

УДК 004.378

О.О. МОСКАЛЕНКО, Т.А. ГРИГОРОВА

Кременчуцький національний університет імені Михайла Остроградського

АЛГОРИТМИ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПОШУКУ ІНФОРМАЦІЇ В СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ

Належний пошук відіграє велику роль в умовах інформаційної насиченості. З кожним днем в пошукових системах індексується все більше інформації, а тому і зростає кількість інформаційного сміття. Для пошуку інформації в мережі інтернет все більш актуальним являється використання алгоритмів штучного інтелекту для видачі більш точного результату та відсіювання зайвої інформації для пошукового запиту. В роботі було досліджено методи і алгоритми штучного інтелекту, які використовуються для пошуку інформації. Для пошуку наукової і навчальної інформації в системі дистанційного навчання було обрано алгоритми колобративної фільтрації, класифікації та задачі пошуку асоціативних правил. Алгоритми було модифіковано і використано для пошуку матеріалу, враховуючи інтереси користувача. Було розширено базу даних дистанційної системи, створено таблиці: історія пошукових запитів, унікальних пошукових запитів (без повторень), ключових слів, зв'язок ключових слів з унікальним пошуковим запитом, зв'язок користувача з ключовими словами. Таблиці дали можливість організувати інтелектуальний пошук за заданим критерієм. В статті наведено алгоритм для зберігання ключових слів та зв'язку ключових слів з пошуковим запитом та з користувачем. Завдяки цьому алгоритму, якщо виконується пошук по заданому критерію, при повторному пошуку з'являється список ключових слів, схожих за змістом на заданий критерій, які були в переглянутих матеріалах. Цей список формується з ключових слів до статті, яка зацікавила користувача, таким чином можна виявити пріоритети, які є у користувача при пошуку наукової і навчальної інформації. При необхідності можна скористатися ключовим словом для пошуку. Користувач може зберігати посилання на той чи інший ресурс, якщо вони були корисними. Ці посилання пропонуються іншому користувачу, який шукає таку саму інформацію. За рахунок використання алгоритму штучного інтелекту при розширенні пошукового модулі, пошукова система значно розширила можливості пошуку. Пошукова система враховує інтереси кожного користувача і є частиною системи дистанційного навчання. Користувачі отримали можливість зручного і швидкого пошуку наукової і навчальної інформації.

Ключові слова: штучний інтелект, машинне навчання, інтелектуальний пошук, класифікація, кластеризація, колаборативна фільтрація.

А.А. МОСКАЛЕНКО, Т.А. ГРИГОРОВА

Кременчугский национальный университет имени Михаила Остроградского

АЛГОРИТМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПОИСКА ИНФОРМАЦИИ В СИСТЕМАХ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Надлежащий поиск играет важную роль в условиях информационной насыщенности. С каждым днем в поисковых системах индексируется все больше информации, а потому и растет количество информационного мусора. Для поиска информации в сети интернет все более актуальным является использование алгоритмов искусственного интеллекта для выдачи более точного результата и отсеивания лишней информации для поискового запроса. В работе были исследованы методы и алгоритмы искусственного интеллекта, которые используются для поиска информации. Для поиска научной и учебной информации в системе дистанционного

обучения были выбраны алгоритмы колаборативной фильтрации, классификации и задачи поиска ассоциативных правил. Алгоритмы были модифицированы и использованы для поиска материала, учитывая интересы пользователя. Расширена база данных дистанционной системы, созданы таблицы: история поисковых запросов, уникальных поисковых запросов (без повторений), ключевых слов, связь ключевых слов с уникальным поисковым запросом, связь пользователя с ключевыми словами. Таблицы позволили организовать интеллектуальный поиск по заданному критерию. В статье приведен алгоритм для хранения ключевых слов и связи ключевых слов с поисковым запросом и с пользователем. При помощи этого алгоритма, если выполняется поиск по заданному критерию, при повторном поиске появляется список ключевых слов, похожих по содержанию на заданный критерий, которые были в просмотренных материалах. Этот список формируется из ключевых слов к статье, которая заинтересовала пользователя, таким образом можно выявить приоритеты, которые имеются у пользователя при поиске научной и учебной информации. При необходимости можно воспользоваться ключевым словом для поиска. Пользователь может сохранять ссылки на тот или иной ресурс, если они были полезными. Эти ссылки предлагаются другому пользователю, который ищет такую же информацию. За счет использования алгоритма искусственного интеллекта при расширении поискового модуля, поисковая система значительно расширила возможности поиска. Поисковая система учитывает интересы каждого пользователя и является частью системы дистанционного обучения. Пользователи получили возможность удобного и быстрого поиска научной и учебной информации.

