Electronics and Information Technologies, 2025, 30, 5–14
http://publications.lnu.edu.ua/collections/index.php/electronics/index

UDC 004.89

DESIGN AND IMPLEMENTATION OF AN INTELLIGENT SEARCH SYSTEM BASED ON NEURAL NETWORKS

Vitalii Bozhovskiy*, *Halyna Klym*
Lviv Polytechnic National University,
12 Stepan Bandera St., Lviv 79013, Ukraine

Bozhovskiy, V., Klym, H. (2025). Design and implementation of an intelligent search system based on neural networks. *Electronics and Information Technologies*, 30, 5–14.
https://doi.org/10.30970/eli.30.1

ABSTRACT

Background. In the digital information era, the ability to retrieve relevant data quickly and accurately is increasingly critical. Traditional search engines such as Google or Bing rely on keyword matching, which can fail in cases of vague queries, multilingual content, or media-based searches. The rapid development of neural networks and AI technologies introduces new opportunities to enhance search systems by understanding context, semantics, and user behaviour. This study aims to develop a search system based on ElasticSearch, integrating multiple neural network modules to improve search precision, personalisation, and flexibility.

Methods. The proposed system includes four main components: ElasticSearch for full-text indexing, a convolutional neural network for image recognition, a graph-based semantic model for query expansion, and a ranking model based on historical user interactions. The backend is developed in Python using Visual Studio, with modular AI components that can be activated or disabled by the user. The semantic model represents terms as graph nodes and semantic proximity as weighted edges, enabling dynamic context-driven query refinement. Additional features include synonym detection, citation filtering, and user-specific ranking.

Results and Discussion. Two key experiments were performed. The first examined system performance by testing search speed across database sizes ranging from 100 to 100,000 records. It was found that even with all neural modules enabled, latency remained minimal, confirming system scalability. The second experiment assessed the impact of training data on the quality of the semantic model. A model trained on low-quality, AI-generated data resulted in incoherent word associations and poor query expansion. In contrast, a model built on human-curated texts produced clear, logical semantic links and significantly improved search relevance. The image search function demonstrated the system's ability to identify relevant visual content based on vague or partial user input, while the context expansion model enhanced result diversity and accuracy even with incomplete or ambiguous queries.

Conclusion. This work presents a hybrid search engine that effectively integrates traditional indexing with AI-powered features. The system offers robust text and image search capabilities, intelligent semantic understanding, and personalised ranking. Experiments confirmed its efficiency, relevance, and adaptability across varying data conditions and resource levels. With modular architecture and advanced context handling, the system addresses limitations of conventional search engines and sets a strong foundation for future development in intelligent information retrieval.

Keywords: Search system, search accuracy, neural networks, ElasticSearch.

INTRODUCTION

Searching for information has become an essential part of modern life. Every day, users unknowingly interact with search engines, whether by searching for documentation on Google when writing code, browsing photos through social media tags, or retrieving emails by keywords [1,2]. With the rapid advancement of artificial intelligence (AI), systems like ChatGPT and Bard [3,4] have emerged, further changing the way we search and interact with information. These AI-based platforms do not simply return search results but engage in semantic understanding, enhancing user experience.

Search engines, while designed with the same overarching goal of retrieving relevant information, differ significantly in their functionality and design. For example, Google helps users find websites, Google Maps helps users navigate to physical locations, and social media search engines focus on locating individuals or tagged content [5]. These differences dictate distinct algorithmic approaches and design choices for each system. Unlike traditional search engines, modern AI systems treat queries as questions posed to a human, offering responses that reflect deeper semantic understanding [6].

Traditional search engines often rely on exact word matches, but this approach is highly impractical in real-world scenarios. For example, users may make spelling errors, use synonyms, or express intent imprecisely, leading to poor results when exact matching is applied. To address this, full-text search engines, such as ElasticSearch, incorporate context, enabling systems to account for minor errors or alternative word choices and improve accuracy. This method improves the search process but still faces challenges when dealing with ambiguous or incomplete queries [7]. Furthermore, global search engines must account for variations in user backgrounds, such as language or cultural expectations, which may influence the results even for the same query [8].

