УДК 303.71:[316.77:004.77](477)
JEL: C46; L82

СТАТИСТИЧНІ МЕТОДИ АНАЛІЗУ ДАНИХ СОЦІАЛЬНИХ МЕДІА

©2020 ТУМАНОВ О. О.

УДК 303.71:[316.77:004.77](477)
JEL: C46; L82

Туманов О. О. Статистичні методи аналізу даних соціальних медіа

Завдяки невпинному поширенню інтернет-технологій двосторонній комунікаційний зв'язок між людьми стає все сильніший і розповсюджується не тільки в конкретному місці, але й у всіх куточках нашого світу. Збільшення використання соціальних медіа генерує великі обсяги даних та нові їх типи, які раніше були не доступні. Поняття «соціальні мережі» та «соціальні медіа» все частіше є частиною соціальних дискусій, організаційної стратегії та наукових досліджень. Зростаючий інтерес до соціальних мереж поєднується із поширенням загальнодоступних мережевих даних, спостерігається суттєве збільшення методів та розуміння того, як аналізувати дані соціальних мереж і де можна використовувати результати таких досліджень. Використання цих даних може бути ключовим у сучасній соціології та мати велике значення у сферах економіки, антропології, біології, демографії, комунікаційних досліджень, географії, історії, інформатики, організаційних досліджень, політології, охорони здоров'я, соціальної психології, дослідження розвитку, соціолінгвістики та інших. Зараз це широко доступний споживчий інструмент, тому зростання інтересу до аналізу та моделювання соціальних даних, а також актуальність цієї теми не можна недооцінювати. У даній статті наведено методи, застосовані для аналізу широкого кола соціальних медіа, що ґрунтуються на родинних зв'язках, діловому співробітництві, політичних альянсах тощо. У цьому контексті структура мережі та внутрішні структури інформації можуть бути визначальними. В одних випадках вони можуть мати позитивний вплив на ефективність економічного зростання, але в інших – мережева структура може призвести до деструктивних наслідків через існування безлічі зв'язків між деякими центральними суб'єктами (наприклад, поширення таких явищ, як епідемії чи фінансові кризи). Ця стаття представляє собою вступ до розгляду статистичних моделей, які спрямовані на вивчення особливостей даних соціальних медіа та наслідків соціальних процесів. Розглянуто підходи до аналізу зібраних даних, а також підіймаються питання надійності та достовірності отриманих результатів.

Ключові слова: *методи аналізу, збір даних, соціальні медіа, аналіз, соціальні мережі.*
DOI:
Рис.: *1.* ***Табл.:*** *1.* ***Бібл.:*** *18.*

Туманов Олексій Олександрович *– здобувач кафедри статистики, обліку та аудиту, Харківський національний університет ім. В. Н. Каразіна (пл. Свободи, 4, Харків, 61022, Україна)*
E-mail: *oleksii.tumanov@gmail.com*

УДК 303.71:[316.77:004.77](477)
JEL: C46; L82

Туманов А. А. Статистические методы анализа данных социальных медиа

Благодаря распространению интернет-технологий двусторонняя коммуникация между людьми становится всё сильнее и распространяется не только в определенном месте, но и во всех уголках нашего мира. Увеличение использования социальных медиа генерирует большие объемы данных и новые их типы, которые ранее были не доступны. Понятия «социальные сети» и «социальные медиа» всё чаще являются частью социальных дискуссий, организационной стратегии и научных исследований. Растущий интерес к социальным сетям сочетается с распространением широкодоступных сетевых данных, наблюдается существенное увеличение методов и понимание того, как анализировать данные социальных сетей и где можно использовать результаты таких исследований. Использование этих данных может быть ключевым в современной социологии и иметь большое значение в сферах экономики, антропологии, биологии, демографии, коммуникационных исследований, географии, истории, информатики, организационных исследований, политологии, здравоохранения, социальной психологии, исследований развития, социолингвистики и других. Сейчас это доступный потребительский инструмент, поэтому рост интереса к анализу и моделированию социальных данных, а также актуальность этой темы нельзя недооценивать. В данной статье представлены методы, применяемые для анализа широкого круга социальных медиа, основанных на родственных связях, деловом сотрудничестве, политических альянсах и т. п. В этом контексте структура сети и внутренние структуры информации могут быть определяющими. В одних случаях они могут иметь положительное влияние на эффективность экономического роста, а в других – сетевая структура может привести к деструктивным последствиям из-за существования множества связей между некоторыми центральными субъектами (например, распространение таких явлений, как эпидемии или финансовые кризисы). Эта статья представляет собой вступление к рассмотрению статистических моделей, направленных на изучение особенностей данных социальных медиа и последствий социальных процессов. Рассмотрены подходы к анализу собранных данных, а также поднимаются вопросы надежности и достоверности полученных результатов.

