DOI: https://doi.org/10.15407/kvt211.01.040

ZOSIMOV V.V., DSc (Engineering), Associate professor,
Professor of the Department of Applied Information Systems,
https://orcid.org/0000-0003-0824-4168, e-mail: zosimovv@gmail.com
Taras Shevchenko National University of Kyiv,
60, Volodymyrska Street, Kyiv, 01033, Ukraine

PROBABILISTIC APPROACH TO RANKING SEARCH RESULTS USING BAYESIAN BELIEF NETWORKS

Introduction. *This paper proposes a probabilistic approach to ranking search results using Bayesian Belief Networks (BBN). The proposed approach utilizes BBN to model the relationships between search queries, web pages, and user feedback, and to calculate the probability of a web page being relevant to a specific query. The approach takes into account various factors, such as keywords, page relevance, domain authority, and user feedback to generate a ranking score for each search result.*

The purpose of the paper *is to conduct an analysis on the feasibility of creating a search engine that uses BBNs and probabilistic ranking methods for improving the accuracy and efficiency of search results.*

Results. *The proposed approach was evaluated on a real-world dataset, and the results showed its effectiveness. Overall, the results suggest that the use of BBNs can provide a promising approach to enhancing search engine performance and user experience. The approach's effectiveness is attributed to its ability to model and reason about uncertainty and dependencies among variables, and its consideration of various factors, such as keywords, page relevance, domain authority, and user feedback.*

Conclusions. *The proposed method has the potential to improve search relevance, reduce user frustration, and increase user satisfaction. However, further research is needed to optimize the proposed approach and to explore its applicability in different contexts. Overall, the study suggests that BBNs can provide a valuable tool for developing more effective and user-friendly search engines. Moreover, the use of Sphinx as a base search system shows promise in enabling the proposed approach to be integrated into practical search systems. Nonetheless, further research is needed to optimize the approach and evaluate its applicability in different contexts.*

Keywords: *search engine, ranking, Bayesian Belief Networks, probabilistic model, information retrieval.*

INTRODUCTION

The exponential growth of digital data and the increasing reliance on the internet as a source of information have made search engines an indispensable tool for navigating the vast and complex landscape of online information. Despite the remarkable progress made in information retrieval and search algorithms, modern search systems still face significant challenges in accurately and efficiently retrieving relevant information. Traditional search algorithms based on keyword matching and frequency analysis have limitations in dealing with the complexity and diversity of search queries and content, leading to inaccurate or incomplete results.

Moreover, search engines are facing issues of ranking and user satisfaction. Ranking, the process of ordering search results based on their relevance to the user's query, has a direct impact on user satisfaction and engagement. Inaccurate ranking, irrelevant results and poor search experience can lead to user frustration, decreased trust in the search engine and ultimately reduced usage. Additionally, the sheer volume of search results and the increasing complexity of search queries make it challenging for users to find what they are looking for, leading to poor user satisfaction and negative feedback.

We explore the potential of BBNs as a probabilistic graphical model that can represent and reason about uncertainty and dependencies among variables, making it a powerful tool for modeling complex systems such as search engines. By leveraging the power of BBN, we propose a novel approach to ranking search results that can take into account multiple factors such as search query relevance, content quality and user behavior, and provide a more accurate and personalized search experience.

The results of this study will help inform future research directions in the field of search engine development and optimization.

Existing research in the field of search engine development and optimization has explored various approaches for improving search result ranking. One approach is to incorporate machine learning techniques such as neural networks, support vector machines and decision trees to model the relationships between different factors influencing search result relevance. These models can learn from large amounts of data and automatically identify important features for ranking search results. However, these models may not capture the uncertainty and dependencies among factors, leading to suboptimal ranking performance.

