https://doi.org/10.15407/intechsys.2026.02.046
UDC 004.912:81'322

O.V.TOKOVA, PhD (Engineering),
Inclusion Educational Assistant,
Wilmington High School,
159 Church St, Wilmington, MA 01887
https://orcid.org/0000-0002-0289-2810
len327@ukr.net

DEVELOPMENT OF AN AUTOMATED TOOL FOR SEARCHING RESEARCH MATERIALS IN THE LEARNING PROCESS

This paper analyses the problem of information overload and the complexity of searching for scientific sources in the modern digital educational environment. It is shown that the availability of electronic resources does not always ensure effective learning due to the insufficient level of students' skills in scientific search and information evaluation.

The principles for constructing an Automated Research Assistant ¾ an automated tool to support educational and research activities — are proposed. A web-based software system for automated search of scientific materials using the open preprint repository arXiv via an API interface has been implemented. The system performs search, primary filtering, and structured storage of results, forming a set of publication metadata (title, summary, and source link). The proposed approach aims to reduce user workload and improve the efficiency of working with scientific information in the learning process. The architecture of the software system provides the possibility for further expansion through integration with additional scientific platforms and data sources.

Keywords: *automated search, Automated Research Assistant, web-based software system, arXiv preprint repository, research efficiency.*

Introduction

Modern students and researchers operate in conditions of rapidly increasing volumes of digital information. The development of open science, electronic libraries, and online platforms has significantly expanded access to scientific knowledge. At the same time, the problem of access to

Cite: Tokova O.V. Development of an Automated Tool for Searching Research Materials in the Learning Process. *Information Technologies and Systems*. 2 (8). 2026. 46—57. https://doi.org/10.15407/intechsys.2026.02.046

information has gradually transformed into the problem of its effective selection and interpretation.

According to recent studies, information overload is one of the key factors that complicate learning activities, reduce concentration, and increase cognitive load [1, 2]. In the digital educational environment, students are required to independently search, evaluate, and systematize a large number of sources, which requires well-developed information literacy and critical thinking skills [3, 4].

Excess unstructured information creates additional extraneous cognitive load that negatively affects learning and knowledge acquisition [3]. Despite the availability of numerous electronic resources, the process of searching for scientific materials remains time-consuming and often leads to excessive time expenditure and decreased learning motivation [5, 6].

The aim of this study is to develop and test an *Automated Research Assistant* — a software tool designed for automated search and primary structuring of scientific materials in educational and research activities.

Problems of Searching for Relevant Information in Research and Academic Activities

The effectiveness of e-learning depends not only on the availability of digital resources but also on the user's ability to navigate large volumes of information. Modern educational platforms, online courses, and open scientific repositories have significantly expanded access to knowledge; however, the mere availability of information does not guarantee its effective assimilation.

Studies in the field of distance learning show that students often face the phenomenon of information overload, where the amount of educational material exceeds their capacity to quickly comprehend and integrate it into knowledge [1]. According to Chen et al., an excess of information in online learning environments reduces concentration and overall satisfaction with the learning process.

The problem of information overload is also considered in a broader context of digitalization. A systematic review by Arnold et al. [2] demonstrates that the rate of growth of information flows exceeds human cognitive capabilities, which necessitates the use of tools for structuring, filtering, and prioritizing information. Similar conclusions are presented in classical studies on information overload, which emphasize its impact on decision-making efficiency and information processing [6, 7].

A psychological explanation of this phenomenon is provided by Cognitive Load Theory, according to which human working memory has limited capacity [3, 8]. Excessive unstructured or irrelevant information creates extraneous cognitive load, hindering the formation of stable knowledge. In the digital learning environment, this issue is intensified by the fragmented presentation of material through hyperlinks, multimedia content, and distributed information sources. In addition, multi-

tasking while working with digital resources negatively affects learning outcomes [9—11].

Another important factor influencing learning effectiveness is the level of users' information literacy. According to UNESCO, information literacy includes the ability to find, critically evaluate, and responsibly use information as a core competency of modern education [4]. A similar perspective is reflected in the Framework for Information Literacy for Higher Education, which views working with information as a complex cognitive process involving the evaluation of source credibility, understanding the context of knowledge creation, and the conscious use of information [5].

