

УДК 004.896:519.687.3

DOI: 10.18372/2073-4751.79.19368

Рябий М.О., к.т.н.,

orcid.org/0000-0002-9651-9135,

e-mail: m.o.ryabyu@gmail.com,

Спірідонов А.Ю.,

orcid.org/0009-0005-4071-1733,

e-mail: 2386932@stud.nau.edu.ua

ОГЛЯД МЕТОДІВ ЗБОРУ ТА АНАЛІЗУ ДАНИХ В СИСТЕМАХ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ІНФОРМАЦІЇ

Національний авіаційний університет

Вступ

За останні десятиліття інтелектуальна обробка інформації відзначилася значним прогресом, що великим чином обумовлено розширенням доступу до великих обсягів даних та стрімким зростанням обчислювальної потужності комп'ютерів. Ця галузь технологій залишається однією з найбільш динамічних і перспективних, маючи потенціал впливати на різноманітні аспекти нашого повсякденного життя. Проте, основною проблематикою в розвитку інтелектуальної обробки інформації (ІОІ) залишається питання чистоти та достатності даних.

Для подолання цих викликів необхідно розвивати ефективні методи збору, очищення та перевірки якості даних. Також важливо розглядати можливості розширення джерел інформації та вдосконалення методів її аналізу для забезпечення максимальної достовірності та репрезентативності результатів інтелектуальної обробки даних.

Викладення основного матеріалу

На сьогодні існує безліч сервісів збору та обробки інформації, які можна успішно використовувати у різних сферах, включаючи маркетинг, юриспруденцію, медицину та інші. З урахуванням сучасної ситуації в країні, технології обробки даних мають потенціал для застосування у військовій сфері. Наприклад, вони можуть бути використані як системи пошуку та аналізу дезінформації та пропаганди, що є критично важливими аспектами в контексті забезпечення національної безпеки.

Військові структури користуються сучасними технологіями обробки даних для забезпечення національної безпеки. Системи аналізу інформації можуть виявляти потенційні загрози та дезінформацію у великому обсязі даних, а також допомагати у прийнятті стратегічних військових рішень [1].

У сучасному маркетингу, рішення для аналізу даних дозволяють підприємствам зрозуміти поведінку своїх клієнтів, виявити тренди на ринку та розробити ефективні стратегії реклами та просування продуктів.

В юриспруденції, інформаційні технології можуть використовуватися для швидкого та ефективного аналізу правової інформації, дослідження судових справ та підготовки адвокатських матеріалів.

У медицині, системи обробки даних допомагають лікарям аналізувати великі медичні набори даних, виявляти патерни у хворобах та лікуванні, а також розробляти індивідуалізовані підходи до лікування пацієнтів [6].

Таким чином, технології обробки даних відіграють важливу роль у багатьох сферах життя, від підприємництва до безпеки нації, і продовжують розвиватися, вносячи значний внесок у суспільний прогрес та ефективність діяльності різних організацій.

Розглянемо рішення які наявні на цей момент (табл. 1).

Для оцінки ефективності збору даних з різних сервісів проведемо експеримент з порівняння можливості отримання даних з різних джерел та реєстрів (рис. 1).

Таблиця 1. Наявні рішення

Тип сервісу	Напрямок застосування	Назва	Джерела	Платний/безкоштовний	Доступність	Опис	Можливість підключення додаткових джерел
Комерційний	Юриспруденція	<i>Youcontrol</i>	Державні реєстри	Платний	SAAS	Пошук даних про юридичну особу	-
		<i>Clarity-project/</i>	Державні реєстри, тендерні майданчики	Платний	SAAS	Пошук даних про юридичну особу	-
	Маркетинг	<i>Looqme.io</i>	Соціальні мережі	Платний	SAAS	пошук згадок у соціальних мережах	-
		<i>Semantic force</i>	Соціальні мережі	Платний	SAAS	пошук згадок у соціальних мережах	-
	Навчання	<i>Scopus</i>	Внутрішня БД+відкриті репозиторії <i>arXiv, SSRN, RePEc</i>	Безкоштовний	SAAS	Пошук наукових статей	-
Державний	Юриспруденція	Державний реєстр ДПС	БД держави	Безкоштовний	SAAS	Пошук даних про юридичну особу	-
		Державний реєстр судових рішень	БД держави	Безкоштовний	SAAS	Пошук даних про юридичну особу	-
		Кадастровий реєстр	БД держави	Безкоштовний	SAAS	Пошук даних про юридичну особу	-
		Реєстр адміністративних послуг	БД держави	Безкоштовний	SAAS	Пошук даних про юридичну особу	-
	Військова справа	Оберіг	Внутрішні БД	Безкоштовний	-	Електронний реєстр прозвонників	-
		Дія	Внутрішні БД	Безкоштовний	-	Система об'єднання різноманітних реєстрів держави працює інтерфейсом внутрішнього продукту "Трембіта"	-