Ключевые слова: *методы анализа, сбор данных, социальные медиа, анализ, социальные сети.*
Рис.: *1.* ***Табл.:*** *1.* ***Бібл.:*** *18.*

Туманов Алексей Александрович *– соискатель кафедры статистики, учета и аудита, Харьковский национальный университет им. В. Н. Каразина (пл. Свободы, 4, Харьков, 61022, Украина)*
E-mail: *oleksii.tumanov@gmail.com*

UDC 303.71:[316.77:004.77](477)
JEL: C46; L82

Tumanov O. O. The Statistical Methods for Analyzing Social Media Data

Due to the spread of Internet technologies, two-way communication between people is becoming stronger and spreading not only in a certain place, but also in all corners of our world. Increased use of social media generates large amounts of data and new types of data that were not previously available. The concepts of «social networks» and «social media» are increasingly part of social discussions, organizational strategy and scientific research. The growing interest in social networks is coupled with the dissemination of widely available network data, there is a significant increase in methods and understanding of how to analyze

social media data and where the results of such researches can be used. The use of this data can be of key nature in modern sociology and be important in economics, anthropology, biology, demographics, communication research, geography, history, informatics, organizational research, political science, health, social psychology, outlook studies, sociolinguistics and others. It is an present-time affordable consumer tool, so the growing interest in the analysis and modeling of social data, as well as the relevance of this topic should not be underestimated. This article presents the methods used to analyze a wide range of social media based on relationship, business cooperation, political alliances, etc. In this context, network structure and internal information structures can be decisive. In some cases, they can have a positive impact on the efficiency of economic growth, while in others, the network structure can have destructive consequences due to the existence of many links between some central actors (e.g., the spread of phenomena such as epidemics or financial crises). This article is an introduction to consideration of the statistical models designed to examine the specifics of social media data and the effects of social processes. Approaches to the analysis of the collected data are considered, as well as issues of reliability and reliability of the results are brought up for discussion.
Keywords: *analysis methods, data collection, social media, analysis, social networks.*
Fig.: *1.* ***Tabl.:*** *1.* ***Bibl.:*** *18.*

Tumanov Oleksii O. *– Applicant of the Department of Statistics, Accounting and Auditing, V. N. Karazin Kharkiv National University (4 Svobody Square, Kharkiv, 61022, Ukraine)*
E-mail: *oleksii.tumanov@gmail.com*

Поява та розвиток соціальних медіа обумовили створення великої кількості інформації, що є потенційним ресурсом для наукових досліджень у різних напрямах соціально-економічного середовища. Аналітики можуть робити більш унікальні та цікаві дослідження з використанням даних соціальних медіа в широкому спектрі науково-дослідницьких дисциплін, таких як соціологія, інформатика, засоби масової інформації та комунікації, політологія та інженерія тощо. Платформи соціальних медіа щодня генерують величезну кількість даних на найрізноманітніші теми, а отже, є ключовим джерелом інформації для тих, хто прагне вивчити суспільство 21 століття. Однак соціальні мережі не є звичними джерелами інформації, до яких вже звикли дослідники та аналітики. Вони містять в собі множину агентів: користувачів, спільнот, різноманітних груп чи сторінок організацій, між якими відбуваються постійні взаємовідносини. Саме тип взаємовідносин і дані про них і мають додаткову велику цінність, яку необхідно вивчати та досліджувати. Ця особливість структури соціальних мереж порушує питання пошуку найбільш ефективних методів їх аналізу.

