

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**МЕЖДУНАРОДНЫЙ ГОСУДАРСТВЕННЫЙ ЭКОЛОГИЧЕСКИЙ
ИНСТИТУТ им. А.Д.САХАРОВА**

Кафедра информационных технологий в экологии и медицине

**РАЗРАБОТКА ПОИСКОВОГО МОДУЛЯ ДЛЯ ТЕКСТОВЫХ
ДОКУМЕНТОВ**

Дипломная работа студента IV курса

ЛУКИЯНЦА Владислава Олеговича

_____ В.О. Лукиянец

«Допустить к защите»
Зав. кафедрой информационных
технологий в экологии и медицине
к.б.н., доцент

Научный руководитель
к.ф.-м.н., доцент

_____ В.В. Журавков
«____» _____ 2022 г.

_____ В.А. Иванюкович

Минск, 2022

РЕФЕРАТ

Дипломная работа, 52 с., 34 рис., 6 источников, 12 приложений.

Ключевые слова: ПОИСКОВЫЙ МОДУЛЬ, ДОКУМЕНТЫ, TOVUS, ELASTIC SEARCH, SOAP, SOA, REST, HTTP.

Цель работы – проектирование и разработка поискового модуля для текстовых документов.

Объект исследования – технологии создания поисковых систем в полнотекстовых документах различных форматов и их метаданных.

Область применения – информационные порталы различных тематических направлений (медицина, юриспруденция, экономика и т.п.), стриминговые сервисы, социальные сети.

Задачи:

1. Изучить особенности модульных и многоуровневых архитектур создания приложений.
2. Изучить особенности поисковых машин Tovus и ElasticSearch.
3. Расширить возможности поисковой машины Tovus и разработать модуль, отвечающий за поиск документов по шаблонам и различным критериям.
4. Разработать решение, позволяющее организовать связи между документами.
5. Разработать двойной поиск по родительским и связанным документам.

Разработанный модуль в настоящий момент используется в нескольких внутренних проектах компании ИООО «ЭПАМ Системз»

РЭФЕРАТ

Дыпломная работа, 52 с., 34 мал., 6 крыніц, 12 прыкладанняў.

Ключавыя словы: ПОШУКАВЫ МОДУЛЬ, ДАКУМЕНТЫ, TOVUS, ELASTIC SEARCH, SOAP, SOA, REST, HTTP.

Мэта работы – прасктаванне і распрацоўка пошукавага модуля для тэкставых дакументаў.

Аб'ект даследавання – тэхналогіі стварэння пошукавых сістэм у паўнатэкставых дакументах розных фарматаў і іх метададзеных.

Вобласць прымянення – інфармацыйныя парталы розных тэматычных напрамкаў (медыцына, юрыспрудэнцыя, эканоміка і да т.п.), стрымінгавыя сэрвісы, сацыяльныя сеткі.

Задачы:

1. Вывучыць асаблівасці модульных і шматузроўневых архітэктур. стварэнне прыкладанняў.
2. Вывучыце магчымасці Tovus і ElasticSearch.
3. Пашырыць магчымасці пошукавай сістэмы Tovus і развіць модуль, які адказвае за пошук дакументаў па шаблонах і розным крытэрыі.
4. Распрацаваць рашэнне, якое дазваляе арганізаваць сувязі паміж дакументы.
5. Распрацаваць падвойны пошук па бацькоўскіх і звязаных дакументах.

Распрацаваны модуль у сапраўдны момант выкарыстоўваецца ў некалькіх унутраных праектах кампаніі ЗТАА «ЭПАМ Сістэмз»

ESSAY

Thesis 52 pages., 34 figures, 6 sources, 12 attachments.

DEVELOPMENT, SEARCH MODULE, DOCUMENTS, TOVUS, ELASTIC SEARCH, SOAP, SOA, REST, HTTP.

Purpose of the work:

Design and development of a search module for text documents.

Object of study:

A set of documents, the Tovus search engine, ElasticSearch.

Scope

In the service sector, on information portals, streaming services, social networks, and so on to obtain relevant information.

Tasks:

1. To study the features of modular and multi-level architectures.
creating applications.
2. Explore Tovus and ElasticSearch.
3. Expand the capabilities of the Tovus search engine and develop
a module responsible for searching for documents by templates and various
criteria.
4. Develop a solution that allows you to organize links between
the documents.
5. Develop a double search for parent and related documents.

The developed module is currently used in several internal projects of EPAM Systems Ltd.