Another pressing challenge is scalability, as information continues to grow exponentially. Search engines must be able to process massive volumes of data without sacrificing speed or accuracy. For this, advanced neural network models, such as convolutional neural networks (CNNs) for image recognition or recurrent neural networks (RNNs) for semantic understanding, are increasingly employed. These models must work efficiently with large datasets, enabling real-time processing of user queries, all while maintaining a high level of personalisation and relevance [9,10].

Additionally, users increasingly expect personalised results, meaning that search systems must not only retrieve relevant content but also learn and adapt to individual preferences over time. This involves integrating data from user interactions and feedback to refine ranking algorithms and query results. While such personalisation has greatly improved search performance, it also raises privacy concerns, which need to be addressed through proper data protection and ethical design [11,12].

Despite these advances, the development of effective search engines remains complex, as these systems must contend with linguistic diversity, multimedia content, varying user expectations, and large-scale data processing. As the demands on search engines continue to evolve, there is a pressing need for systems that can intelligently process and retrieve information while adapting to the needs of individual users [13,14].

The goal of this study is to conduct an in-depth analysis of existing search systems, identifying the tools, techniques, and neural technologies they use. Through this analysis, we aim to identify key challenges and solutions, as well as design and implement a custom search engine tailored for a social network project. Our system will incorporate full-text search capabilities, neural modules for semantic expansion, image-based retrieval, and personalised ranking. The system's performance, scalability, and user experience will then be evaluated through experimental testing to identify improvements over existing methods [15].

METHODS

Designing a search system. In this study, we present a search system that goes beyond the functionality of a traditional search engine by incorporating advanced neural network technologies. The system is designed to provide a more dynamic and user-centric experience by integrating several key components: the search engine itself, a multi-level graph neural network, an image analysis neural network, and a neural network responsible for ranking results based on user behaviour. This section describes the design of the system, emphasizing the interaction between these components and their role in optimising the search process.

Search Engine component is responsible for retrieving results from the database based on user queries. A multi-level graph neural network that processes user queries, refines and expands them by considering the semantic relationships between terms in the query. An image analysis neural network is designed to analyse and interpret images, extracting relevant information that is subsequently stored in the database for future searches. Ranking neural network trained on historical user data that ranks search results based on past interactions and user preferences. These components work in tandem to enhance the accuracy and relevance of search results. The flow of information between these subsystems, as well as the user's role in the interaction, is illustrated in Fig. 1.

The user interacts with the system in a series of steps that involve multiple stages of data processing and refinement. Initially, the user creates a post that may contain text, images, or both. If an image is included, the image analysis neural network is triggered to analyse the content of the image, extracting relevant features and storing them in the system's database. It is important to note that at this stage, the image analysis neural network operates independently of the search engine, performing its task of data extraction.

Once the data is stored, the system is ready to process subsequent user queries. If the query is image-based, it is directly sent to the search engine, which searches the database and returns relevant image results to the user. In the case of a text-based query, the process is more complex. The query is first passed to the multi-level graph neural network, which analyses and refines the query to improve its semantic accuracy. Rather