Ключевые слова: Искусственный интеллект, машинное обучение, интеллектуальный поиск, классификация, кластеризация, колаборативная фильтрация.

O.O. MOSKALENKO, T.A. HRYHOROVA

Kremenchuk Mykhailo Ostohradskyi National University, Kremenchuk, Ukraine

ARTIFICIAL INTELLIGENCE ALGORITHMS FOR SEARCHING INFORMATION IN E-LEARNING SYSTEMS

Proper search plays an important role in the context of information saturation. Every day more and more information is indexed in search engines, and therefore the amount of information waste is increasing. To search for information on the Internet, it is increasingly important to use artificial intelligence algorithms to produce more accurate results and filter out unnecessary information for a search query. The methods and algorithms of artificial intelligence that are used to search for information were investigated during work. The co-filtering algorithm, classification algorithm, and associative rule search algorithm were chosen to search for scientific and educational information in the e-learning system. The algorithms have been modified and used to search for material based on user interests. The remote database was expanded, tables were created: search history, unique searches (no repetitions), keywords, unique keyword searches, user relationship with keywords. The tables gave the opportunity to organize intelligent search by a given criterion. The article provides an algorithm for storing keywords and linking keywords to a search query and to a user. Using this algorithm, if a search is performed according to a specified criterion, a list of keywords appears when searching again. These words are similar in content to a given criterion; they were used in the viewed materials. This list is formed of keywords for the article that interested the user. Thus, it is possible to identify the priorities that the user has when searching for scientific and educational information. You can use the search keyword as needed. The user can save links to a particular resource if they were useful. These links are

offered to another user who is looking for the same information. Through the use of artificial intelligence algorithm in the search module, the search engine has greatly expanded the search capabilities. The search system takes into account the interests of each user and is part of a e-learning system. Users have the opportunity to conveniently and quickly search for scientific and educational information.

Keywords: artificial Intelligence, Machine Learning, Intelligent Search, Classification, Clustering, Co-Filtering

Постановка проблеми

В сучасному суспільстві системи дистанційного навчання стають буденністю, майже всі вищі заклади освіти мають свої системи дистанційного навчання. Пошук навчальної і наукової інформації у цих системах має бути зручним і швидким і задовольняти потребам користувача. Це можливо забезпечити засобами штучного інтелекту для пошукових систем.

Аналіз останніх досліджень і публікацій

Штучним інтелектом можна назвати сукупність технологій, алгоритмів та методів що імітують інтелект людини в деякій моделі реальності. Основними методами штучного інтелекту є наука про дані (Data Science) та машинне навчання (Machine Learning). Зокрема, штучний інтелект використовується для задач, які залежать від багатьох факторів, які потребують складних та важких для розуміння рішень, що складно алгоритмізуються вручну [1].

Машинне навчання – це область штучного інтелекту, в якій застосовуються алгоритми для отримання даних та навчання на їх основі, тобто алгоритм навчається на власному досвіді. Метою машинного навчання є автоматизоване розв'язання різних аналітичних задач. Машинне навчання може бути без вчителя, навчання за допомогою вчителя та комбіноване навчання, в залежності від поставленої задачі.

Можна виділити такі основні типи застосовуваних алгоритмів для машинного навчання: класичне навчання, нейронні мережі та глибоке навчання. Для пошукових систем найчастіше застосовується класичне навчання. Прикладом такого застосування є відображення релевантної реклами в залежності від історії пошукових запитів користувача [2]. До алгоритмів класичного навчання, які використовуються для задач роботи з даними відносяться алгоритм регресії, класифікації, кластеризації, пошуку асоціативних правил.