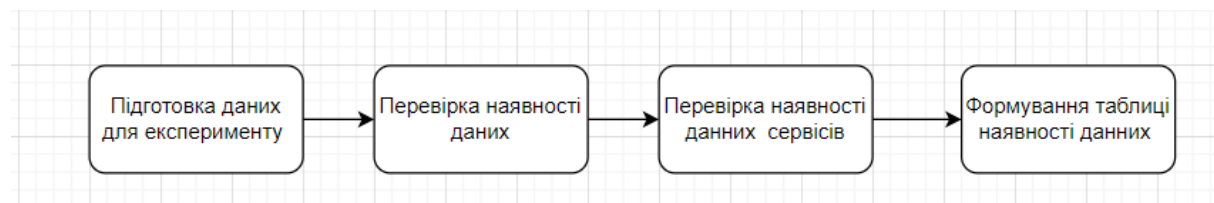


Рис. 1. План проведення експерименту

Вхідними даними слугувала інформація про дві особи, публічна та не публічна: 1) Мустафа Джемільев (далі – Джемільев М.) – політичний та громадський діяч України кримськотатарського походження. Герой України; 2) Спірідонов Антон Юрійович (далі – Спірідонов А.Ю.) – власна фізична особа.

Перелік обраних сервісів для експерименту:

- *youcontrol*;
- державний реєстр судових рішень;
- *clarity-project*;
- *emantic force*;
- *youscan*.

Результати дослідження за напрямком "Юриспруденція" наведено в табл. 2.

Результати дослідження за напрямком "Маркетинг" наведено в табл. 3.

Таблиця 2. Результати за напрямком "Юриспруденція"

Вид даних	Запит, ПІБ	Спірідонов А.Ю.	Джемілев М.	Спірідонов А.Ю.	Джемілев М.	Спірідонов А.Ю.	Джемілев М.
	Сервіс	<i>Youcontrol</i>		Державний реєстр судових рішень		<i>Clarity-project</i>	
Підприємства	Назва	"+"	"+"	"+"	"+"	"+"	"+"
	КВЕД	"+"	"+"	"+"	"+"	"+"	"+"
	Дата реєстрації	"+"	"+"	"+"	"+"	"+"	"+"
	Код ЄДР-ПОУ	"+"	"+"	"+"	"+"	"+"	"+"
	Судові рішення	"+"	"+"	"+"	"+"	"+"	"+"
	Участь у державних закупівлях	"_"	"_"	"_"	"_"	"_"	"_"
Особисті	Прописка	"+"	"+"	"_"	"_"	"+"	"+"
	Телефон	"+"	"_"	"_"	"_"	"+"	"_"
	Нерухомість	"+"		"_"		"+"	
	Транспортний засіб	"+"	"+"	"_"	"_"	"+"	"+"
	Судові рішення	"_"	"_"	"_"	"_"	"_"	"_"

Таблиця 3. Результати дослідження за напрямком "Маркетинг"

Сервіс	Ресурс	Тип даних	Запит	
			Спірідонов А. Ю.	Джемілев М.
<i>Semantic force</i>	<i>FaceBook</i>	Згадки	"+"	"+"
		Публікації	"+"	"+"
		Підписники	"+"	"+"
<i>Youscan</i>	<i>Youtube</i>	Згадки в медіа	"+"	"+"
		Коментарі	"+"	"+"
		Згадки в медіа	"+"	"+"
		Коментарі	"+"	"+"
	<i>FaceBook</i>	Згадки в медіа	"+"	"+"
		Публікації	"+"	"+"
		Підписники	"+"	"+"
		Згадки в медіа	"+"	"+"
		Коментарі	"+"	"+"
		Згадки в медіа	"+"	"+"
<i>Youtube</i>	Коментарі	"+"	"+"	

У рамках огляду (табл. 2-3) також слід відзначити, що більшість інструментів інтелектуальної обробки інформації (IOI) пропонуються у формі *Software as a Service (SaaS)*. Це має свої переваги, зокрема забезпечує зручну доступність та збільшує потужність обробки, проте такий підхід також має свої недоліки [3]. Наприклад, власник даних може втратити контроль над своєю інформацією, оскільки дані передаються третім особам, що може становити проблему з точки зору конфіденційності та безпеки.