Дослідження Інтернету в цілому та соціальних медіа зокрема в нашій країні розпочалося нещодавно. Однак в міру поширення цих двох складників інтерес до їх вивчення зростає з кожним днем. Сьогодні соціальні медіа є одним із величезних джерел даних, котрі можуть бути використані в наукових роботах. Вивченням аналізу даних соціальних медіа займалися такі вітчизняні вчені, як Буда А. Г. [1], Войтович О. П. [1], Головенько В. О. [1], Горчинська О. Ю. [2], Мазуренко В. В. [3], Штовба С. Д. [3]. У своїх роботах вчені розглядають загальні методи вивчення соціальних мереж: теорію графів, структурні еквівалентності, випадкові графи та ін. Однак в міру появи нових джерел соціальних мереж їх аналіз та вивчення контенту потребують подальших досліджень. Дана тенденція обумовлює актуальність обраної теми.

Метою даної статті є обґрунтування вибору статистичних методів дослідження соціальних медіа.

Враховуючи дуже великий набір даних, що міститься в соціальних мережах, головним завданням є визначення типів даних, та підходів до їх аналізу [7].

Соціальні мережі зазвичай містять величезну кількість контенту та ознак, представлених у різних шкалах вимірювання, які можна використовувати для аналізу зв'язків між показниками. Отже, дані можна розділити відповідно на неструктуровані та структуровані, залежно від того, організовані вони заздалегідь визначеним чином (структуровані дані) чи ні (неструктуровані дані). Наприклад, вік, стать користувача, кількість часу, що він провів у мережі, є структурованими даними, тоді як дані про події, засновані на твітах і «лайках», не є структурованими.

Зазвичай структуровані дані соціальної мережі аналізуються за допомогою вбудованих елементів візуалізації, що може бути представлено, наприклад, у вигляді графіка $G = (V, E)$, де V – сукупність вузлів або сутностей (наприклад, людей, організацій та продуктів), а E – набір ребер, або зв'язків, що з'єднують вузли за допомогою моделей взаємодій [1]. Застосування графічної аналітики фокусується на отриманні інформації з взаємопов'язаних даних.

Неструктуровані дані – це дані контенту, що передаються в онлайн соціальній мережі (*OSN* – від англ. *Online Social Network*), також відомі як *UGC* (*Generated User*), тобто дані, що створені кінцевим користувачем. Вони вважаються життєвою ланкою соціальної мережі та включають текст, зображення, відео, твіти, огляди продуктів та інші мультимедійні дані, які, як правило, вивчаються за допомогою контент-аналізу [5; 13], методи якого включають, серед інших, алгоритми структурування даних.

Підходи до аналітики соціальних мереж і розробки контенту відповідають міждисциплінарним принципам штучного інтелекту (*AI* – від англ. *Artificial Intelligence*), статистики та суміжних галузей. До появи штучного інтелекту дослідники намагалися поєднати суперечливе поняття «інтелект» з технічним обладнанням, щоб зрозуміти, як працює світ і придбати подальші можливості для досліджень завдяки простому логічному обчисленню [4; 9].

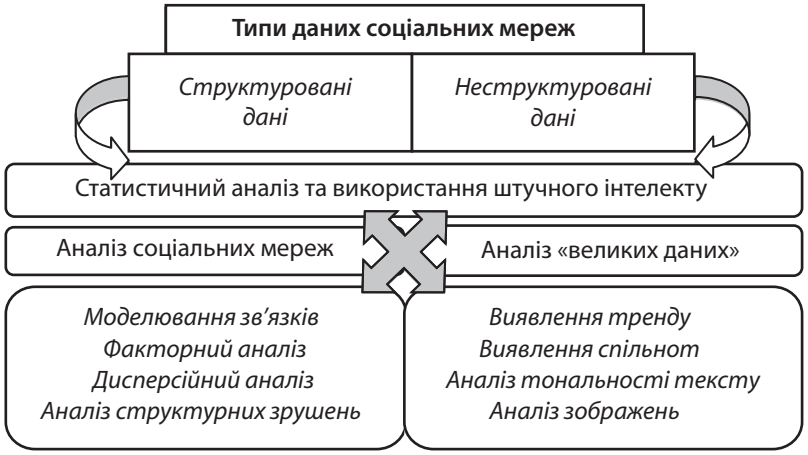

Рис. 1. Типи даних соціальних мереж і методи їх аналізу

На *рис. 1* узагальнено типи даних і відповідні підходи та методи аналізу.

