UDC 004.93'1:004.912:316.77 https://doi.org/10.32515/2664-262X.2025.12(43).2.20-26

Artem Kalancha[1], **Dmytro Uhryn**[1], Prof., DSc., **Oleksandr Dorenskyi**[2], Assoc. Prof., PhD tech. sci., **Vasyl Lytvyn**[3], Assoc. Prof., DSc.
[1] *Yuriy Fedkovych Chernivtsi National University, Cernivtsi, Ukraine,*
[2] *Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine*
[3] *Lviv Polytechnic National University, Lviv, Ukraine*
e-mail: d.ugryn@chnu.edu.ua, kalancha.artem@chnu.edu.ua, dorenskyiop@kntu.kr.ua, vasyl.v.lytvyn@lpnu.ua

Method of Integral Analysis of Relationships between Information Sources based on Temporal-Semantic Metrics

This study is devoted to the analysis of the relationships between information sources in the Ukrainian information space, in particular between news and political Telegram channels. The paper identifies the main shortcomings of existing analysis methods, which are mainly focused on the English-language context and do not take into account the dynamic nature of information sources. As an alternative to these methods, a new method of integral analysis is proposed, which combines several independent metrics at the same time - lexical similarity and temporal-semantic influence. The proposed and implemented approach allows us to assess not only the semantic similarity of sources, but also the nature of their interactions over time, forming a multidimensional matrix of relationships. The results obtained demonstrate the effectiveness of the integral approach for identifying hidden structures of information influence, which can be used for the further development of monitoring systems.
Artificial Intelligence, NLP, Information Flows, Information Sources, Integrated Analysis, Network Structures, Temporal Semantic Influence

Problem Statement. The modern Ukrainian information space is extremely rich, but at the same time fragmented into a large number of information sources. It contains hundreds, and sometimes thousands, of information sources, including anonymous Telegram channels, pages on social networks, and news aggregators, which often do not have clearly identified authors or editorial responsibility. There is an obvious need not only to compare the content of messages, but also to understand the relationships between the sources themselves: who influences whom, what topics become common, and how information signals spread over time.

Despite the growing interest in the analysis of information flows, most existing systems are focused on the foreign context. They mostly focus on large international media or English-language news agencies, leaving out of consideration the Ukrainian segment, which has its own fragmented dynamics, political specifics and social peculiarities. The methods underlying such systems are often one-dimensional, that is, they assess only textual similarity or activity of publications, ignoring temporal aspects, tonality and the complex structure of connections. Such assessments do not reflect the real complexity of interactions between sources.

Considering the above-described shortcomings, there is a need to create approaches that would be adapted to the Ukrainian information environment, taking into account the multidimensional nature of information influence.

Analysis of Recent Research and Publications. In the English-speaking information space, there are already developed systems that specialize in the analysis of information sources, but their effectiveness remains limited due to the specificity of the approaches and the context in which they were created. One example is Ad Fontes Media, a platform that evaluates American media in terms of political preferences and the level of reliability. Its main tool is the “Media Graph”, which shows how much a certain publication gravitates towards the left or right of the political spectrum and how reliable the source itself is [1]. This approach forms a visual representation of the distribution of political positions in the American space.

Similarly, the AllSides project operates, which tries to ensure a comprehensive perception of information, allowing users to see news from different political perspectives [2]. Its value lies in demonstrating how the same event is interpreted by different media. Such systems are primarily focused on the American market, therefore they do not take into account the cultural and sociolinguistic features and structure of the Ukrainian information space.

Analyzing other scientific and technical developments aimed at automated analysis of sources, we also find mostly English-language methods that do not cover the dynamic nature of information influence. The MediaRank system offers a global ranking of thousands of news sites by reputation and semantic characteristics, which allows us to assess their quality and influence, but does not record how one source influences another in time [3]. The NELA ecosystem provides tools for comparing sources across content copying networks using vector representation models, but its analysis is limited to deterministic connections between publications [4]. The Semantic "Echo" method proposes to measure the "echo" of information messages through semantic similarity, calculating the similarity between official releases and publications in social networks, however, such a model does not take into account the temporal relationships between events [5]. One of the promising methods for further research is the approach based on Hawkes processes, which considers the spread of events as a self-exciting process, where each event temporarily increases the probability of the appearance of subsequent ones [6], but again, the impact assessment does not take into account the content similarity of messages, but only the fact of the appearance of the message itself.