Another approach is to incorporate user feedback and behavior into the ranking process. For example, some search engines use click-through rates and dwell time to measure user satisfaction with search results and adjust ranking accordingly. Other search engines allow users to provide explicit feedback on the relevance of search results, which can be used to improve the ranking algorithm. However, incorporating user feedback can be challenging due to the inherent subjectivity and variability of user preferences.

Furthermore, recent research has explored the use of semantic and contextual information to improve search result ranking. By analyzing the meaning and context of search queries and content, these approaches aim to provide more accurate and personalized search results. For example, some search engines use natural language processing techniques to understand the intent behind a query and provide more relevant results. Others use contextual information such as location, time of day, and user history to personalize search results.

Overall, the field of search engine development and optimization is constantly evolving, and researchers are exploring new and innovative approaches for improving search result ranking. The proposed approach of using Bayesian Belief Networks offers a promising direction for incorporating uncertainty and dependencies among factors into the ranking process and providing more accurate and personalized search results [1].

The purpose of the paper is to carrying out an analysis on the feasibility of creating a search engine that uses Bayesian Belief Networks (BBNs) and probabilistic ranking methods for improving the accuracy and efficiency of search results.

SEARCH RESULTS RANKING USING BAYESIAN BELIEF NETWORKS

One of the key advantages of using BBNs for search results ranking is their ability to handle uncertainty and probabilistic reasoning. In traditional ranking methods, the relevance of a search result is typically based on a binary classification of relevant or irrelevant. However, in reality, the relevance of a search result is often uncertain and subjective, depending on various contextual factors and user preferences. BBNs can provide a more nuanced and personalized approach to ranking by considering multiple factors and their interdependencies.

Moreover, BBNs can be used to improve the transparency and interpretability of search results ranking. Traditional ranking methods often rely on black-box algorithms that are difficult to understand and explain. In contrast, BBNs provide a graphical representation of the factors and their relationships, making it easier for users and developers to understand and analyze the ranking process. Additionally, BBNs allow for the incorporation of prior knowledge and domain expertise, enabling the development of more accurate and reliable ranking models.

Bayesian Belief Networks (BBNs) can be used for search results ranking by modeling the relationships between different factors that influence the relevance of a search result to a query. Here are some steps to follow [1]:

- Identify the factors that can influence the relevance of a search result to a query. These may include factors such as keyword matching, page rank, domain authority, user click-through rates and so on.
- Define the network structure by creating a directed acyclic graph that shows the relationships between the different factors. Each factor can be represented as a node in the graph, and the edges represent the dependencies between the factors.
- Assign probabilities to each node in the graph. These probabilities represent the likelihood of a particular factor given its parent nodes.
- As the user enters a query and interacts with the search results, update the probabilities in the network based on the user behavior. For example, if a user clicks on a particular result, this can increase the probability of that result being relevant to the query.
- Rank the results. Use the probabilities in the network to rank the search results based on their relevance to the query.

MATHEMATICAL MODEL OF THE SEARCH RESULTS RANKING SYSTEM

A mathematical model of a search results ranking system can be constructed by using a Bayesian network. The model will consist of several layers that represent the different factors that influence the relevance of a search result to a query.

These factors may include keyword matching, page rank, domain authority, user click-through rates and so on. The model can be represented as a directed acyclic graph in which each node represents a factor and the edges represent the dependencies between the factors.

To construct the Bayesian network, it is necessary to assign probabilities to each node in the graph. These probabilities represent the likelihood of a particular factor given its parent nodes. For example, the probability of a search result being relevant to a query may depend on its page rank, the user's previous search history, and other factors.

As the user interacts with the search results, the probabilities in the network are updated based on the user behavior. For example, if a user clicks on a particular result, this can increase the probability of that result being relevant to the query [3].

Let there be some number of documents that were received from the Internet. Each document has certain characteristics, with various attributes specific to certain documents. In the search system, the necessary keywords are specified, which characterize exactly the information that needs to be found. A set of keywords and criteria is called a search query [2].

