Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
https://doi.org/10.36930/40360320
Article received: 26.03.2026
Article accepted: 28.05.2026
Article published: 29.06.2026
Correspondence author: O. R. Staso Staso.oleh@gmail.com
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)

UDC 004.[912+89]:614.8

О. Р. Стасьо, Н. Є. Бурак
Львівський державний університет безпеки життєдіяльності, м. Львів, Україна

ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ ВИКОРИСТАННЯ НЕСТРУКТУРОВАНИХ ДАНИХ

Проаналізовано неструктуровані дані, сучасні підходи до їх оброблення, а також визначено проблеми і перспективи їх використання в інформаційних системах, зокрема у сфері надзвичайних ситуацій (НС). Виявлено, що неструктуровані дані становлять переважну частину сучасних інформаційних ресурсів і є потенційно цінним джерелом знань для аналізу та прогнозування настання подій. Встановлено, що стрімке зростання обсягів неструктурованих цифрових даних, їх різнорідність і відсутність уніфікованої структури істотно ускладнюють застосування традиційних методів їх оброблення та інтеграції в аналітичні системи. З'ясовано, що класичні підходи до виявлення та відбору знань, орієнтовані на структуровані дані, не забезпечують належної ефективності під час роботи з неструктурованою інформацією. Оцінено вплив ключових проблем на ефективність аналітичного використання неструктурованих даних, серед яких визначено семантичну неоднозначність, інформаційний шум, недостатню якість і актуальність даних, а також обмеженість наявних технологічних рішень. Охарактеризовано особливості використання неструктурованих даних у сфері НС, де вони виступають джерелом оперативної інформації для аналізу, моніторингу та прогнозування критичних подій. Систематизовано основні напрями застосування методів оброблення природної мови, машинного навчання, когнітивного аналізу та логічного моделювання для перетворення неструктурованої інформації у форму, придатну для подальшого аналітичного опрацювання. Узагальнено роль неструктурованих даних у доповненні офіційних джерел інформації, уточненні просторово-часового контексту подій і підвищенні оперативності реагування на загрози. Встановлено, що ефективне використання неструктурованих даних потребує комплексного підходу, який поєднує методи оброблення природної мови, машинного навчання, когнітивного аналізу та логічного моделювання. З'ясовано, що перспективним напрямом подальшого використання неструктурованих даних у сфері НС є створення інтегрованих інформаційних систем, здатних забезпечити збирання, оброблення та структуризацію даних з різнорідних джерел у режимі реального часу, що підвищує ефективність роботи систем підтримки прийняття рішень (СППР) у сфері НС. Обґрунтовано доцільність проведення подальших етапів дослідження, спрямованих на розроблення масштабованих архітектур інформаційно-аналітичних систем і гібридних аналітичних підходів до оброблення неструктурованих даних для використання неструктурованих даних у критично важливих сферах.

Ключові слова: великі дані; оброблення інформації; виявлення знань; інтелектуальний аналіз даних; системи підтримки прийняття рішень; надзвичайні ситуації.

Вступ / Introduction

Проблематика роботи з неструктурованими даними набуває дедалі більшої актуальності у світовій науковій спільноті через стрімке зростання обсягів цифрової інформації, що надходить із різнорідних джерел, зокрема – соціальних мереж, сенсорних систем, текстових звітів і мультимедійних потоків. За оцінюваннями дослідників [12], понад 80 % усіх наявних даних є неструктурованими, що створює істотні виклики для їх інтеграції в аналітичні системи, орієнтовані переважно на структуровані дані. У науковій літературі активно досліджують підходи до оброблення великих даних, зокрема – методи інтелектуального аналізу, машинного навчання та оброблення природної мови, однак, їх застосування до неструктурованої інформації потребує попереднього очищення, семантичної нормалізації та перетворення у форму, придатну для подальшого машинного аналізу. Тому використання методів інтелектуального аналізу даних, машинного навчання та оброблення природної мови для роботи з неструктурованою інформацією передбачає етап попереднього її оброблення, який охоплює очищення, нормалізацію, виділення релевантних ознак і перетворення початкових даних у структурований або частково структурований формат.

Попри значний прогрес у розвитку технологій аналізу даних, досі недостатньо систематизованими залишаються проблеми оброблення, структуризації, інтеграції та аналітичного використання неструктурованих даних у системах підтримки прийняття рішень (СППР) у сфері надзвичайних ситуацій (НС). Зокрема, залишаються відкритими питання інтеграції різнорідних потоків даних, забезпечення їхньої якості, а також підвищення ефективності аналітичних систем за умов значної невизначеності та інформаційного шуму. Це визначає актуальність проведення дослідження, спрямованого на

Стасьо Олег Романович, аспірант, кафедра інформаційних технологій та систем електронних комунікацій. **Email:** Staso.oleh@gmail.com; https://orcid.org/0009-0005-6049-6161

Бурак Назарій Євгенович, канд. техн. наук, доцент, начальник кафедри інформаційних технологій та систем електронних комунікацій. **Email:** n.burak@ldubgd.edu.ua; https://orcid.org/0000-0002-3880-4077

Цитування за ДСТУ: Стасьо О. Р., Бурак Н. Є. Проблеми та перспективи використання неструктурованих даних. Науковий вісник НЛТУ України. 2026, т. 36, № 3. С. 188–196.

Citation APA: Staso, O. R., & Burak, N. Ye. (2026). Challenges and prospects of using unstructured data. *Scientific Bulletin of UNFU*, 36(3), 188–196. https://doi.org/10.36930/40360320

узагальнення сучасних підходів до оброблення, структуризації та аналітичного використання неструктурованих даних і виявлення ключових обмежень у цій сфері дослідження. Водночас, використання таких даних у сфері НС потребує врахування безпекових обмежень, пов'язаних із конфіденційністю, захистом геолокаційної інформації та технологічною готовністю оперативних штабів до роботи з різнорідними потоками даних у режимі реального часу.

Об'єкт дослідження – оброблення та аналіз великих обсягів цифрових даних у інформаційних системах.

Предмет дослідження – методи і засоби оброблення, структуризації й аналітичного використання неструктурованих даних в інформаційних СППР, які дадуть змогу сформувати системне уявлення про чинники, які знижують ефективність їх аналітичного застосування у сфері НС.

Мета роботи – дослідити проблеми та перспективи використання неструктурованих даних у СППР під час НС, що дасть змогу на підставі їх узагальнення здійснювати миттєвий моніторинг ситуації, створювати карти руйнувань, аналізувати настрої населення та бути джерелом актуальної інформації, а також прогнозувати розвиток подій.