These studies and systems cover a wide range of approaches to source assessment, but they all focus on unidimensional indicators and neglect multidimensional interaction between sources. For the Ukrainian information space, which is characterized by rapid changes in narratives and heterogeneity of sources, such a simplified logic of analysis is insufficient. The tasks related to the implementation and use of modern information technologies, particularly artificial intelligence, in the sphere of state information security are examined in scholarly work [12]. At the same time, studies [13, 14] present successful implementations of AI applications across various sectors.

Problem Definition. The aim of this study is to develop an approach that allows for an integrated analysis of information sources based on a combination of several independent source metrics (or their comparative metrics). Unlike traditional one-dimensional approaches, the proposed methodology involves the coordinated use of different types of similarity: semantic and temporal-semantic with the subsequent formation of a single resulting integral matrix. The desired approach allows not only to quantitatively assess the similarity between sources, but also to identify hidden links of influence that are formed over time.

This study is based on the idea of multidimensional comparison, in which each source is considered according to a single metric, but as part of a complex system of interactions. Within this system, it is important not only to compare the similarity of the text, but also to analyze the synchronicity of information waves, the patterns of occurrence of coincidences and the sequence of their distribution between sources. An integrated approach [7] creates the prerequisites for a more accurate reflection of the information space, identification of interconnected groups of sources, and construction of further models for assessing the impact within the Ukrainian information environment.

Main Results. This study used data obtained from eight information sources, in particular Telegram channels belonging to the news and political segment of the Ukrainian media space [8]. The collection period covers a six-month interval – from June 12, 2024 to December 12, 2024 inclusive. For each message, the basic message attributes are saved: channel identifier, message text and date of publication. This minimum required data format provides the possibility of further aggregation, thematic classification and temporal and semantic analysis. The channels were selected based on the principle of their activity, audience reach and belonging to a common information space focused on news and political topics.

After determining the set of sources and data format, a stage of pre-processing of texts was carried out, which is of key importance for ensuring the reliability of further results. Taking into account the specifics of the content of Telegram channels, messages underwent several consecutive cleaning steps: removing special characters, numeric values, hyperlinks, formatting elements, as well as abbreviations and technical markers. Then, the texts were normalized by tokenization and removing stop words. Special attention was paid to language unification: all Russian-language messages were automatically translated into Ukrainian to preserve the semantic integrity of the corpus and avoid biasing the results when comparing lexical features. Detailed aspects of preprocessing text messages, adapted to the specifics of Telegram sources, are described in the author's previous study devoted to the methodology for normalizing and standardizing text data from open information streams [9].

The cosine similarity method is used to measure the similarity between lexical fields of information sources [10]. For each channel, a dictionary of words is formed with a count of the number of their occurrences in messages for the period under study. Words that occur less frequently than a set threshold (for example, less than 20 occurrences) are removed as statistically insignificant. Based on the dictionaries from all channels, a global dictionary is created that defines the feature space (words), where each channel is represented as a vector where the number of occurrences of a word forms its coordinates in the dimension.

To assess the similarity between channels, the cosine similarity formula is used, which defines the angle between two vectors in a multidimensional space:

$$cs(s_i, s_j) = \frac{z(s_i)^T z(s_j)}{\|z(s_i)\|_2 \, \|z(s_j)\|_2} \in [-1,1],$$

where $z(s_{i,j})$ – source lexical vector.

