

Тушницький Р. Б.,  
Макар В. М.

# РОЗРОБКА ПРОГРАМНОЇ СИСТЕМИ ДЛЯ АВТОМАТИЗОВАНОГО ФОРМУВАННЯ ПОШУКОВИХ ЗАПИТІВ СКЛАДНО-СТРУКТУРОВАНИХ ДАНИХ

В роботі запропоновано алгоритм формування запитів пошуку складно-структурованих даних на основі операторів та ключових слів. На основі цього алгоритму розроблено програмне забезпечення, яке дає змогу автоматизувати процес формування пошукових запитів складно-структурованих даних. Даний підхід покладено в основу програмної підсистеми пошуку бібліографічних описів публікацій інформаційно-аналітичної системи «ScienceLP».

**Ключові слова:** алгоритм пошуку, бібліографічний опис, ключові слова, оператори пошуку.

## 1. Вступ

Використання пошукових машин дає змогу користувачам швидко знаходити потрібну інформацію. Усі запити для пошукових систем умовно можна поділити на три типи:

- *інформаційні запити* — користувач шукає інформацію, не знаючи, на якому саме веб-ресурсі вона буде перебувати;
- *навігаційні запити* — користувач попередньо шукає веб-ресурс, на якому має бути інформація, що його цікавить;
- *транзакційні запити* — користувач у формулюванні запиту виражає готовність зробити якусь дію, наприклад «купити телевізор».

Різні пошукові системи використовують різні алгоритми й механізми пошуку, по-різному оцінюють ступінь відповідності знайдених посилань запиту користувача. Деяким з існуючих пошукових систем можна навіть ставити запити у вигляді звичайного питання [1–3]. Сучасні пошукові системи здатні самостійно визначити, які слова й словосполучення є ключовими.

Пошук із використанням ключових слів в базі даних, яка не вимагає знання її схеми є складним і актуальним завданням. Маючи набір ключових слів, результуючий набір даних може формуватися об'єднанням декількох таблиць. Використання ключових слів разом з операторами дає змогу користувачу максимально точно сформулювати запит. Актуальним є і розроблення візуальних засобів формування запитів пошуку, які відкривають можливість проведення ефективного пошуку незалежно від наявного досвіду користування програмним забезпеченням.

## 2. Аналіз літературних даних та постановка проблеми

Для більш швидкого й успішного пошуку в пошукових системах крім ключових слів [4, 5] використовують різні логічні оператори [6, 7]. Завдяки цьому, можна конструювати запит так, що будуть знайдені не тільки ресурси на тему, яка цікавить користувача, а й конкретні сторінки й навіть окремі документи. Для автоматиза-

ції формування запиту в роботі [8] розроблено власну бібліотеку із проміжною мовою представлення запиту.

Правила складання складних запитів в одній пошуковій системі, можуть відрізнятися від таких в іншій, але в кожному разі будуть використовуватися основні логічні оператори й синтаксичні вирази [9] (табл. 1).

Таблиця 1

Логічні оператори й синтаксичні вирази

Назва	Функція	Приклад використання
<i>I (AND)</i>	Поєднує два або більше слів так, щоб вони всі були присутні в шуканому документі (часто замість <i>I</i> використовують & або +)	За запитом «Червона І Шапочка» будуть знайдені документи, що містять і те, й інше слово
<i>АБО (OR)</i>	Забезпечує пошук за кожним зі слів групи	За запитом «освіта АБО навчання» будуть знайдені документи, що містять слово освіта або слово навчання
<i>Логічні дужки</i>	Управляє порядком проходження логічних операторів	За запитом «Довженко АБО (Михайло І Васильович)» будуть знайдені документи, що містять слово Довженко або словосполучення Михайло Васильович
<i>НЕ (NOT)</i>	Виключає яке-небудь ключове слово з результатів пошуку	За запитом «Ссавці НЕ хижакі» будуть знайдені документи, що містять слово ссавці і не містять слово хижакі
<i>БЛИЗЬКО (NEAR)</i>	Дає змогу вказати, на якій відстані одне від одного можуть розташовуватися слова в документі	За запитом «музика NEAR скачати безкоштовно» будуть видані документи, у яких ці слова перебувають недалеко одне від одного, тобто між ними не повинно бути більше десяти слів
<i>Ланки</i>	Забезпечують дослівний пошук виразу або словосполучення	За запитом «відношення долара до євро» будуть видані документи, у яких дослівно міститься ця фраза

Для спрощення завдання формування складних запитів використовують спеціальні форми, за допомогою яких виконується розширений пошук.

