

Review of Data Collection and Analysis Methods in Intelligent Information Processing Systems*

Myroslav Ryabyy^{1,2,†}, Anton Spiridonov^{2,†}, Nataliia Korshun^{3,†} and Roman Kyrychok^{3,*†}

¹ State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, 3/6 M. Zaliznyaka str., 03142 Kyiv, Ukraine

² National Aviation University, 1 Liubomyra Huzara ave., 03058 Kyiv, Ukraine

³ Borys Grinchenko Kyiv Metropolitan University, 18/2 Bulvarno-Kudriavska str., 04053 Kyiv, Ukraine

Abstract

This paper examines modern methods of data collection and analysis in intelligent information processing (IIP) systems, emphasizing the growing importance of integrating non-traditional data sources into analytical frameworks. A comprehensive review of available services is conducted, highlighting their features, advantages, and limitations across various domains, including marketing, jurisprudence, medicine, and military applications. The study particularly focuses on the integration of data from messengers and public communication channels, given their increasing role in disseminating real-time information. The proposed approach combines traditional data collection services with alternative sources such as Telegram, Facebook, and YouTube, offering a more holistic and representative information environment. This integration facilitates the automated detection of patterns, trends, and anomalies, thereby enhancing decision-making processes in dynamic and data-intensive sectors. The results of experimental research confirm the viability of utilizing even unconventional information sources in analytical systems, demonstrating their effectiveness in detecting disinformation, forecasting potential threats, and automating routine analytical tasks. Furthermore, the paper introduces a conceptual model that integrates hyper-automation technologies with IIP to optimize data collection, preprocessing, and analysis. The model leverages robotic process automation (RPA) and artificial intelligence (AI)-driven classification techniques to enhance efficiency and scalability. Experimental validation of the proposed model demonstrates its potential for real-world implementation, particularly in scenarios requiring rapid adaptation to evolving information landscapes. The findings underscore the significance of hyper-automated systems in addressing contemporary challenges in data intelligence. By improving the accuracy, speed, and adaptability of information processing, such systems hold substantial promise for applications in cybersecurity, regulatory compliance, business intelligence, and public sector decision-making. The study concludes with insights into the future development of hyper-automated data processing frameworks and their role in shaping next-generation analytical capabilities.

Keywords

intelligent information processing, hyper-automation, data analysis, monitoring systems, integration, automation, disinformation

1. Introduction

In recent decades, intelligent information processing has made significant progress, largely driven by increased access to large volumes of data and rapid growth in computational power. This field remains one of the most dynamic and promising, with the potential to impact various aspects of everyday life. However, a major challenge in the development of intelligent information processing (IIP) remains the issue of data purity and sufficiency.

To overcome these challenges, it is necessary to develop effective methods for data collection, cleansing, and quality verification. It is also important to explore the possibilities of expanding information sources and improving analysis methods to ensure the maximum accuracy and representativeness of data processing results.

*CPITS 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2025, Kyiv, Ukraine

†Corresponding author.

†These authors contributed equally.

✉ m.o.ryabyy@gmail.com (M. Ryabyy); 2386932@stud.nau.edu.ua (A. Spiridonov); n.korshun@kubg.edu.ua (N. Korshun); r.kyrychok@kubg.edu.ua (R. Kyrychok)

ORCID 0000-0002-9651-9135 (M. Ryabyy); 0009-0005-4071-1733 (A. Spiridonov); 0000-0003-2908-970X (N. Korshun); 0000-0002-9919-9691 (R. Kyrychok)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Main content

Today, numerous services for data collection and processing can be successfully used in various fields, including marketing, law, medicine, and others. Considering the current geopolitical situation, data processing technologies have significant potential for military applications. For example, they can be utilized in systems for detecting and analyzing disinformation and propaganda, which are critical aspects of national security.

Military organizations use modern data processing technologies to ensure national security. Information analysis systems can identify potential threats and disinformation in vast data sets and assist in strategic military decision-making [1–3].

In modern marketing, data analysis solutions enable businesses to understand customer behavior, identify market trends, and develop effective advertising and product promotion strategies [9].