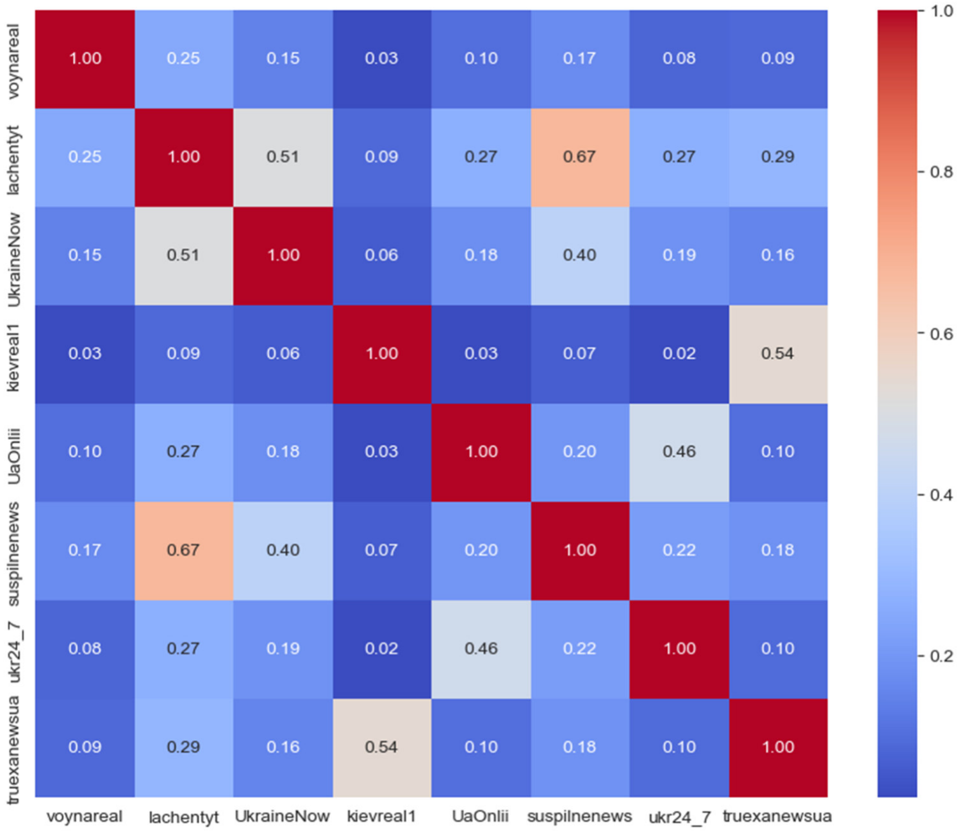

Figure 1 – Cosine similarity matrix between lexical fields of information sources

Source: visualized by the authors

This method is effective because it takes into account the proportions of vocabulary use, and not only its absolute frequencies, which allows you to correctly compare sources with different activity. Based on this, a similarity matrix is formed with values from 0 to 1, where larger values indicate greater proximity of lexical fields between channels.

Using input data and the cosine similarity method, similarity estimates between their lexical fields were calculated [11]. For each pair of sources, the similarity coefficient between the word frequency vectors was calculated and integrated into the similarity matrix (Fig. 1).

The Timed Semantic Influence (TSI) algorithm is designed to quantify the mutual influence between two information sources based on the temporal and semantic proximity of their messages. For each pair of channels A and B, messages from one channel are sequentially compared with messages from the other within a given time horizon, for example ±8 hours. If there is a high semantic similarity between two messages (above a threshold s), such a pair is considered potentially related.

The key step of the algorithm is to calculate the TSI indicator, which takes into account both the degree of semantic similarity between the messages and the time difference between their publication. For this, the formula is used:

$$TSI = sm(m_a, m_b) * exp(-\alpha \times (\Delta t/60)),$$

where the first factor reflects the semantic similarity between two messages, and the second – the exponential decrease in influence with increasing time distance between them, controlled by the coefficient α. After determining all relevant pairs of messages, the total indicator of influence between channels is calculated, which is recorded in the matrix of mutual influences.