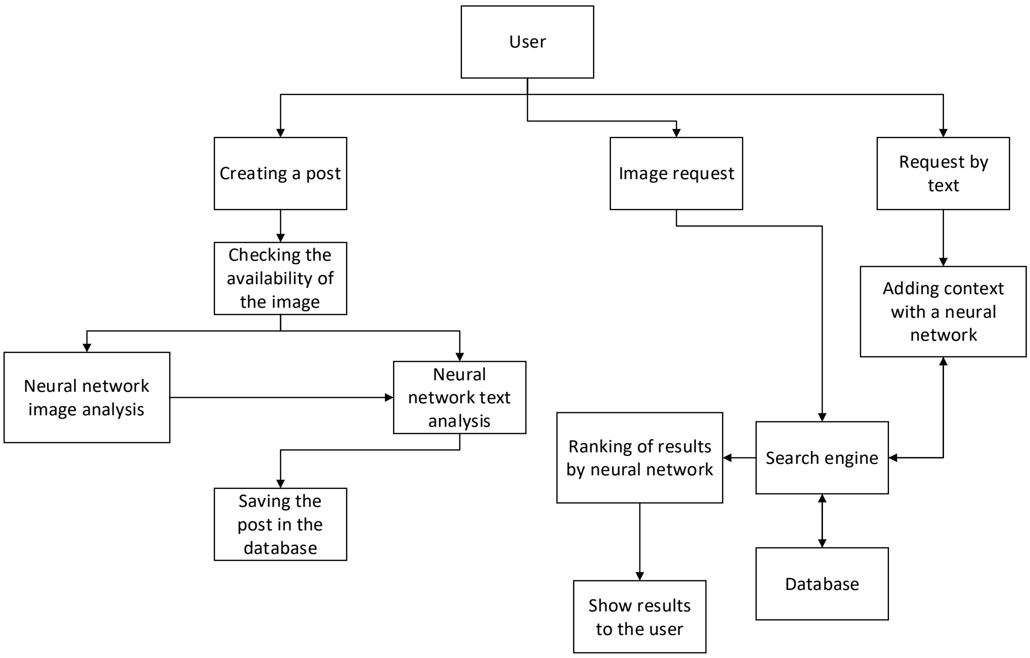


Fig. 1. The process of user interaction with the system.

than simply transmitting the modified query, this neural network actively interacts with the query to ensure it is more precise, helping the search engine to retrieve more relevant results.

Once the search engine returns the results, the ranking neural network takes over. It evaluates the results based on historical data, ranking them according to the user's past behaviour and preferences. The ranked results are then presented to the user. This approach allows the system to provide results that are not only accurate but also personalised, based on the individual user's search history and interests.

A key advantage of our system is its flexibility, allowing users to control the extent to which neural networks are involved in the search process. At any point in the system where neural networks are used, users have the option to disable them. This feature ensures that users can choose a more traditional, non-AI-driven search experience if they prefer. In some cases, users may find that the neural networks' refinement of queries and ranking of results do not align with their expectations, and the ability to turn these features off provides a tailored experience. This level of customisation is a significant advantage over other search systems, which often offer limited control over the search process.

Fusion of information for a multisensory system. The next phase of development for our search system involves the implementation of a neural network model designed to enhance the user's search query. The primary task of this model is to expand the query by utilising a range of techniques aimed at improving the accuracy and breadth of search results. The expansion process follows a systematic approach. The neural network begins by analysing the user's original query. If the initial query results in very few matches or limited synonyms, the model attempts to generate and integrate synonyms to improve the query's reach. Based on the newly generated words or phrases, the model expands the query and sends it to the search engine for further refinement and result retrieval. If the model identifies corresponding words or synonyms within the system, it ensures that relevant results are returned to the user.

This process works seamlessly when the user does not manually alter the query. For example, in a query like "ignorant [girl]," the neural network identifies the word "girl" within brackets and uses it as a cue for modification. The system then searches for syntactically close words to refine the query further. This enables the model to send new queries to the search engine until it produces the most accurate and relevant results, if possible.