In law, information technologies can be used for the rapid and efficient analysis of legal information, legal case studies, and the preparation of legal documentation.

In medicine, data processing systems assist doctors in analyzing large medical datasets, identifying disease patterns, and developing individualized treatment approaches [6].

Thus, data processing technologies play a crucial role in many areas of life, from business to national security, and continue to evolve, making a significant contribution to social progress and the efficiency of various organizations.

3. Existing solutions

A review of available solutions (Table 1) was conducted to evaluate the effectiveness of data collection services from various sources. An experiment was carried out to compare data retrieval capabilities from different registries and services (Fig. 1).

Table 1
Available Solutions

Service Type	Application Area	Name	Sources	Paid/Free	Availability	Description
Commercial	Jurisprudence	Youcontrol	State registers	Paid	SaaS	Search for data about a legal entity
		Clarity-project	State registers, tender platforms	Paid	SaaS	Search for data about a legal entity
	Marketing	Looqme.io	social networks	Paid	SaaS	Search for mentions in social networks
		Semantic force	social networks	Paid	SaaS	Search for mentions in social networks
State	Teaching	Scopus	Internal database + open repositories	Free	SaaS	Search for scientific articles
	Jurisprudence	State Register of State Tax Service	DB of the state	Free	SaaS	Search for data about a legal entity
		State Register of Court Decisions	DB of the state	Free	SaaS	Search for data about a legal entity
		Cadastral	DB of the state	Free	SaaS	Search for data

	register				about a legal entity
	Register of administrative services	DB of the state	Free	SaaS	Search for data about a legal entity
Soldiery	Oberig	Internal DB	Free	—	Electronic register of conscripts
	Dia	Internal DB	Free	—	The system for unifying various state registers operates as an interface to the internal product “Trembita”



Figure 1: Experiment plan

Experimental Research

The study involved two individuals, one public and one private:

1. Mustafa Dzhemilev—a political and public figure of Ukraine of Crimean Tatar descent, Hero of Ukraine.
2. Anton Spiridonov—a private person.

Selected services for the experiment:

1. YouControl
2. State Register of Court Decisions
3. Clarity-project
4. Semantic Force
5. YouScan.

The research findings in the legal and marketing sectors are presented in Table 2 and Table 3, respectively.

Table 2

Legal Data Retrieval Result

Type of data	Request, full name	Spiridonov	Dzhemilev	Spiridonov	Dzhemilev	Spiridonov	Dzhemilev
	Service	Youcontrol	State Register of Court Decisions		Clarity-project		
Enterprises	Name	+	+	+	+	+	+
	KVED	+	+	+	+	+	+
	Date of registration	+	+	+	+	+	+
	VAT code	+	+	+	+	+	+
	Court decisions	+	+	+	+	+	+
	Participation in public procurement	—	—	—	—	+	+

Personal	Registration	+	+	–	–	+	+
	Phone	+	–	–	–	+	–
	Real estate	+	–	–	–	+	–
	Vehicle	+	+	–	–	+	+
	Court decisions	–	–	–	–	–	–

Table 3
Marketing Data Retrieval Results

Service	Resource	Data type	Request	
			Spiridonov A.	Dzhemilev M.
Semantic force	FaceBook	Mentions	“+”	“+”
		Publications	“+”	“+”
		Subscribers	“+”	“+”
		Mentions in the media	“+”	“+”
		Comments	“+”	“+”
Youscan	Youtube	Mentions in the media	“+”	“+”
		Comments	“+”	“+”
		Mentions in the media	“+”	“+”
		Comments	“+”	“+”
		Mentions in the media	“+”	“+”
	FaceBook	Publications	“+”	“+”
		Subscribers	“+”	“+”
		Mentions in the media	“+”	“+”
		Comments	“+”	“+”
		Mentions in the media	“+”	“+”
	Youtube	Comments	“+”	“+”

4. Key findings and proposed model

The study highlighted that most intelligent information processing (IIP) tools are offered as Software as a Service (SaaS). While SaaS solutions provide convenience and enhanced processing power, they also have drawbacks [3]. For example, data owners may lose control over their information as data is transferred to third parties, raising concerns about confidentiality and security.