Для досягнення зазначеної мети визначено такі основні *завдання дослідження*:

- проаналізувати сучасний стан досліджень у сфері оброблення неструктурованих даних, що дасть змогу визначити рівень розробленості проблеми та виявити напрями, які потребують подальшого наукового опрацювання;
- узагальнити основні підходи до виявлення та відбору знань у великих даних, що дасть змогу встановити їхні переваги й обмеження під час роботи з неструктурованою інформацією;
- визначити ключові проблеми та обмеження під час роботи з неструктурованими джерелами, що дасть змогу сформувати системне уявлення про чинники, які знижують ефективність їх аналітичного застосування у сфері НС;
- визначити основні технологічні та безпекові виклики використання неструктурованих даних у сфері НС, зокрема – проблеми конфіденційності, захисту геолокаційної інформації та інтеграції таких даних у роботу оперативних штабів;
- оцінити можливості використання неструктурованих даних у сфері НС, що дасть змогу обґрунтувати їхню роль у підвищенні ефективності роботи СППР.

Аналіз останніх досліджень та публікацій. Використання неструктурованих даних (тексти, соцмережі, фото, відео) у системах підтримки прийняття рішень (СППР) під час НС є критично важливим, тому й має низку серйозних можливостей [16]: неструктуровані дані у СППР для НС; миттєвий моніторинг ситуації, коли аналіз публікацій у реальному часі дає можливість виявити епіцентр НС до офіційних звітів служб порятунку; створення карт руйнувань, коли БПЛА засобами комп'ютерного зору (англ. *Computer Vision*) може автоматично оцінювати масштаби пошкоджень; аналіз настроїв населення, коли засоби технології NLP (оброблення природної мови) допомагають оцінити рівень паніки та потреби людей у гуманітарній допомозі; краудсорсинг інформації, де місцеві громадяни стають "живими джерелами", передаючи точні координати завалів, пожеж чи затоплень; прогнозування розвитку НС, коли мультимодальні моделі штучного інтелекту здатні об'єднувати метеодані, пости людей та відеопотоки для прогнозування поширення лиха.

У дослідженні [12] проаналізовано засоби і методи опрацювання неструктурованих даних та зазначено, що такі дані становлять значну частину сучасного цифрового інформаційного середовища. Автор акцентує увагу на тому, що неструктурована інформація потребує спеціалізованих підходів до пошуку, інтерпретації, структурування та подальшого використання. Водночас, у роботі основну увагу зосереджено на загальних засобах аналізу неструктурованих даних, тоді як питання їх застосування у сфері НС потребує подальшого уточнення.

У роботі [2] автори запропонували концепцію Knowledge Discovery in Databases, у межах якої процес виявлення знань подано як послідовність етапів відбирання даних, попереднього оброблення, трансформації, інтелектуального аналізу даних та інтерпретації отриманих результатів. Цей підхід до виявлення знань у базах даних є фундаментальним для розуміння процесів перетворення даних у знання. Однак, зазначена концепція виявлення знань у базах даних орієнтована переважно на структуровані набори даних, що обмежує її пряме застосування для аналізу неструктурованої інформації без попереднього семантичного оброблення та нормалізації.

У дослідженні [3] автори розглянули основні поняття, методи й аналітичні підходи до оброблення великих даних, пов'язані з великими даними, та виокремили ключові характеристики великих даних, зокрема – обсяг, швидкість надходження, різноманітність і складність оброблення. Важливим для цього дослідження є висновок про те, що різнорідність даних істотно ускладнює їх інтеграцію в аналітичні системи. Водночас, у роботі недостатньо деталізовано проблеми саме неструктурованих джерел у контексті оперативного прийняття управлінських рішень.

У роботі [15] досліджено можливості використання великих мовних моделей для роботи зі структурованими та неструктурованими даними. Автори показали, що великі мовні моделі можуть підвищити ефективність вилучення, класифікації та узгодження інформації з різнорідних джерел. Водночас, використання великих мовних моделей не усуває повністю проблем достовірності, пояснюваності результатів, інформаційного шуму та інтеграції отриманих даних у спеціалізовані СППР.

У дослідженні [7] автори розглянули застосування нейронних мереж для аналізу відеопотоків з камер спостереження під час оцінювання евакуаційних потоків. Ця робота демонструє практичний потенціал інтелектуальних методів аналізу даних у завданнях безпеки та реагування на небезпечні події. Однак, дослідження зосереджене переважно на одному типі інформаційного джерела інформації, тоді як за реальних умов НС інформація надходить одночасно з текстових, сенсорних, мультимедійних та відкритих інформаційних джерел.

У роботі [8] автори проаналізували використання даних соціальних мереж і методів оброблення природної мови для дослідження природних катастроф. Вони показали, що повідомлення користувачів можуть містити оперативну інформацію про місце, характер і наслідки небезпечних подій. Водночас, для повідомлень соціальних мереж, як джерела даних, характерні неповнота, неоднорідність, суб'єктивність, дублювання повідомлень і високий рівень інформаційного шуму, що потребує попередньої фільтрації, класифікації та семантичної інтерпретації.

У дослідженні [13] автори розглянули підходи до поєднання структурованої, напівструктурованої та неструктурованої інформації в межах семантичних вікі-систем. Запропоновані підходи до інтеграції структурованої, напівструктурованої та неструктурованої інформації є важливими для розуміння проблеми інтегрування з різнорідних джерел інформації в єдине середовище. Однак, такі рішення мають переважно прикладний характер і не повністю враховують вимоги до оперативного оброблення інформації за умов невизначеності, обмеженого часу та потреби швидкого реагування.

У роботі [11] автори запропонували підхід до вилучення та аналізу неструктурованих і напівструктурованих даних для систем управління знаннями. У дослідженні розглянуто послідовність перетворення неструктурованої інформації у форму, придатну для подальшого машинного опрацювання та подання знань. Водночас, запропонований підхід до вилучення та аналізу неструктурованих і напівструктурованих даних потребує подальшого розвитку для застосування у сфері НС, де важливими є не тільки вилучення знань, а й швидкість оброблення, достовірність інформаційних джерел, інтеграція потокових даних і підтримка прийняття управлінських рішень.

Отже, аналіз останніх досліджень і публікацій показує, що в науковій літературі достатньо ґрунтовно розглянуто деякі аспекти оброблення великих даних, виявлення знань, застосування машинного навчання, оброблення природної мови та інтеграції різнорідних інформаційних джерел. Водночас, недостатньо систематизованими залишаються проблеми комплексного використання неструктурованих даних у сфері НС, зокрема – питання їх попереднього оброблення, семантичної нормалізації, інтеграції з іншими джерелами та застосування в СППР. Це обґрунтовує потребу проведення цього дослідження, спрямованого на узагальнення сучасних підходів до оброблення, структуризації та аналітичного використання неструктурованих даних, визначення ключових обмежень і формування цілісного уявлення про перспективи використання неструктурованих даних у безпекових завданнях.