Аналіз технологій свідчить про наявність багатьох сервісів, які спеціалі-

зуються на конкретних галузях, але не можуть ефективно взаємодіяти між собою. Одночасно, в сучасному світі набуває популярності передача інформації через месенджери та публічні канали в цих месенджерах, що стає важливим фактором у взаємодії з аудиторією та розповсюдженні контенту [4]. Відсутність даних з різноманітних месенджерів та публічних каналів пов'язана з архітектурними особливостями (рис. 2) систем збору, в класичному вигляді вона виглядає наступним чином:

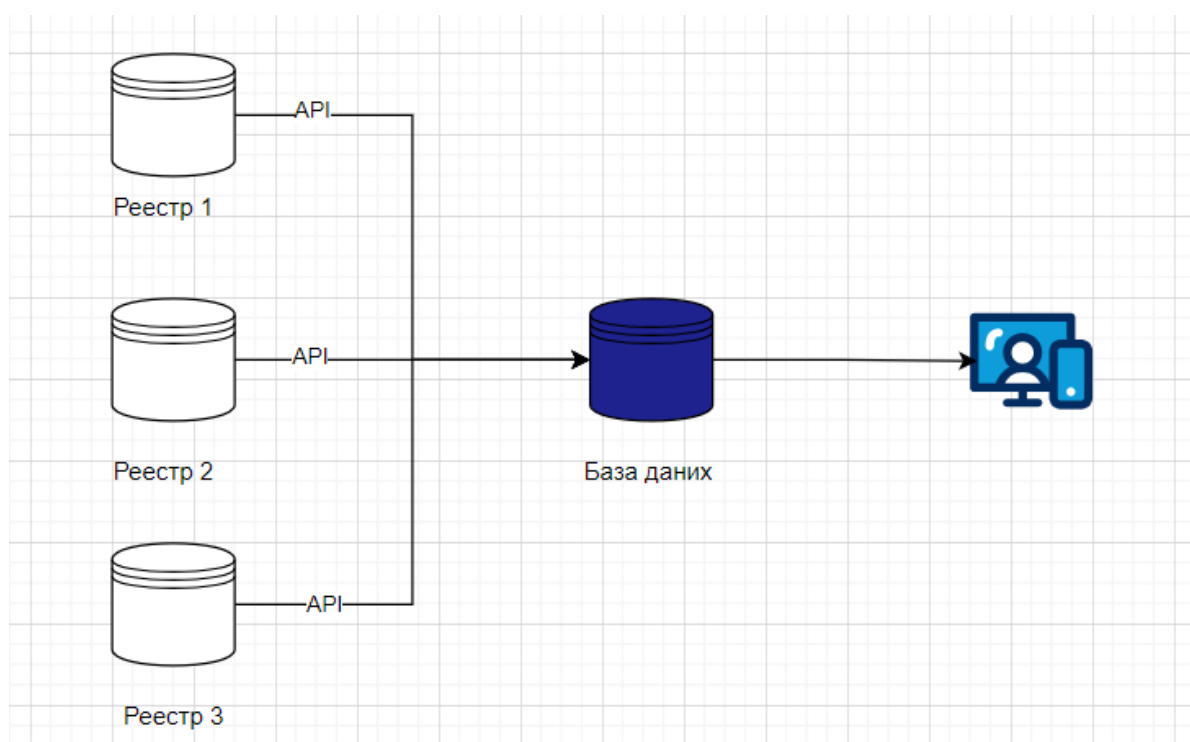


Рис. 2. Приклад стандартної архітектури систем обробки даних

У випадку концептуальної моделі система може мати додаткові джерела вхідних даних, що побудовані на принципах гіперавтоматизації, та проміжні блоки очищення і обробки даних [5] (рис. 3).