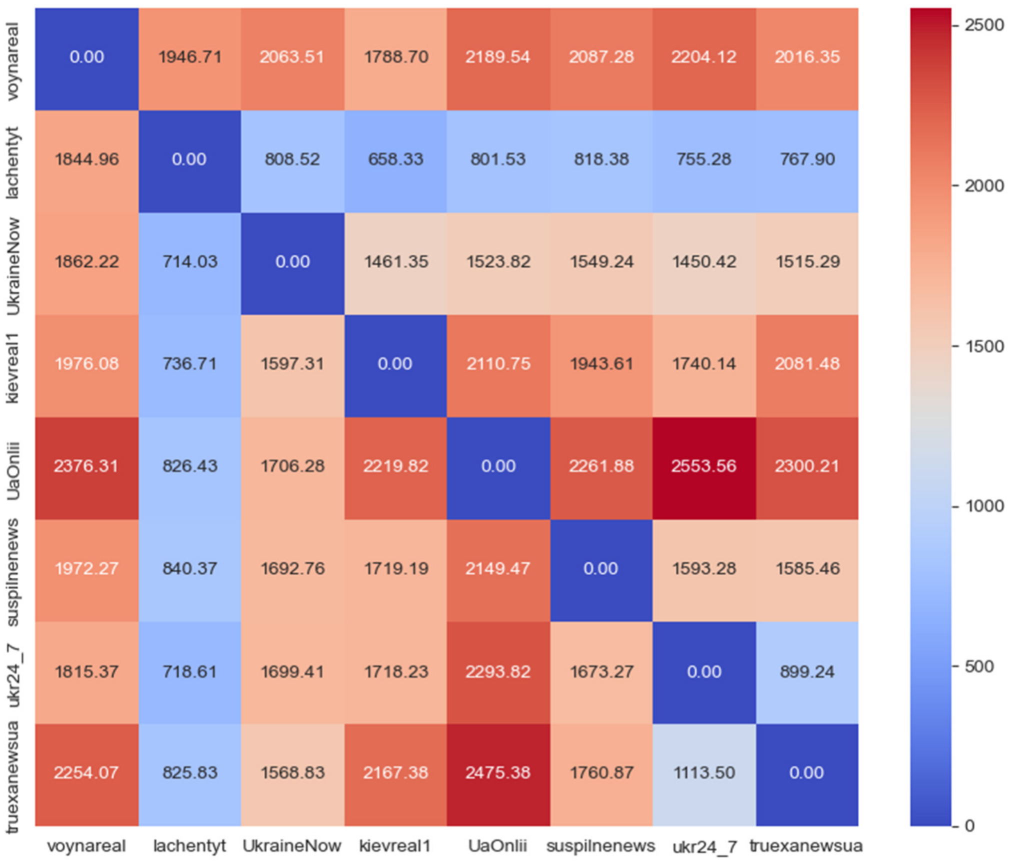

Figure 2 – Influence matrix between information sources based on TSI score

Source: visualized by the authors

Based on the prepared data, the TSI algorithm was applied to assess the mutual influence between channels. The TSI value was calculated for each pair of sources. As a result, a similarity matrix was formed (Fig. 2), where each element reflects the level of influence of one channel on another in absolute values.

Now, using the lexical similarity matrix and the mutual influence matrix between the channels, a resulting relationship matrix was constructed, which simultaneously takes into account the lexical and temporal characteristics of the sources and their messages. To ensure the correct combination, both matrices were reduced to a single scale. The influence matrix was normalized using the Min-Max normalization.

After normalization, the average value between the two corresponding metrics (matrices) was calculated for each pair of sources, forming an integral similarity matrix (Fig. 3). This approach allows you to simultaneously take into account both the similarity of the lexical field and the temporal-semantic influence between the sources, providing a more complete assessment of the relationships in the information space.