In educational practice, both at the secondary and higher education levels, there is often an insufficient development of skills related to searching for and critically evaluating scientific information. When completing academic tasks, students frequently rely on general-purpose search engines, which are effective for everyday queries but do not provide the level of source quality required for academic work. As a result, several common difficulties arise:

- significant time spent searching for sources;
- accumulation of unsystematized materials;
- difficulty in assessing the scientific relevance of publications;
- decreased learning motivation;
- increased cognitive and emotional load.

Thus, the modern paradox of e-learning lies in the fact that easier access to information is accompanied by increased complexity in its effective use. Without the development of information literacy and the use of tools for automated knowledge structuring, e-learning may create additional cognitive load and reduce the overall effectiveness of the educational process [6].

Principles of Building the Automated Research Assistant

Analysis of modern educational challenges shows that the main difficulty lies not in the lack of information, but in the complexity of its search, selection, and systematization. This became the basis for the development of the Automated Research Assistant.

The idea of the tool is to simplify the initial stages of working with scientific information, namely searching, reviewing, and the preliminary selection of sources. This allows the user to move more quickly from unsystematic browsing of materials to their meaningful analysis and use in learning or research.

The system operates based on the following principles:

1. *Centralized scientific search*. In its current implementation, the system uses the open scientific repository arXiv, which provides access to up-to-date research publications across various fields. This ensures that users work with academic rather than random sources.

2. *Automated preliminary filtering*. The system generates a list of results based on a given query while excluding previously processed articles. This helps avoid duplication and reduces the need for manual review.

3. *Structured presentation of results*. For each retrieved publication, key metadata are generated, including the title, summary, and source link (URL). This facilitates quick familiarization with the material and supports the selection of relevant sources.

4. *Reduction of information overload*. The tool helps users focus on the content of scientific materials by reducing the time spent on searching and navigating between sources.

Table 1. presents a comparative overview of the process of searching and selecting scientific sources before and after using the Automated Research Assistant.

In a broader educational context, the Automated Research Assistant can be considered a tool for developing a new type of **information literacy**, where technology does not replace thinking but supports it. This approach aligns with modern views on effective learning in a digital environment, where the student is not left alone with the information flow but receives support at key stages of working with knowledge.

Thus, the Automated Research Assistant can be viewed as a tool that helps organize the process of searching for scientific sources and reduce user workload. Automating the initial selection of materials allows the transition from time-consuming and unstructured searching to a more organized approach to working with information. This creates conditions in which students and researchers can focus more on analyzing materials and forming their own conclusions, rather than spending most of their time searching for sources.

Fig. 1 presents a generalized model of the impact of the Automated Research Assistant on the learning process and Fig. 2 demonstrated the stages of this process, illustrating the relationship between key challenges

Table 1. **Comparative characteristics of the scientific source search process**

Before using the Automated Research Assistant	After using the Automated Research Assistant
The student is assigned a topic (coursework, master's thesis, or research project) and independently attempts to find relevant sources. The search is typically conducted using general-purpose search engines, which often results in a large number of irrelevant or random materials. As a result, information accumulates without a clear structure, making further work more difficult.	The student formulates a topic or keywords and uses the Automated Research Assistant. The system automatically generates a list of scientific publications with brief descriptions and links. This enables the user to quickly select relevant materials and proceed to their analysis and use in academic work.