Для підтвердження можливостей концептуальної моделі виконали створення прототипу для збору даних с

декількох сервісів, а саме: з сервісу *Youcontrol* та каналу месенджера *Telegram* (ТСН новини). Для реалізації прототипу використовувалися мова програмування *python*, та стек технологій *RPA UIPATH* [2].

Для отримання вхідних параметрів використано *WEB* інтерфейс на рис. 4.

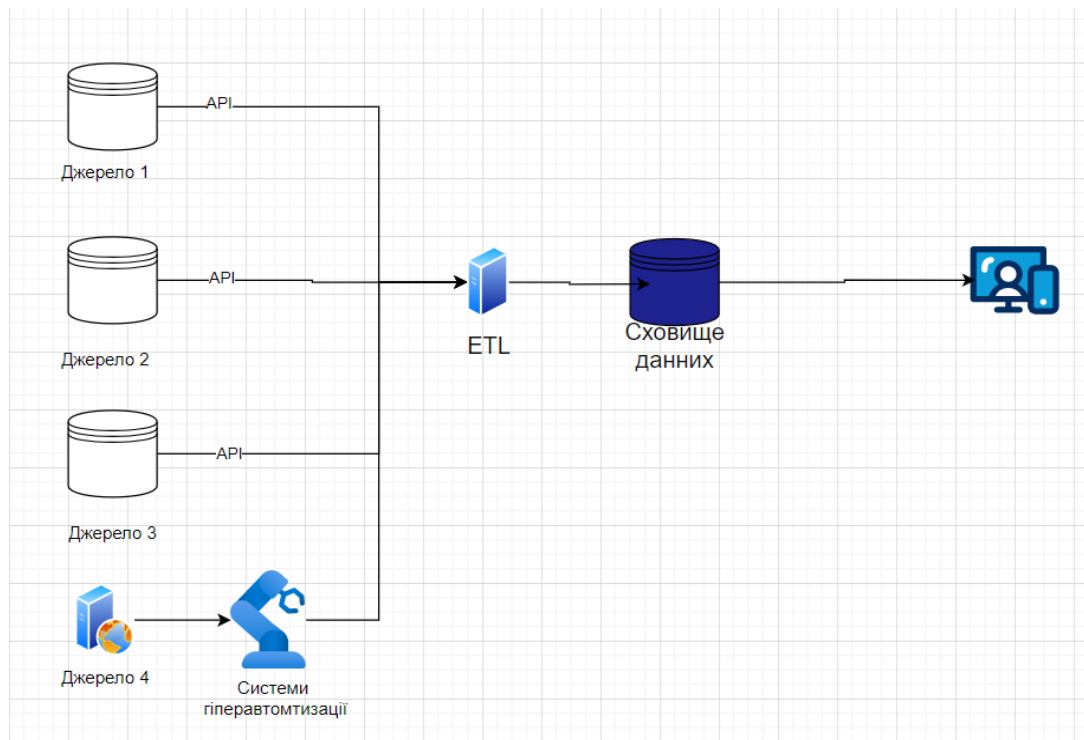


Рис. 3. Запропонована архітектура концептуальної моделі

Рис. 4. Вхідний інтерфейс для початку пошуку інформації

За допомогою *API* отримаємо інформацію з сервісу *Youcontrol* у вигляді *JSON*.

Приклад запити:

<https://api.youscore.com.ua/v1/usr/14308500?showCurrentData=false&apiKey={apiKey}>

Приклад відповіді:

```

{
  "$schema": "http://json-schema.org/draft-07/schema",
  "$id": "/v1/usr/{contractorCode}",
  "type": "object",
  "title": "Єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців. Фізична особа-підприємець",
  "[
    {
      "name": {

```

```

"fullName": "ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ \"ІНТЕРСПОРТ\",
"shortName": "ТОВ \"ІНТЕРСПОРТ\"",
},
"nameInEnglish": null,
"code": "30045171",
"legalPersonName": "\"ІНТЕРСПОРТ\"",
"legalForm": "ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ",
"registrationViaReformation": null,
"branches": null,
"economicActivities": [
{
"code": "68.20",
"description": "Надання в оренду й експлуатацію власного чи орендованого нерухомого майна (основний)"
},
{
"code": "77.39",
"description": "Надання в оренду інших машин, устаткування та товарів. н. в. і. у."
},
{
"code": "93.11",
"description": "Функціонування спортивних споруд"
},
{
"code": "93.29",
"description": "Організування інших видів відпочинку та розваг"
},
{
"code": "46.90",
"description": "Неспеціалізована оптова торгівля"
},
{
"code": "56.10",
"description": "Діяльність ресторанів, надання послуг мобільного харчування"
}
],
"authorityInfo": "Оболонська районна в місті Києві державна адміністрація",
"managingGovernmentAuthority": null,
"founders": [

```

При обробці запиту телеграм система повертає повідомлення з ключовими словами в форматі *JSON*.