Матеріали та методи дослідження. Методологічну основу роботи становили методи аналізу, систематизації, порівняння та узагальнення наукових джерел з оброблення неструктурованих даних, виявлення знань, застосування машинного навчання, оброблення природної мови та інформаційних СППР. Матеріалами дослідження були наукові публікації, матеріали конференцій та аналітичні праці, що висвітлюють проблеми інтеграції, структуризації й використання неструктурованих даних у сфері НС та НС.

Результати дослідження та їх обговорення / Research results and their discussion

Здійснений аналіз наукових джерел і сучасних підходів до роботи з неструктурованими даними дав змогу систематизувати їхню роль у сучасних інформаційних системах, визначити особливості процесів виявлення та відбору знань, а також окреслити ключові проблеми, що ускладнюють їх ефективне використання. Отримані під час аналізу наукових джерел результати структуровано за трьома основними напрямами: роль неструктурованих даних у сфері НС, процеси виявлення та відбору знань, а також проблеми та обмеження під час роботи з неструктурованими даними.

Роль неструктурованих даних у сфері НС. У сфері НС інформація формується з великої кількості різнорідних джерел, серед яких можуть бути сенсорні системи, програмні комплекси моніторингу, журнали реєстрації подій, текстові повідомлення, мультимедійні матеріали та дані з відкритих інформаційних ресурсів. Частина таких даних має визначену структуру, наприклад записи автоматизованих систем або детекторів, тоді як інша частина надходить у довільній формі, зокрема – у вигляді текстових описів, повідомлень очевидців, службових записів або публікацій у соціальних мережах. Здебільшого дані, що надходять під час надзвичайних подій, мають неструктурований або змішаний формат, що ускладнює їх безпосереднє використання в аналітичних системах [6].

Особливість неструктурованих даних, що формуються під час НС, полягає в тому, що вони відображають події в реальному або близькому до реального часі, однак, часто для них характерні неоднорідність, неповнота, семантична неоднозначність та різний рівень достовірності. Через відсутність єдиного формату їх важко безпосередньо інтегрувати в традиційні інформаційні системи, побудовані переважно для роботи зі структурованими даними. Це створює потребу в попередньому обробленню, семантичній нормалізації, логічному тегуванні та подальшій трансформації інформації у формат, придатний для автоматизованого аналізу [8].

Узагальнення сучасних підходів до оброблення, структуризації та аналітичного використання неструктурованих даних показує, що використання неструктурованих даних у сфері НС має значний практичний потенціал. Такі дані можуть доповнювати офіційні структуровані джерела інформації, підвищувати оперативність виявлення загроз, уточнювати просторово-часовий контекст подій і забезпечувати додаткову інформацію для прогнозування розвитку кризових ситуацій. Водночас, їх ефективне застосування можливе тільки за умови створення інформаційної системи, здатної виконувати три ключові функції: збирати дані з різнорідних джерел без наперед визначеної структури; видобувати релевантну інформацію з неструктурованих або частково структурованих потоків; організовувати отримані відомості в узгоджену структуру, придатну для подальшого аналізу.

Отже, результати аналізу сучасних підходів до використання неструктурованих даних у сфері НС свідчать про те, що неструктуровані дані можуть відігравати важливу роль у підвищенні ефективності роботи СППР у сфері НС. Їх опрацювання дає змогу розширити інформаційну базу для моніторингу, аналізу та прогнозування небезпечних подій, однак, потребує поєднання методів попереднього оброблення, семантичної інтерпретації, машинного навчання та інтеграції даних у єдине аналітичне середовище.

Технологічні та безпекові виклики використання неструктурованих даних у сфері НС. Використання неструктурованих даних у сфері НС створює не тільки аналітичні можливості, а й низку технологічних та безпекових викликів. Текстові повідомлення, фото-, відео- та аудіоматеріали, публікації у соціальних мережах і дані з відкритих джерел можуть містити персональні дані постраждалих, очевидців, рятувальників, службову інформацію, точні просторові координати або непрямі ознаки місця події. Особливої ваги ця проблема набуває

за умов воєнного стану, коли оприлюднення геолокації НС, фото з місця події, відео із впізнаваними об'єктами або метаданих цифрових файлів може створювати додаткові загрози для населення та підрозділів реагування. Зокрема, така інформація може бути використана для уточнення місця події, визначення напрямів роботи рятувальних підрозділів, ідентифікації критичної інфраструктури або планування повторного ураження об'єкта.

Тому інтеграція неструктурованих даних у СППР у сфері НС має передбачати не тільки їх збирання та аналітичне опрацювання, а й попередню перевірку, фільтрацію та захист чутливої інформації. Доцільним є застосування механізмів анонімізації персональних даних, видалення або приховування геолокаційних метаданих, обмеження доступу до первинних матеріалів, перевірки достовірності джерел і пріоритезації повідомлень за рівнем критичності. Це дає змогу зменшити ризики неконтрольованого поширення чутливої інформації та підвищити безпеку використання неструктурованих даних у процесах моніторингу, аналізу й підтримки прийняття рішень.

Окремим обмеженням є технологічний розрив між потенційними можливостями сучасних СППР та реальними умовами роботи оперативного штабу НС. Сучасні інформаційно-аналітичні системи можуть підтримувати автоматичну класифікацію повідомлень, аналіз текстових і мультимедійних потоків, геопросторову прив'язку подій та виявлення критичних сигналів у режимі, близькому до реального часу. Однак ефективність їх практичного застосування залежить від наявності стабільного зв'язку, захищених каналів передавання даних, сумісності інформаційних систем, достатніх обчислювальних ресурсів, актуальних протоколів обміну інформацією та належного рівня цифрової підготовки персоналу. За відсутності таких умов навіть технологічно розвинені СППР можуть не забезпечити миттєвого зв'язку між джерелами даних, аналітичними модулями та оперативним штабом НС, що знижує практичну цінність неструктурованих даних для прийняття управлінських рішень у реальному часі.

Для підвищення наукової обґрунтованості аналізу доцільно систематизувати основні джерела неструктурованих даних, які можуть використовуватися у сфері НС, з урахуванням їхньої аналітичної цінності, ризиків і потрібних етапів попереднього оброблення. Така систематизація джерел дає змогу перейти від загального опису потенціалу неструктурованих даних до визначення конкретних напрямів їх практичного використання у СППР (таблиця). Неструктуровані дані у сфері НС – це текстові повідомлення, фото, відео та аудіозаписи, які не мають чіткого формату, але містять до 80 % усієї корисної інформації під час криз. Вони дають змогу рятувальникам діяти швидше, хоча й потребують складного оброблення через велику кількість "сміття" та фейків.

