних складається з отримання, класифікації та збереження інформації, при цьому контент може бути представлений у різних формах, а саме якісній (текстові дані, зображення), кількісній (числові дані) та комбінованій. Перевести якісні дані у цифровий формат та в подальшому маніпулювати з інформацією у матричній формі дає змогу контент-аналіз.

Таким чином, контент-аналіз – це дуже гнучкий за своєю природою метод. Його можна застосовувати до різноманітних типів неструктурованої інформації.

**Аналіз останніх досліджень і публікацій.** Серед зарубіжних науковців сьогодні існує багато дослідників у галузі контент-аналізу. Так, одним із перших визначив контент-аналіз як відтворюваний та дієвий метод для формування висновків із контексту інформаційних джерел Клаус Кріппендорф [1]. В подальших дослідженнях він відзначив, що у ході еволюції контент-аналізу зростає не тільки рівень технічної та наукової складності процесу, але й те, що цей підхід набуває все більшого поширення в різноманітних галузях науки та практики, які використовують будь-які дані для аналізу, у тому числі соціально-економічної та психологічної спрямованості [2].

У своєму дослідженні Стівен Лейсі, Деніел Ріфф, Стейсі Стоддарт, Х'ю Мартін та Куанг-Куо Чанг [3] запропонували методіку контент-аналізу друкованих щоденних ЗМІ. На основі твердження, що медіа-контент, як і актуальність інформації, безперервно змінюється, визначено оптимальні проміжки часу та обсяги вибірки даних для проведення кількісного аналізу на основі текстових даних.

У статті Льоса Олдерінга та Ренса Флігентхарта [4] досліджується застосування комп'ютерного аналізу контенту для вивчення шаблонів політичного лідерства. У статті пропонується методологія вимірювання дванадцяти таких шаблонів або образів політичного

лідерства у висвітленні ЗМІ на основі широкомасштабного кількісного контент-аналізу голландських ЗМІ.

У монографії Філіп Мейрінг [5] критикує методологічну дихотомізацію якісного та кількісного дослідження та визначає якісний аналіз даних як підхід змішаних методів (містить як якісні, так і кількісні етапи аналізу), а також пропонує покрокову модель процесу (якісно-кількісного) дослідження контенту.

Удо Кукарц [6] визначає тематичний аналіз як один із найбільш часто використовуваних методів аналізу якісних даних та надає основи цього систематичного методу, висвітлює ключові характеристики та типовий аналітичний процес. Особлива увага у роботі приділяється визначенню та формуванню категорій.

Пол Скальські, Кімберлі Нойендорф та Джулі Каджігас [7] розглядають вплив інтерактивних медіа на методіку аналізу контенту через те, що соціальні медіа швидко замінюють традиційні засоби масової інформації та способи спілкування, такі як газети, журнали, телебачення і навіть традиційний телефон.

Загалом українські науковці не приділяють питанню контент-аналізу достатньо уваги, але слід виділити О.В. Іванова, який у роботі [8] визначає відмінності між концептуально-методологічними засадами методу аналізу контенту та методиками й техніками кількісної обробки текстових даних, обґрунтовує необхідність оновлення вимог до контент-аналітичного дослідження із застосуванням комп'ютерних технологій, а у дослідженні [9] розглядає методологічну цінність, переваги та недоліки контент-аналізу з використанням комп'ютерних програм.

У роботі Д.Ю. Трепачко [10] досліджується специфіка застосування методу контент-аналізу під час дослідження новинних повідомлень засобів масової інформації.

У публікаціях згаданих авторів немає чітко сформованої моделі процесу отримання, оброблення, систематизації та аналізу даних, тому подальше дослідження має ґрунтуватись на мисленні, технічних рішеннях та стратегії великих даних, які дають можливість керувати даними та поширювати їх на більший швидкості та у більших масштабах.

**Мега статті** полягає в розробленні методології аналізу контенту електронних засобів масової інформації для таргетування цільової аудиторії, визначення вподобань щодо змісту публікацій, підвищення їх привабливості та утримання читача.