Задача класифікації в машинному навчанні застосовується тоді, коли є множина об'єктів, які відносяться до деякого класу і необхідно визначити класову належність для невідомих об'єктів.

Задача кластеризації використовується коли є необхідність використовувати машинне навчання без вчителя. В задачі вибірки даних (об'єктів) розбиваються на групи (кластери) в яких дані складаються зі схожих об'єктів. Об'єкти різних кластерів значно відрізняються між собою. В задачі кластеризації необхідно використовувати всі дані та передбачити відповідність об'єктів до кластерів [3].

В задачі пошуку асоціативних правил необхідно знаходити закономірності між подіями, в яких є зацікавленість, у великій базі даних. Це теж відноситься до задач машинного навчання без вчителя. В задачі асоціативних правил сенсом є визначення наборів об'єктів, що часто зустрічаються у великій множині таких наборів [4]. Для розв'язання задач пошуку асоціативних правил використовуються такі основні алгоритми: алгоритм Apriori та алгоритм FP-Growth (Frequent Pattern) [5].

До специфічних задач відноситься задача колоборативної фільтрації (Collaborative Filtering). В ній поєднуються елементи класифікації, кластеризації та

заповнення пропущених даних. Ця задача застосовується для прогнозування переваг користувачів на основі їх попередніх переваг та переваг схожих користувачів.

Мета дослідження

Метою роботи є розробка для системи дистанційного навчання інтелектуального пошуку наукового матеріалу в залежності від інтересів користувача. Алгоритм пошуку повинен враховувати актуальні інтереси користувача та в залежності від них формувати результат пошуку. Пошук наукового матеріалу повинен здійснюватися в наукових журналах.

На кафедрі інформатики та вищої математики Кременчуцького національного університету використовується система дистанційної освіти на базі системи з відкритим кодом E-Front. Для даної системи було розроблено модуль повнотекстового пошуку на базі технології Sphinx, а також підключені API для пошуку в Google, у системі IEEE та Scopus [6]. Цей модуль можливо використовувати і для інших систем дистанційного навчання. В модулі є можливості введення додаткових параметрів, що обмежують або розширюють пошук [7], а також можливості автоматичного перекладу та пошуку на англійській мові. Недоліком такого пошуку є те, що система не враховує інтереси користувачів (пошук за зацікавленістю).

Алгоритм пошуку повинен враховувати належність користувача до кафедри, факультету та формувати групу інтересів в залежності від попередніх запитів пошуку та історії відкритих вкладок. Кожна стаття містить ключові слова, тому група інтересів повинна формуватися з ключових слів статей, які зацікавили користувача. Для перевірки актуальності груп інтересів користувача алгоритм повинен перевіряти поточні пошукові запити на відповідність до минулих груп інтересів, якщо інтереси користувача змінились, необхідно сформувати нову групу інтересів для користувача. Також необхідно формувати свою базу знань (посилання на статтю та короткий опис) в залежності від того, яка інформація з результату пошуку була корисною для користувача, та в подальшому виводити ці результати іншим користувачам по даному пошуковому запиту. Для вирішення цих задач необхідно використовувати алгоритми та методи штучного інтелекту. Всі ці задачі є специфічними, і відносяться відразу до декількох типів задач машинного навчання, вони включають елементи задачі колоборативної фільтрації, класифікації та задачі пошуку асоціативних правил. Для пошукової системи задачею пошуку асоціативних правил є відображення схожих пошукових запитів. Формування груп інтересів користувача відноситься до задачі класифікації. Наприклад, якщо об'єктом є наукова стаття, тоді в задачі класифікації необхідно визначити, до яких наук дана стаття відноситься. Задачею колоборативної фільтрації є формування корисних посилань для подальшої рекомендації цих посилань іншим користувачам.

Для реалізації поставлених задач необхідно розширити існуючу базу даних дистанційної системи. Необхідно створити таблицю, яка буде містити в собі історію пошукових запитів, таблицю для унікальних пошукових запитів (без повторень), таблицю для зберігання ключових слів, таблицю для зв'язку ключових слів з унікальним пошуковим запитом, таблицю для зв'язку користувача з ключовими словами. Також необхідно створити таблицю для зберігання корисних посилань користувача. Таблиця унікальних пошукових запитів повинна мати зв'язок «один до багатьох» з таблицею для зберігання корисних посилань.