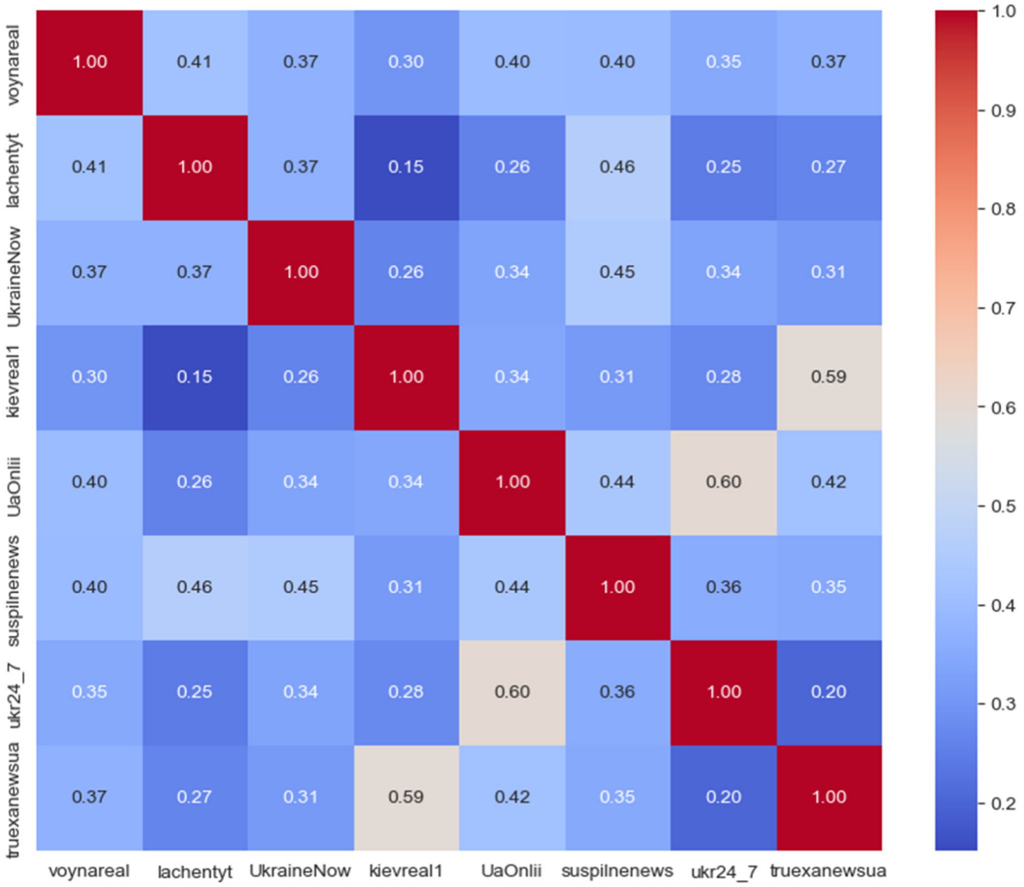

Figure 3 – Integral source comparison matrix

Source: visualized by the authors

Analysis of the obtained results showed that the greatest integral similarity is demonstrated by the pairs of channels truexanewsua – kievreal1, urk24_7 – UaOnlii, suspilnenews – lachentyt and suspilnenews – UkraineNow. This indicates a close connection between their information flows. At the same time, it is observed that the leaders by the integral metric differ from those determined separately by cosine similarity or by TSI. It can be noted that cosine similarity takes into account the similarity of the general lexical field of the source. While the TSI algorithm takes into account the similarity of two messages in a limited time space (8 hours).

Conclusions. This study demonstrated the problem of one-dimensional analysis of information sources, when only one type of similarity is assessed: temporal or, for the most part, only lexical. In response to this, an integral assessment method was proposed, which combines several metrics simultaneously. The proposed approach was implemented and tested on a limited set of Telegram channels, which allowed us to demonstrate its effectiveness and flexibility in real conditions.

The method showed that the combination of semantic and temporal characteristics allows us to more accurately identify the structure of relationships between sources. In particular, the pairs of channels truexanewsua – kievreal1 and urk24_7 – UaOnlii achieved the highest connection indicators precisely during the integral analysis, although they had mediocre values in individual dimensions. This indicates the importance of a comprehensive approach that reflects the multidimensional nature of the information source.

In the further development of this direction, it is advisable to improve the method by introducing a weighted integral analysis, where each metric will have its own coefficient. This will allow you to regulate the influence of individual metrics on the final result and increase the accuracy of assessing the relationships between information sources.