Останні кілька років спостерігається швидкий прогрес у вирішенні давніх, складних проблем при використанні штучного інтелекту, і зараз він широко розповсюджений серед найпопулярніших сервісів Інтернету [6; 17]. З іншого боку, статистика передбачає менш складні процедури, які акцентують увагу на статистичних моделях на шляху до кращого розуміння процесу генерування даних. Аналіз на основі контенту в соціальних медіа вивчається за допомогою аналітики великих даних, і його увага зосереджена на витягуванні знань з контенту, створеного та поширеного в мережі.

Аудіо-, чи мовленнєвий аналіз, проводиться за допомогою безперервного розпізнавання мови онлайн-словником або фонетичного підходу до отримання інформації з неструктурованих аудіоданих [13]; аналіз контенту відео включає різні методи контролю, аналізу та вилучення значної інформації з відеопотоків [13]; методи аналізу зображень варіюються від простих до складних залежно від завдання аналізу.

Останнім часом велику увагу привертають методи розпізнавання облич [10] та вилучення настроїв [16] на основі даних соціальних медіа.

Аналіз даних соціальних медіа – термін, який охоплює описовий і структурний аналіз [11]. Для того, щоб зрозуміти структуру мережі, отримати уявлення про те, як вона «працює» та приймати рішення щодо неї, потрібно досліджувати характеристики вузла/зв'язку (наприклад, центральність) або переглядати показники всієї згуртованості мережі (наприклад, щільність) [11]. Порівняння мереж, відстеження змін у мережі з плином часу, виявлення спільнот і важливих вузлів, а також визначення відносного положення осіб і кластерів у межах мережі є деякими загальними процедурами [14]. Вони включають або статичний, або динамічний аналіз. Перший передбачає, що соціальна мережа поступово змінюється з часом, і аналіз всієї мережі можна проводити в пакетному режимі. І навпаки, динамічний аналіз, який є більш заплутаним, охоплює потокові дані, що розвиваються в часі з високою швидкістю.

Динамічний аналіз часто знаходиться в області взаємодій між сутностями, тоді як статичний аналіз стосується таких властивостей, як зв'язність, щільність, ступінь, діаметр і геодезична відстань.

З огляду на різноманітність та потенційний розмір даних соціальних медіа розробляються нові та динамічні підходи до існуючих кількісних та якісних методик дослідження.

Методологія дослідження соціальних медіа (*табл. 1*) охоплює широкий спектр кількісних та якісних методів, кожен із яких має свої переваги та недоліки.

Кількісні підходи надають змогу отримати широкий діапазон знань при вимірах частоти дискретних або категорійних змінних у наборах даних. До них належить, перш за все, кількісний аналіз – це найпростіший спосіб перегляду обсягів будь-яких даних асоційованих з певними групами (наприклад, користувачі платформи з певними демографічними характеристиками) або з обсягами згадок про певне ключове слово.

Наступним є аналіз стосунків: враховуючи обсяги інформації, необхідно також дивитися на взаємодію між користувачами, що враховуватиме кількість фіксованих стосунків у користувача (наприклад, друзів у Facebook, підписників в Instagram тощо) або кількість відповідей на публікацію (наприклад, ретвіти / цитати в Twitter, коментарі на Tumblr / YouTube тощо). Це корисно для аналізу взаємодії та часто є ключовою інфографікою, що відображається на платформах аналізу соціальних медіа.

Доцільно порівняти традиційні та нові методи дослідження, які можуть бути корисними аналітичними інструментами для різноманітних типів даних.