Таблиця. Джерела неструктурованих даних у сфері НС, їх аналітична цінність, ризики та методи попереднього оброблення / Sources of unstructured data in the field of national security, their analytical value, risks and pre-processing methods

Джерело неструктурованих даних	Приклади даних	Аналітична цінність для СППР у сфері НС	Основні ризики та обмеження	Потрібне попереднє оброблення
Соціальні мережі та месенджери	Пости, коментарі, повідомлення очевидців, фото з місця події	Оперативне виявлення події, уточнення місця, оцінювання потреб населення	Інформаційний шум, фейки, дублювання повідомлень, ризик розкриття геолокації	Фільтрація, верифікація джерел, видалення чутливих даних, класифікація повідомлень
Фото- та відеоматеріали очевидців	Фото руйнувань, відео пожеж, затоплень, завалів	Візуальне оцінювання масштабу події, уточнення типу та наслідків НС	Персональні дані, EXIF-метадані, впізнавані об'єкти, ризик повторного ураження	Видалення метаданих, анонімізація, комп'ютерний зір, геопросторова перевірка
Відео з БПЛА	Аерофотознімання, відеопотоки з дронів, знімки територій	Оцінювання масштабу руйнувань, пошук осередків небезпеки, підтримка координації підрозділів	Розкриття позицій, маршрутів і місць роботи рятувальників, обмеження передавання великих файлів	Сегментація зображень, розпізнавання об'єктів, захищене передавання, обмеження доступу
Аудіозаписи та дзвінки	Виклики на лінії екстрених служб, голосові повідомлення	Швидке виявлення критичних потреб, класифікація звернень, визначення пріоритетності реагування	Низька якість звуку, емоційність повідомлень, персональні дані	Перетворення мовлення в текст, очищення шуму, анонімізація, тематична класифікація
Службові текстові звіти	Опис події, польові нотатки, оперативні повідомлення	Формалізація досвіду реагування, уточнення причин і наслідків події	Різний стиль опису, неповнота, неоднозначність термінів	Нормалізація термінів, виділення сутностей, структурування, логічне тегування
Сенсорні та IoT-потоки	Дані давачів диму, температури, рівня води, сейсмічні або метеодані	Раннє виявлення загроз, моніторинг динаміки події, прогнозування розвитку НС	Нестабільність зв'язку, різні формати, пропуски даних, технічні збої	Очищення, синхронізація, агрегація, виявлення аномалій
Відкриті веб-джерела та медіа	Новини, вебповідомлення, карти, відкриті бази даних	Додаткова контекстна інформація, перевірка повідомлень, уточнення масштабу події	Застарілість, неповнота, різна достовірність джерел	Перевірка актуальності, зіставлення з офіційними даними, семантична класифікація

Як видно з таблиці, найбільшу аналітичну цінність для СППР у сфері НС мають ті джерела, які забезпечують оперативність, просторово-часову прив'язку та можливість уточнення фактичного стану події. Водночас, саме такі джерела часто створюють найбільші безпекові ризики, пов'язані з конфіденційністю, геолокацією, достовірністю та технічною сумісністю даних. Тому ефективне використання неструктурованих даних потребує поєднання методів фільтрації, верифікації, анонімізації, семантичної нормалізації та інтеграції в єдине аналітичне середовище. Для систем підтримки прийняття рішень (СППР) у сфері надзвичайних ситуацій (НС) найціннішими є джерела, які дають найшвидшу та найчіткішу інформацію тут і зараз.Коли стається

аварія, пожежа чи повінь, рятувальникам потрібно діяти негайно. Звичайні паперові звіти чи застарілі довідники тут не допоможуть.

Процеси виявлення та відбір знань у великих даних. Великі дані можуть містити значні обсяги прихованої корисної інформації, яку потрібно виявити, інтерпретувати та подати у формі, придатній для подальшого аналізу. У цьому контексті важливу роль відіграють два взаємопов'язані процеси: виявлення знань (англ. *Knowledge Discovery*) та відбір знань (англ. *Knowledge Extraction*). Виявлення знань передбачає пошук нових, значущих і потенційно корисних закономірностей у масивах даних. Для неструктурованої інформації цей процес є особливо складним, оскільки передбачає не тільки аналіз змісту, а й попередню інтерпретацію контексту, структури повідомлень і семантичних зв'язків. У роботі [2] процес KDD подано як послідовність етапів: вибірка (англ. *Selection*), попереднє оброблення (англ. *Preprocessing*), трансформація (англ. *Transformation*), добування даних (англ. *Data Mining*), інтерпретація та оцінювання результатів (англ. *Interpretation/Evaluation*). Узагальнену схему переходу від початкових даних до знань наведено на рис. 1. Процес виявлення знань у базах даних (його часто називають KDD) – це пошук прихованих, нових і корисних закономірностей у великих обсягах інформації. Головна мета цієї технології – перетворити сирі факти на зрозумілі людині знання для прийняття управлінських рішень.

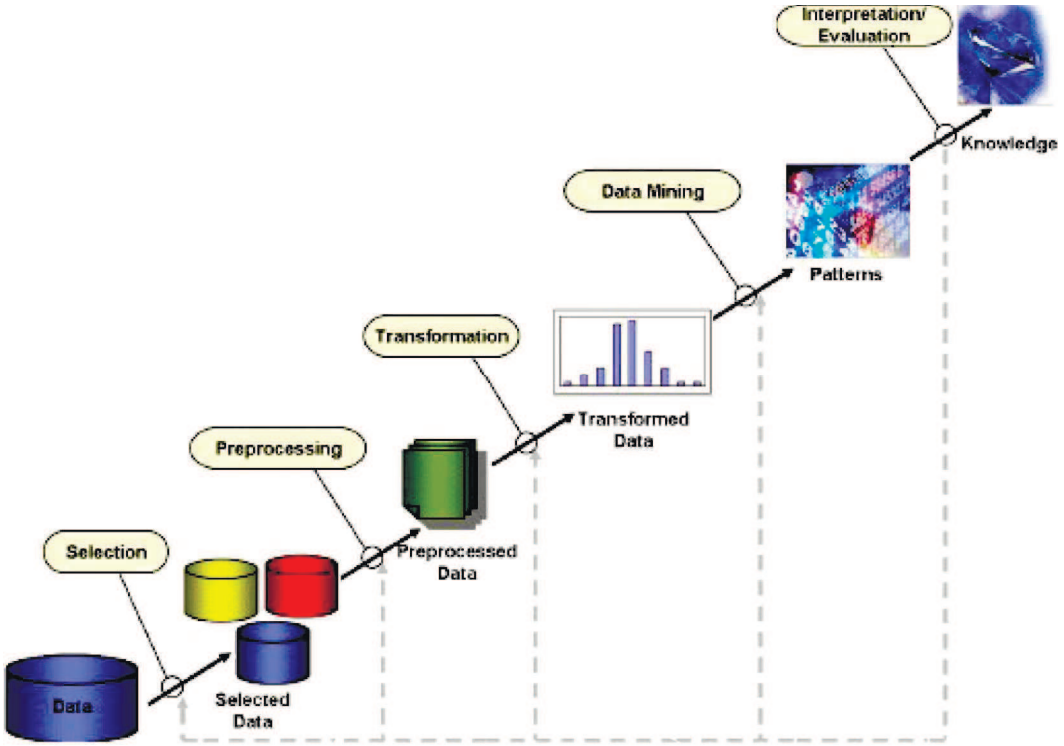

Рис. 1. Основні етапи процесу виявлення знань у базах даних (KDD): вибірка, попереднє оброблення, трансформація, добування даних та інтерпретація результатів / Main stages of the Knowledge Discovery in Databases (KDD) process: selection, preprocessing, transformation, data mining, and interpretation of results [4]