The degree of relevance of each particular document to the query is called the relevance of the query document. The task of the search engine is to provide the user with the most relevant results that match his request as much as possible.

Thus, it is necessary to construct an intelligent system that allows you to assess the degree of relevance of each document of a given request. The relevance measure is a number, according to its value, documents can be sorted by relevance. This number should have the following parameters:

1) non-negative real number;
2) the degree of relevance will be as high as the relevance of the query itself;
3) limiting the relevance measure from above.

The Bayesian network is used to model subject areas characterized by uncertainty. This uncertainty may be due to insufficient understanding of the subject area or a combination of given factors.

Bayesian networks are also called Bayesian Belief Networks (BBNs) or simply trust networks. A BBN is a graph in which the nodes are connected by directed edges, with a probability function assigned to each node. A network in BBN is a directed acyclic graph in which there is no directed route that starts and ends at the same nodes.

If a node has no edges pointing to it, it will contain a table of unconditional probabilities of its states. In the case of a discrete node, such a table contains the distribution of probabilities between all possible states of this nodes. If the nodes are parents (one or more edges directed to it), then such a node contains a conditional probability table (CPT), each cell of which contains the conditional probability of the nodes being in a certain state for the case of a certain configuration of the states of all its parents. Thus, the number of cells in the table of conditional probabilities of a discrete BBN node is equal to the product of the number of possible states of this node by the product of the number of possible states of all its parent nodes [3].

A trivial BBN reflects a cause-and-effect relationship between two elements of some subject area — A and B. The presence of a cause-and-effect relationship from A to B means that if A is in a certain state, it affects the state of B.

Table 1. Conditional probabilities for BBN nodes

B	$P(b_i\|a_1)$	$P(b_i\|a_2)$
b_1	1	0,6
b_2	0	0,2
b_3	0	0,2

Table 2. Conditional probabilities for BMD nodes

A	$P(a_i)$
a_1	0,5
a_2	0,5

The random discrete variable represented by the node A can be in one of two states, as shown in Table 1, a_1 or a_2. Node B has three possible states: b_1, b_2, b_3.

Since node A has no parents, the probabilities of its states are not dependent. For node B, on the contrary, the probabilities of states depend on the state of its parent node A:

— If A is in state a_1, then B is in state b_1;

— If A is in state a_2, then the probability that B is in state b_1 is 0.6, and in states b_2 and b_3 — 0.2.

If node A has no parents, then unconditional probabilities $P(A)$ are automatically used instead of conditional probabilities.

The process of calculating probabilities is the basis for decision-making in conditions of uncertainty based on BBNs [4]. Disclosure of uncertainty (dealing with uncertainty) is carried out in the BBN by calculating the probabilities of the states of the nodes, that interest us, based on the available information about the values (parts) of other nodes of the network. The mathematical basis for this process is determined by the Bayesian approach to uncertainty analysis and the corresponding apparatus of classical probability theory.

The top of the BBN represents the basis of the Bayesian approach, which constitutes the concept of conditional probability $P(A|B) = x$, which means that, given the occurrence of B (and everything else that is not related to B), the probability of occurrence of A is equal to x. The joint probability of occurrence of A and B is determined by the formula of total probability

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A) \tag{1}$$

Equation (1) is the fundamental principle of probability calculation and the basis for Bayes' theorem:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} \tag{2}$$

Bayes' theorem is used if there is information on the dependent variables in the given order, and the essence of the study is to determine the probability of the initial variables [5]. So, let the conditional probability $P(B|A)$ of the occurrence of some event B be known under the condition that event A occurs. Then the Bayes theorem gives a solution to the inverse problem — what is the probability of occurrence of event A, event B has occurred.