Таблиця 1

Методи дослідження соціальних медіа

Кількісні		**Якісні**
Одиниці кількості та частоти: кількість послідовників / друзів, кількість користувачів, темпи використання та взаємодії, пошуки. ***Кількість реакцій:*** перегляди, коментарі, лайки / схвалення, ретвіти / цитати. ***Обсяги за одиницю часу*** ***Оцінки / Інші порядкові рейтинги*** ***Видалення***	***Біографічні дані:*** вік, ім'я, стать, національність, місце проживання, професія або кваліфікація, діяльність, життя чи інтереси. ***Місцезнаходження:*** широта / довгота, поселення / адреса ***Текстова семантика:*** вміст ключових слів із постів, коментарі до основних публікацій, хештеги. ***Вплив:*** закономірності реакції	***Візуальний та аудіоконтент:*** теги для фотографій, медіа-тон і зміст. ***Тон і почуття:*** емоції та почуття, тон і думка. ***Вплив:*** теми обговорення / пошуку
Методи дослідження		
Регресійний аналіз Кореляційний аналіз Дисперсійний аналіз Описові статистичні тести Статистичне моделювання ГІС	Аналіз мережі Семантичний аналіз ГІС Псевдоексперименти	Семантичний аналіз Тематичний аналіз Кодифікація Етнографічне спостереження Активні дослідження

Порівнюючи набір даних соціальних медіа з іншим набором даних або якоюсь іншою незалежною змінною (наприклад, місцеположення, вік), можна провести дисперсійний аналіз. Це є основою використання даних соціальних медіа як орієнтовного або прогнозного інструменту.

Використовуючи пов'язані дані, можна побудувати моделі для прогнозування значень залежних змінних. Це забезпечує корисний інструмент інформування. Наприклад, такий інструмент дозволить спрогнозувати зростання вірусу при дослідженні твітів про симптоми з метою інформування населення, спрямованого на зменшення випадків захворювання.

Іншим варіантом кількісного підходу є кластеризація. Цей інструмент, по суті, є кількісною версією сегментації. Він використовує алгоритм для призначення даних кластера, де всі елементи мають подібні характеристики. Це корисно при перегляді демографічних характеристик користувачів, які обговорюють обрані теми.

Також до уваги необхідно взяти просторовий аналіз. Географічні інформаційні системи (ГІС) використовують просторовий елемент, який часто надається даними соціальних медіа (IP-адреси з комп'ютерів / GPS-локацій під час публікації з мобільних пристроїв). Це означає, що набори даних можуть бути використані для надання в реальному часі чи історичному аспекті інформації про поширення події (наприклад, протест, спалах хвороби тощо).

Якісні підходи. Враховуючи якісний характер багатьох даних, отриманих на платформах соціальних медіа, зрозуміло, що якісні методи можуть надати цілий спектр аналітичних підходів:

- *сегментація / ідентифікація групи:* дослідники також можуть активно брати участь у даних соціальних медіа як додатковому джерелі, яке доповнює існуючі якісні дослідження. Зазвичай такого роду аналіз може бути досягнутий через платформу аналізу соціальних медіа. Одним із прикладів такого підходу є визначення важкодоступних груп для опитування або додаткового опитування;
- *тематичний аналіз:* дані соціальних медіа можуть бути закодовані та тематично проаналізовані для виявлення емоційного характеру контенту;
- *аналіз почуттів:* існуючі алгоритми можуть бути використані для автоматизованого аналізу настроїв і виявлення, чи є текст позитивним чи негативним поглядом [8]. Цей підхід усе ще обмежений у своїх можливостях оцінювання настрою під час опрацювання складних предметів чи неоднозначного, непослідовного або культурно-специфічного матеріалу, такого як сарказм;
- *графічний медіа-аналіз:* зображення та відеовміст є все більш важливою формою взаємодії в Інтернеті та може надати важливі дані про сфери, які представляють інтерес. Окрім прямого семіотичного аналізу такого матеріалу, спосіб та причини його обміну потребують більш глибокого аналізу, особливо з огляду на ризик їх витіснення з початкового контексту.

Комбінування підходів. Соціальні медіа можна вважати «якісними даними в кількісному масштабі» [12]. Таким чином, традиційні методологічні межі все більше розмиваються при розгляді найбільш підходящих інструментів вирішення дослідницького питання. Біографічні дані, соціальні характеристики, такі як заняття користувачів або інтереси, спосіб життя, часто

можуть бути чітко визначені та статистично співвіднесені з певними моделями поведінки. Однак будь-які висновки, отримані в результаті такого аналізу, можуть бути посилені шляхом рандомізованої якісної перевірки сенсу того, як користувачі вводять такі дані.

