

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики

Кафедра програмного забезпечення комп'ютерних систем

«До захисту допущено»

Науковий керівник кафедри

І.А. Дичка

_____ (підпис)

“ _____ ” _____ 2017 р.

Дипломний проект
на здобуття ступеня бакалавра

з напрямку підготовки 6.050103 “Програмна інженерія”

на тему ПРОГРАМНІ ЗАСОБИ ПОВНОТЕКСТОВОГО ПОШУКУ У
НЕРЕЛЯЦІЙНИХ БАЗАХ ДАНИХ

Виконав: студент 4 курсу, групи КП-31

Романченко Вікторія Вікторівна

_____ (підпис)

Керівник доц., доц., к.т.н. Петрашенко А.В.

_____ (підпис)

Консультант з нормоконтролю старший викладач Онай М.В.

_____ (підпис)

Рецензент доц. кафедри СПСКС факультету прикладної
математики, доц., к.т.н. Замятін Д.С.

_____ (підпис)

Засвідчую, що у цьому дипломному
проекті немає запозичень з праць інших
авторів без відповідних посилань.

Студент _____
(підпис)

Київ – 2017

АНОТАЦІЯ

Даний дипломний проект присвячений програмним засобам повнотекстового пошуку у нереляційних базах даних, реалізований шляхом створення web-ресурсу для повнотекстового пошуку по науковим статтям.

Шляхом порівняльного аналізу програмних засобів повнотекстового пошуку у нереляційних базах даних було вирішено використовувати пошуковий сервер Elasticsearch. Реалізований web-ресурс являє собою сайт, на якому можна шукати по науковим статтям у режимі реального часу. Сайт містить наступні сторінки: сторінка для ознайомлення з описом даних для пошуку, сторінка для заповнення пошукового запиту, сторінки для видачі результатів запиту та сторінки для обробки користувацьких або серверних помилок. Також реалізовані алерти, що підказують користувачеві які дані для пошуку дадуть результат. Користувач може виконувати повнотекстовий пошук по назвам статей та по їх вмісту англійською мовою. Пошукова видача є відсортованою за відповідністю до користувацького запиту. Основними складовими ПЗ є модуль повнотекстового пошуку, модуль керування та статичні сторінки для відображення.

У даному дипломному проекті розроблено: архітектуру web-ресурсу, процедуру повнотекстового пошуку даних, графічні елементи, дизайн web-сторінок, інформаційну сторінку про входні дані для пошуку, обробка виключних ситуацій.

ABSTRACT

This diploma project is dedicated to software tools for full-text search in non-relational databases, implemented by creating web-resource for full-text search for scientific articles.

By means of a comparative analysis of software tools for full-text search in non-relational databases, it was decided to use the search engine ElasticSearch. Web-resource is implemented as a site where user can search for scientific articles in real time. The site contains the following pages: the data describing page, the search page, the query results page, page for handling server or user-made errors. Alerts that prompt the user to fill data for search correctly are also implemented. User can perform a full-text search by article content or title in English. Search results are sorted by relevance to the users' query. The main components of the software is full-text search module, control module and static pages for display.

In this project developed: the architecture of the web-site, full-text search engine, graphical elements, web-pages design, exceptions handling.

СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. ElasticSearch guide [Електронний ресурс] // elasticsearch.co. — 2016. — Режим доступу до ресурсу : <https://goo.gl/woZORy>. — Дата доступу : листопад 2016. — Назва з екрана.
2. MongoDB [Електронний ресурс] // mongo. — 2016. — Режим доступу до ресурсу : <https://goo.gl/V9NpTU>. — Дата доступу : листопад 2016. — Назва з екрана.
3. Apache Lucene [Електронний ресурс] // arachelucene. — 2015. — Режим доступу до ресурсу : <https://goo.gl/eVTxid>. — Дата доступу : листопад 2016. — Назва з екрана.
4. Apache Solr [Електронний ресурс] // arachesolr. — 2016. — Режим доступу до ресурсу : <https://goo.gl/dPFYZA>. — Дата доступу : листопад 2016. — Назва з екрана.
5. ElasticSearch [Електронний ресурс] // elasticsearch.co. — 2016. — Режим доступу до ресурсу : <https://goo.gl/MtuwF8>. — Дата доступу : листопад 2016. — Назва з екрана.
6. Apache Solr vs ElasticSearch [Електронний ресурс] // arachesolr. — 2016. — Режим доступу до ресурсу : <https://goo.gl/zLUazc>. — Дата доступу : листопад 2016. — Назва з екрана.
7. Java (programming language) [Електронний ресурс] // java. — 2016. — Режим доступу до ресурсу : <https://goo.gl/A8amqa>. — Дата доступу : листопад 2016. — Назва з екрана.
8. Java API (elasticsearch) [Електронний ресурс] // javaapi. — 2016. — Режим доступу до ресурсу : <https://goo.gl/PkBrH1>. — Дата доступу : листопад 2016. — Назва з екрана.

9. JavaScript (programming language) [Електронний ресурс] // javaapi. — 2016. — Режим доступу до ресурсу : <https://goo.gl/bYZCхq>. — Дата доступу : листопад 2016. — Назва з екрана.
10. JavaScript API (elasticsearch) [Електронний ресурс] // javaapi. — 2016. — Режим доступу до ресурсу : <https://goo.gl/E52Tdp>. — Дата доступу : листопад 2016. — Назва з екрана.
11. Python [Електронний ресурс] // python. — 2016. — Режим доступу до ресурсу : <https://goo.gl/IEVGF1>. — Дата доступу : листопад 2016. — Назва з екрана.
12. М. Лутц, Изучаем Python (4-е издание) [Текст] / М. Луц — СПб. : Символ-Плюс, 2010. — 43 с.
13. Python & Java: A Side-by-Side Comparison [Електронний ресурс] // python. — 2016. — Режим доступу до ресурсу : <https://goo.gl/c4R9MA>. — Дата доступу : листопад 2016. — Назва з екрана.
14. Python Elasticsearch Client [Електронний ресурс] // PythonElasticsearch. — 2016. — Режим доступу до ресурсу : <https://goo.gl/SWiVgW>. — Дата доступу : листопад 2016. — Назва з екрана.
15. Flask (веб-фреймворк) [Електронний ресурс] // Flask. — 2016. — Режим доступу до ресурсу : <https://goo.gl/h0hKDk>. — Дата доступу : листопад 2016. — Назва з екрана.
16. HTML [Електронний ресурс] // HTML. — 2016. — Режим доступу до ресурсу : <https://goo.gl/duyGDV>. — Дата доступу : листопад 2016. — Назва з екрана.
17. Bootstrap (фреймворк) [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/n61uTw>. — Дата доступу : листопад 2016. — Назва з екрана.
18. HTTP 404 [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/9T93qJ>. — Дата доступу : листопад 2016. — Назва з екрана.

19. Список кодів состояния HTTP [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/PUImRJ>. — Дата доступу : листопад 2016. — Назва з екрана.
20. Пошуковий індекс [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/zLUazc>. — Дата доступу : листопад 2016. — Назва з екрана.
21. Основы Elasticsearch [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/u46fnJ>. — Дата доступу : листопад 2016. — Назва з екрана.
22. Anatomy of Elasticsearch Clusrter [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/MSzvvV>. — Дата доступу : листопад 2016. — Назва з екрана.
23. CSS [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/Mrzqv>. — Дата доступу : листопад 2016. — Назва з екрана.
24. Консенсус [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/gf657h>. — Дата доступу : листопад 2016. — Назва з екрана.
25. Тестування [Електронний ресурс] // Вікіпедія. — 2016. — Режим доступу до ресурсу : <https://goo.gl/GRfd5d>. — Дата доступу : листопад 2016. — Назва з екрана.