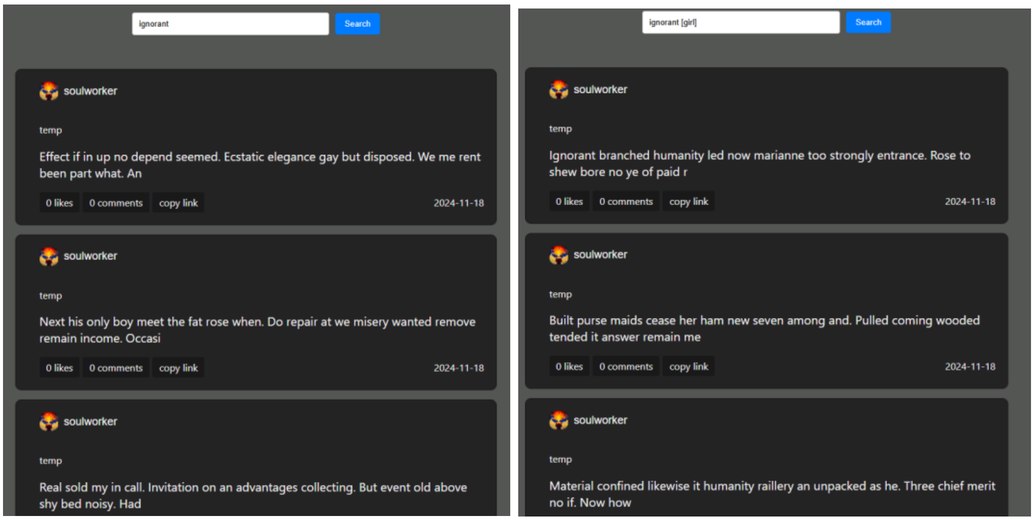
The core of this neural network is based on representing words as vectors within a multi-dimensional space, where semantically similar words are closer to each other. This is achieved by considering the words as nodes within a graph, where edges connect semantically related words, and each edge is assigned a numerical value that quantifies the proximity of those words in the semantic space. By leveraging these connections, the model can generate expanded search queries with a higher likelihood of returning accurate and relevant results. Fig. 2 illustrates how the neural network operates.

These modifications enhance the model’s ability to find more relevant results by broadening the scope of the original search while maintaining the semantic integrity of the query.

To provide users with more control over the search process, we offer the option to disable the neural network model. By using the “-nomodel” modifier, users can opt out of the query expansion process and return to a more traditional search experience. This feature is valuable for users who may prefer a simpler search approach or believe that the neural network's modifications might not always align with their specific needs.

RESULTS AND DISCUSSION

Impact of Data Quality on Model Performance. During the development of the search system, a custom neural network was created and trained to enhance query interpretation and improve search precision. One of the most critical factors influencing the effectiveness of such a model is the quality of the training data. The performance of neural



(a) (b)

Fig. 2. An example of a search query without modification(a), where the search engine processes the query as it is presented by the user and the query after modification (b), where the neural network has expanded and refined the original request by adding semantically related terms and synonyms.

networks, particularly in natural language processing tasks, is heavily dependent not only on the quantity but more importantly on the relevance and coherence of the data. Leading organisations in the field, such as OpenAI, emphasise the importance of high-quality user-generated content for training large-scale language models.

To empirically assess the role of data quality, an experiment was conducted involving the training of two distinct models using datasets of varying quality. The first model was trained on a large dataset consisting of randomly selected, low-quality texts, many of which were generated by other neural networks. Although the dataset comprised several thousand samples, the training process yielded suboptimal results. The semantic graph produced by this model, presented in Fig. 3a, reveals a high degree of disorder. Most nodes (words) are connected with high-weight edges, indicating an artificial and misleading semantic proximity between unrelated terms. This lack of structural coherence renders the model ineffective for search-related tasks, as it fails to capture meaningful linguistic relationships.

In contrast, the second model was trained on a smaller but carefully curated corpus of texts authored by real users. Despite the inherent variability in user-generated content, the selection process prioritised syntactic clarity, topical relevance, and contextual consistency. The resulting semantic graph (Fig. 3b) displays a logical and interpretable network of word associations. Connections between terms are no longer arbitrary, reflecting a semantically sound structure suitable for query expansion and refinement in the search process.

This experiment demonstrates that data quality significantly outweighs quantity when training models for semantic tasks. Excessive data without proper filtering may introduce noise, reduce model robustness, and impair downstream performance. Furthermore, these findings help explain why generative neural networks often struggle to improve through self-generated content, which lacks the depth and structure of well-formed human language. Accordingly, the developed search system integrates the second model, as it provides a stable semantic foundation aligned with real user needs and expectations.

