

Tsiutsiura Mykola

DSc (Eng.), Professor, Professor of the Department of Software Engineering and Cybersecurity,

<https://orcid.org/0000-0003-4713-7568>

Kyiv National University of Trade and Economics, Kyiv

Yerukaiev Andrii

PhD (Eng.), Associate Professor, Associate Professor of the Department of Software Engineering and

Cybersecurity, <https://orcid.org/0000-0002-9956-3713>

Kyiv National University of Trade and Economics, Kyiv

Khorolska Karyna

Senior Lecturer of the Department of Software Engineering and Cybersecurity,

<https://orcid.org/0000-0003-3270-4494>

State University of Trade and Economics

Bebeshko Bohdan

Senior Lecturer of the Department of Software Engineering and Cybersecurity,

<https://orcid.org/0000-0001-6599-0808>

State University of Trade and Economics

DATA COLLECTION AND ACCUMULATION OF MEDIA RESOURCES USING ARTIFICIAL INTELLIGENCE

Abstract. *In this difficult time in which our country found itself, information technology has become the force that is successfully used against the aggressive actions of the enemy. The armed forces, in addition to the necessary physical assistance, are quite significantly helped in the information sphere, in particular, from the side of the scientific community. Different directions and branches of computer science are combined to get an effective way to achieve the desired victory and the long-awaited peace and prosperity. In this article, the authors offer their vision in the implementation of technology for evaluating information coming from various media sources. For this, one of the most widely used areas of information technology – artificial intelligence – was involved in the research. As you know, AI includes a sufficiently wide and complete set of models and methods for working out various practical problems. In particular, the authors focused on "soft" calculations, that is, the use of neural networks, fuzzy sets, and genetic algorithms in combination with methods from another area of AI – intelligent data analysis. All the above-mentioned ideas are suggested to be considered in the rationale before creating a scientific project. In this article, foreign and domestic sources will be analyzed in detail in order to identify common and distinctive features with the author's ideas. Particular attention will be paid to those components that are not disclosed in other works and are proposed for development in this project. The object, the subject of research, the goal will also be indicated, the scientific and practical novelty of the implementation of the ideas disclosed in this work will be revealed. In addition, the advantages of using the proposed methods for solving not only the task of processing and evaluating information from various media sources will be described, but also the benefit of their application in other critical military tasks will be shown. Thus, the authors set themselves the goal of helping our defenders implement an approach based on the application of artificial intelligence. This article covers only proposals with their detailed scientific justification. So, summarizing all of the above, the work reveals a theoretical-applied approach at the level of revealing the basic ideas for creating a project. In the following works by the authors, after the presentation of the research base, more attention will be paid to already mathematical methods of description and coverage of the conducted experimental studies.*

Keywords: *artificial Intelligence; intelligent data analysis; fuzzy systems; evolutionary programming*

Introduction

To date, the country has found itself in a difficult state, and not at all through its own fault. Everyone faced new tasks, which citizens were forced to solve jointly in

order to steadily bring the current dangerous state to the indicators of 2013, and in parallel to this, to imitate the world level of free, strong, democratic, developed, independent countries as much as possible. The group of authors of this project is invited to contribute to this

desirable and expected goal for every Ukrainian. The proposed approach does not cover something global, abstract, but proposes to focus on one of the concrete, current, urgent problems, namely, intellectual analysis and processing of media data as active assistance to intelligence services. Of course, this is not enough for the needs of the army, but the authors are deeply convinced that the processing of significant volumes of data surrounding each of us from all sides of available media sources using the most promising scientific and applied direction - artificial intelligence, will be a non-passive help to the state in order to expel the enemy as soon as possible and ensure peace and economic development on the territory of Ukraine.

The purpose of the article

The fundamental problem of the research is the expansion of scientific knowledge of society in terms of the practical use of intelligence data of the work of special services.

The object of the scientific work is the information model of data processing from various media resources, which is planned to be studied with the help of practical, and as an addition to them, theoretical, state-of-the-art methods of artificial intelligence and intelligent data analysis.