Let A_1, A_2, ..., A_n be a complete group of incompatible mutually exclusive events (or alternative hypotheses). Then the posterior probability $P(A_j|B)$ of each of the events A_j, $j = 1..n$, provided that event B has occurred, is expressed by the a priori probability A_j [6]:

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \tag{3}$$

EVALUATING DOCUMENT RELEVANCE

To evaluate the relevance of a document, we need to define the factors that determine its relevance. These factors may include keyword matching, page rank, domain authority, user click-through rates and so on. We can model these factors as nodes in the Bayesian network, with the relevance of the document as the final node in the network. Each node represents a factor that affects the relevance of the document, and the edges between nodes represent the dependencies between these factors.

To calculate the probability of the document being relevant given the values of the other nodes, we need to define conditional probability tables for each node in the network. These tables describe the probability of a node being in a certain state given the states of its parent nodes. For example, the probability of a document being relevant given that it has a high page rank and keyword matching may be higher than if it only has one of these factors.

Once the Bayesian network is constructed and the conditional probability tables are defined, we can use it to evaluate the relevance of a document for a given query. By propagating the probabilities through the network, we can calculate the probability of the document being relevant given the query. We can then rank the documents based on these probabilities and present the most relevant documents to the user [9].

One of the tasks in which BBNs are successfully used is the task of classification. The so-called naive Bayesian classifier, which is a simple Bayesian network, is one of the most effective classifiers. This approach assumes that the relevance assessment task can also be considered as a classification task. Each document belongs to one of two non-intersecting areas: C_1 — relevant documents, C_2 — irrelevant documents [7].

In this case, the task of assessing the relevance of the request document is presented as the task of assigning it to one of two classes. In this case, the document belonging to the first class allows testifying that this document is relevant to the request.

Let's solve this problem with the help of a Bayesian network, matching the concept of "document" to the top of the network. This node can be in two states: c_1 — "the document is relevant" and c_2 — "the document is not relevant". We set the a priori probabilities of these states equal to 0.5, which corresponds to the concept of uncertainty in probabilistic analysis:

$$P(c_1) = P(c_2) = 0{,}5. \tag{4}$$

If, as a result of calculations, we get that the probability of finding this node in the "document is relevant" state is equal to 0.9, it means that with a probability of 0.9 this document belongs to class C_1.

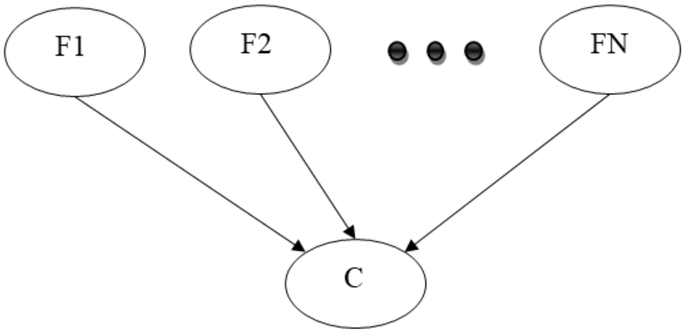

Fig. 1. Bayesian network for evaluating the relevance of a query document

Next, let $F = \{Fi\}$, $i = 1..n$ be the set of factors affecting the relevance of the document. Consider, for example, such a factor as the presence of a query keyword in the title of the document. Obviously, the presence of a keyword in the title increases the relevance of the document. Then we will enter the node F_1 corresponding to the event "keyword in the title of the document" into the network. This node will have two states: f_{11} - "the keyword is found in the title of the document" and f_{12} - "the keyword is not found in the title of the document". Conditional probabilities $P(f_{1j} \mid c_j)$, $i, j = 1..2$, then for this there is a table of conditional probabilities for the node F_1, and you can calculate the probabilities $P(c_i \mid f_{1i})$, $i, j = 1..2$.

To assign document D to the class of relevant when the state is known f_{1j}, an obvious rule is used: if $P(c_1 \mid f_{1j}) > P(c_2 \mid f_{1j})$, then $D \in C_1$.