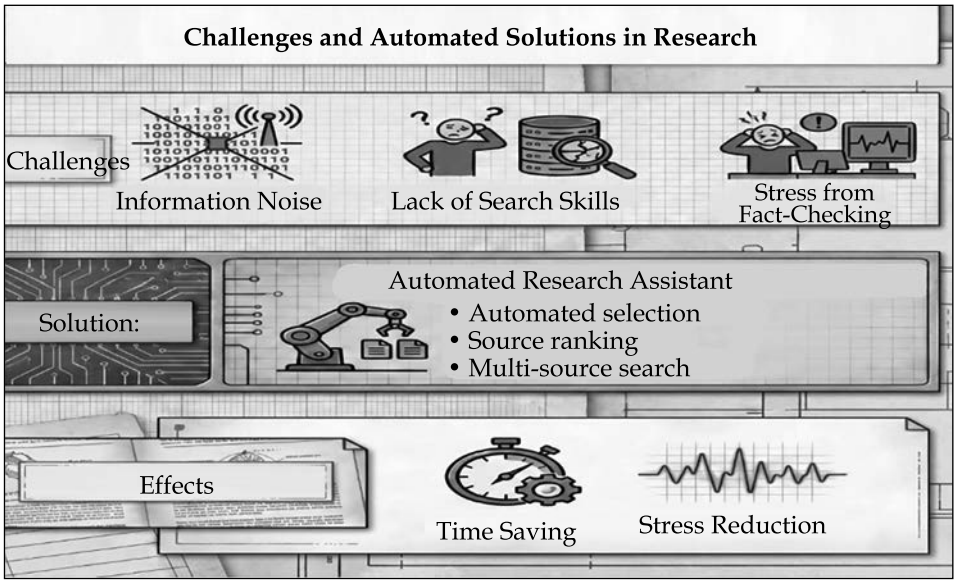

Fig. 1. Model of the impact of the Automated Research Assistant on the educational process

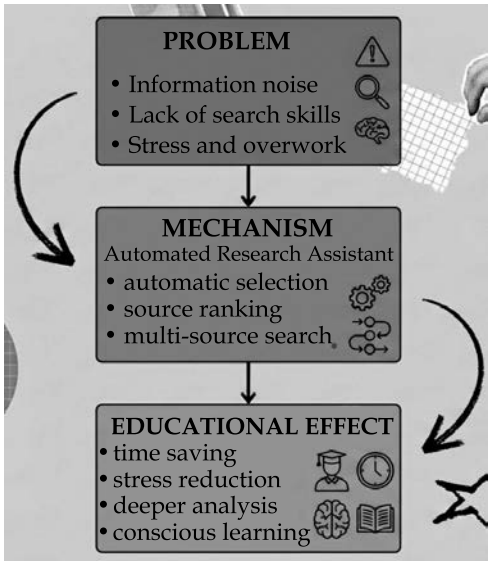

Fig. 2. Stages of information search of the educational process

in information search, the functional capabilities of the tool, and the resulting educational outcomes.

Table 2 presents a generalized scheme illustrating the relationship between the key challenges of modern learning, the functional capabilities of the **Automated Research Assistant**, and the resulting educational outcomes.

The table demonstrates how the use of the tool transforms the process of working with information from unstructured searching and overload to a more organized, manageable, and effective learning process.

Table 2. **Time spent on initial analysis →**
→ Structured results output → Time savings

Problem of Modern Learning	Solution in the Automated Research Assistant	Educational Outcome
Information overload	Automated selection and filtering of sources	Reduction of cognitive load
Unstructured search using general-purpose search engines	Scientific search within an academic repository	Higher relevance of materials
Inability to assess the quality of sources	Structured results output	Improved academic quality
Time spent on initial analysis	Support for systematic work and process control	Time savings in the learning process
Decreased motivation and increased stress	Ranking by relevance and impact	Increased confidence and engagement

Thus, the Automated Research Assistant can be considered a tool that simplifies working with scientific information at the initial stages. Its use reduces the randomness of the search process, making it more structured and focused on content rather than on the technical difficulties of accessing information.

Future Development of the Automated Research Assistant

At the current stage, the Automated Research Assistant is implemented as a tool for automated search and initial structuring of scientific materials. The system simplifies access to relevant sources and reduces the time required for searching and reviewing them.

At the same time, the functionality of the tool can be further expanded in several directions.

One of the promising areas of development is the integration of Natural Language Processing (NLP) methods. This would allow the transition from keyword-based search to content-based analysis of scientific texts. In the future, the system may be extended with the following capabilities:

- analysis of abstracts and full-text articles;
- identification of thematic similarity between sources;
- grouping of materials by research areas;
- extraction of key ideas and approaches.