The subject of the research work is the structure of the model of the system under investigation, that is, the special processing of the received data (for example, their clustering by a neural network) according to clearly established indicators, its so-called filtering, and the patterns of development of the system model in the environment of a military situation.

Main material

In the process of analyzing the results obtained by the authors of the project by direction, problem, topic, object and subject of research, the following markers were established, which determine what exactly the contribution of the mentioned scientists is and why their work needs to be continued, supplemented, and improved [1; 14]:

1. The research problem and task were formed: the definition of a specific problem or question that needs to be studied and analyzed (improvement of the algorithm for collecting and analyzing information from media resources using artificial intelligence).

2. Conducted a detailed literature review: analysis of existing research and works in the relevant area, including information technology, artificial intelligence and media studies.

3. Research methodology is presented: determination of the methodological approach to conducting research (choice of data collection and analysis methods, use of machine learning algorithms, etc.).

4. Data collection and analysis was carried out: implementation of the data collection process from media resources (website parsing, API data collection sessions, creation of algorithms for collecting text, video and audio information from media resources, development of own tools for information accumulation) ; analysis and processing of collected data (use of analytical methods, statistics, machine learning to obtain valuable information from data).

5. The development and improvement of intelligent data processing algorithms has been proven: work has been carried out on artificial intelligence algorithms that will improve the collection and analysis of media data (creation of new algorithms for text processing, video classification, image analysis).

In addition, an analysis of the results obtained by other scientists in recent years was carried out with reference to specific publications [3-12], and:

1. The authors have identified an innovative and interdisciplinary approach for conducting project research, which combines elements of information technology, machine learning, and media research. This approach can be called an information-intellectual approach to the analysis of media data using artificial intelligence. The novelty of this approach lies in the fact that it combines important aspects of the modern information ecosystem - media, artificial intelligence and data - to solve complex social and technological tasks. It allows not only to make the media more accessible and informative, but also helps to solve socio-cultural issues related to the consumption of media content and its impact on society.

2. The experiment and validation are clearly presented, for which the experiments and validation of the developed algorithms were carried out (determining the efficiency and accuracy of the results, comparing them with existing approaches).

3. The authors did not ignore the ethical aspect, namely, they paid special attention to the consideration of ethical issues related to the collection and use of media data, in particular, from the point of view of confidentiality, privacy and secrecy of data (state secrets).

Therefore, it is possible to draw such an intermediate conclusion that the formulation of the conclusions of these authors based on the results of the research they provided, when providing recommendations for practical implementation or further research, have a good scientific basis, but insufficient for implementation.

As part of the performed studies of identification of parameters of influence on the processes of data collection and accumulation, scientific results based on ideas and working hypotheses are expected [14]:

1. Apply the method of successive concessions to determine the priorities of the impact parameters in the fuzzy multi-criteria assessment of real and forecasted data.

2. At the analysis stage, develop a SWOT analysis model in combination with fuzzy systems methods to obtain a fuzzy model for evaluating the selection of a set of data.

3. To build a mathematical model in tree notation for making optimal decisions under conditions of entropy minimization based on Data Mining methods

4. Build a model of an expert system based on production rules in the description of the genetic algorithm method.

5. To improve the method of genetic evaluation to replace the opinions of experts in the formation of a fuzzy model of the evaluation of parameters influencing the selection of a set of data.

Thus, the authors see the goal of the project in the following way: the development of intelligent information technology for the collection and accumulation of media resource data, which will allow to increase the efficiency of the processes of using data from unmanned aerial vehicles and improve the quality of decision-making regarding the selection of the most suitable structured and correct data in the conditions of martial law and recovery of Ukraine in the post-war period.

The authors defined the main tasks to be solved by the project [2; 13 – 14]:

1. Development of a general structural and functional scheme of software for collecting and accumulating big data for unmanned systems of domestic production.

2. Justification of the composition of technical means for collecting and monitoring the environment by unmanned systems.

3. Conduct an analysis of existing information technologies used to determine the suitability of specific media resources for existing information systems.

4. To carry out a systematic formalization of various parameters of influence on the choice of a technological approach to working with information.