System Quality and Performance Testing. To evaluate the efficiency and scalability of the developed search system, a performance testing experiment was conducted under controlled conditions. Before testing, a utility was implemented to populate the database with synthetic entries, and computational resources were intentionally constrained to

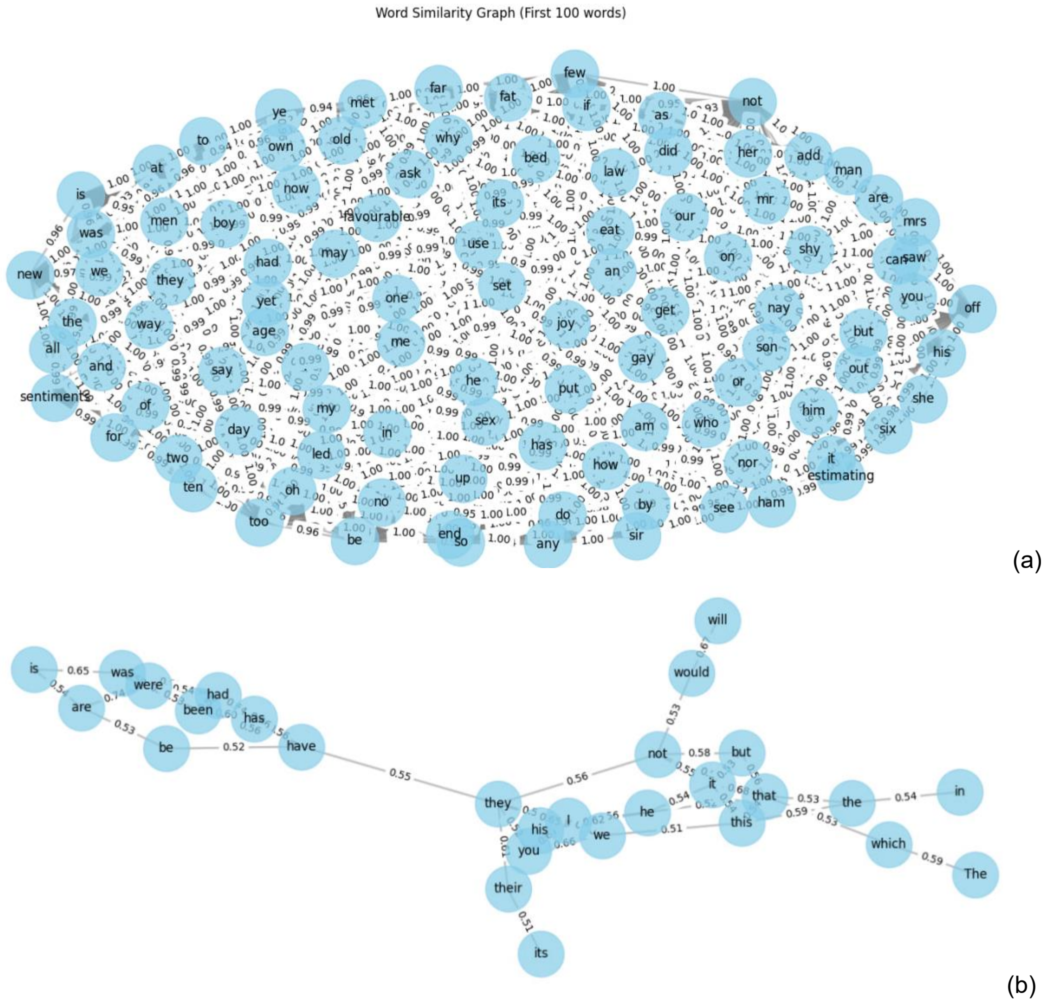


Fig. 3. Graph structure derived from low-quality (a) and high-quality (b) training data.

simulate realistic usage conditions. The objective was to assess the system's response time when processing search queries across varying dataset sizes.

Four distinct dataset sizes were selected for the experiment: 100, 1,000, 10,000, and 100,000 records. These quantities were deemed sufficient to demonstrate performance trends, especially under resource-limited conditions. Three types of search operations were tested: a standard text query, a query involving image analysis and a query processed through all integrated neural networks. The execution times for each scenario were recorded and visualised in the resulting performance graph (Fig. 4).