Another direction of development is the implementation of personalized recommendations. In the future, the system may take into account previous user queries and selected materials, which would enable it to:

- suggest relevant topics for further study;
- adapt search results to the user's level of expertise;
- assist in refining or narrowing research queries.

Another promising direction is the expansion of search sources and the integration of additional scientific platforms, which would improve the completeness and diversity of results.

A separate area of development involves analytical functions related to identifying scientific trends. In the future, the system could analyze changes in publication topics, identify emerging and relevant research areas, and support the selection of topics for coursework, master's theses, and PhD research.

Thus, the further development of the Automated Research Assistant is associated not only with technical improvements but also with expanding its role in supporting learning and research processes. In the future, this will make working with scientific information more structured, clear, and efficient for users.

Based on the proposed principles, a software tool for searching scientific information has been implemented. Its functionality is described in more detail in the following section.

Software Implementation of the Automated Research Assistant

The *Automated Research Assistant* is implemented as a software system and operates as a web-based application. The system is developed using the Python programming language and web tools (in particular, Streamlit), which enable interaction with external scientific databases through application programming interfaces (APIs).

The architecture of the software system is based on a modular approach, which allows the integration of multiple scientific sources into a unified environment and provides the possibility for further functional expansion without changing the core logic of the system.

In its current implementation, the Automated Research Assistant performs a search for scientific materials within the open preprint repository arXiv. This approach allows access to up-to-date scientific publications based on user-defined keywords and enables the formation of an initial set of relevant sources.

The software tool provides automated search, initial filtering, and structured presentation of results. For each retrieved source, key metadata are generated, including the publication title, summary, and link to the source.

The retrieved results are also stored as separate text files, which allows users to perform further analysis outside the system interface.

For the initial evaluation of the tool, thematic queries related to educational technologies and interdisciplinary research areas were used. For each query, the system returned a limited number of results selected based on keywords and relevance. The obtained results were analyzed in terms of their alignment with learning objectives and their usability in the educational process.