5. Development of a complex sensor for monitoring and measuring the volume of useful information for UAVs;

6. Creation of information and analytical subsystem software using artificial intelligence and remote data processing technologies to collect and display important information.

7. Development and integration of algorithms for the protection of information in the subsystem of collection, transmission and accumulation of data on the results of monitoring the parameters of the state of the UAV.

8. Develop integrated models and apply information technology research methods to the development of a system for identifying parameters of influence on the accumulation of media resource data using the analytical apparatus of fuzzy sets according to the implemented scheme of strategic development.

9. To conduct experimental studies of the operation of the automated system for identifying the parameters of influence on the selection of these media resources under the conditions of its operation by the means of information, analytical and software developed in the work.

As the events of recent years show, almost the entire territory of Ukraine is a risk zone from the point of view of the emergence of natural and man-made emergencies. An important issue is the development of new information technologies that will allow remote and prompt expert assessment of the risks of emergency situations due to the lack of accumulated data.

Innovative and interdisciplinary approaches combining elements of information technology, machine learning, and media research can be identified for project research. This approach can be called an information-intellectual approach to the analysis of media data using artificial intelligence.

The novelty of this approach lies in the fact that it combines important aspects of the modern information ecosystem - media, artificial intelligence and data - to solve complex social and technological tasks. It allows not only to make the media more accessible and informative, but also helps to solve socio-cultural issues related to the consumption of media content and its impact on society.

The main components and structural features of the research project should include:

1. Formulation of the problem and research task: definition of a specific problem or question that needs to be studied and analyzed (improvement of the algorithm for collecting and analyzing information from media resources using artificial intelligence).

2. Literature review: analysis of existing research and works in the relevant field, including information technology, artificial intelligence and media studies.

3. Research methodology: determination of the methodological approach to conducting research (choice of data collection and analysis methods, use of machine learning algorithms, etc.).

4. Data collection and analysis: implementing the process of data collection from media resources (parsing websites, API data collection sessions, creating algorithms for collecting text, video and audio information from media resources, developing own tools for information accumulation); analysis and processing of collected data (use of analytical methods, statistics, machine learning to obtain valuable information from data).

5. Development and improvement of algorithms for intelligent data processing: work on artificial intelligence algorithms that will improve the collection and analysis of media data (creation of new algorithms for text processing, video classification, image analysis).

6. Experiment and validation: conducting experiments and validating the developed algorithms

(determining the efficiency and accuracy of the results, comparing them with existing approaches).

7. Ethical aspect: consideration of ethical issues related to the collection and use of media data, in particular, from the point of view of confidentiality, privacy and secrecy of data (state secrets).

8. Conclusions and recommendations: formulating conclusions based on research results and providing recommendations for practical implementation or further research.

9. Publication of research results: publication of scientific articles, report abstracts, and/or report reflecting research results (interaction with academic and professional communities to exchange information and receive feedback).

The material and technical base of the State University of Trade and Economics meets international standards for conducting scientific research (computer equipment, software, laboratory equipment). Auditoriums are equipped with demonstration equipment, there are 50 computer rooms. The library complex "SMART-library" is functioning with computers, virtual reality helmets and personal gadgets. The University has registered on the Research4Life platform, which provides the opportunity for scientific and pedagogical staff of the University and students to get free access to the resources of leading publishers of scientific literature, including Elsevier, Springer Nature, etc.

At the department where the project is planned to be developed, specialized laboratories "Development and testing of software" and "Cyber security" have been set up. The laboratories include 39 modern computers and specialized equipment: Controlled switch VM-1100S 25 Rm; Managed router D-Link DES-3200-52; MikroTik hAP ac lite router; Cloner Clone 2 Bay HDD Docking Station; Wireless video content management device; IP video camera, etc. Installed special software: Cisco Packet Tracer, PRTG Network Monitor, RAdmin, Recover my files, Recovery tools, USB data recovery, Easeus data recovery, VPN, Virtual Box, HDD Low Level Format Tool.

Detailed content of expected results:

1. The model being created must process data from various media resources, i.e. this includes audio and video data obtained from the appropriate media;

2. Data of different formats will be brought to a common format for the purpose of their direct processing;

3. The processing itself will be performed using methods of artificial intelligence and intelligent data analysis;

4. Depending on the applied method, the output will be clusters or classes of data groups that will have a priority scale from the highest to the lowest value for attention in the further work of intelligence services.