Therefore, to determine the relevance, all the factors that make up the set F must be selected, and the tables of conditional probabilities for each factor must be specified. Each of the factors is calculated accordingly for each query keyword.

Thus, the vertices of the network in this case are factors affecting the probability of a given "main" node responsible for the relevance of the document as a whole. The Bayesian network for this problem looks as shown in Fig. 1.

C is the top of the network, representing the probability that the document is relevant to the request, and F_1, F_2 ... Fn are factors taken into account when calculating this probability. An essential point is the direction of cause-and-effect relationships in the network. So, the arrows leave the node C and enter the node F_i. Here, the Bayesian network performs reverse logical inference — it determines the probability of each state of the node C given the known states of the vertices F_i [8].

The F_1 factor considered above had one of two values. However, factors can have a different nature - they can take several values, or they can not be discrete at all. In the general case, we consider a certain range of changes in the values of each factor. Let some factor Fi take the value, then the value of this factor is normalized in the range [-1; 1] using the formula:

$$\widetilde{x} = \frac{x - \frac{x_{\min} + x_{\max}}{2}}{x_{\max} - \frac{x_{\min} + x_{\max}}{2}} \qquad (5)$$

and the calculated probabilities for the corresponding node are assumed to be equal

$$\begin{aligned}\widetilde{P}(f_i \mid c_1) &= \frac{1-[1-2P(f_i \mid c_1)]\widetilde{x}}{2}, \\ \widetilde{P}(f_i \mid c_2) &= \frac{1-[1-2P(f_i \mid c_2)]\widetilde{x}}{2}, i=1...n\end{aligned} \quad (6)$$

where $P(f_i \mid c_1)$ is an element of the table of conditional probabilities for the i-th network node, which shows with what probability the factor F_i takes the maximum value $x = x_{max}$ in the relevant document; $P(f_i \mid c_1)$ — is the probability with which the factor F_i takes the maximum value $x = x_{max}$ in an irrelevant document.

The resulting calculated probabilities $\widetilde{P}(f_i \mid c_1)$ and $\widetilde{P}(f_i \mid c_2)$ can now be used in the Bayes formula:

$$P(c_1 \mid f_i) = \frac{P(f_i \mid c_1)P(c_1)}{P(f_i \mid c_1)P(c_1) + P(f_i \mid c_2)P(c_2)},\ i=1...n \quad (7)$$

Thus, the above-described scheme allows taking into account both discrete and continuous values of factors affecting the overall relevance of the document. At the same time, if the increase in the value of the factor corresponds to a decrease in relevance (for example, the number of days that have passed since the publication of the announcement), then it is enough to repeat these considerations for the case when the element of the table of conditional probabilities $P(f_i \mid c_1)$ shows with what probability the factor F_i takes the minimum value $x = x_{min}$ in the relevant document [9].

In this case, the network is sufficiently trivial that the calculation of probabilities can be performed by sequential application of Bayes' theorem. Such a calculation is possible only if a strong assumption about the conditional independence of the vertices of the network works. The conditional independence of the vertices of the Bayesian network means blocking the influence between these vertices. Variables (sets of variables) F_1 and F_2 are independent when the state of variable A is known, if

$$P(F_1,A)=P(F_1|A, F_2). \quad (9)$$

This means that if the state of node A is known, then no information about F_1 changes the probability of F_2. This is represented by the absence of causal relationships between all factors of the set F.

In fact, this assumption is obviously completely unrealistic (which is why classifiers of this structure are called "naive"). At the same time, violation of this assumption in real-world conditions does not have a significant effect on the final result. It turns out that such a sequential approach is an advantage in this case, as it sharply reduces the computational complexity and, accordingly, the speed of the algorithm.

Attesting to obtaining numerical values for tables of conditional probabilities, it should be noted that conceptually two approaches are distinguished for solving this task [10]:

- receiving information from experts in the subject area;
- obtaining information on the basis of data.