Для прикладу використовували запит з ключовим словом “Bradley”.

Приклад відповіді:

```

{
"name": "ТСН новини / ТСН.ua",
"type": "public_channel",
"id": 1305722586,
"messages": [
"text_entities": [
{

```

```

"type": "bold",
"text": "👁️ БМП Bradley проти російського танку Т-80 – хто переможе?"
},
{
"type": "plain",
"text": "\n\nЗвісно перемогли бійці 47-ї ОМБр, які ефектно "
},
{
"type": "bold",
"text": "знищили ворожу техніку"
},
{
"type": "plain",
"text": " за допомогою протитанкової ракети TOW.\n\n"
},
{
"type": "text_link",
"text": "Сайт",
"href": "https://tsn.ua/"
},
{
"type": "text_link",
"text": "Facebook",
"href": "https://www.facebook.com/tsn.ua"
},
{
"type": "plain",
"text": " | "
},
{
"type": "text_link",
"text": "YouTube",
"href": "https://www.youtube.com/tsn"
},
{
"type": "plain",
"text": ""
}
]

```

У відповіді, що наведена вище, відображено результати запитів за ключовими словами, що підтверджує можливість збору і обробки даних з месенджерів. Аналіз результатів показує, що навіть такі різноманітні джерела інформації, як публічні канали в месенджерах, можуть бути інтегровані у систему для побудови більш повної та репрезентативної картини. Це дозволяє виявляти тренди, патерни та аномалії у потоках інформації, що, в свою чергу, є

критично важливим для прийняття ефективних управлінських рішень в умовах швидко змінюваних ринків, правового середовища чи національної безпеки. Таким чином, можна досягти охоплення більшої кількості даних для аналізу та формування розширеної бази даних внутрішнього реєстру системи обробки інформації.

Розширення джерел даних та застосування гіперавтоматизації дозволяє не тільки прискорити процес збору та

обробки інформації, але й підвищити якість аналітики. Інтеграція технологій роботизованої автоматизації процесів з інтелектуальною обробкою інформації (IOI) створює можливості для автоматичного виявлення взаємозв'язків між даними, що раніше було неможливо при традиційних підходах до аналізу. Завдяки цьому можна автоматизувати складні завдання, такі як виявлення дезінформації або прогнозування потенційних загроз у сфері національної безпеки, що значно підвищує рівень готовності та реакції на нові виклики.

Отже, експеримент підтверджує перспективність дослідження в галузі розробки інформаційних технологій на основі поєднання IOI та гіперавтоматизації, що має великий потенціал для створення нових інноваційних рішень. Ці рішення можуть значно поліпшити різні сфери життя та діяльності людей, такі як бізнес, юриспруденція, охорона здоров'я та державне управління, забезпечуючи більш точну і оперативну обробку інформації та прийняття рішень.

Висновки

Після аналізу наявних технологій інтелектуальної обробки інформації (IOI) виявлені як їх сильні, так і слабкі сторони. Спираючись на можливе поєднання технологій IOI та гіперавтоматизації, можна зробити висновок про значний потенціал для подальших досліджень в області розробки інформаційних систем, що базуються на гіперавтоматизації.

Для успішної реалізації поставленої задачі пропонується комплексне дослідження наступних технологій:

1. Системи гіперавтоматизації (зокрема *RPA*): Роботизований процес автоматизації (*RPA*) дозволяє автоматизувати рутинні завдання та процеси, звільняючи людські ресурси для виконання більш складних завдань. Це може бути корисно для автоматизації певних операцій в процесі обробки даних та їх інтегрування в систему.