Під даними у цьому контексті розуміють сукупність зафіксованих фактів про об'єкти, події або процеси, а під шаблоном даних – модель або вираз, що описує певну вибірку. Відбір шаблонів полягає у встановленні структури в даних, побудові моделі або формуванні високорівневого опису досліджуваного набору. У межах KDD-компонент data mining виконує функцію безпосереднього виявлення закономірностей, тоді як інші етапи забезпечують підготовку даних, їх трансформацію, інтерпретацію та подальше використання результатів.

Проблеми та обмеження під час роботи з неструктурованими даними. У сучасному інформаційному середовищі термін *"неструктуровані дані"* використовують дедалі частіше, однак, досі не існує єдиного, загальноприйнятого визначення цього поняття. Згідно з підходом до неструктурованих даних, запропонованим у дослідженні [13], до них належить інформація, яка або не має чітко визначеної моделі, або не вписується в реляційні таблиці. Найчастіше такі дані подано у вигляді тексту, проте також можуть містити часові мітки, числові значення, зображення, відео та інші форми, що не піддаються простому табличному уявленню.

Неструктурованість даних призводить до таких проблем, як двозначність інтерпретацій, відсутність уніфікованих форматів, труднощі з пошуком релевантної інформації. Це значно ускладнює використання традиційного програмного забезпечення, призначеного для роботи зі структурованими базами даних чи анотованими наборами документів.

Окрему складність становить оброблення неструктурованих потоків даних у режимі реального часу. Здебільшого швидкість генерування інформації перевищує обчислювальні можливості систем, що змушує звертатися до гібридних обчислювальних архітектур інформаційно-аналітичних систем. Такі архітектури поєднують семантичний аналіз, логічні моделі та статистичні методи оброблення [3]. До прикладів таких підходів до оброблення та інтерпретації неструктурованих потоків даних належать методи когнітивного аналізу, які об'єднують лінгвістичну інтерпретацію з машинним навчанням на основі попередніх сценаріїв.

Ще одним викликом є обмежена релевантність або застарілість неструктурованих джерел. Це означає, що інформація у звичайних текстах, листах чи відео часто є

непотрібною або старою. Через відсутність структури такі дані важко шукати та перевіряти. За відсутності значної якості даних, аналітичні моделі можуть помилково трактувати кореляції як причинно-наслідкові зв'язки або пропустити важливі патерни. Це створює ризики за використання неструктурованої інформації у СППР, особливо у чутливих сферах, таких як безпека.

Спроби виявлення причинно-наслідкових зв'язків у неструктурованих джерелах даних, особливо за умов значного "інформаційного шуму", призвели до широкого застосування методів текстового їх добування та автоматичної класифікації. Ці методи дають змогу виявляти ключову інформацію навіть у хаотичних масивах, як-от пости із соціальних мереж або польові звіти в екстремальних ситуаціях.

У контексті Інтернету речей (IoT) неструктуровані дані є особливо поширеними. Потоки від сенсорів, аудіо- та відеообладнання, зазвичай, не мають фіксованої структури. Це ускладнює їх інтеграцію в централізовані системи моніторингу та аналізу. Для вирішення цієї проблеми застосовують спеціалізовані програмні рішення, зокрема – платформа Talend, що забезпечує гнучкі інструменти трансформації, маршрутизації та уніфікації даних з різних джерел [1].

Останні досягнення в галузі оброблення природної мови та великих мовних моделей (LLM) відкривають нові можливості для роботи з неструктурованою інформацією. Такі моделі здатні автоматично виділяти сутності, класифікувати повідомлення та створювати структуроване подання з хаотичних даних. Їх застосування особливо перспективне для оброблення текстів із соціальних мереж, звітів з польових давачів та внутрішньої документації [15].

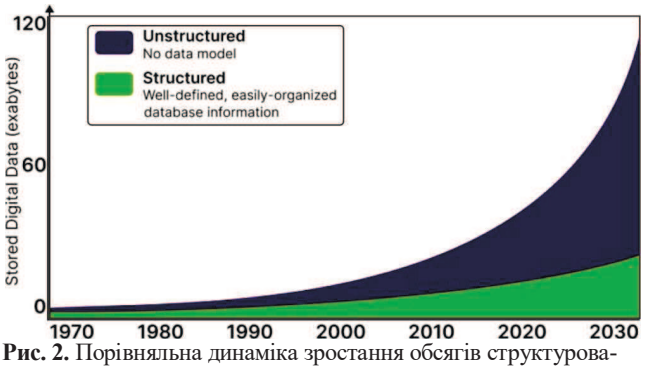

Рис. 2. Порівняльна динаміка зростання обсягів структурованих (зелений колір) і неструктурованих (фіолетовий колір) цифрових даних у період 1970-2030 рр. / Comparative dynamics of the growth of structured (green) and unstructured (purple) digital data volumes in 1970-2030 [10]

Невпинне зростання обсягів цифрових даних, зокрема – неструктурованих, стає одним із критичних викликів сучасного інформаційного середовища. За оцінюваннями дослідників, частка неструктурованих даних становить від 80 до 90 % усієї доступної інформації, що значно перевищує частку структурованих [12]. Подібна динаміка посилює потребу в адаптивних підходах до їхнього зберігання, оброблення та інтерпретації. Дослідження [1, 15] вказують на те, що традиційні моделі аналізу даних не справляються з новими обсягами і форматами інформації, що потребує інтеграції гібридних рішень та когнітивних технологій. Зокрема, важливо розробити системи, здатні ефективно поєднувати оброблення структурованих, напівструктурованих і неструктурованих джерел у межах єдиної аналітичної архітектури [13]. Порівняння динаміки зростання структурованої та неструктурованої цифрової інформації свідчить про істотне переважання обсягів неструктурованих джерел у сучасному інформаційному середовищі, що узагальнено на рис. 2.