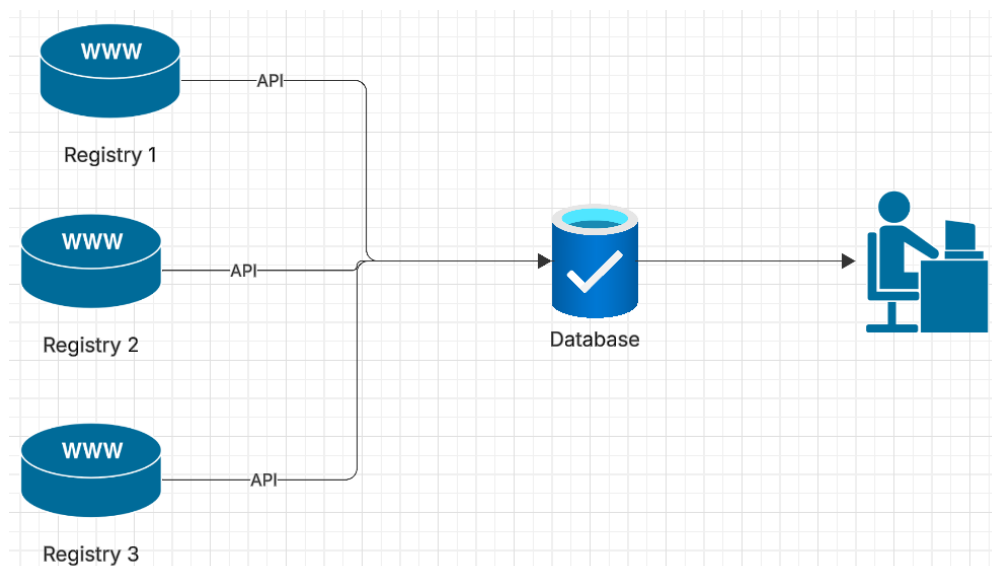


Figure 2: Standard Data Processing System Architecture

The analysis of technologies shows that many services specialize in specific domains but lack efficient interaction with each other [4]. At the same time, the increasing use of messengers and public channels as information transmission tools emphasizes the need for integrating data from these sources.

A conceptual model with additional data sources based on hyperautomation principles and intermediate data cleansing and processing blocks has been proposed [5] (Fig. 3).

To validate the capabilities of the conceptual model, a prototype was developed for data collection from multiple services, including YouControl and the Telegram news channel (TSN News). The prototype was implemented using Python and the RPA UiPath [2] technology stack.