Tables of conditional probabilities are most often generated on the basis of data using statistical methods. However, it is worth noting that the fundamentally subjective Bayesian approach does not require the "objectivity" of probabilities, and therefore allows the formation of tables of conditional probabilities to be based on the subjective assessments of experts. Conditional probabilities, the numerical values that are used for calculation, are obtained on the basis of merging the results of statistical studies and expert assessments. A statistical analysis of a set of relevant and irrelevant documents for various queries from various sources of information was carried out and these values were entered into the table of conditional probabilities of the network.

PROCESSING OF SEARCH REQUESTS

Since developing a search engine from scratch is a very time-consuming task, the Sphinx system was chosen as the base module for the search engine. Sphinx is a free, open-source search engine designed for exceptionally fast text searches [11]. For example, in a real database consisting of about 300,000 strings and five indexed columns, where each column contains about 15 words, Sphinx can return the result of a search for "any of these words" in one-hundredth of a second (at server with an AMD Opteron processor with a frequency of 2 GHz and 1 GB of memory, running Debian Linux).

Before starting the search, you need to check the query for errors, otherwise the result will be incorrect or nothing will be found at all. For this task, this block uses a conditional operator that checks the correctness of a regular expression. The structure of the expression is presented in the form:

```
([A-Za-z_]+):([0-9.]+):([a-z]+);([A-Za-z_]+):([0-9.]+):([a-z]+);([A-Za-z_]+):
([0-9.]+):([a-z]+);([A-Za-z_]+):([0-9.]+):([a-z]+)
```

This expression prevents the possibility of a false query.

In case of mismatch - returns an error.

After checking the request, in order to conduct a search on the specified sites, you first need to download them to the server and carry out certain processing for a more convenient search. This method is called ranking.

Ranking is performed by the method of sorting through all available text content on the site and processing according to the rank of importance:

- ranking of websites by Meta-tags;
- website ranking through backlinks;
- ranking of websites by content quality.

CONCLUSIONS

The proposed method has the potential to improve search relevance, reduce user frustration, and increase user satisfaction. However, further research is needed to optimize the proposed approach and to explore its applicability in different contexts. Overall, the study suggests that Bayesian Belief Networks can provide a valuable tool for developing more effective and user-friendly search engines. Moreover, the use of Sphinx as a base search system shows promise in enabling the proposed approach to be integrated into practical search systems. Nonetheless, further research is needed to optimize the approach and evaluate its applicability in different contexts. Overall, BBNs can provide a flexible and powerful way to model the complex relationships between the different factors that influence search results ranking. However, creating an accurate and effective BBN for search results ranking can be a complex task that requires expertise in both search engine technology and probabilistic modeling.