Поєднання різних методів може допомогти встановити більш широке контекстне значення. Наприклад, використання хештегів Twitter як критеріїв вибірки призводить до самостійного вибору випадків, оскільки будуть вивчені лише користувачі, які розміщують певну фразу. У цьому випадку такі підходи, як мережевий аналіз чи додаткові якісні дослідження, можуть допомогти встановити змінення використання певного хештегу з часом та між різними групами або розвинути розуміння використання певного хештегу [18]. Багато алгоритмів машинного навчання будуються для реалізації того, що традиційно виконується людиною із використанням багатьох аналітичних підходів, як кількісних, так і якісних. Важливо розуміти, що хоча згадка про «машинне навчання» та «алгоритми» може вказувати на кількісні методи – аналітичний результат усе ще є якісним. Використовувані методи та процеси, що застосовуються для їх застосування, слід розглядати окремо.

Достовірність та надійність. Хоча список не є вичерпним, деякі ключові атрибути даних соціальних медіа, які можуть мати наслідки для достовірності та надійності, перераховано нижче:

- *користувачі соціальних медіа іноді не є типовими представниками населення* [15]. Таким чином, існуватимуть упередження та може бути складно зробити висновки для загальної сукупності. Однак ця характеристика може бути корисною, якщо дослідження зосереджено на групі, яка особливо активно працює на платформі соціальних медіа. Платформи можуть мати численні автоматизовані «боти» та професійно керовані акаунти. Тому при проведенні дослідження науковці повинні намагатися відфільтрувати результати з таких аномальних джерел під час аналізу;
- *«органічні» дані в реальному часі:* дані соціальних медіа рідко створюються для наукових цілей. Це означає, що велика кількість даних може бути нерелевантною або у важкому для аналізу форматі, але вона має перевагу в усуненні упередженості дослідників, проблемах з навантаженням на учасників. Користувачі соціальних медіа також беруть участь у перформативній соціальній взаємодії. На індивідуальному рівні це може стосуватися проблеми репутації, тоді як користувачі, що мають спільні інтереси, можуть спільно просувати матеріали;
- *поведінка в Інтернеті порівняно з поведінкою в режимі офлайн:* важко зробити висновок про те, наскільки відображається поведінка користувача в Інтернеті щодо його поведінки в режимі офлайн, без інформації про нього з інших джерел. Зазвичай вважається, що як позитивні, так і негативні почуття в Інтернеті завищені, і що інтерес до теми може насправді не перетворюватися на подальші дії (розрив вартості та дії). Частково це пов'язано з «ефектом еха», який виробляється на платформах соціальних медіа. Наприклад, користувачу пропонуються матеріали, які переглядають його друзі, тим самим заохочуючи групову думку. Подальші поглиблені дослідження користувачів можуть надати більшу ясність у цьому питанні;
- *приватна власність на платформи та дані:* доступ до даних регулюється компаніями, які володіють даними та їх угодами щодо конфіденційності з користувачами. Багато компаній не розповсюджують деталі про взаємодію соціальних медіа, яка відбувається на їх платформах. Тому, хоча взаємодії можна спостерігати та аналізувати, важливі нюанси чи контекст можуть бути відсутніми. Крім того, існує непрозорість у створенні наборів даних. Платформи регулярно змінюють функціональність, налаштування та популярність, що впливає на спосіб збирання та аналізу даних. Хоча часто є позитивні зрушення в можливостях, доступних для наборів даних, забезпечення узгодженості даних досліджень у більш тривалих часових рамках може бути проблематичним.