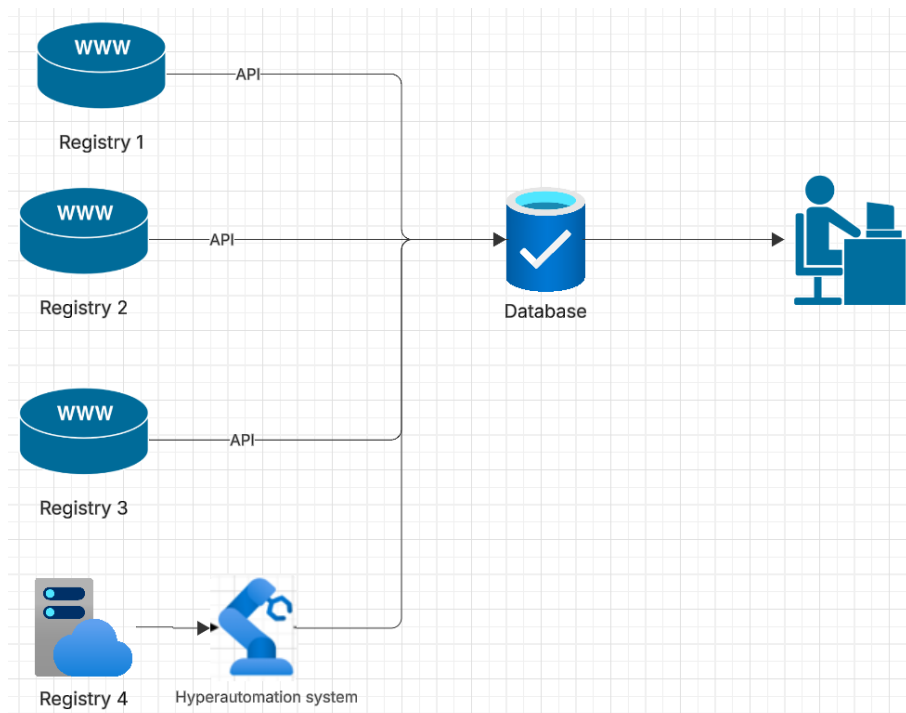


Figure 3: Proposed conceptual model architecture

Figure 4: Input interface to start searching for information

An API request retrieved information from YouControl in JSON format.

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

Example request:

```

{
  "$schema": "http://json-schema.org/draft-07/schema",
  "$id": "/v1/usr/{contractorCode}",
  "type": "object",
  "title": "Unified State Register of Legal Entities and Individual Entrepreneurs. Individual Entrepreneur",

```

```

" [
  {
    "name": {
      "fullName": "LIMITED LIABILITY COMPANY \"INTERSPORT\"",
      "shortName": "LLC \"INTERSPORT\""
    },
    "nameInEnglish": null,
    "code": "30045171",
    "legalPersonName": "\"INTERSPORT\"",
    "legalForm": "LIMITED LIABILITY COMPANY",
    "registrationViaReformation": null,
    "branches": null,
    "economicActivities": [
      {
        "code": "68.20",
        "description": "Renting and operating of own or leased real estate (main activity)"
      },
      {
        "code": "77.39",
        "description": "Rental of other machinery, equipment, and tangible goods not
elsewhere classified"
      },
      {
        "code": "93.11",
        "description": "Operation of sports facilities"
      },
      {
        "code": "93.29",
        "description": "Other amusement and recreation activities"
      },
      {
        "code": "46.90",
        "description": "Non-specialized wholesale trade"
      },
      {
        "code": "56.10",
        "description": "Restaurant activities and mobile food service activities"
      }
    ],
    "authorityInfo": "Obolon District State Administration in Kyiv",
    "managingGovernmentAuthority": null,
    "founders": [

```

When processing a Telegram query, the system returns messages containing keywords in JSON format. For example, a query was made using the keyword “Bradley”.

Example response:

```

{
  "name": "TSN News / TSN.ua",
  "type": "public_channel",
  "id": 1305722586,

```

```

"messages": [
  "text_entities": [
    {
      "type": "bold",
      "text": "BMP Bradley vs. Russian tank T-80 — who will win?"
    },
    {
      "type": "plain",
      "text": "\n\nOf course, the fighters of the 47th Mechanized Brigade won, who
spectacularly "
    },
    {
      "type": "bold",
      "text": "destroyed enemy equipment"
    },
    {
      "type": "plain",
      "text": " using a TOW anti-tank missile.\n\n"
    },
    {
      "type": "text_link",
      "text": "Website",
      "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": ""
    }
  ]
}

```

The response presented above displays the results of queries by keywords, confirming the feasibility of collecting and processing data from messengers. The analysis of results indicates that even diverse information sources, such as public channels in messengers, can be integrated into a system to construct a more comprehensive and representative information model. This enables the identification of trends, patterns, and anomalies in data streams, which is crucial for making effective management decisions in rapidly changing markets, legal environments, or national

security contexts. Thus, a broader data coverage can be achieved for analysis and the development of an extended internal database for an information processing system.

Expanding data sources and applying hyper-automation not only accelerates the process of data collection and processing but also enhances analytical quality. The integration of robotic process automation (RPA) technologies with intelligent information processing (IIP) enables the automatic detection of relationships between data that were previously unattainable using traditional analytical approaches. This allows for the automation of complex tasks such as detecting disinformation or predicting potential threats to national security, significantly improving readiness and response to new challenges [7–11].