**Виклад основного матеріалу.** Для аналізу змісту медіаконтенту важливо відзначити, що користувачі сучасних електронних ЗМІ – це більше, ніж просто одержувачі чи споживачі інформації. Вони відіграють активну роль у адаптації, зміні та навіть створенні контенту. Інтерактивна медіареволюція, яка почалася з відеоігор у 1970-х роках, продовжилася з домашніми комп'ютерами у 1980-х роках та Інтернетом у 1990-х роках, а на початку XXI століття виростає і переросла у Web 2.0 [7].

Відповідно, контент-аналіз має розглядатись як багатогранний системний процес, який значно ширше за інструменти для отримання, класифікації та збереження великих даних.

Так, контент-аналіз визначається нами як метод із чітко визначеними етапами та процесами (рис. 1).

Перший процес – це створення та щоденне доповнення бази даних повнотекстової інформації. Цей

процес починається з автоматизованого збору даних та отримання контенту з вебсайтів ЗМІ (вебскрапінг). Скрапінг – це послідовний перебір сторінок сайту та їх завантаження, для чого нами використовуються відповідні програмні засоби, які включають модулі визначення корисного контенту (парсінг) та його подальшого збереження у базі даних.

З кожним роком парсінг вебсторінок стає складнішим. На відміну від перших вебсторінок, коли вони були простими конструкціями HTML, сучасні сайти мають багато різноманітних, а інколи і динамічних елементів, які не можуть бути зафіксовані за допомогою функції збереження сторінки у браузері, тому в програмному забезпеченні доводиться використовувати різні варіанти парсінгу сторінок, які поки ще працюють із різним ступенем ефективності залежно від типу сторінки та її вмісту.

Окрім парсінгу та збереження основної текстової інформації, ми зберігаємо теги та інші мета-дані сторінки, які в подальшому можуть бути корисними як інструменти для аналізу контенту. Коментарі до публікації, їх кількість та емоційне забарвлення додають можливість у подальшому проводити аналіз тональності тексту. Такий контент створюється окремим користувачем або редактором сайту і допомагає визначити подібний вміст для подальшого кодування.

Наприклад, цифрові камери вбудовують у фотографії мета-дані, окрім часу, вони можуть включати тип камери, налаштування камери та дані геолокації. Так само ЗМІ для кожної публікації можуть створювати теги, щоб допомогти класифікувати свої публікації на тематичні групи, даючи змогу читачам або контент-аналітикам легко орієнтуватися в їхньому вмісті.

Другий процес – це безпосередньо аналіз даних, отримання результатів дослідження змісту контенту та його тональності.

Етап кількісного аналізу має особливе значення, коли потрібні узагальнення, агрегування результатів для підрахунку частоти появи слів (частотний аналіз), категорій, тегів і виразів, а також їх спільного використання у тексті. Такий аналіз легко адаптувати до специфіки великих даних, оскільки підрахунок частотності не обов'язково потребує участі людини.

Кількісний контент-аналіз виділяє такі одиниці [1], які також можна використовувати для якісного контент-аналізу:

- одиниця кодування, що визначає найменший компонент матеріалу, який можна оцінити, а також мінімальну частину тексту, яка може входити в одну категорію;
- блок контексту, що визначає найбільший текстовий компонент, який може входити в одну категорію;
- блок запису, що визначає, які частини тексту стикаються з однією системою категорій.

На етапі кількісного аналізу важливо визначити, як певний контент повторюється в подібній формі і з певною частотою. Однак у системі контент-аналізу реєстрація того, як часто зустрічається та чи інша категорія або тег, також може надати додаткову вагу їх значенню та важливості. Звичайно, у кожному випадку це має бути належним чином обґрунтовано. Подальше точне якісне присвоєння категорій певному матеріалу також може бути доповнене більш складними методами статистичної оцінки, якщо вони відповіда-