As illustrated in Fig. 4, standard queries consistently demonstrated the fastest response times. The most noticeable change in latency occurs between the dataset sizes of 1,000 and 10,000 records. Beyond this point, performance degradation becomes more gradual. Although queries involving image analysis and neural network processing introduce some additional delays, the increase in execution time remains within acceptable margins, amounting to only a few milliseconds per additional increment of records.

In addition to efficiency, the quality of search results was also tested and analysed. Used a benchmark set of 50 representative queries. These queries were executed in all three search modes, and standard information retrieval metrics were computed to assess

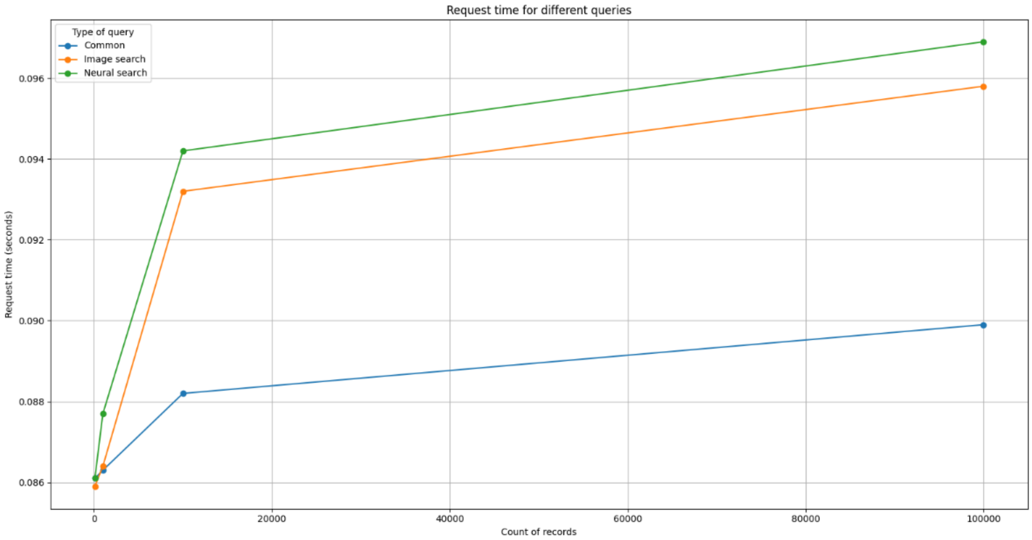


Fig. 4. Requested time as a function of the number of records.

rankings quality: Mean Reciprocal Rank (MRR) and Normalised Discounted Cumulative Gain at rank 10 (NDCG@10). The results are presented in Table 1.

As represented in Table 1, the neural search significantly increased quality performance compared to common search across metrics, which indicates improving relevance and ranking of search results. As a result, this test shows that neural search improves the user experience with the search engine.

These results suggest that while the inclusion of neural networks marginally affects search latency, it does not significantly hinder user experience. The search engine remains responsive and efficient, even as the volume of data increases. This outcome indicates that the system architecture and neural integration are sufficiently optimised to handle large-scale queries without compromising performance.

Table 1. **Quality metrics representation.**

Metric	Common search	Image Search	Neural Search
MRR	0.55	0.6	0.76
NDCG@10	0.53	0.58	0.74

CONCLUSION

As a result of the research conducted, an intelligent search system was developed based on modern technologies for information processing and analysis, in particular, artificial neural networks. At the initial stage, a comprehensive analysis of the search engine market was carried out, which revealed the main shortcomings of existing solutions. Special attention was paid to persistent issues that remain unresolved by current systems, which substantiated the need to create a new tool capable of addressing these challenges.

During the research, the key aspects that should be considered when developing a custom solution were identified. A suitable technological framework was selected, enabling the implementation of a search system with extended functionality. A central feature of the system is the integration of a proprietary neural network that refines user queries and improves result relevance. Additionally, the system provides the ability to activate or deactivate individual modules, ensuring adaptability to diverse user needs.