У дослідженні [9] зазначено, що навіть великі компанії, які визнають свою залежність від неструктурованих джерел, не мають усталених практик і ефективних рішень для їхнього оброблення. Більшість організацій тільки починають усвідомлювати масштаб проблеми, яка стосується управління розподіленими неструктурованими даними у межах їхніх систем.

З огляду на це, наразі доступні кілька основних стратегій для покращення керування такими даними [1]:

1. Оновлення процесів управління даними, враховуючи інтеграцію неструктурованих джерел у наявні бази або розроблення нових типів сховищ.
2. Інвестиції в технології та інфраструктуру, що дають можливість удосконалити аналіз тексту, збільшити підтримку тегування, а також генерувати корисні метадані.
3. Розширення можливостей індексування та логічного зв'язування, що дає змогу пришвидшити пошук інформації та створити перехресні зв'язки з іншими даними.

Проте навіть за наявності зазначених рішень ефективне впровадження гальмується такими факторами, як високі витрати, нестача фахівців, низьке усвідомлення критичної ролі неструктурованих даних в організаційних процесах та обмеження поточних технологій.

Відбір знань з неструктурованих даних. Традиційні методи відбору знань переважно орієнтовані на роботу зі структурованими наборами даних або потребують значної участі експерта під час інтерпретації отриманих результатів. У разі неструктурованих джерел ключовою проблемою є перетворення початкової інформації у форму, придатну для подальшого машинного аналізу та логічного опрацювання. З огляду на це, у дослідженнях [5, 13] запропоновано послідовність етапів, що охоплює вилучення даних, синтаксичний і семантичний аналіз, класифікацію, застосування правил логічного висновку та подання результатів у структурованому форматі:

Неструктуровані дані → Вилучення даних → Синтаксичний та Семантичний аналізи → Класифікація даних → Застосування правил логічного висновку → Подання даних у структурованому форматі (*у вигляді XML документів чи таблиць даних*)

- Вхідними даними є неструктуровані джерела інформації, зокрема –вебсторінки, документи, повідомлення або записи баз даних.
- Вилучення даних передбачає ідентифікацію корисної інформації в початковому джерелі та її перенесення у формат, придатний для подальшого оброблення. Наприклад, вебсторінки можуть містити не тільки основний текст, а й посилання, зображення, навігаційні елементи та службові метадані.
- Синтаксичний аналіз формує дерево розбору, яке відображає граматичні зв'язки між основними компонентами речення.
- Класифікація даних забезпечує розподіл елементів інформації за попередньо визначеними категоріями. Для цього можуть застосовуватися, зокрема, метод *k*-найближчих сусідів (*k*-nearest neighbors, KNN) і наївний баєсівський класифікатор.
- Правила логічного висновку дають змогу формувати висновки на основі класифікованих даних із збереженням їхніх семантичних властивостей. Результати такого опрацювання можуть подаватися у вигляді таблиць даних або XML-документів.

Важливим аспектом цього процесу є забезпечення сумісності між різними форматами даних. За реальних умов інформація часто надходить одночасно з різнорідних джерел: структурованих (реляційні БД), напівструктурованих (JSON, XML) та неструктурованих (вільний текст, зображення). Це вимагає попередньої трансформації в уніфіковану форму [5, 11].

Після перетворення даних на аналітичному рівні застосовують методи інтелектуального аналізу (Data Mining), орієнтовані на конкретні завдання. Після того як дані очистили та підготували, їх починають ретельно досліджувати. Для цього комп'ютер використовує спеціальні методи Data Mining, щоб знайти приховані підказки чи закономірності. Наприклад [14]:

- асоціативні правила – виявлення частотних залежностей між ознаками у великих вибірках;
- кластеризація – групування об'єктів на основі подібності;
- регресія – побудова моделей для прогнозування кількісних значень.

Кожен із підходів до інтелектуального аналізу даних має свої переваги та обмеження. Так, асоціативне навчання здатне генерувати надмірну кількість правил, що ускладнює інтерпретацію результатів. Кластеризація є чутливою до вибору метрик відстані та кількості кластерів, а регресійні моделі можуть вимагати попередньої нормалізації даних для досягнення точності [14].

Отже, відбір знань із неструктурованих даних вимагає комплексного підходу до оброблення, структуризації та аналітичного використання неструктурованої інформації, що поєднує мовне оброблення, машинне навчання, логічні моделі та трансформацію у формат, придатний для автоматизованого аналізу.

Обговорення результатів дослідження. Звіти про пожежі та зображення на місці містять багаті, але неструктуровані знання з пожежної безпеки, які потребують явної формалізації для підтримки прийняття рішень, що потребують знань. Великий обсяг та мультимодальний характер таких даних ускладнюють їх перетворення на обчислювальні та повторно використовувані подання знань.

Отримані наші результати дослідження узгоджуються з положеннями роботи [6], у якій наголошено на тому, що для великих даних характерні значні обсяги, різнорідність форматів, висока швидкість надходження та складність оброблення. У межах цього дослідження зазначені характеристики розглянуто саме в контексті неструктурованих даних, які створюють додаткові труднощі для інформаційних систем через відсутність уніфікованої структури, семантичну неоднозначність і високий рівень інформаційного шуму. Це підтверджує, що ефективне використання неструктурованої інформації потребує не тільки технічних засобів зберігання та оброблення, а й попередньої структуризації, фільтрації та аналітичної інтерпретації.

Результати проведеного аналізу проблем та перспектив використання неструктурованих даних також узгоджуються з дослідженням [1], у якому розглянуто механізми керування структурованими та неструктурованими даними в середовищі Інтернету речей. Для сфери НС це має особливе значення, оскільки інформація може надходити одночасно із сенсорних систем, аудіо- та відеообладнання, текстових повідомлень, службових звітів і відкритих інформаційних ресурсів. На відміну від звичайних інформаційних систем, СППР у сфері НС мають працювати з різними типами даних за умов обмеженого часу, неповноти відомостей і потреби швидкого реагування.

Важливим для інтерпретації отриманих результатів є положення роботи [5], у якій розглянуто процес перетворення неструктурованих і напівструктурованих даних у знання. Отримані наші результати підтверджують, що робота з неструктурованою інформацією не може обмежуватися тільки її збиранням або збереженням. Для подальшого використання в аналітичних системах така інформація має пройти етапи вилучення релевантних даних, синтаксичного й семантичного аналізу, класифікації, логічного опрацювання та подання у структурованому форматі. Саме така послідовність дій дає змогу перетворити початкові неструктуровані повідомлення на знання, придатні для підтримки прийняття управлінських рішень.