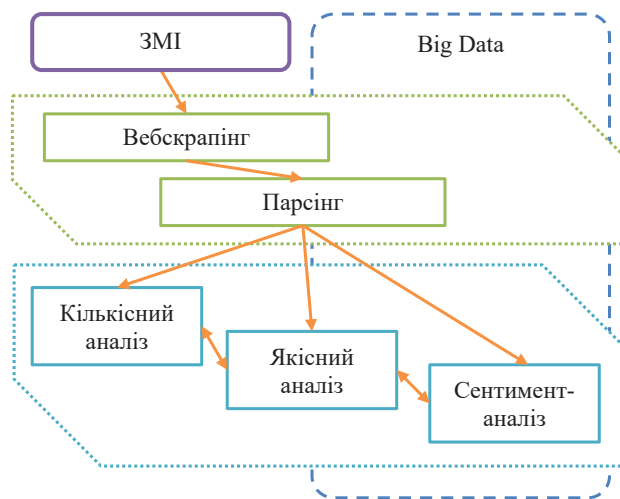


Рис. 1. Аналіз контенту засобів масової інформації

Джерело: розроблено авторами

ють меті аналізу та підходять для залученого об'єкта. У цьому разі поєднання якісного та кількісного етапу аналізу надають особливого змісту всьому процесу [5].

Якісні дані набагато різноманітніші та складніші, ніж кількісні дані, тому що вихідною точкою якісного аналізу виступає окремий текстовий компонент, який необхідно проаналізувати (наприклад, його текстовий контекст). Проведення оцінювання текстового фрагменту починається зі встановлення його відношення до інших текстових компонентів і, як правило, має на меті узагальнення матеріалу. Метою такого аналізу є скорочення матеріалу таким чином, щоб основний зміст залишився. Так само якісний аналіз може надавати додатковий матеріал про окремі компоненти тексту (терміни, визначення) задля покращення розуміння, пояснення, інтерпретації конкретного уривку тексту.

Однак найчастіше якісний аналіз застосовується для структурування, коли потрібно відфільтрувати окремі аспекти текстових даних, надати поперечний розріз контенту відповідно до заздалегідь визначених критеріїв упорядкування або оцінити його за певними критеріями.

Останнім етапом, але не менш важливим є аналіз тональності текстів (сентимент-аналіз), тому що автори текстів (або користувачі, які залишають коментарі) активно використовують різні засоби задля створення потрібної картини тієї чи іншої події в уяві читача та виклику відповідних емоцій.

Існує значна кількість систем автоматичного оцінювання емоційного забарвлення тексту, що базуються на таких методах:

- 1) метод, що ґрунтується на правилах, полягає у використанні наборів правил на основі аналізу предметної галузі;
- 2) метод, що ґрунтується на застосуванні тональних словників;
- 3) навчання з учителем, за якого алгоритм тренується на заздалегідь налаштованому наборі текстів;
- 4) навчання без учителя, за якого алгоритм тренується на довільному наборі текстів і самостійно визначає вагу окремих слів у контексті їх застосування.

Більшість систем для проведення сентимент-аналізу є потужними комерційними продуктами (напри-

клад, Semantrum) [11], але існують системи у вільному доступі, які дають змогу проводити аналіз тональності невеликих текстових масивів. Однією з таких програм є система сентимент-аналізу Semantria for Excel [12]. Система є надбудовою для програми Microsoft Excel і дає змогу класифікувати тональність повідомлень багатьма мовами, але підтримка української мови відсутня. Однією з переваг цієї системи є те, що загальна тональність тексту визначається не просто як позитивна, негативна та нейтральна, а у діапазоні від -1,5 до +1,5.

**Висновки.** В сучасному інформаційному суспільстві особливого значення набуває аналіз потоку інформації ЗМІ. У рамках цього дослідження були узагальнені погляди на використання контент-аналізу.

Основне застосування цього методу полягає у переведенні текстової інформації в кількісний формат, який дає можливість більш глибокого аналізу з використанням сучасних методів біг-дата-аналітики та штучного інтелекту. Це є необхідною вимогою під час роботи з

матеріалами ЗМІ, які мають різну тональність текстів стосовно однакових об'єктів дослідження, при цьому аналіз медіаконтенту значною мірою може змінюватись від суб'єктивної інтерпретації кожного дослідника.

Процес аналізу контенту необхідно будувати на поєднанні якісних та кількісних методів, а головне завдання полягає в тому, щоб визначити ті точки або аспекти аналітичного процесу, до яких можна ввести кількісні показники.