Існуючі бібліотеки та підходи не враховують спеціфіки застосування пошуку бібліографічних описів публікацій інформаційно-аналітичної системи «ScienceLP»,

а також не надають гнучких засобів для автоматизації процесу формування пошукових запитів складно-структурованих даних. Тому виникла необхідність розробити власний алгоритм формування запитів пошуку складно-структурованих даних на основі операторів та ключових слів.

### 3. Об'єкт, ціль та задачі дослідження

Об'єктом дослідження є процес автоматизованого формування пошукових запитів складно-структурованих даних на основі операторів та ключових слів.

Метою досліджень є розробка методу та його програмна реалізація для задачі пошуку складно-структурних даних із використанням ключових слів та операторів.

Для досягнення поставленої мети вирішувалися такі задачі:

- визначити необхідні оператори та ключові слова;
- автоматично згенерувати на основі аналізу операторів та ключів пошуковий запит для бази даних;
- розробити програмне забезпечення на основі запропонованого підходу.

### 4. Матеріали та методи дослідження

В якості складно-структурованих даних обрано бібліографічний опис публікацій. Експериментальні дослідження ефективності розробленого підходу проведено для задачі пошуку бібліографічних описів публікацій у системі звітності про наукову-дослідну діяльність Національного університету «Львівська політехніка» — інформаційно-аналітичній системі «ScienceLP» [10].

Система «ScienceLP» містить інформацію про бібліографічні описи таких видів публікацій: монографії, підручники, навчальні посібники, публікації у періодичних наукових виданнях, публікації у неперіодичних наукових виданнях, участь у виставках, доповіді на наукових семінарах. Крім основних даних, що містяться в бібліографічному описі, кожна публікація характеризується роком наукового звіту та унікальним ідентифікатором в системі. Крім цього, наукові видання містять інформацію про приналежність до фахового видання, наукометричних баз даних тощо.

Практичними застосуваннями методу пошуку бібліографічних описів публікацій в системі звітності підрозділів є наступні:

- 1) простий пошук публікацій в базі даних за ключовою фразою;
  - 2) розширений пошук публікацій в базі даних за ключовою фразою, ключовими словами та операторами.
- Прикладами запитів можуть бути такі:
- *Знайти всі публікації автора Довженко за 2015 рік звіту, які опубліковані у наукових виданнях, що містяться в наукометричних базах даних.*
  - *Знайти всі публікації, назва яких містить фразу «математична модель», і опубліковані у фахових виданнях за 2015 рік.*

### 5. Розробка операторів та ключів пошуку

Введемо оператори, які складуть мінімальну множину потрібних операторів, які можуть бути задіяні у пошуковому запиті:

«+» — оператор «логічне і»;

«|» — оператор «логічне або»;

«[« — відкрита обгортка для ключа;

«]» — закрита обгортка для ключа;

«:» — сепаратор для ключа.

В табл. 2 подано перелік розроблених ключів та їхні властивості.

Таблиця 2

Перелік ключів та їх властивості

Ключ	Значення	Тип даних	Може містити оператори в значенні?
author_surname	прізвище автора	стрічка	так
publishing_house	назва видавництва	стрічка	так
year	рік публікації	число	ні
page_count_from	кількість сторінок від	число	ні
page_count_to	кількість сторінок до	число	ні
publication_name	ім'я публікації	стрічка	так
report_year	рік звіту	число	ні
journal_name	ім'я журналу (вісника або збірки тез)	стрічка	так
journal_VAK	журнал є фаховим	логічне значення	ні
journal_NMBD	журнал входить до переліку наукометричних баз даних	логічне значення	ні
description	частина тексту в анотації монографії	стрічка	так
conference_name	частина тексту в назві конференції	стрічка	так

На рис. 1 наведено послідовність дій пошуку із використанням операторів та ключових слів.