Отримані наші результати доповнюють положення дослідження [4], у якому процес виявлення знань розглянуто як складну багатоетапну процедуру, що потребує підтримки на різних рівнях оброблення даних. У контексті нашої роботи це дає підстави стверджувати, що використання неструктурованих даних у сфері НС потребує не окремого програмного інструмента, а цілісної інформаційної технології, здатної забезпечити послідовне збирання, попереднє оброблення, структуризацію, інтерпретацію та інтеграцію інформації в аналітичне середовище. Такий підхід є важливим для зменшення впливу інформаційного шуму, дублювання повідомлень і контекстної неоднозначності.

Практичний аспект отриманих результатів узгоджується з висновками дослідження [9], у якому зазначено, що організації часто визнають цінність неструктурованих даних, однак не мають достатньо сформованих практик і технологічних рішень для їх ефективного використання. У сфері НС ця проблема є особливо актуальною, оскільки несвоєчасне або неповне опрацювання інформації може негативно впливати на якість управлінських рішень. Тому результати нашої роботи акцентують увагу на потребі створення інтегрованих інформаційно-аналітичних систем, які поєднують роботу з офіційними структурованими даними та неструктурованою інформацією з різнорідних інформаційних джерел.

Положення роботи [10] підтверджують важливість використання сучасних технологій оброблення великих даних для підвищення ефективності аналітичних процесів. Водночас, результати цього дослідження показують, що для сфери НС важливим є не тільки застосування технологій великих даних загалом, а й адаптація таких технологій до умов невизначеності, різнорідності інформаційних потоків і потреби оперативного прийняття рішень. Це дає змогу розглядати неструктуровані дані не як допоміжний інформаційний ресурс, а як важливу складову частину інформаційного забезпечення систем моніторингу, прогнозування та реагування.

У роботі [16] запропоновано гібридну систему підтримки прийняття рішень, графи знань та науку про дані KG-DS (англ. *Knowledge Graph and Data Science*), яка формалізує знання з пожежної безпеки за допомогою графа знань про пожежні інциденти FIKG (англ. *Fire Incident Knowledge Graph*) та інтегрує його з мультимодальним мисленням на основі візуально-мовної моделі VLM (англ. *Vision-Language Model*) для підтримки прийняття рішень у разі НС під час, наприклад, пожежі.

Система складається із двох модулів: моделі вилучення спільних трійок, орієнтованих на відношення RO-JTEM (англ. *Relation-Oriented Joint Triples Extraction Model*), яка слугує механізмом формалізації знань шляхом вилучення "суб'єкт-відношення-об'єкт" трійки подій з текстів розслідування пожеж для побудови структурованої FIKG; модулі міркування на основі VLM, яка поєднує пошук графів з мультимодальним сприйняттям для формування інтерпретованих циклів "сприйняття-думка-дія" для автономного генерування стратегій подальших дій. Отримані результати демонструють, що поєднання явних подань інженерних знань з мультимодальним обчислювальним міркуванням забезпечує надійний та зрозумілий механізм для підтримки прийняття знань у разі НС під час пожежі.

У контексті подальшого розвитку етапів дослідження доцільним є поєднання методів оброблення неструктурованої інформації з імовірнісно-статистичними методами інтелектуального аналізу даних, що розглядалися у роботі [14]. Такий напрям є перспективним, оскільки дає змогу не тільки структурувати початкові повідомлення, а й надалі використовувати отримані дані для оцінювання ризиків, прогнозування розвитку небезпечних подій і підтримки прийняття управлінських рішень. Отже, результати проведеного дослідження формують підґрунтя для подальшого розроблення прикладних моделей і масштабованих інформаційно-аналітичних архітектур для роботи з неструктурованими даними у сфері НС.

Водночас, потрібно зазначити, що дослідження має певні обмеження, пов'язані з його оглядово-аналітичним характером, різнорідністю проаналізованих інформаційних джерел, швидкою зміною технологій оброблення даних і відсутністю експериментальної перевірки всіх розглянутих підходів у межах однієї статті. Тому перспективним напрямом подальших етапів дослідження є розроблення та тестування прикладних моделей оброблення неструктурованих даних за реальних умов, зокрема – для інформаційно-аналітичних СППР у сфері НС та реагування на них.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – отримала подальший розвиток методика узагальнення технологічних підходів до оброблення неструктурованих даних у сфері НС, яка, на відміну від наявних методик, враховує семантичну неоднозначність таких даних, інформаційний шум та потребу їх інтеграції з різних джерел, що дає можливість здійснити попередню структуризацію інформації у джерелах її використання СППР.

Практична значущість результатів дослідження – можна використати під час обґрунтування вимог до інтегрованих інформаційно-аналітичних систем, призначених для збирання, попереднього оброблення, структуризації та аналітичного використання неструктурованих даних у процесах підтримки прийняття управлінських рішень у сфері НС та реагування на них.

Висновки / Conclusions

Досліджено проблеми та перспективи використання неструктурованих даних у СППР під час НС, що дає можливість на підставі їх узагальнення здійснювати миттєвий моніторинг ситуації, створювати карти руйнувань, аналізувати настрої населення, бути джерелом/краудсорсингом інформації, а також прогнозувати розвиток подій. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Проаналізовано сучасний стан досліджень у сфері оброблення неструктурованих даних і встановлено, що такі дані становлять значну частину сучасних цифрових інформаційних ресурсів, однак їх використання ускладнюється різнорідністю форматів, відсутністю уніфікованої структури, семантичною неоднозначністю та високим рівнем інформаційного шуму.
2. Систематизовано основні підходи до виявлення та відбору знань у великих даних, зокрема – методи інтелектуального аналізу даних, оброблення природної мови, машинного навчання, семантичної інтерпретації та логічного моделювання, що дають змогу перетворювати неструктуровану інформацію у форму, придатну для подальшого аналітичного опрацювання.
3. З'ясовано, що традиційні методи аналізу даних, орієнтовані переважно на структуровані набори, мають обмежену ефективність під час роботи з неструктурованими джерелами, тому потребують доповнення засобами попереднього оброблення, семантичної нормалізації, фільтрації інформаційного шуму та інтеграції різнорідних інформаційних потоків.
4. Встановлено, що у сфері НС неструктуровані дані можуть доповнювати офіційні структуровані джерела інформації, підвищувати оперативність виявлення загроз, уточнювати просторово-часовий контекст подій і розширювати інформаційну основу для підтримки прийняття управлінських рішень. Проведена систематизація джерел неструктурованих даних дала змогу визначити їхню аналітичну цінність для СППР, а також основні ризики та обмеження, пов'язані з достовірністю, конфіденційністю, геолокаційною чутливістю, інформаційним шумом і потребою попереднього оброблення.
5. Встановлено, що перспективи використання неструктурованих даних у сфері НС мають розглядати разом із безпековими та технологічними обмеженнями, зокрема – необхідністю анонімізації даних, захисту геолокаційної інформації, перевірки достовірності джерел і подолання технологічного розриву між сучасними СППР та реальними умовами роботи оперативних штабів НС.
6. З'ясовано, що подальші етапи дослідження доцільно спрямувати на розроблення та експериментальну перевірку прикладних моделей оброблення неструктурованих даних, на створення масштабованих інформаційно-аналітичних архітектур для їх використання за реальних умов безпекової діяльності та реагування на НС, а також на кількісне оцінювання ефективності різних джерел неструктурованої інформації за критеріями оперативності, достовірності, повноти, безпечності та придатності для інтеграції у СППР.