Як інструменти сентимент-аналізу потрібно використовувати потужніші україномовні системи, які дають змогу застосовувати нейро-нечітку логіку та більш точно налаштування параметрів (тюнінг) (наприклад, NLTK у середовищі Python).

Підсумовуючи, зауважимо, що запропонована в дослідженні методологія є основою для реалізації сучасних засобів контент-аналізу з використанням сучасних інформаційних технологій, методів і моделей штучного інтелекту.

### Список використаних джерел:

1. Krippendorff K. Validity in content analysis. *Computerstrategien fur die kommunikationsanalyse*. Frankfurt, Germany : Campus, 1980. P. 69–112. URL: [http://repository.upenn.edu/asc\\_papers/291](http://repository.upenn.edu/asc_papers/291) (дата звернення: 06.12.2012).
2. Krippendorff K. Content analysis. *International encyclopedia of communication*. New York, NY : Oxford University Press, 1989. Vol. 1. P. 403–407. URL: [http://repository.upenn.edu/asc\\_papers/226](http://repository.upenn.edu/asc_papers/226) (дата звернення: 06.12.2012).
3. Lacy S., Riffe D., Stoddard S., Martin H., Chang K. Sample Size for Newspaper Content Analysis in Multi-Year Studies. *Journalism & Mass Communication Quarterly*. 2001. Vol. 78. No. 4. P. 836–845. DOI: <https://doi.org/10.1177/107769900107800414>
4. Aaldering L., Vliegthart R. Political leaders and the media. Can we measure political leadership images in newspapers using computer-assisted content analysis? *Qual Quant*. 2016. Vol. 50. P. 1871–1905. DOI: <https://doi.org/10.1007/s11135-015-0242-9>
5. Mayring P. Qualitative content analysis: theoretical foundation, basic procedures and software solution. *Klagenfurt*. 2014. URL: <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-395173> (дата звернення: 08.12.2012).
6. Kuckartz U. Qualitative Text Analysis: A Systematic Approach. *Compendium for Early Career Researchers in Mathematics Education, ICME-13 Monographs*. 2019. DOI: [https://doi.org/10.1007/978-3-030-15636-7\\_8](https://doi.org/10.1007/978-3-030-15636-7_8)
7. Skalski P.D., Neuendorf K.A., Cajigas J.A., Content Analysis in the Interactive Media Age. *The Content Analysis Guidebook*. 2017. URL: <https://academic.csuohio.edu/kneuendorf/SkalskiVita/SkalskiNeuendorfCajigas17.pdf> (дата звернення: 12.12.2012).
8. Іванов О.В. Класичний контент-аналіз та аналіз тексту: термінологічні та методологічні відмінності. *Вісник Харківського національного університету імені В.Н. Каразіна*. 2013. № 1045. С. 69–74. URL: [http://ekmair.ukma.edu.ua/bitstream/handle/123456789/7244/VKHSID\\_2013\\_1045\\_30\\_13.pdf](http://ekmair.ukma.edu.ua/bitstream/handle/123456789/7244/VKHSID_2013_1045_30_13.pdf) (дата звернення: 16.12.2012).
9. Іванов О.В. Комп'ютерний контент-аналіз: проблеми та перспективи вирішення. *Методологія, теорія та практика соціологічного аналізу сучасного суспільства*. 2009. Вип. 15. С. 335–340. URL: <http://ekmair.ukma.edu.ua/bitstream/handle/123456789/2407/Ivanov%20komjuteruy%20analiz.pdf> (дата звернення: 16.12.2012).
10. Трєпачко Д.Ю. Особливості застосування контент-аналізу в дослідженнях новинних повідомлень ЗМІ. *Вісник Національного технічного університету України «Київський політехнічний інститут»*. 2010. № 1(5). С. 87–91. URL: <https://ela.kpi.ua/bitstream/123456789/6048/1/10-1%285%29-15.pdf> (дата звернення: 16.12.2012).
11. Semantrum. Онлайн-сервіс для моніторингу традиційних та Інтернет-ЗМІ. URL: <https://promo.semantrum.net/uk/golovna> (дата звернення: 22.12.2012).
12. Semantria for Excel. Text and sentiment analysis for surveys, social media, and reviews. URL: <https://www.lexalytics.com/semantria/excel> (дата звернення: 22.12.2012).