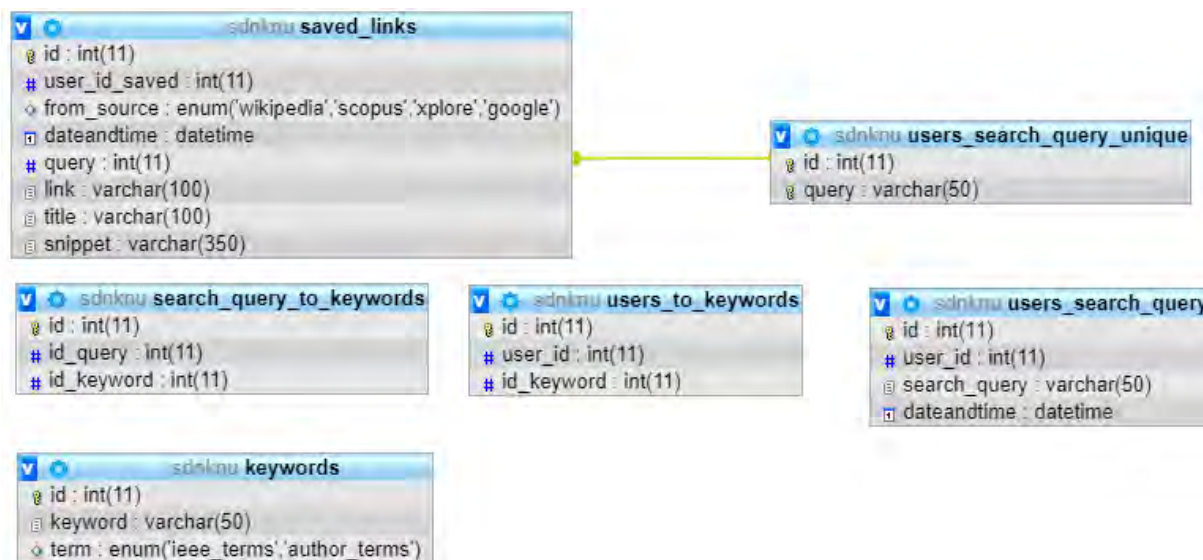


Рис. 1. Модель даних для інтелектуального пошуку.

В системі дистанційного навчання підключені API IEEE та Scopus для пошуку наукового матеріалу.