References

1. Azad, P., Navimipour, N. J., & Rahmani, A. M. (2020). The role of structured and unstructured data managing mechanisms in the Internet of Things. *Cluster Computing*, 23, 1185–1198. https://doi.org/10.1007/s10586-019-02986-2
2. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230
3. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
4. Guerra-Hernandez, A., Mondragon-Becerra, R., & Cruz-Ramirez, N. (2008). Explorations of the BDI multi-agent support for the

knowledge discovery in databases process. *Research in Computing Science*, 39, 221–238. URL: https://www.researchgate.net/publication/236373188_Explorations_of_the_BDI_Multi-agent_support_for_the_Knowledge_Discovery_in_Databases_Process
5. Halcu, I., Neculoiu, G., & Marinescu, V. (2013). Converting unstructured and semi-structured data into knowledge. In *Proceedings of RoEduNet International Conference*, pp. 1–4. https://doi.org/10.1109/RoEduNet.2013.6511736
6. Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In 2013 *Sixth International Conference on Contemporary Computing* (IC3) (pp. 404–409). IEEE. https://doi.org/10.1109/IC3.2013.6612229
7. Khlevnoi, O., Burak, N., Borzov, Y., & Raita, D. (2023). Neural network analysis of evacuation flows according to video surveillance cameras. In: Babichev, S., & Lytvynenko, V. (Eds.). *Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making*. ISDMCI 2022 (Vol. 149, pp. 349–361). Cham: Springer. https://doi.org/10.1007/978-3-031-16203-9_35
8. Ma, Z., Li, L., Mao, Y., Wang, Y., Patsy, O. G., Bensi, M. T., Hemphill, L., & Baecher, G. B. (2024). Surveying the use of social media data and natural language processing techniques to investigate natural disasters. *Natural Hazards Review*, *25*(4), article ID 03124003. https://doi.org/10.1061/NHREFO.NHENG-2047
9. McKendrick, J. (2011). Survey on unstructured data. Unisphere Research. Retrieved July 1, 2025, URL: http://www.ciosummits.com/media/pdf/solution_spotlight/marklogic_2011-survey.pdf
10. Oleshchenko, L. M. (2021). Technologies for processing big data. Kyiv: KPI im. Ihoria Sikorskoho. [In Ukrainian]. URL: https://ela.kpi.ua/items/4ad484d3-aa21-4f7e-be09-a6d8e1d1e8bc
11. Onwujekwe, G., Osei-Bryson, K.-M., & Ngwum, N. (2020). A framework for capturing and analyzing unstructured and semi-structured data for a knowledge management system. In: Wyld, D. C., & Nagamalai, D. (Eds.). *Proceedings of the 9th International Conference on Advanced Information Technologies and Applications* (ICAITA 2020), Computer Science & Information Technology, 10(9), 83–94. https://doi.org/10.5121/csit.2020.100907
12. Rohushyna, Y. V. (2019). Means and methods for analyzing unstructured data. *Programming Problems*, 1, 57–77. [In Ukrainian]. URL: https://pp.isofts.kiev.ua/ojs1/article/view/348/346
13. Sint, R., Shaffert, S., Stroka, S., & Ferstl, R. (2009). Combining unstructured, fully structured and semi-structured information in semantic wikis. In *Proceedings of the 4th Semantic Wiki Workshop* (SemWiki 2009) *at the 6th European Semantic Web Conference* (ESWC) (pp. 73–80). Hersonissos, Greece. URL: https://ceur-ws.org/Vol-464/paper-14.pdf
14. Staso, O. R., & Burak, N. Ye. (2023). Probabilistic and statistical methods of intelligent data analysis. In Achievements of 21st Century Scientific Community: Proceedings of the I International Scientific and Practical Internet Conference (pp. 426–429). WayScience, September 14–15, 2023. [In Ukrainian]. URL: https://www.wayscience.com/wp-content/uploads/2023/09/Conference-Proceedings-September-14-15-2023.pdf
15. Tan, W. C. (2023). Unstructured and structured data: Can we have the best of both worlds with large language models? arXiv. https://doi.org/10.48550/arXiv.2304.13010
16. Tao, J., Wang, P., Jiang, H., & Han, Y. (2026). A knowledge-formalization and vision-language-model system for fire emergency decision support. *Advanced Engineering Informatics*, vol. 71, part C, article ID 104409. https://doi.org/10.1016/j.aei.2026.104409

O. R. Staso, N. Ye. Burak

Lviv State University of Life Safety, Lviv, Ukraine

CHALLENGES AND PROSPECTS OF USING UNSTRUCTURED DATA

This study examines the nature of unstructured data, the challenges associated with its processing, and its potential role in supporting decision-making in emergency management contexts. The study applies methods of analysis, systematization, comparison, and synthesis of recent scientific publications, technical reports, and practical solutions in the field of big data analytics. The study systematically explores how heterogeneous data sources – such as textual reports, sensor data streams, and social media content – contribute to the growing volume of unstructured information and evaluates their analytical limitations. The conducted analysis indicates that conventional data mining and knowledge discovery techniques, which are primarily designed for structured datasets, have limited effectiveness in extracting, structuring, and analytically using unstructured information due to its semantic ambiguity, lack of standardized formats, and high levels of informational noise. Moreover, the study identifies key barriers to effective use of unstructured data in analytical systems, including insufficient data quality, scalability constraints, and the lack of unified processing frameworks. For example, emergency response systems often rely on fragmented and inconsistently formatted data streams, which, without prior structuring, normalization, and the use of specialized process and data analysis methods, complicates operational processing and may reduce the justification of managerial decisions. Consequently, the research highlights the need for integrated approaches to processing, structuring, and analytically using unstructured data that combine natural language processing, machine learning, and logical reasoning techniques to transform raw data into actionable knowledge. Furthermore, recent advances in large language models are shown to significantly enhance the extraction, classification, and structuring of information from unstructured sources. As a result, the benefits of this study lie in establishing a conceptual foundation for developing adaptive analytical systems capable of handling complex, heterogeneous data environments. The study concludes that future research should focus on designing scalable architectures for information-analytical systems and hybrid approaches to processing unstructured data to improve the efficiency and reliability of decision support systems in safety-critical domains.

Keywords: big data analytics; knowledge extraction; decision support systems; emergency management; data processing.