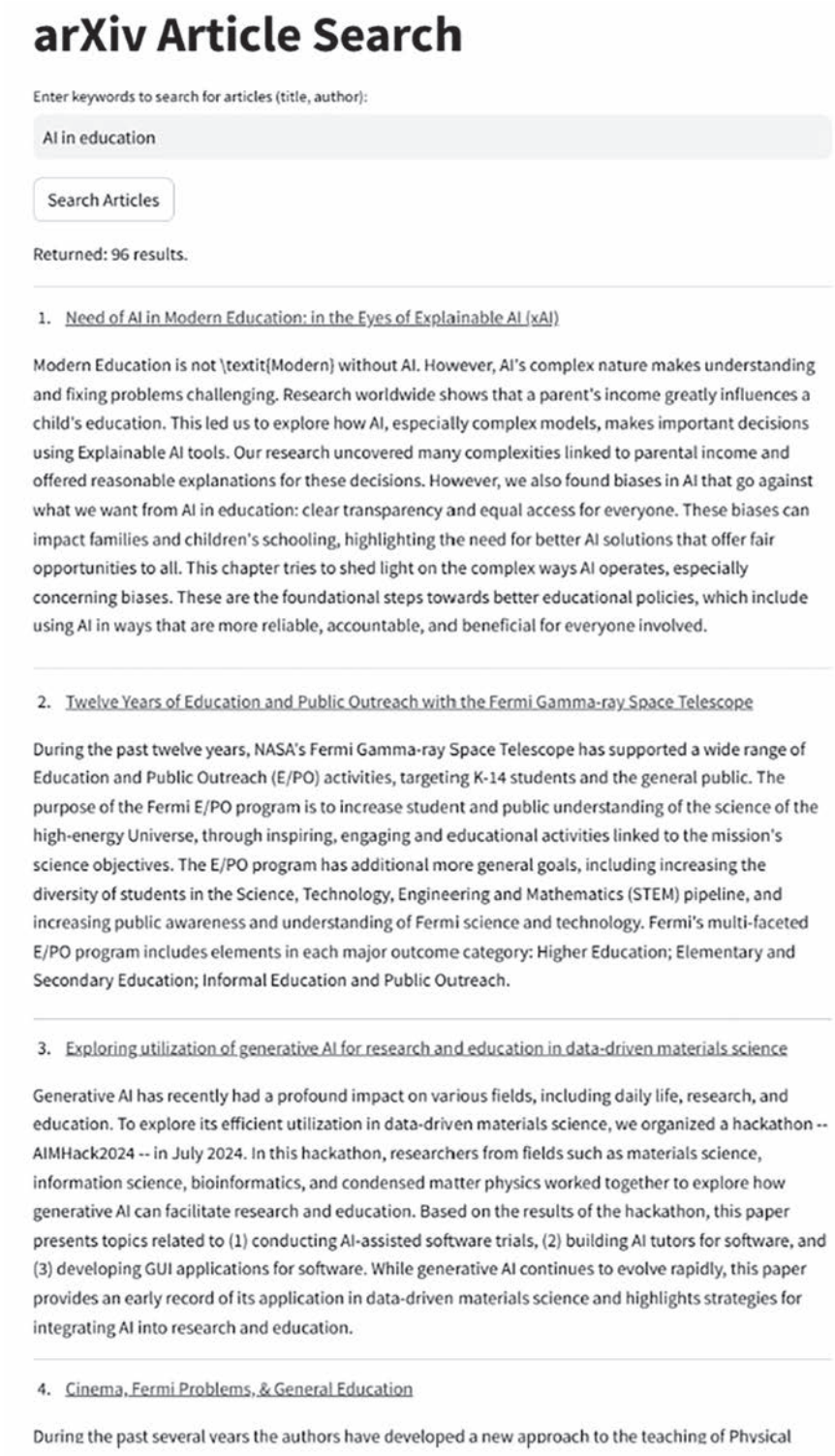

Fig. 3. Example of executing the search query "AI in education" in the Automated Research Assistant interface

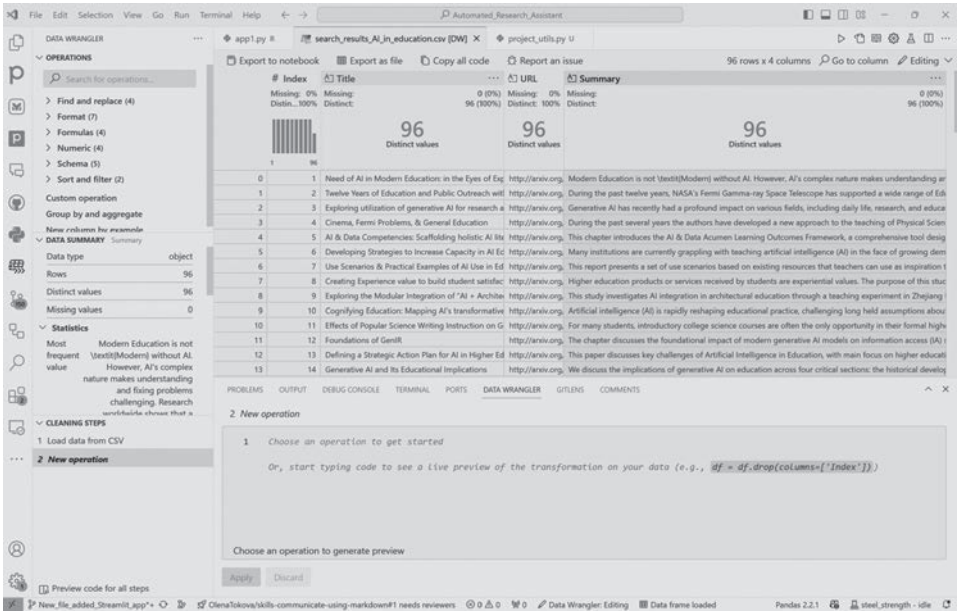

Fig. 4. Storage of search results in the Automated Research Assistant system

The current implementation is limited to the arXiv repository, which may restrict the diversity and coverage of retrieved sources.