Today, in the field of data processing by methods of artificial intelligence and intelligent data analysis, or, as

they are also called, methods of machine learning, the following approaches are most widespread: the use of artificial neural networks with various variants of their architecture, fuzzy logical sets of various types of orders, methods of evolutionary programming. Each of the above methods can also be combined with another method in order to eliminate the disadvantages of each one and strengthen the advantages of the joint combination. This also includes classification methods (decision trees, the support vector method, various regression options), clustering (hierarchical, k-means), ensemble construction, and associative rules. And this is far from a complete list. The authors suggest using exactly fuzzy sets to represent fuzzy logical derivation in combination with genetic algorithm methods. And to determine priorities, use the classification method with extensive use of the technology of building forests on binary search trees.

Since the methods of artificial intelligence will be used for the work, which have a clear, rigorous and proven mathematical apparatus in their work, their totality proposed by the authors will not require such a justification. Much more interested in the benefits of use in practice. Therefore, the authors of this study will focus on this. Indeed, not only the processing of data from media resources will benefit from the proposed approach. After all, data surrounds you from everywhere. Areas such as area assessments for military operations, assessment of safe routes for delivery or shipment of cargo, etc., will also benefit from the model proposed in this project.

The scientific novelty of the results of the scientific work of this project is represented by the following criteria:

a) obtained for the first time: synthesis of artificial intelligence methods and intelligent data analysis, namely fuzzy sets and evolutionary programming;

b) improved: classification method in machine learning based on entropy for determining scale priority;

c) developed: a combination of the first and second points for application in the subject area of data processing from various media resources.

The scientific and technical development in this project is the following idea of scientific and technical activity proposed by the authors, which is based on the knowledge obtained as a result of the use of the researched information model in increasing the effectiveness of intelligence operations by means of the analysis and evaluation of media data, which are selected as a result of research in this in the subject area, and gradually bringing the acquired knowledge to the stage of their practical use in further scientific works and projects.

After analyzing existing analogs in world science in specific publications, the authors reached the following conclusions [6; 7 – 10]:

a) the vast majority of publications contain only general features, the information provided does not provide specific practical examples;

b) the focus is mainly on testing the results, but the articles do not specify which model these results were based on;

c) in general, the presented results do not have a specific implementation. All calculations given in the available minimum of models are provided to solve some abstract theoretical-scientific problems.

Conclusion

The information technology of collecting and accumulating data of media resources with the use of artificial intelligence is of great value for the development of the country and the world community, contributing to the improvement of the quality of life, ensuring security and the development of science and technology, namely:

1. Improving decisions in the management of the country: with the help of collected and processed data, it is possible to analyze the socio-economic situation in the country, identify trends and forecast possible development scenarios.

2. Ensuring national security: the collection and analysis of data on media resources is important for identifying possible threats to national security, including disinformation, cyber attacks, and other threats related to information warfare.

3. Improving the quality of life of citizens: Information technology allows collecting data on air quality, pollution levels, traffic, meteorological conditions and other factors that affect the lives of citizens.

4. Development of science and technology: the development and implementation of such technologies creates new opportunities for research into artificial intelligence, data analysis and development of new algorithms.

5. Global collaboration and innovation: The collection and analysis of data at the global level helps countries to jointly address global challenges such as climate change, public health and economic development.

6. Effective crisis and emergency management: Collecting and analyzing data from media resources can be critical for managing crisis situations such as natural disasters, epidemics, terrorist acts, etc.