List of References

1. Ad Fontes Media. Interactive Media Bias Chart. 2025. URL: https://adfontesmedia.com/interactive-media-bias-chart/ (дата звернення: 02.09.2025).
2. AllSides. AllSides Media Bias Ratings. 2024. URL: https://www.allsides.com/ (дата звернення: 03.09.2025).
3. Ye J., Skiena S. MediaRank: Computational ranking of online news sources. *arXiv preprint.* 2019. arXiv:1903.07581. DOI: https://doi.org/10.48550/arXiv.1903.07581.
4. Horne B. D., Dron W., Khedr S., Adalı S. Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. *Companion Proceedings of The Web Conference 2018 (WWW'18).* 2018. P. 235–238. DOI: https://doi.org/10.1145/3184558.3186987.
5. Cann T. J. B., Dennes B., Coan T., O'Neill S., Williams H. T. P. Using semantic similarity and text embedding to measure the social media echo of strategic communications. *arXiv preprint.* 2023. arXiv:2303.16694. DOI: https://doi.org/10.48550/arXiv.2303.16694.
6. Worrall J., Browning R., Wu P., Mengersen K. Fifty years later: New directions in Hawkes processes. *SORT.* 2022. Vol. 46, № 1. P. 3–38. DOI: https://doi.org/10.2436/20.8080.02.116.
7. Bazeley P. Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research.* 2009. Vol. 3, № 3. P. 203–207. DOI: https://doi.org/10.1177/1558689809334443.
8. Telegram. Telegram APIs. 2025. URL: https://core.telegram.org/ (дата звернення: 01.09.2025).
9. Kalancha A., Uhryn D. Optimizing text data storage for improved natural language processing: A comprehensive analysis. *Advances in Computer Science for Engineering and Education VII (ICCSEEA 2024).* Vol. 242. 2025. P. 294–303. DOI: https://doi.org/10.1007/978-3-031-84228-3_25.
10. Magara B. M., Ojo S. O., Zuva T. A comparative analysis of text similarity measures and algorithms in research paper recommender systems. *Conference on Information Communications Technology and Society (ICTAS).* 2018. P. 1–5. DOI: https://doi.org/10.1109/ICTAS.2018.8368766.
11. Uhryn D., Kalancha A., Zakharov M. Analysis of information sources based on the cosine similarity algorithm, affinity propagation and spectral clustering methods. *Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025).* Vol. 1416. 2025. P. 239–249. DOI: https://doi.org/10.1007/978-981-96-6441-2_21.
12. Доренський О. П., Улічев О. С., Задорожний К. О., Коваленко А. С., Дрєєва Г. М. Концептуальна модель системи інформаційного протиборства координаційного центру з питань національної безпеки і оборони. *Центральноукраїнський науковий вісник. Технічні науки.* 2024. Вип. 10(41), ч. 2. С. 23–31. URL: https://doi.org/10.32515/2664-262X.2024.10(41).2.23-31.
13. Korniienko O., Kozub N., Dorenskyi O. Method and Technological Solution of an AI-Based Adaptive Investor Survey Service for Determining an Individual Risk Profile. *Central Ukrainian Scientific Bulletin. Technical Sciences.* 2025. Issue 11(42), Part II. P. 3–10. DOI: https://doi.org/10.32515/2664-262X.2025.11(42).1.3-10.
14. Шелехов І. В., Прилепа Д. В., Хібовська Ю. О., Шамонін К. Є., Доренський О. П. Інформаційно-екстремальна технологія інтелектуального аналізу якості освітнього контенту в закладах вищої освіти. *Центральноукраїнський науковий вісник. Технічні науки.* 2025. Вип. 12(43), ч. 1. С. 58–72. DOI: https://doi.org/10.32515/2664-262X.2025.12(43).1.58-72.