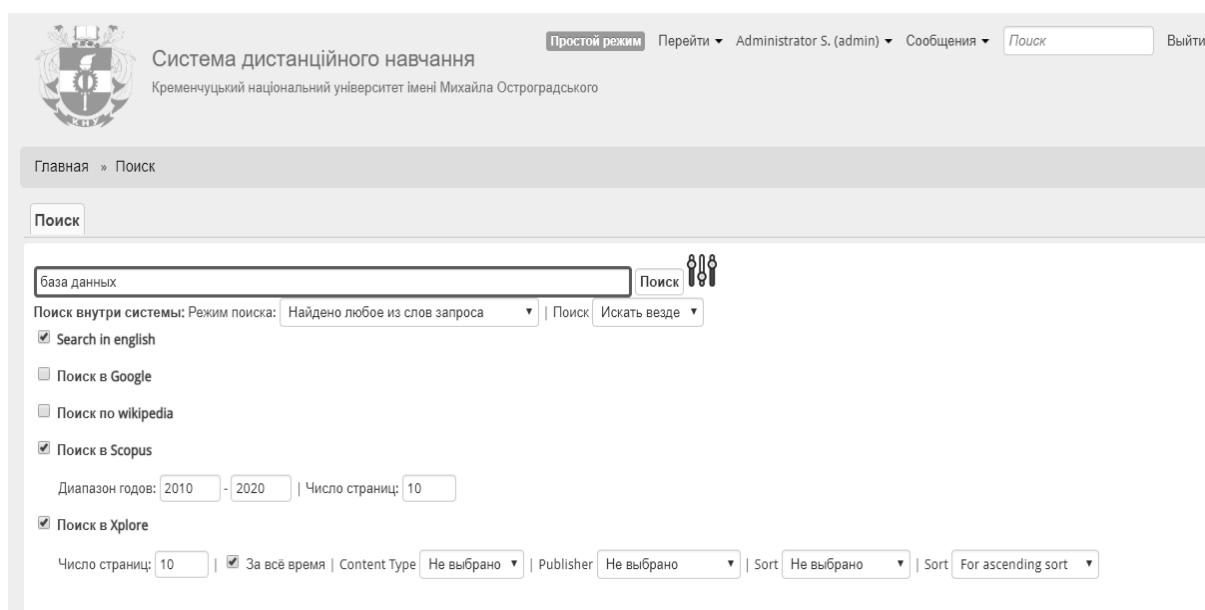


Рис. 2. Пошук в наукових журналах IEEE та Scopus.

Кожна стаття має ключові слова (рис. 3) з яких можна формувати групи інтересів користувача, в залежності від відкритих та збережених сторінок.

Для зберігання ключових слів та зв'язку ключових слів з пошуковим запитом та з користувачем будемо використовувати наступний алгоритм (рис. 4).

Із результатів пошукового запиту в IEEE та Scopus користувач переходить по зацікавленому посиланні. Якщо ми отримали ключові слова статті – створюємо масив ключових слів, зберігаємо ключові слова в таблицю *keywords*, пов'язуємо ключове слово з пошуковим запитом в таблиці *search_query_to_keywords* та пов'язуємо ключове слово з користувачем в таблиці *users_to_keywords*.

<https://ieeexplore.ieee.org/document/8510719/keywords#keywords>

Keywords

IEEE Keywords

Structured Query Language, Data models, Relational databases, Security, Europe

INSPEC: Controlled Indexing

ISO standards, learning (artificial intelligence), legislation, organisational aspects, risk management

INSPEC: Non-Controlled Indexing

organisational aspects, learning (artificial intelligence), ISO standards, risk management, legislation, Internet, educational institutions, content management, computer based training, game theory

Author Keywords

NoSQL, SQL, MongoDB, MySQL, evaluation, analysis, comparison

Рис. 3. Приклад ключових слів статті в IEEE.

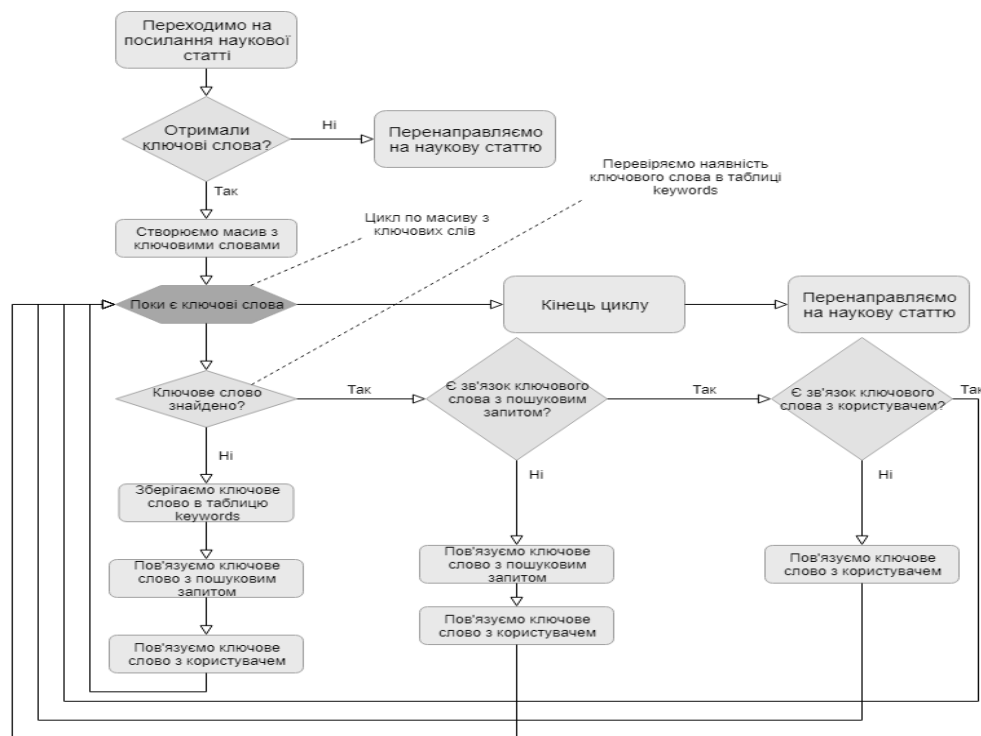


Рис. 4. Алгоритм зберігання та зв'язку ключових слів.