ВИСНОВКИ

У даній статті розглянуто методи аналізу даних соціальних медіа залежно від їх типу. Приділено увагу кількісному та якісному підходам при вивченні та аналізі даних із соціальних мереж. Так, до кількісного аналізу належить аналіз кількості підписок, лайків, постів у соціальних мережах тощо. Аналіз даної інформації зазвичай надається в обраних платформах. Тоді як вивчення думок і почуттів, ставлення до запису в соціальній мережі потребує більш складних методів аналізу. Акцентовано увагу на проблемі достовірності та надійності даних, отриманих із соціальних медіа. Вивчення соціальних медіа надає низку можливостей для досліджень нового рівня, однак це також зумовлює виникнення проблем достовірності та надійності одержаних досліджень. Завдяки сучасним технологіям аналіз соціальних медіа здебільшого виконується автоматично, за допомогою «машинного навчання», проте все ще існує необхідність у якісному аналізі отриманих варіантів людиною, що зумовлює подальше вивчення цього питання. ■

ЛІТЕРАТУРА

1. Войтович О., Буда А., Головенько В. Дослідження методів аналізу соціальних мереж як середовища інформаційних війн // VI Міжнародна науково-практична конференція «Методи та засоби кодування, захисту й ущільнення інформації» : тези доповідей (м. Вінниця, 24–25 жовтня 2017 р.). Вінниця : ВНТУ, 2017. С. 76–80. URL: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/27958/76-86ycc0hnc6o8o3xgkr97hrynqd5m0obr.pdf?sequence=1&isAllowed=y
2. Горчинская О., Ривкин А. Анализ данных социальных сетей. *Открытые системы*. 2015. Вып. 03. С. 22–23. URL: http://www.fors.ru/upload/smi/Open-Systems_Gorchinskaya-Rivkin_09-2015.pdf
3. Мазуренко В. В., Штовба С. Д. Огляд моделей аналізу соціальних мереж. *Вісник Вінницького політехнічного інституту*. 2015. № 2. С. 62–74. URL: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/11950/813-812-1-PB.pdf?sequence=1&isAllowed=y
4. Adeli H., Siddique N. Introduction to Computational Intelligence // Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing. Oxford, UK : John Wiley & Sons Ltd, 2013. P. 1–17.
5. Aggarwal C. (ed.) Social Network Data Analytics. Boston, MA : Springer US, 2011. URL: http://www.charuaggarwal.net/socialtoc.pdf
6. Amodei D., Christiano P., Mané D., Olah C., Schulman J., Steinhardt J., Concrete Problems in AI Safety. 2016. URL: https://arxiv.org/pdf/1606.06565.pdf
7. Armour F., Espinosa J., Kaisler S., Money W. Big Data: Issues and Challenges Moving Forward // 46th Hawaii International Conference on System Sciences, 2013. P. 995–1004. DOI: 10.1109/HICSS.2013.645
8. Armstrong M., Bachrach Y., Sharma V., Volkova S. Inferring Latent User Properties from Texts Published in Social Media // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015. P. 4296–4297. URL: https://www.cs.jhu.edu/~svitlana/papers/VBAS-aaai15-demo.pdf
9. Cambria E., Hussain A. Introduction in Sentic Computing. Springer International Publishing, 2015. P. 1–21.
10. Chen B.-W., Du X., Jiang F., Rho S., Zhao D. Face hallucination and recognition in social network services. *The Journal of Supercomputing*. 2015. Vol. 71. No. 6. P. 2035–2049. DOI: 10.1007/s11227-014-1257-z
11. Csárdi G., Kolaczyk E. Statistical Analysis of Network Data with R. *Journal of Statistical Software*. 2015. Vol. 66. Book review 1. DOI: 10.18637/jss.v066.b01
12. D'Orazio F. The future of social media research: or how to re-invent social media listening in 10 steps. 2013 // Pulsar. URL: https://www.pulsarplatform.com/resources/the-future-of-social-media-research/
13. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information *Management*. 2015. Vol. 35. Issue 2. P. 137–144. DOI: https://doi.org/10.1016/j.ijinfomgt.2014.10.007
14. Hansen D., Shneiderman B., Smith M. Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, 2011. 304 p. URL: https://www.sciencedirect.com/book/9780123822291/analyzing-social-media-networks-with-nodexl#book-info
15. Ruths D., Pfeffer J. Social media for large studies of behavior. *Science*. 2014. Vol. 346. Issue 6213. P. 1063–1064. URL: https://people.cs.umass.edu/~brenocon/smacss2015/papers/Science-2014-Ruths-1063-4.pdf
16. Lin H., Meng J., Yu Y., Zhao Z. Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks. *Algorithms*. 2016. Vol. 9. No. 2. P. 41–52. DOI: 10.3390/a9020041
17. AI Is Transforming Google Search. The Rest of the Web is Next / WIRED, 2016. URL: https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/
18. Tufecki Z. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls // ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, 2014. P. 876–894. URL: https://arxiv.org/abs/1403.7400