### References:

1. Krippendorff, K. (1980) Validity in content analysis. *Computerstrategien fur die kommunikationsanalyse*. Frankfurt, Germany: Campus, pp. 69–112. Available at: [http://repository.upenn.edu/asc\\_papers/291](http://repository.upenn.edu/asc_papers/291) (accessed 6 December 2021).
2. Krippendorff, K. (1989) Content analysis. *International encyclopedia of communication*. New York, NY: Oxford University Press, vol. 1, pp. 403–407. Available at: [http://repository.upenn.edu/asc\\_papers/226](http://repository.upenn.edu/asc_papers/226) (accessed 6 December 2021).
3. Lacy, S., Riffe, D., Stoddard, S., Martin, H., and Chang, K. (2001) Sample Size for Newspaper Content Analysis in Multi-Year Studies. *Journalism & Mass Communication Quarterly*, vol. 78, no. 4, pp. 836–845. DOI: <https://doi.org/10.1177/107769900107800414>
4. Aaldering, L., Vliegthart, R. (2016) Political leaders and the media. Can we measure political leadership images in newspapers using computer-assisted content analysis? *Qual Quant*, vol. 50, pp. 1871–1905. DOI: <https://doi.org/10.1007/s11135-015-0242-9>
5. Mayring, P. (2014) Qualitative content analysis: theoretical foundation, basic procedures and software solution. *Klagenfurt*. Available at: <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-395173> (accessed 8 December 2021).
6. Kuckartz, U. (2019) Qualitative Text Analysis: A Systematic Approach. *Compendium for Early Career Researchers in Mathematics Education, ICME-13 Monographs*. DOI: [https://doi.org/10.1007/978-3-030-15636-7\\_8](https://doi.org/10.1007/978-3-030-15636-7_8)
7. Skalski, P.D., Neuendorf, K.A., Cajigas, J.A. (2017) Content Analysis in the Interactive Media Age. *The Content Analysis Guidebook*. Available at: <https://academic.csuohio.edu/kneuendorf/SkalskiVita/SkalskiNeuendorfCajigas17.pdf> (accessed 12 December 2021).
8. Ivanov, O.V. (2013) Klasychnyi kontent-analiz ta analiz tekstu: terminolohichni ta metodolohichni vidminnosti [Classical content analysis and text analysis: terminological and methodological differences]. *Visnyk Kharkivskoho natsionalnoho uni-*

versytetu imeni V.N. Karazina, no. 1045, pp. 69–74. Available at: [http://ekmair.ukma.edu.ua/bitstream/handle/123456789/7244/VKhISD\\_2013\\_1045\\_30\\_13.pdf](http://ekmair.ukma.edu.ua/bitstream/handle/123456789/7244/VKhISD_2013_1045_30_13.pdf) (accessed 16 December 2021).

9. Ivanov, O.V. (2009) Kompiuternyi kontent-analiz: problemy ta perspektyvy vyryshennia [Computer Content Analysis: Problems and Prospects]. *Metodolohiia, teoriia ta praktyka sotsiologichnoho analizu suchasnoho suspilstva*, vol. 15, pp. 335–340. Available at: <http://ekmair.ukma.edu.ua/bitstream/handle/123456789/2407/Ivanov%20komputernyi%20analiz.pdf> (accessed 16 December 2021).

10. Trepachko, D.Iu. (2010) Osoblyvosti zastosuvannia kontent-analizu v doslidzhenniakh novynnykh povidomlen ZMI [Features of the application of content analysis in the study of news media reports]. *Visnyk Natsionalnoho tekhnichnoho universytetu Ukrainy "Kyivskiy politekhnichnyi instytut"*, no. 1(5), pp. 87–91. Available at: <https://ela.kpi.ua/bitstream/123456789/6048/1/10-1%285%29-15.pdf> (accessed 16 December 2021).

11. Semantrum. AI-platform for media analytics and reputation management. Available at: <https://promo.semantrum.net/en/main> (accessed 22 December 2021).

12. Semantria for Excel. Text and sentiment analysis for surveys, social media, and reviews. Available at: <https://www.lexalytics.com/semantria/excel> (accessed 22 December 2021).