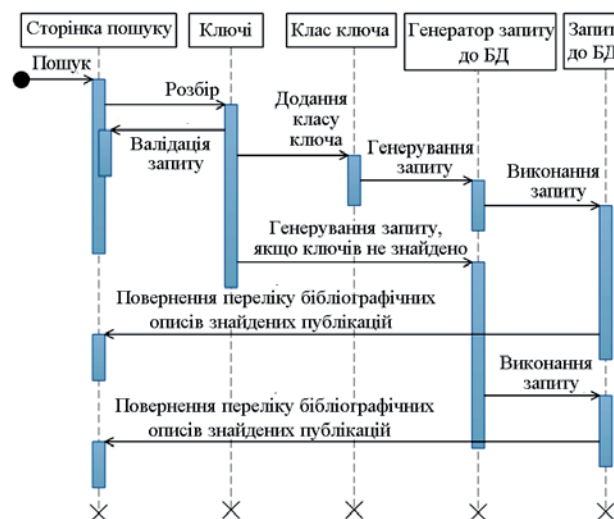


Рис. 1. UML-діаграма послідовностей дій під час пошуку

Опишемо алгоритм пошуку.

1. Користувач відправляє запит на сервер застосування, що може містити в собі оператори та ключові слова.
2. Сервер застосування обробляє вхідний запит, здійснюючи перевірку коректності запиту та у разі успіху викликає метод розбору пошукового запиту.

3. Пошуковий запит розбивається на окремі складові, серед яких здійснюється пошук операторів, ключів та звичайних слів.

4. На основі створеної колекції за типізованими структурними елементами генеруються запит до бази даних, який запам'ятовує унікальні ідентифікатори описів публікацій.

5. Формується новий запит на основі унікальних ідентифікаторів, який повертає колекцію записів з обчисленою кількістю співпадінь по словам та ключам.

6. Результуюча колекція бібліографічних описів сортується за кількістю співпадінь.

7. Отримана вибірка повертається користувачу.

### 6. Результати експериментальних досліджень пошуку із використанням операторів та ключових слів

На основі розробленого тематичного забезпечення створено програмний засіб для дослідження процесу пошуку інформації.

На рис. 2 подано діаграму компонент, що демонструє всі компоненти, які задіяні для здійснення пошуку із використанням ключових слів та операторів.

Програмне застосування розроблено із використанням мови програмування C#, скриптові мови javascript, технології AJAX для асинхронного обміну даними між клієнтом та сервером, фреймворка ASP.NET MVC 5, системи управління базами даних Microsoft SQL Server 2014.

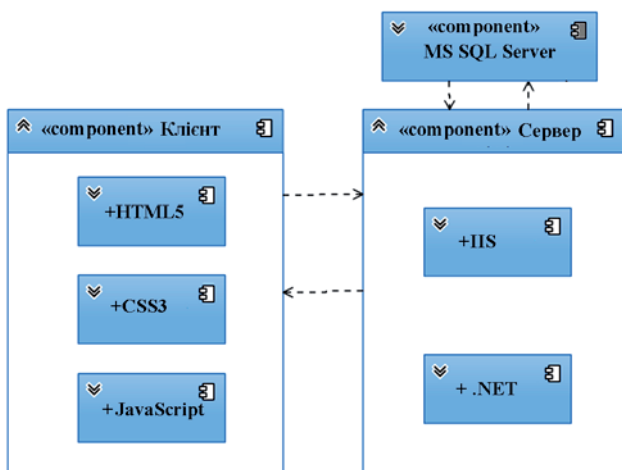


Рис. 2. UML-діаграма компонент розробленого програмного застосування

Розроблено такі 3 режими пошуку:

«Звичайний режим» — дає змогу користувачу заповнити поле пошуку і не вникати в ключові поля та принцип запитів, а пошуковий двигун сам розпізнає текстові поля та числові, і виконає запит за відповідними полями.

«Текстовий професійний» режим — дає змогу користувачу, який зможе вникнути в принцип пошуку, виконати пошук по ключам, і побачити більш точний результат.

«Візуальний професійний режим» — дає змогу користувачу виконати пошук по ключам, не вникаючи в них та в принцип запиту і пошуку. Тобто виконати пошук так, як в текстовому режимі, тільки без вводу запиту в текстове поле, а за допомогою форми, в якій наявні ключові поля та метод пошуку. Таким чином, користувач виконає пошук в текстовому професійному режимі завдяки згенерованому запиту з форми професійного пошуку.

На рис. 3 наведено вигляд графічного інтерфейсу користувача для режиму пошуку «візуальний професійний», який дає змогу користувачу вказати ключову фразу для пошуку та обрати відповідні ключі і логічні оператори.