Practical Application of the Automated Research Assistant

As part of the study, a practical evaluation of the developed Automated Research Assistant software was carried out during the preparation of educational and research materials.

To demonstrate the system's functionality, queries related to educational technologies, artificial intelligence in education, and interdisciplinary research were used. The user entered a search query through the web interface of the application, after which the system automatically retrieved data from the connected scientific source.

As a result, the system generated a structured list of scientific articles, including publication titles, summaries, and direct links to the sources.

For demonstration purposes, the search query "AI in education" was used. After entering the keywords, the system queried the open-access arXiv repository and generated a set of relevant scientific publications.

Fig. 2 shows the application interface during the search process. As a result of processing the query, the system returned 96 scientific articles corresponding to the specified topic.

The retrieved results are automatically stored as structured files (Fig. 3), containing key metadata such as the title of the publication, a short summary, and a link to the source. This approach allows users to further analyze the materials without repeating the search and supports the accumulation of a personal research dataset.

Structure of stored search results (Title, Summary, URL) in the Automated Research Assistant system is shown in Fig. 3.

The search results are automatically stored in a structured format, allowing users to further analyze sources without repeating the query.

The use of the tool demonstrated that the time required for the initial search of sources is reduced from several hours to a few minutes, while the structured presentation of results enables users to proceed directly to the analytical stage of their work.

Conclusion

The modern educational environment is characterized by the continuous growth of scientific information, which complicates its effective use in learning and research. Under these conditions, it is important not only to have access to information but also to be able to quickly find, select, and analyze relevant sources.

Within this study, the Automated Research Assistant software system was proposed and implemented as a tool for automating the search and initial selection of scientific materials. The results demonstrate that the use of such a tool reduces the time required for information search, organizes the obtained results, and facilitates further work with sources.

The use of the tool does not replace the analytical work of a student or researcher but allows them to focus on the content of materials and the development of their own conclusions. This makes the learning process more structured and manageable.

Further development of the system is associated with the expansion of search sources. In its current implementation, the tool operates with the arXiv repository; however, future integration with other scientific platforms such as CrossRef, OpenAlex, and GitHub is planned. This will expand the range of sources and improve the quality of search results.

Thus, the proposed approach can be used as a supporting tool in educational and research processes and has strong potential for further development.

DECLARATION

Declaration of Competing Interest. No potential conflict of interest is reported by the author.

Funding. The author declare that no funds, grants, or other support were received during the preparation of this manuscript.

Use of AI. The author declare that text was edited using language tools such as Google translate, Gramarly and Language Tool to improve readability and grammatical accuracy.

REFERENCES

1. Chen C. Y., Pedersen S., Murphy K. L. Learners' perceived information overload in online learning via computer-mediated communication. *Research in Learning Technology*, 2011, *19* (2). https://doi.org/10.3402/rlt.v19i2.10345

2. Arnold M., Goldschmitt M., Rigotti T. Dealing with information overload: a comprehensive review. *Frontiers in psychology*, 2023, Vol. 14, Article 1122200. https://doi.org/10.3389/fpsyg.2023.1122200
3. Skulmowski A., Xu K. M. Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational psychology review*, 2022, Vol. 34 (1), 171-196. https://doi.org/10.1007/s10648-021-09624-7
4. Information Literacy. URL: https://www.unesco.org/en/ifap/information-literacy?utm_source=chatgpt.com [Accessed 15 Jul. 2025].
5. Framework for Information Literacy for Higher Education. URL: https://www.ala.org/acrl/standards/ilframework?utm_source=chatgpt.com [Accessed 01 Dec. 2025].
6. Eppler M. J., Mengis J. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 2004, Vol. 20 (5), 325–344. https://doi.org/10.1080/01972240490507974
7. Bawden D., Robinson L. The dark side of information: Overload, anxiety and other paradoxes and pathologies. J*ournal of Information Science*, 2009, Vol. 35(2), 180–191. https://doi.org/10.1177/0165551508095781
8. Sweller J. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 1988, Vol. 12 (2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
9. Sweller J., van Merriënboer J.J.G., Paas F. Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 2019, Vol. 31, 261–292. https://doi.org/10.1007/s10648-019-09465-5
10. Sana F., Weston T., Cepeda N.J. Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, 2013, Vol. 62, 24-31. https://doi.org/10.1016/j.compedu.2012.10.003
11. Kirschner P. A., De Bruyckere P. The myths of the digital native and the multitasker. *Teaching and Teacher education,* 2017, Vol. 67, 135-142. https://doi.org/10.1016/j.tate.2017.06.001

