

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І.І.МЕЧНИКОВА

(повне найменування вищого навчального закладу)

Факультет математики, фізики та інформаційних технологій

(повне найменування інституту, назва факультету (відділення))

Кафедра математичного забезпечення комп'ютерних систем

(повна назва кафедри (предметної, циклової комісії))

Дипломна робота

на здобуття освітньо-кваліфікаційного рівня «бакалавр»

(освітньо-кваліфікаційний рівень)

на тему Електронний визначник класифікаційного коду наукової публікації
по набору ключових слів та анотації

Виконав: студент денної форми навчання
напряму підготовки 6.050102 – Комп'ютерна інженерія
(шифр і назва напряму підготовки, спеціальності)

Присяжний Михайло Борисович

(прізвище, ім'я, по-батькові)

Керівник ст. викладач Трубіна Наталія Федорівна
(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рецензент ст. викладач Орлов Сергій Вячеславович
(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:

Протокол засідання кафедри

№ ___ від «___» _____ 2019 р.

Завідувач кафедри

(підпис)

Н.Ф.Трубіна
(прізвище, ініціали)

Захищено на засіданні ЕК № ____

протокол № __ від «__» _____ 2019 р.

Оцінка _____ / _____ / _____
(за національною шкалою, шкалою ECTS, бали)

Голова ЕК

(підпис)

О.О. Арсірій
(прізвище, ініціали)

АННОТАЦИЯ

Предлагается построение определителя универсального десятичного кода (УДК) научной публикации на базе унифицированного подхода к построению определителей. Проведен анализ доступных на данный момент веб-систем, связанных с УДК. Сформулированы требования к архитектуре определителя, приведены проектные решения, описаны используемые технологии.

Реализована система, которая помогает правильно классифицировать публикацию по ключевым словам и терминам из аннотации. Система дает возможность поиска по ключевым словам и аннотации по всем категориям книг и определение наиболее подходящей категории. Архитектура системы соответствует шаблону проектирования MVC.

Данные системы сохраняются в базе, созданной с помощью СУБД PostgreSQL. Серверная часть разработана с помощью языка Java. Клиентская часть реализована с помощью HTML 5, CSS 3, JavaScript, Vuejs, Vuetify и npm. Поиск по ключевым словам и аннотации осуществляется с помощью средства Elasticsearch.

АНОТАЦІЯ

Пропонується побудова визначника універсального десяткового коду (УДК) наукової публікації на базі уніфікованого підходу до побудови визначників. Проведено аналіз доступних на даний момент веб-систем, зв'язаних з УДК. Сформульовано вимоги до архітектури визначника, наведені проектні рішення, описані використовувані технології.

Реалізована система, яка допомагає правильно класифікувати публікацію за ключовими словами і термінам із анотації. Система дає можливість пошуку за ключовими словами і анотації по всім категоріям книг та визначення найбільш придатної категорії. Архітектура системи відповідає шаблону проектування MVC.

Дані системи зберігаються в базі, створеної за допомогою СУБД PostgreSQL. Серверна частина розроблена за допомогою мови Java. Клієнтська частина реалізована за допомогою HTML 5, CSS 3, JavaScript, Vuejs, Vuetify і прм. Пошук за ключовими словами і анотації здійснюється за допомогою засобу Elasticsearch.

ABSTRACT

It is proposed to construct a determinant of the universal decimal code (UDC) of a scientific publication based on a unified approach to the construction of determinants. The analysis of available web-systems that are somehow correlated with determination of UDC has been done. The requirements for the determinant architecture are formulated, the design decisions are given, the used ones are described.

The system which helps to correctly classify the publication of keywords and terms from the annotation was developed. The system provides the ability to search for keywords and annotations across all categories of books and determine the most appropriate category. The system architecture corresponds to the MVC design pattern.

The system data are stored in a database created by PostgreSQL. The server part is developed using Java language. Client part was implemented using HTML 5, CSS 3, JavaScript, Vuejs, Vuetify and npm. Search by keywords and annotation is done using the ElasticSearch tool.

ЗМІСТ

ВСТУП	7
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПРОГРАМНИХ ПРОДУКТІВ	9
1.1 Універсальна десяткова класифікація.....	9
1.2 Огляд програмних продуктів.....	12
1.3 Уніфікований підхід до побудови моделі даних визначника.....	16
1.4 Висновки	18
2 ПРОЕКТУВАННЯ СИСТЕМИ.....	19
2.1 Архітектура системи.....	19
2.2 Засоби реалізації	20
2.2.1 Springboot Embedded Tomcat Server	20
2.2.2 СКБД PostgreSQL.....	21
2.2.3 ElasticSearch	21
2.2.4 Засоби для зображення клієнтської частини.....	23
2.3 Специфікація вимог до бази даних	23
2.3.1 Визначення типів сутностей	24
2.3.2 Визначення типів зв'язків	24
2.3.3 Визначення атрибутів сутностей.....	25
2.3.4 Визначення доменів атрибутів	25
2.4 Створення схеми концептуальної моделі.....	27
2.5 Модель даних ElasticSearch	27
2.5.1 Визначення типів сутностей ElasticSearch	28
2.5.2 Визначення доменів атрибутів ElasticSearch	29
3 РЕАЛІЗАЦІЯ СИСТЕМИ.....	30
3.1 Проектування додатку користувача.....	30
3.2 Дизайн та тестування системи.....	30
3.3 Опис програмних компонентів.....	33
3.3.1 Серверна частина	33
3.3.2 Клієнтська частина.....	38

3.4 Опис роботи системи.....	41
ВИСНОВКИ.....	43
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	44
ДОДАТОК А ІЄРАРХІЯ КЛАСІВ СЕРВЕРНОЇ ЧАСТИНИ	46
ДОДАТОК Б ІЄРАРХІЯ КОМПОНЕНТІВ КЛІЄНТСЬКОЇ ЧАСТИНИ	47

ВСТУП

Важливим поняттям при роботі з інформацією є класифікація об'єктів – система розподілу об'єктів по класах відповідно до певної ознаки.