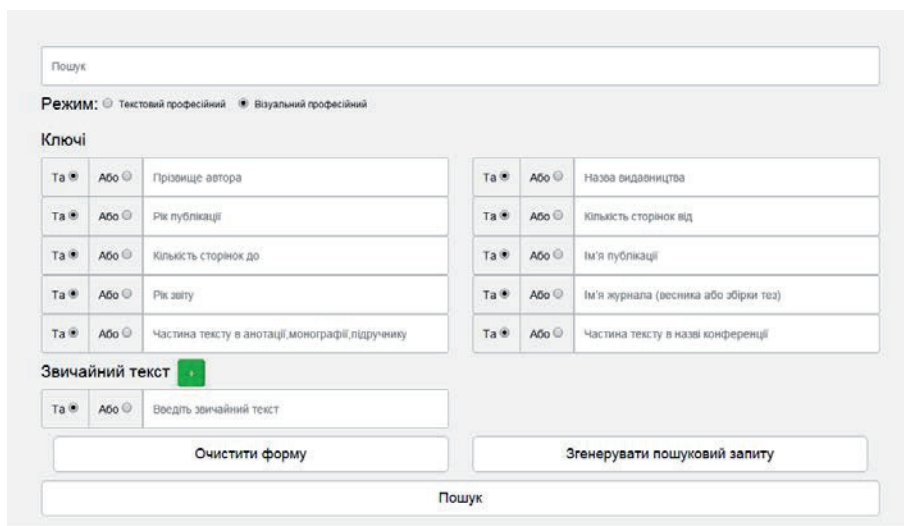


Рис. 3. Інтерфейс користувача для режиму пошуку «візуальний професійний»

### 7. Обговорення результатів експериментальних досліджень

В табл. 3 подано приклади пошукових запитів та отримувані результати.

Таблиця 3

Приклади пошукових запитів

Тип запиту	Запит	Результат
Звичайний	2014	Всі публікації, де зустрічається 2014
Звичайний	Довженко 2014	Всі публікації, де зустрічається «Довженко 2014»
З оператором	Довженко + 2014	Всі публікації, де в зустрічається «Довженко» і «2014»
З ключем	year:[2015]	Всі публікації, де рік публікації 2015
З ключем та з оператором	author_surname:[ Довженко ]   2015	Всі публікації, де автор Довженко, або де зустрічається 2015
З ключем та з оператором та з оператором в ключі	author_surname:[ Довженко + Шевченко ]   2015	Всі публікації, де автори Довженко та Шевченко, або де зустрічається 2015
З помилкою	+2015	Інформаційне повідомлення «Першим значенням має бути слово, а не оператор»

Як видно з табл. 3, введені оператори та ключові слова дають змогу користувачу максимально точно задати пошуковий запит. Крім цього, у разі помилкового запиту користувачу відображається інформаційне повідомлення із ймовірним вирішенням помилки.

Реалізований підхід характеризується простотою завдання пошукового запиту кінцевим користувачем системи.

Використання ключів та операторів дає змогу здійснити ранжування результатів пошуку за кількістю співпадінь.

## 8. Висновки

В результаті досліджень розроблено мінімальний набір операторів та групу ключових слів, які призначені для покращення пошукового запиту. Набір операторів і групу ключових слів можна доповнити новими елементами.

На основі аналізу операторів та ключів розроблено алгоритм автоматичного генерування результуючого пошукового запиту для бази даних.

Розроблене на основі запропонованого підходу програмне забезпечення дає змогу здійснити швидкий пошук інформації за заданими словами, ключами та операторами. Для здійснення пошуку розроблено 3 режими — звичайний, текстовий професійний та візуальний професійний, які дають змогу користувачу сформулювати пошуковий запит залежно від наявного досвіду користування системою пошуку.

Подальші дослідження включають в себе розробку нових операторів та ключів, а також методів пришвидшення пошуку в базі даних.