Upon completion of the development phase, a series of experiments was conducted to evaluate the performance and effectiveness of the system. Testing included measuring technical performance, analysing the impact of data quality and volume on model structure, and assessing the contribution of neural networks to query refinement. The results showed that high-quality and semantically rich data lead to logical and useful relationships within the model, while large volumes of low-quality data result in chaotic and less effective knowledge structures. It was also demonstrated that neural networks significantly enhance search relevance while only minimally affecting request processing speed, indicating a high degree of system optimisation.

So, a competitive search engine has been created that addresses both current and longstanding problems in the field. It introduces innovative solutions, demonstrates high operational efficiency, and holds considerable potential for future commercialisation and scalability.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, [V.B., H.K.]; methodology, [V.B.]; investigation, [V.B.]; writing – original draft preparation, [V.B., H.K.]; writing – review and editing, [V.B., H.K.]; visualization, [V.B.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] Mansour, E., & Alsaud, A. B. F. B. M. (2025). Advanced Information Retrieval Techniques in the Big Data Era: Trends, *Challenges, and Applications. Metallurgical and Materials Engineering*, 31(4), 466-483. https://doi.org/10.63278/1474

[2] Reinanda, R., Meij, E., & de Rijke, M. (2020). Knowledge graphs: An information retrieval perspective. *Foundations and Trends® in Information Retrieval*, 14(4), 289-444. http://dx.doi.org/10.1561/1500000063

[3] Legashev, L., Shukhman, A., Badikov, V., & Kurynov, V. (2025). Using Large Language Models for Goal-Oriented Dialogue Systems. *Applied Sciences*, 15(9), 4687. https://doi.org/10.3390/app15094687

[4] Ram, B., & Verma, P. (2023). Artificial intelligence AI-based chatbot study of ChatGPT, Google AI Bard and Baidu AI. *World Journal of Advanced Engineering Technology and Sciences*, 8(01), 258-261. https://doi.org/10.30574/wjaets.2023.8.1.0045

[5] Aggarwal, C. C., & Aggarwal, C. C. (2018). Information retrieval and search engines. *Machine Learning for Text*, 259-304. https://doi.org/10.1007/978-3-319-73531-3_9

[6] Hambarde, K. A., & Proenca, H. (2023). Information retrieval: recent advances and beyond. *IEEE Access,* 11, 76581-76604. https://doi.org/10.1109/ACCESS.2023.3295776

[7] Walter-Tscharf, F. F. W. V. (2022). Indexing, clustering, and search engine for documents utilizing Elasticsearch and Kibana. *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2022*, 897-910. https://doi.org/10.1007/978-981-19-2069-1_62

[8] Zhang, F., Zhang, Z., Ao, X., Gao, D., Zhuang, F., Wei, Y., & He, Q. (2022, June). Mind the gap: Cross-lingual information retrieval with hierarchical knowledge enhancement. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4), 4345-4353. https://doi.org/10.1609/aaai.v36i4.20355

[9] Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. Advanced Machine Learning Technologies and Applications: *Proceedings of AMLTA 2020*, 599-608. https://doi.org/10.1007/978-981-15-3383-9_54
[10] Choi, M., Kim, H., Han, B., Xu, N., & Lee, K. M. (2020). Channel attention is all you need for video frame interpolation. *Proceedings of the AAAI conference on artificial intelligence*, 34(7), 10663-10671. https://doi.org/10.1609/aaai.v34i07.6693
[11] Liu, G., Sun, X., Li, Y., Li, H., Zhao, S., & Guo, Z. (2023). An automatic privacy-aware fFramework for text data in online social network based on a multi-deep learning model. *International Journal of Intelligent Systems*, 2023(1), 1727285. https://doi.org/10.1155/2023/1727285
[12] Karamiyan, F., Mahootchi, M., & Mohebi, A. (2024). A personalized ranking method based on inverse reinforcement learning in search engines. *Engineering Applications of Artificial Intelligence*, 136, 108915. https://doi.org/10.1016/j.engappai.2024.108915
[13] Shi, F., Chen, L., Han, J., & Childs, P. (2017). A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design*, 139(11), 111402. https://doi.org/10.1115/1.4037649
[14] Gupta, V., Sharma, D. K., & Dixit, A. (2021). Review of information retrieval: Models, performance evaluation techniques and applications. *International Journal of Sensors Wireless Communications and Control*, 11(9), 896-909. https://doi.org/10.2174/2210327911666210121161142
[15] Zhao, S., Liang, Z., Wen, J., & Chen, J. (2022). Sparsing and smoothing for the seq2seq models. *IEEE Transactions on Artificial Intelligence*, 4(3), 464-472. https://doi.org/10.1109/TAI.2022.3207982