Ієрархічні класифікації використовуються у багатьох галузях знань, зокрема, в інформаційно-бібліотечній діяльності. З 2017 року, згідно з постановою Кабінету Міністрів України №177, від 22 березня усі бібліотечні заклади України для організації фондів та довідково-пошукового апарату бібліотек повністю перейшли на універсальну десяткову класифікацію (УДК).

Універсальна десяткова класифікація є бібліографічною класифікацією, яка представляє систематичну структуру галузей людського знання, організовану як систему, у якій поля знань пов'язані між собою. УДК - це система класифікації, що містить лексику та синтаксис, що дозволяє здійснювати потужне індексування вмісту та пошук інформації у великих масштабах.

У зв'язку з дуже великим обсягом таблиць УДК, завдання правильного визначника коду УДК для конкретної публікації ж досить складним. Зазвичай, визначити цей код допомагають працівники бібліотек. При цьому вони опираються на описи певних категорій. Здається доцільним формулювати цей процес. В природних науках для визначення місця конкретного об'єкта по його ключовим ознакам в системі класифікації використовуються так звані визначники. Для визначення кожна наукова публікація супроводжується анотацією та переліком ключових слів. Якщо ключові слова та важливі терміни з анотацій зіставити з категоріями УДК, їх можна було б використати в якості ключових ознак для визначення коду УДК конкретного документа.

Метою дипломної роботи є створення електронного визначника класифікаційного коду наукової публікації по набору ключових слів та анотації. Дана система дасть користувачеві можливість швидко визначити різні потрібні йому види документів.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

- 1) провести огляд предметної області;
- 2) провести огляд існуючих програмних аналогів;
- 3) сформулювати вимоги до створюваної системи;
- 4) провести проектування бази даних інформаційної системи;
- 5) провести проектування додатка;
- 6) вибрати інструментальне середовище розробки і виконати програмну реалізацію інформаційної системи.

ВИСНОВКИ

В ході виконання даної дипломної роботи розроблена веб-система, яка дозволяє проводити пошук класифікаційного коду наукової публікації за набором ключових слів та анотації.

Під час виконання даної роботи організована, описана і нормалізована база даних. Обрані, на мою думку, найбільш актуальні для даної теми веб-технології, які дозволили досягнути головної мети цього проекту.

Виконані наступні завдання:

- 1) здійснений пошук по ключовим словам та анотації;
- 2) здійснений доступ до всіх категорій на одній сторінці за допомогою розгортання у дереві УДК;
- 3) пошук по ключовим словам та анотації розділені у різні вкладки, замість переходу на інші інтернет-ресурси;
- 4) зроблена панель адміністратора для додання ключових слів для можливості додання у майбутньому адміністратора як користувача, котрий матиме до неї доступ.

У майбутньому можна розробити призначений для користувача режим і застосувати штучний інтелект при пошуку потрібної категорії. Також, можна зробити версію визначення коду УДК по тексту книги та створити інтерактивний механізм опитування користувача з приводу ефективності пошуку з метою навчання нейромережі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Універсальна десяткова класифікація [Електронний ресурс] – Режим доступу: http://www.ukrbook.net/UDC/UDC_skoroch_table.html
2. Систематизація документів [Електронний ресурс] – Режим доступу: <http://dnpgov.gov.ua/wp-content/uploads/2017/06>
3. UDC – Universal Decimal Classification [Електронний ресурс] – Режим доступу: <http://www.lisbdnet.com/universal-decimal-classification-udc/>
4. УДК – Наукова бібліотека Одеського національного університету [Електронний ресурс] – Режим доступу: <http://lib.onu.edu.ua/ukrayinska-udk/>
5. УДК mat.net.ua [Електронний ресурс] – Режим доступу: <http://mat.net.ua/yuliya-article/udk-2015.htm>
6. УДК triumph [Електронний ресурс] – Режим доступу: <https://www.triumph.ru/udk-klassifikator.html>
7. УДК ГПНТБ [Електронний ресурс] – Режим доступу: <http://library.gpntb.ru/cgi/irbis64r/gpntb/>
8. T.Petrushina, N.Trubina, Quality analysis of the computer identifier based on a unified approach, PROCEEDINGS of the 3d International Conference on Computer Algebra and Information Technologies, August 20 – 25, 2018 Odessa, Ukraine, p.188-192
9. Трирівнева архітектура БД. [Електронний ресурс] – Режим доступу: <http://um.co.ua/4/4-17/4-179510.html>
10. Патерн MVC. [Електронний ресурс] – Режим доступу: https://www.tutorialspoint.com/design_pattern/mvc_pattern.htm
11. Embedded Tomcat Server [Електронний ресурс] – Режим доступу: <https://dzone.com/articles/spring-boot-embedded-tomcat-example>
12. Elasticsearch basics. [Електронний ресурс] – Режим доступу: <https://habr.com/ru/post/280488/>

13.API [Електронний ресурс] – Режим доступу:

<https://habr.com/ru/sandbox/52599/>

14.Vuejs. [Електронний ресурс] – Режим доступу:

<https://metanit.com/web/vuejs/>

15.Малахов Є.В. Організація баз даних: конспект. – О.: ОНУ, 2014. - 169 с.

16.Spring Boot. [Електронний ресурс] – Режим доступу:

<https://spring.io/guides>.

17.Восточно-Европейский журнал [Електронний ресурс] - Режим доступу:

<http://journals.uran.ua/eejet/issue/archive>