Thus, the experiment confirms the potential for research in the field of information technology development based on the combination of IIP and hyper-automation, which holds great promise for the creation of new innovative solutions. These solutions can significantly enhance various aspects of life and business, such as commerce, law, healthcare, and government administration, ensuring more accurate and timely information processing and decision-making [12–14].

Conclusions

The research confirms the relevance of developing information technologies that integrate IIP with hyper-automation, unlocking significant potential for new innovative solutions. These solutions can improve various fields such as business, law, healthcare, and public administration by enabling more accurate and timely data processing and decision-making [8].

For successful implementation, the following technologies should be explored:

- Hyperautomation Systems (RPA): Automates routine tasks and processes, freeing human resources for more complex activities. Useful for automating data processing operations and integration.
- Interconnected Databases: Facilitates fast data exchange across systems and applications, forming the foundation for integrating diverse services into a unified global solution.
- Intelligent Information Processing (IIP): Enables automated analysis, classification, and interpretation of large data volumes, ensuring efficient collection and processing of information from messengers and other sources.

A comprehensive approach leveraging these technologies will allow the creation of a robust and efficient system capable of collecting, processing, and analyzing information from diverse sources while automating routine operations to enhance productivity and decision-making quality.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

- [1] S. Lenkov, et al., Conceptual scheme of an intelligent data processing system, Collection of scientific works of the Military Institute of the Taras Shevchenko National University of Kyiv, 46, 2014, 181–190.
- [2] G. Council, Injecting (artificial) intelligence into robotic process automation. URL: <http://www.datacenterjournal.com/injecting-artificialintelligence-robotic-process-automation/>
- [3] V. Sytnyk, M. Krasnyuk, Intelligent data analysis (data mining): Textbook, Kyiv, 2007.
- [4] M. Bazzel, Open source intelligence techniques: Resources for searching and analyzing online information, 2012.

- [5] P. Bornet, I. Barkin, J. Wirtz, Intelligent automation: learn how to harness artificial intelligence to boost business & make our world more human, 1st ed., 2020.
- [6] V. Mayer-Schönberger, K. Cukier, Big Data: A revolution that will transform how we live, work, and think, 1st ed., New York: Houghton Mifflin Harcourt, 2013.
- [7] O. Solomentsev, et al., Efficiency of operational data processing for radio electronic equipment, *Aviation* 23(3) (2020) 71–77. doi:10.3846/aviation.2019.11849
- [8] A. Shyian, et al., Development of a software module for modeling the process of semi-automatic recognition of objects on aerial images, in: *Proceedings of the Institute of Applied Mathematics and Mechanics*, 2023.
- [9] O. Zolotukhina, V. Titarenko, Machine learning and forecasting in cyber-physical systems, in: *Control, Optimisation and Analytical Processing of Social Networks*, vol. 2392, 2019, 121–130.
- [10] S. Gnatyuk, Critical aviation information systems cybersecurity, meeting security challenges through data analytics and decision support, *NATO Science for Peace and Security Series, D: Information and Communication Security*, IOS Press Ebooks, vol. 47(3), 2016, 308–316.
- [11] M. Zaliskyi, et al., Method of traffic monitoring for DDoS attacks detection in e-health systems and networks, in: *Informatics & Data-Driven Medicine*, vol. 2255, 2018, 193–204.
- [12] Y. Deville, L. T. Duarte, A. Deville, A taxonomy of relationships between information processing, machine learning and quantum physics: Quantum-inspired, quantum-assisted, quantum-targeted and related approaches, in: *IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, 2024, 361–365. doi:10.1109/M2GARSS57310.2024.10537558
- [13] V. Ivanov, et al., The removal of phosphorus from reject water in a municipal wastewater treatment plant using iron ore, *J. Chem. Technol. Biotechnol.* 84(1) (2009) 78–82. doi:10.1002/jctb.2009
- [14] R Zhang, et al., Exploration of the methods of artificial intelligence computer processing information, in: *3rd Int. Conf. for Innovation in Technology (INOCON)*, 2024, 1–4. doi:10.1109/INOCON60754.2024.10511463