2. Суміжні бази даних: Використання суміжних баз даних забезпечить доступність та швидкість обміну даними між

різними системами та додатками. Це стане основою для інтеграції різних сервісів у єдине глобальне рішення та оптимізації процесів збору, зберігання і обробки інформації.

3. Інтелектуальна обробка інформації (IOI): Ця технологія дозволяє автоматизувати процеси аналізу, класифікації та інтерпретації великих обсягів даних. Вона забезпечує ефективний збір та обробку інформації з месенджерів та інших джерел, а також виявлення патернів і трендів для подальшого використання у прийнятті рішень.

Комплексне використання цих технологій дозволить створити потужну та ефективну систему, яка забезпечить збір, обробку та аналіз інформації з месенджерів та інших джерел, а також автоматизує рутинні операції для підвищення продуктивності та якості прийняття рішень.

Література

1. Ленков С. В. та ін. Концептуальна схема системи інтелектуальної обробки даних. *Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка*. 2014. Вип. 46. С. 181–190.

2. Council G. Injecting (artificial) intelligence into robotic process automation. URL : <http://www.datacenterjournal.com/injecting-artificialintelligence-robotic-process-automation/>.

3. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг) : навч. посібник. Київ : КНЕУ, 2007. 376 с.

4. Bazzel M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. 2012. 264 p.

5. Bornet P., Barkin I., Wirtz J. Intelligent Automation: Learn How to Harness Artificial Intelligence to Boost Business & Make Our World More Human. 1st ed. 2020. 432 p.

6. Mayer-Schönberger V., Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. 1st ed. New York : Houghton Mifflin Harcourt, 2013. 256 p.

Рябий М.О., Спірідонов А.Ю.

ОГЛЯД МЕТОДІВ ЗБОРУ ТА АНАЛІЗУ ДАНИХ В СИСТЕМАХ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ІНФОРМАЦІЇ

У статті розглянуто сучасні методи збору та аналізу даних у системах інтелектуальної обробки інформації (ІОІ). Проведено огляд доступних сервісів, їхніх особливостей, переваг і недоліків у різних галузях, таких як маркетинг, юриспруденція, медицина та військова справа. Особливу увагу приділено інтеграції даних із месенджерів і публічних каналів. У статті запропоновано підхід, що передбачає об'єднання класичних сервісів збору даних із джерелами, такими як Telegram, Facebook, YouTube, для побудови повнішої та репрезентативнішої інформаційної картини. Результати експериментального дослідження показали, що навіть нестандартні джерела інформації можуть бути ефективно інтегровані в аналітичні системи для автоматизованого виявлення патернів, трендів і аномалій. Запропоновано концептуальну модель, що інтегрує гіперавтоматизацію з ІОІ для покращення ефективності збору, обробки та аналізу даних. Експериментальне дослідження підтвердило доцільність впровадження таких рішень, зокрема для виявлення дезінформації, прогнозування загроз і автоматизації рутинних завдань. Стаття підкреслює перспективи розвитку гіперавтоматизованих систем в умовах швидко змінюваного інформаційного середовища.

Ключові слова: інтелектуальна обробка інформації; гіперавтоматизація; аналіз даних; системи моніторингу; інтеграція; автоматизація; дезінформація.

Ryabyu M.O., Spiridonov A.Y.

OVERVIEW OF DATA COLLECTION AND ANALYSIS METHODS IN INTELLIGENT INFORMATION PROCESSING SYSTEMS

This article examines modern methods of data collection and analysis in intelligent information processing (IIP) systems. A review of available services, their features, advantages, and disadvantages across various fields such as marketing, jurisprudence, medicine, and military affairs is provided. Special attention is paid to the integration of data from messengers and public channels. The article proposes an approach that combines traditional data collection services with sources such as Telegram, Facebook, and YouTube to build a more comprehensive and representative informational framework.

The results of experimental research demonstrate that even non-standard information sources can be effectively integrated into analytical systems for automated detection of patterns, trends, and anomalies. A conceptual model is proposed, which integrates hyper-automation with IIP to improve the efficiency of data collection, processing, and analysis. Experimental research confirms the feasibility of implementing such solutions, particularly for detecting disinformation, forecasting threats, and automating routine tasks.

The article highlights the prospects for developing hyper-automated systems in the context of a rapidly evolving information environment.

Keywords: intelligent information processing; hyper-automation; data analysis; monitoring systems; integration; automation; disinformation.