ПРОЕКТУВАННЯ ТА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ ПОШУКОВОЇ СИСТЕМИ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Віталій Божовський*, *Галина Клим*
Національний університет «Львівська політехніка», вул. Степана Бандери 12, м. Львів, 79013 Україна

АНОТАЦІЯ

Вступ. У цифрову епоху інформації здатність швидко та точно знаходити релевантні дані стає дедалі важливішою. Традиційні пошукові системи, такі як Google або Bing, базуються на співставленні ключових слів, що може бути неефективним у випадках нечітких запитів, багатомовного контенту або пошуку за медіаданими. Швидкий розвиток нейронних мереж та технологій штучного інтелекту відкриває нові можливості для вдосконалення пошукових систем шляхом розуміння контексту, семантики та поведінки користувачів. Це дослідження спрямоване на розроблення пошукової системи на базі ElasticSearch з інтеграцією кількох модулів нейронних мереж для підвищення точності пошуку, персоналізації та гнучкості.

Методи. Запропонована система включає чотири основні компоненти: ElasticSearch для повнотекстового індексування, згорткову нейронну мережу для розпізнавання зображень, семантичну модель на основі графів для розширення запитів і модель ранжування, побудовану на основі історії взаємодії користувача. Серверна частина реалізована на мові Python з використанням середовища Visual Studio, а модулі штучного інтелекту мають модульну структуру й можуть бути увімкнені або вимкнені користувачем. Семантична модель представляє терміни у вигляді вузлів графа, а семантичну близькість — у вигляді зважених ребер, що дозволяє динамічно уточнювати запити з урахуванням контексту. Додаткові функції включають виявлення синонімів, фільтрацію за цитованістю та персоналізоване ранжування результатів.

Результати. Було проведено два ключові експерименти. Перший досліджував продуктивність системи шляхом вимірювання швидкості пошуку при розмірі бази даних від 100 до 100 000 записів. Результати показали, що навіть за увімкнення всіх нейронних модулів затримка залишалася мінімальною, що підтверджує масштабованість системи. Другий експеримент оцінював вплив навчальних даних на якість семантичної моделі. Модель, навчена на неякісних текстах, згенерованих штучним інтелектом, продемонструвала хаотичні зв'язки між словами та низьку ефективність у розширенні запитів. Натомість модель, побудована на текстах, відібраних людиною, формувала чіткі логічні семантичні зв'язки та суттєво підвищувала релевантність результатів пошуку. Функція пошуку за зображенням засвідчила здатність системи знаходити релевантний візуальний контент навіть за нечіткими або частковими запитами користувача, а модель контекстного розширення забезпечувала різноманітність і точність результатів навіть за неповних чи неоднозначних запитів.

Висновки. У даній роботі представлено гібридну пошукову систему, яка ефективно поєднує традиційне індексування з можливостями, що надаються штучним інтелектом. Система забезпечує надійний пошук як текстової, так і візуальної інформації, демонструє інтелектуальне розуміння семантики та персоналізоване ранжування результатів. Експериментальні дослідження підтвердили її ефективність, релевантність та адаптивність за різних умов обсягу даних і обмежених ресурсів. Завдяки модульній архітектурі та розширеній обробці контексту система вирішує обмеження традиційних пошукових механізмів і закладає міцну основу для подальшого розвитку інтелектуального пошуку інформації.

Ключові слова: Система пошуку, точність пошуку, нейронні мережі, ElasticSearch.

Received / Одержано	Revised / Доопрацьовано	Accepted / Прийнято	Published / Опубліковано
07 May, 2025	19 May, 2025	23 May, 2025	26 June, 2025