Была осуществлена разработка поисковой системы по сайтам – приложения, которое помогает его пользователям быстро находить нужную информацию, используя поле поиска. Работа осуществляется с помощью индексирования страниц и осуществления по ним быстрого поиска.

Существует много поисковых систем, лучшей из которых, по мнению большинства разработчиков и пользователей, является Яндекс.Сервер. Но она прекратила своё существование как отдельный сервис. Создание собственной поисковой системы поможет ускорить поиск и выдачу релевантной, т.е. соответствующей запросу, информации за счёт ограниченного списка сайтов-источников информации. Принципы работы такой системы можно менять и развивать.

Поисковая система представляет из себя Spring-приложение (JAR-файл, запускаемый на любом сервере или компьютере), работающее с локально установленной базой данных MySQL, имеющее простой веб-интерфейс и API, через который им можно управлять и получать результаты поисковой выдачи по запросу.

Данная система разрабатывалась на фреймворке Spring. Фреймворк — это программная платформа, которая упрощает разработку программного продукта, определяет структуру проекта и помогает удобно объединять в нём разные компоненты. Spring — это популярный фреймворк для разработки на Java, который используют для создания корпоративных приложений, например CRM. Иногда Spring называют фреймворком фреймворков, поскольку он состоит из отдельных модулей: Struts, Hibernate, Tapestry, EJB, JSF и других. Каждый модуль — это набор инструментов для решения технических задач, объединённых общей логикой Spring. Таким образом, Spring — это набор взаимосвязанных мини-фреймворков, созданных для работы над разными частями приложения. Подключать их можно по отдельности, в зависимости от задач. В частности Spring содержит Spring Data - модуль для работы с реляционными и нереляционными базами данных: MySQL, Redis, Microsoft Azure Cosmos DB и другими. Включает в себя набор интерфейсов для работы с данными через JPA Entity.

Данная система отражает следующие принципы работы:

1. В конфигурационном файле перед запуском приложения задаются адреса сайтов, по которым система должна осуществлять поиск.
2. Поисковая система должна самостоятельно обходить все страницы заданных сайтов и индексировать их (создавать так называемый индекс) так, чтобы потом находить наиболее релевантные страницы по любому поисковому запросу.
3. Пользователь присылает запрос через API движка. Запрос — это набор слов, по которым нужно найти страницы сайта.
4. Запрос определённым образом трансформируется в список слов, переведённых в базовую форму. Например, для существительных — именительный падеж, единственное число.
5. В индексе ищутся страницы, на которых встречаются все эти слова.
6. Результаты поиска ранжируются, сортируются и отдаются пользователю.

Данная система состоит из двух частей – видимой пользователю и невидимой ему. Вторая часть реализовывается на языке Java в среде разработки IntelliJ Idea. Первая часть – это интерфейс системы.

Веб-интерфейс (frontend-составляющая) проекта представляет собой одну веб-страницу с тремя вкладками:

● Dashboard. Эта вкладка открывается по умолчанию. На ней отображается общая статистика по всем сайтам, а также детальная статистика и статус по каждому из сайтов (статистика, получаемая по запросу /api/statistics).

● Management. На этой вкладке находятся инструменты управления поисковым движком — запуск и остановка полной индексации (переиндексации), а также возможность добавить (обновить) отдельную страницу по ссылке:

● Search. Эта страница предназначена для тестирования поискового движка. На ней находится поле поиска, выпадающий список с выбором 1 сайта для поиска, а при нажатии на кнопку «Найти» выводятся результаты поиска (по API-запросу /api/search): Вся информация на вкладки подгружается путём запросов к API вашего приложения. При нажатии кнопок также отправляются запросы.

Для корректной оценки работы данной поисковой системы, необходимо сравнивать её со следующими критериями:

* Все три страницы открываются.
* На странице DASHBOARD выводится статистика по сайтам: список сайтов, а по каждому сайту — информация о количестве проиндексированных страниц и лемм, соответствующая реальности.
* На странице MANAGEMENT работают запуск и остановка индексации.
* На странице SEARCH работает поиск как по всем сайтам, так и по каждому отдельному. В поисковой выдаче выводится список найденных страниц, соответствующих запросу. Каждая страница выводится в виде её заголовка и сниппета, содержащего одно или несколько упоминаний слов из поискового запроса. Сниппеты выводятся в виде понятных текстов, не содержащих HTML-теги. При клике по любому результату в новой вкладке открывается соответствующая веб-страница на том сайте, к которому относится этот результат.