References

1. Ad Fontes Media. (2025). Interactive Media Bias Chart. URL: https://adfontesmedia.com/interactive-media-bias-chart/ (accessed: 02.09.2025).
2. AllSides. (2024). AllSides Media Bias Ratings. URL: https://www.allsides.com/ (accessed: 03.09.2025).
3. Ye, J., & Skiena, S. (2019). MediaRank: Computational ranking of online news sources. arXiv preprint, arXiv:1903.07581. https://doi.org/10.48550/arXiv.1903.07581.
4. Horne, B. D., Dron, W., Khedr, S., & Adali, S. (2018). Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. *Companion Proceedings of The Web Conference 2018 (WWW'18)*, 235–238. https://doi.org/10.1145/3184558.3186987.
5. Cann, T. J. B., Dennes, B., Coan, T., O'Neill, S., & Williams, H. T. P. (2023). Using semantic similarity and text embedding to measure the social media echo of strategic communications. arXiv preprint, arXiv:2303.16694. https://doi.org/10.48550/arXiv.2303.16694.
6. Worrall, J., Browning, R., Wu, P., & Mengersen, K. (2022). Fifty years later: New directions in Hawkes processes. *SORT*, 46(1), 3–38. https://doi.org/10.2436/20.8080.02.116.
7. Bazeley, P. (2009). Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research*, 3(3), 203–207. https://doi.org/10.1177/1558689809334443.
8. Telegram. (2025). Telegram APIs. URL: https://core.telegram.org/ (accessed: 01.09.2025).
9. Kalancha, A., & Uhryn, D. (2025). Optimizing text data storage for improved natural language processing: A comprehensive analysis. *Advances in Computer Science for Engineering and Education VII (ICCSEEA 2024)*, 242, 294–303. https://doi.org/10.1007/978-3-031-84228-3_25.
10. Magara, B. M., Ojo, S. O., & Zuva, T. (2018). A comparative analysis of text similarity measures and algorithms in research paper recommender systems. *Conference on Information Communications Technology and Society (ICTAS)*, 1–5. https://doi.org/10.1109/ICTAS.2018.8368766.
11. Uhryn, D., Kalancha, A., & Zakharov, M. (2025). Analysis of information sources based on the cosine similarity algorithm, affinity propagation and spectral clustering methods. *Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025)*, 1416, 239–249. doi.org/10.1007/978-981-96-6441-2_21.
12. Dorenskyi, O. P., Ulichev, O. S., Zadorozhnyi, K. O., Kovalenko, A. S., & Drieieva, H. M. (2024). Conceptual model of information confrontation system of coordination center for national security and defense. *Tsentralnoukrainskyi naukovyi visnyk. Tekhnichni nauky*, 10(41) II, 23–31. https://doi.org/10.32515/2664-262X.2024.10(41).2.23-31 [in Ukrainian].
13. Korniienko, O., Kozub, N., & Dorenskyi, O. (2025). Method and technological solution of an AI-based adaptive investor survey service for risk profile identification. *Central Ukrainian Scientific Bulletin. Technical Sciences*, 11(42) II, 3–10. https://doi.org/10.32515/2664-262X.2025.11(42).1.3-10.
14. Shelekhov, I. V., Prylepa, D. V., Khibovska, Yu. O., Shamonin, K. Ye., & Dorenskyi, O. P. (2025). Information-extreme technology of intelligent quality analysis of educational content in universities. *Tsentralnoukrainskyi naukovyi visnyk. Tekhnichni nauky*, 12(43) I, 58–72. https://doi.org/10.32515/2664-262X.2025.12(43).1.58-72 [in Ukrainian].

А. Д. Каланча[1], **Д. І. Угрин**[1], проф., д-р техн. наук, **О. П. Доренський**[2], доц., канд. техн. наук,
В. В. Литвин[3], проф., д-р техн. наук
[1]*Чернівецький національний університет імені Юрія Федьковича, м. Чернівці, Україна,*
[2]*Центральноукраїнський національний технічний університет, м. Кропивницький, Україна,*
[3]*Національний університет "Львівська Політехніка", м. Львів, Україна*

Метод інтегрального аналізу зв'язків між джерелами інформації на основі часово-семантичних метрик

Праця присвячена дослідженню взаємозв'язків між інформаційними джерелами в українському інформаційному просторі, зокрема між новинно-політичними каналами в Telegram. У роботі визначено основні недоліки наявних методів аналізу, що переважно орієнтовані на англомовний контекст і не враховують динамічну природу інформаційних джерел. Як альтернативу цим методам запропоновано новий метод інтегрального аналізу, який поєднує одночасно кілька незалежних метрик: лексичну подібність та часово-семантичний вплив. Запропонований та реалізований підхід дозволяє оцінювати не лише змістову схожість джерел, а й характер їхніх взаємодій у часі, формуючи багатовимірну матрицю зв'язків. Отримані результати демонструють ефективність інтегрального підходу для виявлення прихованих структур інформаційного впливу, що можуть бути використані для подальшого розвитку систем моніторингу.

штучний інтелект, NLP, інформаційні потоки, інформаційні джерела, інтегральний аналіз, мережеві структури, часово-семантичний вплив

Одержано (Received) 20.10.2025 *Прорецензовано (Reviewed) 07.11.2025*
Прийнято до друку (Approved) 25.11.2025