References

1. Bebeszko, B. (2023). Training of an artificial neural network based on data for evaluating the performance and risks of investing in digital assets. *Cyber security: education, science, technology*, 3(19), 135–145. DOI: <https://doi.org/10.28925/2663-4023.2023.19.135145>.
2. Khorolska, K. (2022). The potential of applying different methods of artificial intelligence in the problem of drawing recognition and transformation 2D→3D. *Cyber security: education, science, technology*, 1(17), 21-30. DOI: <https://doi.org/10.28925/2663-4023.2022.17.2130>.
3. Bäck, Thomas, Kursawe, Frank. (1998). Evolutionary Algorithms for Fuzzy Logic: A Brief Overview URL: https://www.researchgate.net/publication/2509441_Evolutionary_Algorithms_for_Fuzzy_Logic_A_Brief_Overview, DOI: 10.1142/9789812830753_0001.
4. Maan, Afathi. (2021). Implementation of new hybrid evolutionary algorithm with fuzzy logic control approach for optimization problems. URL: <https://journals.uran.ua/ejet/article/view/245222>, DOI: <https://doi.org/10.15587/1729-4061.2021.245222>.
5. Alipouri, Yousef, Poshtan, Javad, Alipour, Hasan. (2015). Global minimum routing in evolutionary programming using fuzzy logic. *Information Sciences*, 292, 162-174, DOI: <https://doi.org/10.1016/j.ins.2014.08.061>.
6. Krömer, Pavel, Platoš, Jan, Snášel, Václav, Ajith, Abraham. (2011). Fuzzy classification by evolutionary algorithms. *Proc. IEEE International Conference on Systems, Man, and Cybernetics*. URL: <https://ieeexplore.ieee.org/document/6083684/authors#authors>, DOI: 10.1109/ICSMC.2011.6083684.
7. Castiello, Ciro. (2007). Evolutionary neuro-fuzzy systems and applications. URL: https://www.academia.edu/13411136/Evolutionary_neuro_fuzzy_systems_and_applications.
8. Matthews, Stephen G., Gongora, Mario A., Hopgood, Adrian A. (2010). Evolutionary algorithms and fuzzy sets for discovering temporal rules. *International Journal of Applied Mathematics and Computer Science*, 23, 4, 855-868. URL: <https://eudml.org/doc/262453>.
9. Edmondst, A. N., Burkhardt, Diana, Osei, Adjei. (2017). Genetic Programming of Fuzzy Logic Production Rules. URL: <https://sci2s.ugr.es/keel/pdf/specific/congreso/ieeee49.pdf>.
10. Biaobiao, Zhang, Yue, Wu, Jiabin, Lu, and Du, K.-L. (2011). Evolutionary Computation and Its Applications in Neural and Fuzzy Systems. *Applied Computational Intelligence and Soft Computing*. URL: <https://www.hindawi.com/journals/acisc/2011/938240/>, DOI: <https://doi.org/10.1155/2011/938240>.
11. Kroeske, Jens, Ghandar, Adam, Michalewicz, Zbigniew and Neumann, Frank. (2009). Learning Fuzzy Rules with Evolutionary Algorithms – an Analytic Approach. URL: https://cs.adelaide.edu.au/~frank/papers/ppsn_fuzzy.pdf.
12. Shen, Qiang, Galea, Michelle. (2004). Evolutionary approaches to fuzzy modelling for classification. URL: <https://typeset.io/pdf/evolutionary-approaches-to-fuzzy-modelling-for-lyuih8r49q.pdf>, DOI: 10.1017/S0269888904000189.

13. Tsiutsiura, M., Kostyshyna, N., Yerukaiev, A., Danylyshyn, S., Honcharenko, Y. and Tao, L. (2022). Research of Housing Comfort Using Linguistic Variables. 2022 International Conference on Smart Information Systems and Technologies (SIST). 208–211. DOI: 10.1109/SIST54437.2022.9945736.

14. Kostiuk, Y., Kryvoruchko, O., Tsiutsiura, M., Yerukaiev, A. and Rusan, N. (2022). Research of Methods of Control and Management of the Quality of Butter on the Basis of the Neural Network. 2022 International Conference on Smart Information Systems and Technologies (SIST), 307-313, DOI: 10.1109/SIST54437.2022.9945764.