Науковий керівник – Чала Т. Г., кандидат економічних наук, доцент кафедри статистики, обліку та аудиту Харківського національного університету ім. В. Н. Каразіна

REFERENCES

"AI Is Transforming Google Search. The Rest of the Web is Next". WIRED, 2016. https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/

Adeli, H., and Siddique, N. "Introduction to Computational Intelligence". In *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*, 1-17. Oxford, UK: John Wiley & Sons Ltd, 2013.

Amodei, D. et al. "Concrete Problems in AI Safety". 2016. https://arxiv.org/pdf/1606.06565.pdf

Armour, F. et al. "Big Data: Issues and Challenges Moving Forward". *46th Hawaii International Conference on System Sciences*, 2013. 995-1004. DOI: 10.1109/HICSS.2013.645

Armstrong, M. et al. "Inferring Latent User Properties from Texts Published in Social Media". Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015. https://www.cs.jhu.edu/~svitlana/papers/VBAS-aaai15-demo.pdf

Cambria, E., and Hussain, A. *Introduction in Sentic Computing*. Springer International Publishing, 2015.

Chen, B.-W. et al. "Face hallucination and recognition in social network services". *The Journal of Supercomputing,* vol. 71, no. 6 (2015): 2035-2049. DOI: 10.1007/s11227-014-1257-z

Csardi, G., and Kolaczyk, E. "Statistical Analysis of Network Data". R. *Journal of Statistical Software*, vol. 66 (2015), book review 1. DOI: 10.18637/jss.v066.b01

D'Orazio, F. "The future of social media research: or how to re-invent social media listening in 10 steps. 2013". Pulsar. https://www.pulsarplatform.com/resources/the-future-of-social-media-research/

Gandomi, A., and Haider, M. "Beyond the hype: Big data concepts, methods, and analytics". International Journal of Information Management, vol. 35, no. 2 (2015): 137-144.
DOI: https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gorchinskaya, O., and Rivkin, A. "Analiz dannykh sotsialnykh setey" [Social Network Data Analysis]. Otkrytyye sistemy. 2015. http://www.fors.ru/upload/smi/Open-Systems_Gorchinskaya-Rivkin_09-2015.pdf

Hansen, D., Shneiderman, B., and Smith, M. "Analyzing Social Media Networks with NodeXL: Insights from a Connected World". Morgan Kaufmann, 2011. https://www.sciencedirect.com/book/9780123822291/analyzing-social-media-networks-with-nodexl#book-info

Lin, H. et al. "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks". *Algorithms*, vol. 9, no. 2 (2016): 41-52.
DOI: 10.3390/a9020041

Mazurenko, V. V., and Shtovba, S. D. "Ohliad modelei analizu sotsialnykh merezh" [Overview of Models for Social Network Analysis]. Visnyk Vinnytskoho politekhnichnoho instytutu. 2015. https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/11950/813-812-1-PB.pdf?sequence=1&isAllowed=y

Ruths, D., and Pfeffer, J. "Social media for large studies of behavior". Science. 2014. https://people.cs.umass.edu/~brenocon/smacss2015/papers/Science-2014-Ruths-1063-4.pdf

"Social Network Data Analytics". Boston, MA : Springer US, 2011. http://www.charuaggarwal.net/socialtoc.pdf

Tufecki, Z. "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls". ICWSM'14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, 2014. https://arxiv.org/abs/1403.7400

Voitovych, O., Buda, A., and Holovenko, V. "Doslidzhennia metodiv analizu sotsialnykh merezh yak seredovyshcha informatsiinykh viin" [Analyses Methods Research of Social Networks as a Environment of Warfare]. Metody ta zasoby koduvannia, zakhystu i ushchilnennia informatsii. 2017. https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/27958/76-86ycc0hnc6o8o3xg-kr97hrynqd5m0obr.pdf?sequence=1&isAllowed=y