Для отримання ключових слів пошукового запиту для користувача будемо використовувати SQL запит:

```

SELECT keywords.keyword FROM keywords INNER JOIN
users_search_query_unique ON users_search_query_unique.id = $search_query_unique_id
INNER JOIN search_query_to_keywords ON keywords.id =
search_query_to_keywords.id_keyword AND users_search_query_unique.id =
search_query_to_keywords.id_query INNER JOIN users_to_keywords ON

```

users_to_keywords.id keyword = keywords.id AND users_to_keywords.user_id = \$user_id
GROUP BY keywords.keyword

Де, «\$search_query_unique_id» id поточного пошукового запиту, а «\$user_id» id поточного користувача.

Для прикладу, виконаємо пошук по пошуковому запиту «database» в науковому журналі. Після відкриття зацікавлених нас сторінок із результату пошуку виконаємо такий самий пошуковий запит. Після цього, окрім результату пошуку, пошукова система буде пропонувати виконати пошук по ключовим словам для пошукового запиту «database» (рис. 5).



Прикладне програмування в комп'ютерних мережах

- Урок : Особливості роботи програми у локальній та глобальній мережах
Матеріал урока:
 - PHP
... з базами даних (Open **Database** Connectivity Standard — ODBC ...
- Урок : Різноманітні підходи до мережевого програмування
Матеріал урока:
 - Web-додатки з модулями розширення клієнтської частини
Web-додатки з модулями розширення клієнтської частини Для ...
 - Дворівневі Web-додатки
Дворівневі Web-додатки При дворівневій архітектурі ...



Сучасні технології програмування

- Урок : Zend Framework
Матеріал урока:
 - Фільтрація даних у полях форми відповідають за введення пароля
... , \$password); mysql_select_db(\$ database); ?>

Ключевые слова по вашему запросу:

- [Bayes methods](#)
- [Computer architecture](#)
- [Data mining](#)
- [Data warehouses](#)
- [Database systems](#)
- [Detection algorithms](#)
- [Feature extraction](#)
- [Probability distribution](#)
- [Relational databases](#)
- [Security](#)
- [Task analysis](#)

Рис. 5. Відображення ключових слів в результатах пошуку.

Для зберігання корисних посилань, після перегляду посилання пошукова система пропонує зберегти посилання, якщо воно було корисним. Для прикладу виконаємо пошук по запиту «mysql» та збережемо корисні для нас посилання. Після цього зйдемо в систему з іншого акаунту та виконаємо такий самий пошуковий запит (рис. 6).