REFERENCES

1. Baeza-Yates, R., & Ribeiro-Neto, B. Modern Information Retrieval: The Concepts and Technology behind Search. Addison-Wesley Professional. 2011.
2. Sattari P. Bayesian deep reinforcement learning: A survey. *Journal of Machine Learning Research*. JMLR.org. 2020, Vol. 21, pp. 1–35.
3. Agichtein, E., Brill, E., & Dumais, S. Improving Web Search Ranking: Beyond the Query-Document Similarity. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2006, Vol. 1(1), pp. 1–136.
4. Chau M. Spidering and Filtering Web Pages for Vertical Search Engines. *Proceedings of The Americas Conference on Information Systems*. AMCIS 2002 Doctoral Consortium, Dallas, TX, USA, 2002.
5. Zosimov V.V., Bulgakova O.S., Pozdeev V.O. Complex internet data management system. *Advances in Intelligent Systems and Computing*. AISC. 2021, Vol.1246, pp. 639–652.
6. Pelt M. Uncertainty quantification in deep learning using Bayesian convolutional neural networks. *Journal of Computer Vision*. 2019, Vol. 126, pp. 617–635.
7. Zosimov. V.V., Bulgakova. O.S. Calculation the Measure of Expert Opinions Consistency Based on Social Profile Using Inductive Algorithms. Advances in Intelligent Systems and Computing. 2020. Vol. 1020. pp. 622–636.
8. Bendersky, M., Croft, W. B., & Zhang, J. Predicting query performance via classification. *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*. 2010, pp. 79–88.
9. Hron J. Probabilistic programming for deep learning: A review. *Machine Learning Research*. 2018, Vol. 19, pp 1–41.
10. Gallego C. A review of Bayesian deep learning techniques and their application to computer vision problems. *Big Data Analytics*, IGI Global. 2018, pp. 11-25.
11. Guo C. Deep Bayesian active learning for neural networks. *Journal of Machine Learning Research*, JMLR.org. 2017, Vol. 18, pp. 1–47.
12. Sattari P. Bayesian deep reinforcement learning: A survey. *Journal of Machine Learning Research*, JMLR.org. 2020, Vol. 21, pp. 1–35.
13. Nalisnick M. Deep Bayesian neural networks with many irrelevant inputs. *Proceedings of the 35th International Conference on Machine Learning*. 2019, Vol. 97, pp. 1748–1757.
14. Official Sphinx search system site. URL: Sphinx http://www.sphinxsearch.com/

Received 23.01.2023

Зосімов В.В., доктор техн. наук, доцент,
Професор кафедри прикладних інформаційних систем,
e-mail: zosimovvv@gmail.com

Київський національний університет імені Тараса Шевченка,
60, Володимирська вул., Київ, 01033, Київ, Україна

ІМОВІРНІСНИЙ ПІДХІД ДО РАНЖУВАННЯ РЕЗУЛЬТАТІВ ПОШУКУ З ВИКОРИСТАННЯМ БАЙЄСІВСЬКИХ МЕРЕЖ ДОВІРИ

Вступ. *В статті запропоновано імовірнісний підхід до ранжування результатів пошуку за допомогою байєсівських мереж довіри (БМД). Запропонований підхід використовує БМД для моделювання зв'язків між пошуковими запитами, веб-сторінками та відгуками користувачів, а також для розрахунку ймовірності відповідності веб-сторінки конкретному запиту. Цей підхід враховує різні фактори, такі як ключові слова, релевантність сторінки, авторитет домену та відгуки користувачів для того, щоб створити оцінку рейтингу для кожного результату пошуку.*

Мета статті *— проведення аналізу доцільності створення пошукової системи, що використовує байєсівські мережі довіри (БМД) та ймовірнісні методи ранжирування для підвищення точності та ефективності результатів пошуку.*

Результати. *Запропонований підхід було оцінено на наборі реальних даних, і результати показали його ефективність. Загалом результати свідчать про те, що використання байєсівських мереж довіри може забезпечити ефективний підхід до підвищення продуктивності пошукової системи та взаємодії з користувачем. Ефективність підходу пояснюється його здатністю моделювати та надавати висновки про невизначеність і залежності між змінними, а також враховувати різні фактори, такі як ключові слова, релевантність сторінки, авторитет домену та відгуки користувачів.*

Висновки. *Запропонований метод може підвищити релевантність пошуку, зменшити розчарування користувачів і підвищити їхнє задоволення. Проте є необхідними подальші дослідження для оптимізації запропонованого підходу та вивчення його застосовності в різних контекстах. Загалом дослідження свідчить про те, що Bayesian Belief Networks може стати цінним інструментом для розроблення більш ефективних і зручних пошукових систем. Крім того, використання Sphinx як базової пошукової системи є перспективним у тому, що запропонований підхід може бути інтегровано у практичні пошукові системи. Тим не менш, необхідно проведення подальших досліджень для оптимізації підходу та оцінювання його застосовності в різних контекстах.*

Ключові слова: *пошукова система, ранжування, байєсівські мережі довіри, імовірнісна модель, пошук інформації.*