## Література

1. Bosker, B. Siri Rising: The Inside Story Of Siri's Origins (And Why She Could Overshadow The iPhone) [Electronic resource] / B. Bosker // The Huffington Post (online). — 24.01.2013. — Available at: \www/URL: [http://www.huffingtonpost.com/2013/01/22/siri-do-engine-apple-iphone\\_n\\_2499165.html](http://www.huffingtonpost.com/2013/01/22/siri-do-engine-apple-iphone_n_2499165.html)
2. Warren, T. Microsoft's Cortana is coming to iOS and Android [Electronic resource] / T. Warren // The Verge. — 13.03.2015. — Available at: \www/URL: <http://www.theverge.com/2015/3/13/8207361/microsoft-cortana-ios-android>
3. Ravichandran, D. Surface Text Patterns for a Question Answering System [Text] / D. Ravichandran, E. Hovy // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002. — P. 41–47.
4. Fakhraee, S. DBSemSxplorer: semantic-based keyword search system over relational databases for knowledge discovery [Text] / S. Fakhraee, F. Fotouhi // Proceedings of the Third International Workshop on Keyword Search on Structured Data — KEYS '12. — Association for Computing Machinery (ACM), 2012. — P. 54–62. doi:10.1145/2254736.2254748
5. Shimizu, T. Kikori-KS: An Effective and Efficient Keyword Search System for Digital Libraries in XML [Text] / T. Shimizu, N. Terada, M. Yoshikawa // Digital Libraries: Achievements, Challenges and Opportunities. — Springer Science + Business Media, 2006. — P. 390–399. doi:10.1007/11931584\_42
6. Mei, Y. ZStream: a cost-based query processor for adaptively detecting composite events [Text] / Y. Mei, S. Madden // Proceedings of the 35th SIGMOD international conference on Management of data — SIGMOD '09. — Association for Computing Machinery (ACM), 2009. — P. 193–206. doi:10.1145/1559845.1559867
7. Oehler, M. A Conjunction for Private Stream Searching [Text] / M. Oehler, D. S. Phatak // 2013 International Conference on Social Computing. — Institute of Electrical & Electronics Engineers (IEEE), 2013. — P. 441–447. doi:10.1109/socialcom.2013.69
8. Gan, K. H. A query transformation framework for automated structured query construction in structured retrieval environment [Text] / K. H. Gan, K. K. Phang // Journal of Information Science. — 2014. — Vol. 40, № 2. — P. 249–263. doi:10.1177/0165551513519240
9. Truncation and Wildcard Operators [Electronic resource] / Earl K. Long Library. The University of New Orleans. — 2008. — Available at: \www/URL: <http://library.uno.edu/helpfiles/handouts/LIB-HANDOUT-keyword-boolean.pdf>
10. Макар, В. Інформаційно-аналітична система для автоматизації підготовки наукових звітів підрозділів Львівської політехніки [Текст] / В. Макар, Р. Тушницький // Матеріали 6-ї науково-практичної конференції «Інноваційні комп'ютерні технології у вищій школі». — Львів, 2014. — С. 177–182.

## РАЗРАБОТКА ПРОГРАММНОЙ СИСТЕМЫ ДЛЯ АВТОМАТИЗИРОВАННОГО ФОРМИРОВАНИЯ ПОИСКОВЫХ ЗАПРОСОВ СЛОЖНО-СТРУКТУРИРОВАННЫХ ДАННЫХ

В работе предложен алгоритм формирования запросов поиска сложно-структурированных данных на основе операторов и ключевых слов. На основе этого алгоритма разработано программное обеспечение, позволяющее автоматизировать процесс формирования поисковых запросов сложно-структурированных данных. Данный подход положен в основу программной подсистемы поиска библиографических описаний публикаций информационно-аналитической системы «ScienceLP».

**Ключевые слова:** алгоритм поиска, библиографическое описание, ключевые слова, операторы поиска.

*Тушницький Руслан Богданович, кандидат технічних наук, доцент, кафедра програмного забезпечення, Національний університет «Львівська політехніка», Україна, e-mail: [ruslan.tushnytskyi@gmail.com](mailto:ruslan.tushnytskyi@gmail.com).*

*Макар Володимир Мирославович, кандидат технічних наук, доцент, кафедра програмного забезпечення, Національний університет «Львівська політехніка», Україна.*

*Тушницький Руслан Богданович, кандидат технических наук, доцент, кафедра программного обеспечения, Национальный университет «Львовская политехника», Украина.*

*Макар Владимир Мирославович, кандидат технических наук, доцент, кафедра программного обеспечения, Национальный университет «Львовская политехника», Украина.*

*Tushnytskyi Ruslan, Lviv Polytechnic National University, Ukraine, e-mail: [ruslan.tushnytskyi@gmail.com](mailto:ruslan.tushnytskyi@gmail.com).*

*Makar Volodymyr, Lviv Polytechnic National University, Ukraine*