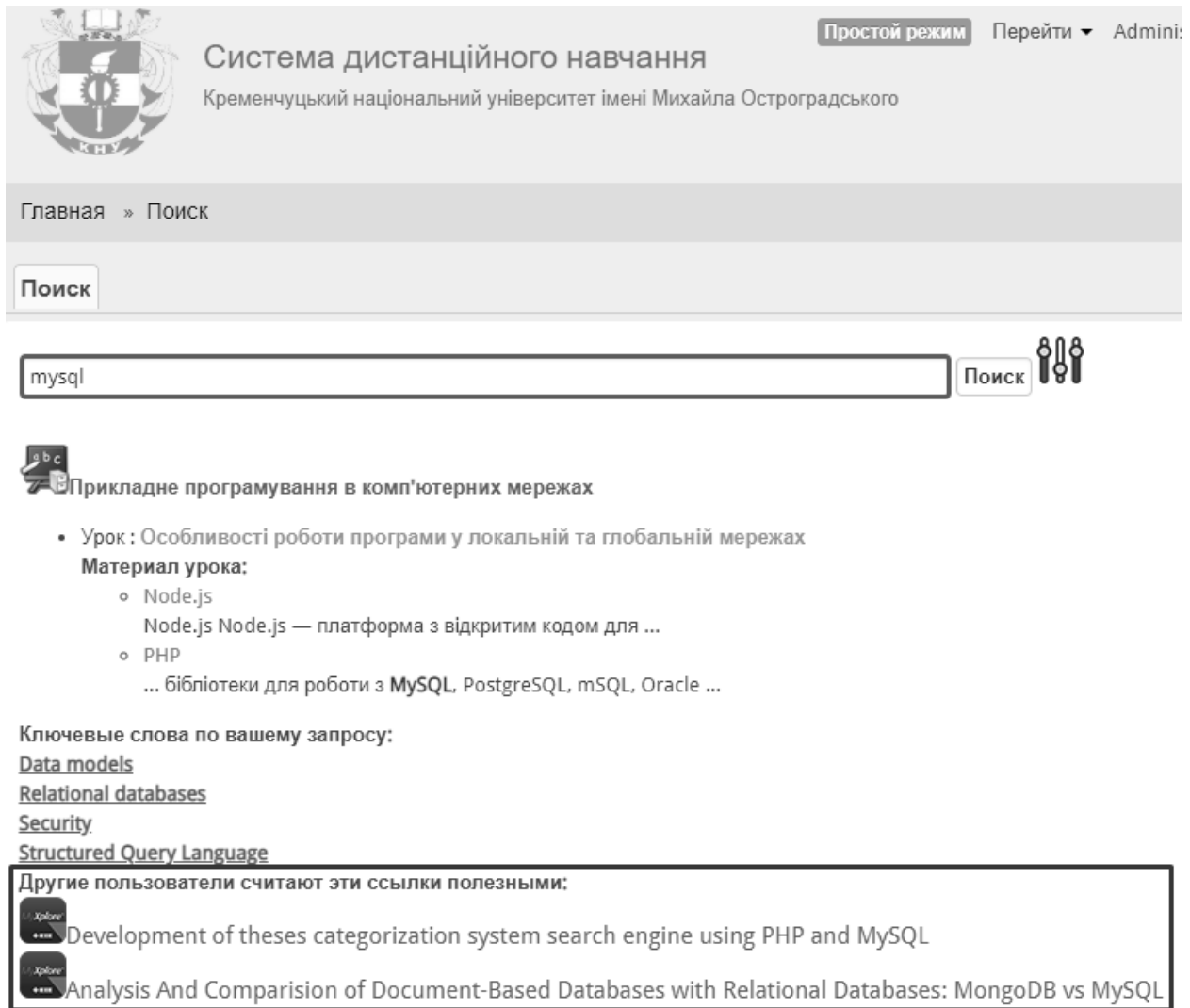


Рис. 6. Відображення корисних посилань в результатах пошуку.

Для відображення схожих пошукових запитів будемо використовувати SQL запит:

```
SELECT users_search_query.search_query FROM users_search_query WHERE
users_search_query.search_query LIKE '%search query%' AND
users_search_query.search_query <> 'search query' GROUP BY
users_search_query.search_query
```

Перед пошуком будемо перевіряти історію самих повторюваних пошукових запитів користувача:

```
SELECT `search_query` COUNT(*) AS `count` FROM `users_search_query` WHERE
`search_query` IS NOT NULL AND user_id = $user_id GROUP BY `search_query` ORDER
BY `count` DESC LIMIT 2
```

Якщо по даним пошуковим запитах є ключові слова або корисні посилання, тоді будемо відображати їх перед пошуком.

Висновки

В даній роботі було проаналізовано засоби та методи штучного інтелекту, які використовують для пошуку. Були розглянуті алгоритми машинного навчання і їх використання. Для того щоб розширити пошуковий модуль таким чином, щоб пошук враховував актуальні інтереси користувача та в залежності від них формував результат пошуку було модифіковано алгоритми колоборативної фільтрації та класифікації, а також використано алгоритм пошуку асоціативних правил. В результаті пошукова система значно розширила можливості пошуку наукового і навчального матеріалу.