Received 28.12.2025
Accepted 05.05.2026
Published 01.06.2026

О.В. ТОКОВА, канд. техн. наук,
Асистент з питань інклюзивної освіти,
Середня школа Вілмінгтона,
159 Черч-стріт, Вілмінгтон, Массачусетс 01887
https://orcid.org/0000-0002-0289-2810
len327@ukr.net

РОЗРОБЛЕННЯ ІНСТРУМЕНТА АВТОМАТИЗОВАНОГО ПОШУКУ ДОСЛІДНИЦЬКИХ МАТЕРІАЛІВ У ПРОЦЕСІ НАВЧАННЯ

Вступ. Сучасні студенти і дослідники діють в умовах стрімкого зростання обсягів цифрової інформації. Розвиток відкритої науки, електронних бібліотек і онлайн-платформ призвів до суттєвого розширення доступу до наукових знань. Водночас проблема доступу до інформації поступово трансформувалася у проблему її ефективного відбору та інтерпретації. За результатами сучасних досліджень, інформаційне перевантаження є одним із ключових чинників, що ускладнюють навчальну діяльність, знижують концентрацію уваги та підвищують пізнавальне навантаження користувачів. У цифровому освітньому середовищі студенти змушені самостійно здійснювати пошук, оцінювання та систематизацію великої кількості джерел, що потребує сформованих навичок інформаційної грамотності та критичного мислення. Надлишок неструктурованої інформації створює додаткове зовнішнє когнітивне навантаження, яке негатив-

но впливає на процес навчання та засвоєння знань. Незважаючи на значну кількість доступних електронних ресурсів, процес пошуку наукових матеріалів залишається трудомістким і часто супроводжується перевитратами часу та зниженням навчальної мотивації.

Мета. Розробити та апробувати програмний інструмент *Automated Research Assistant*, призначений для автоматизованого пошуку та первинного структурування наукових матеріалів у процесі навчальної й дослідницької діяльності.

Методи. *Automated Research Assistant* реалізований як програмний комплекс і функціонує як веборієнтований застосунок. Програмний комплекс розроблено з використанням мови програмування *Python* та вебінструментів (зокрема *Streamlit*), що забезпечують взаємодію з зовнішніми науковими базами даних через програмні інтерфейси (*API*).

Результати. У роботі проаналізовано проблему інформаційного перевантаження та складності пошуку наукових джерел у сучасному цифровому освітньому середовищі. Показано, що доступність електронних ресурсів не завжди забезпечує ефективність навчання через недостатній рівень сформованості у студентів навичок наукового пошуку та оцінювання інформації. Запропоновано принципи побудови *Automated Research Assistant* — автоматизованого інструмента підтримки навчальної та дослідницької діяльності. Реалізовано веборієнтований програмний комплекс автоматизованого пошуку наукових матеріалів із використанням відкритого репозиторію препринтів *arXiv* через програмний інтерфейс *API*. Система здійснює пошук, первинний відбір та структуроване збереження результатів, формуючи набір метаданих публікацій (назва, короткий опис, посилання на джерело).

Висновки. Запропонований підхід спрямовано на зменшення навантаження на користувача та підвищення ефективності роботи з науковою інформацією в процесі навчання. Архітектура програмного комплексу передбачає можливість подальшого розширення із підключенням додаткових наукових платформ та різних джерел даних.

Ключові слова: *автоматизований пошук, Automated Research Assistant, веборієнтований програмний комплекс, репозиторій препринтів arXiv, підвищення ефективності роботи.*