Стаття надійшла до редколегії 20.10.2023

Цюцюра Микола Ігорович

Доктор технічних наук, професор, професор кафедри інженерії програмного забезпечення та кібербезпеки,
<https://orcid.org/0000-0003-4713-7568>

Державний торговельно-економічний університет

Єрукаєв Андрій Віталійович

Кандидат технічних наук, доцент, доцент кафедри інженерії програмного забезпечення та кібербезпеки,
<https://orcid.org/0000-0002-9956-3713>

Державний торговельно-економічний університет

Хорольська Карина Вікторівна

Старший викладач кафедри інженерії програмного забезпечення та кібербезпеки, <https://orcid.org/0000-0003-3270-4494>
Державний торговельно-економічний університет

Бєбешко Богдан Тарасович

Старший викладач кафедри інженерії програмного забезпечення та кібербезпеки, <https://orcid.org/0000-0001-6599-0808>
Державний торговельно-економічний університет

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ЗБИРАННЯ ТА АКУМУЛЯЦІЇ ДАНИХ МЕДІАРЕСУРСІВ
ІЗ ЗАСТОСУВАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ**

Анотація. У цей складний час, у якому опинилася наша країна, інформаційні технології стали тією силою, яка успішно використовується проти загарбницьких дій ворога. Збройним силам, окрім необхідної фізичної допомоги, доволі суттєво допомагають в інформаційній сфері, зокрема і з боку наукової спільноти. Різні напрями та галузі комп'ютерних наук поєднуються, аби отримати дієвий спосіб для досягнення бажаної перемоги та довгоочікуваного миру і добробуту. У статті автори пропонують своє бачення в реалізації технології для оцінки інформації, що лунає з різних медіаджерел. Для цього в дослідженнях було задіяно один з найбільш широко використовуваних напрямів інформаційних технологій – штучний інтелект. Як відомо, AI включає в себе доволі широкий, повноцінний набір моделей та методів для опрацювання різноманітних практичних задач. Зокрема, автори зупинилися на «м'яких» обчисленнях, тобто використанні нейронних мереж, нечітких множин та генетичних алгоритмів у поєднанні з методами ще одного напрямку AI – інтелектуального аналізу даних. Усі вищезгадані ідеї пропонуються розглянути в обґрунтуванні до створення наукового проєкту. У статті детально проаналізовано зарубіжні і вітчизняні джерела з метою виявлення спільних та відмінних рис з авторськими ідеями. Особливу увагу звернено саме на ті компоненти, які не розкриті в інших роботах і пропонуються до розроблення у цьому проєкті. Також зазначено об'єкт, предмет дослідження, мету, розкрито наукову та практичну новизну від реалізації ідей запропонованої роботи. Окрім цього, описано переваги використання запропонованих методів для розв'язання не тільки задачі опрацювання та оцінки інформації з різних медіаджерел, а й наведено користь від їх застосування в інших критичних задачах військової справи. Отже, автори поставили собі за мету долучитися до допомоги нашим захисникам у реалізації підходу, що заснований на застосуванні штучного інтелекту. У цій статті висвітлено лише пропозиції з детальним їх науковим обґрунтуванням. Отже, підсумовуючи усе вищезазначене, в роботі наведено теоретико-прикладний підхід на рівні розкриття базових ідей до створення проєкту. У наступних роботах авторами після представлення бази дослідження буде приділено більше уваги вже математичним методам опису і висвітленню проведених експериментальних досліджень.

Ключові слова: штучний інтелект; інтелектуальний аналіз даних; нечіткі системи; еволюційне програмування

Посилання на публікацію

APA Tsiutsiura, M., Yerukaiev A., Khorolska K., & Bebeshko B., (2023). Data collection and accumulation of media resources using artificial intelligence. *Management of Development of Complex Systems*, 56, 50–55, [dx.doi.org/10.32347/2412-9933.2023.56.50-55](https://doi.org/10.32347/2412-9933.2023.56.50-55).

ДСТУ Цюцюра М. І., Єрукаєв А. В., Хорольська К. В., Бєбешко Б. Т. Інформаційна технологія збирання та акумуляції даних медіаресурсів із застосуванням штучного інтелекту. *Управління розвитком складних систем*. Київ, 2023. № 56. С. 50 – 55, [dx.doi.org/10.32347/2412-9933.2023.56.50-55](https://doi.org/10.32347/2412-9933.2023.56.50-55).