Список використаної літератури

1. Naeem M. Asif, Noreen Asif. A Web Smart Space Framework for Intelligent Search Engines. *International Journal of Emerging Sciences*. 2011. Vol. 1. Issue 1. P. 1–10.
2. Micarelli A., Gasparetti F., Biancalana C. Intelligent Search on the Internet. In: Stock O., Schaerf M. (eds) *Reasoning, Action and Interaction in AI Theories and Systems. Lecture Notes in Computer Science*. Vol. 4155. Springer, Berlin, Heidelberg, 2006. P. 247–264.
3. Щербakov Д. Как искусственный интеллект повлиял на поисковые системы. URL: <https://www.uplab.ru/blog/artificial-intelligence/>
4. Джуматов Е.К., Вишня А.С., Филиппов С.А. Применение алгоритмов ассоциативных правил для выявления рекомендуемых к продаже товаров строительной отрасли. *Теория. Практика. Инновации*. 2018. № 1. С. 1–15. URL: <http://www.tpinauka.ru/2018/02/Jumatov.pdf>
5. Artificial Intelligence Applications in Search Engines. URL: <https://medium.com/aimarketingassociation/artificial-intelligence-applications-in-search-engines-437c57f8b265>
6. Hryhorova T., Moskalenko O.O. Use of Information Technologies to Improve Access to Information in E-Learning Systems. *Recent Developments in Data Science and Intelligent Analysis of Information*. Springer Nature Switzerland AG, 2019. P. 206–215. Retrieved from https://doi.org/10.1007/978-3-319-97885-7_21.
7. Москаленко О.О., Григорова Т.А. Особливості пошуку навчальної та наукової інформації в системах електронного навчання. *Системи та засоби штучного інтелекту: матеріали міжнародної наукової молодіжної школи*, (м. Київ, 18 жовтня 2018). Київ, 2018. С. 80–85.

References

1. Naeem, M. Asif, & Noreen, Asif. (2011). A Web Smart Space Framework for Intelligent Search Engines. *International Journal of Emerging Sciences*. 1, 1, 1–10.
2. Micarelli, A., Gasparetti, F., & Biancalana, C. (2006). Intelligent Search on the Internet. In: Stock O., Schaerf M. (eds) *Reasoning, Action and Interaction in AI Theories and Systems. Lecture Notes in Computer Science*. Vol. 4155. Springer, Berlin, Heidelberg, pp. 247–264.
3. Sherbakov, D. Kak iskusstvenny intellekt povliyal na poiskovye sistemy. Retrieved from <https://www.uplab.ru/blog/artificial-intelligence/>
4. Djumatov, E. K., Vishnya, A. S., & Filippov, S. A. (2018). Primenenie algoritmov asociativnyh pravil dlya vyyavleniya rekomenduemyh k prodage tovarov stroitelnoyi otrasli. *Teoriya. Praktika. Inovacii*. 1, 1–15. Retrieved from <http://www.tpinauka.ru/2018/02/Jumatov.pdf>
5. Artificial Intelligence applications in search engines. Retrieved from <https://medium.com/aimarketingassociation/artificial-intelligence-applications-in-search-engines-437c57f8b265>

6. Hryhorova, T., & Moskalenko, O. O. (2019). Use of Information Technologies to Improve Access to Information in E-Learning Systems. *Recent Developments in Data Science and Intelligent Analysis of Information*. Springer Nature Switzerland AG, pp. 206–215. Retrieved from https://doi.org/10.1007/978-3-319-97885-7_21.
7. Moskalenko, O. O., & Hryhorova T. A. Osoblyvosti poshuku navchalnoi ta naukovoï informacii v systemah elektronnoho navchannya. Proceedings of the *Systemy ta zasoby shtuchnogo intelektu*: Materialy mignarodnoi naukovoï konferencii molodignoi shkoly. (Ukraine, Kyiv, October 18, 2018). Kyiv, pp. 80 – 85.

Москаленко Олександр Олександрович – аспірант кафедри інформатики і вищої математики Кременчуцького національного університету імені Михайла Остроградського, e-mail: alexsashamosk@gmail.com.

Григорова Тетяна Альбертівна – к.т.н., доцент, доцент кафедри інформатики і вищої Кременчуцького національного університету імені Михайла Остроградського, e-mail: grital0403@gmail.com, ORCID: 0000-0002-4371-8624.