

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.775:34

Глацкевич
Олег Вячеславович

**ПРОГРАММНЫЙ КОМПЛЕКС ОПТИМИЗАЦИИ ПОИСКА В
ИНФОРМАЦИОННО-ПРАВОВОЙ СИСТЕМЕ**

АВТОРЕФЕРАТ

диссертации на соискание степени магистра

по специальности 1-40 80 01 Компьютерная инженерия. Хранение и
обработка данных

Научный руководитель
Одинец Дмитрий Николаевич
кандидат технических наук,
доцент кафедры ЭВМ, БГУИР

Минск 2021

ВВЕДЕНИЕ

Одной из важнейших задач, которые стоят перед государством, является обеспечение правовой информированности граждан, предоставление и распространение правовой информации. Чтобы у каждого гражданина своей страны была возможность доступа к правовой информации.

Применение информационно-поисковых систем (далее - ИПС) для получения доступа к правовым актам и иной правовой информации стало естественным процессом в работе не только государственных органов, организаций, но и физических лиц. В Республике Беларусь Национальным центром правовой информации Республики Беларусь (далее – НЦПИ) создана и развивается государственная система правовой информации, предназначенная, прежде всего, для информационно-правового обеспечения принятия решений.

ИПС правовой информации применяется для обеспечения доступа пользователей к электронной копии эталонного банка данных правовой информации Республики Беларусь и иных правовых банков данных, предоставляя многофункциональную систему быстрого поиска нормативной и иной информации, необходимой для практической, научной либо учебной деятельности, а также для систематизации и кодификации законодательства.

К основным задачам, возложенным на НЦПИ, отнесены формирование и ведение эталонного банка данных правовой информации Республики Беларусь, государственных информационно-правовых ресурсов, среди которых информационно-поисковая система «ЭТАЛОН-ONLINE».

ИПС «ЭТАЛОН-ONLINE» предназначена для работы с эталонным и иными банками данных правовой информации посредством сети Интернет с помощью браузера или мобильных приложений для смартфонов и планшетов, работающих под операционными системами Android и iOS.

Актуальность задачи совершенствования информационного поиска в системе «ЭТАЛОН-ONLINE» обусловлена как неуклонно возрастающими ролью и объемами использования электронных информационно-правовых ресурсов во всех сферах нормотворческой и правоприменительной деятельности, которые являются одними из важнейших составляющих государственного управления в целом, так и возрастанием правовой культуры населения страны и процессами формирования правового государства и информационного общества, что предъявляет новые требования к данному ресурсу, как с точки зрения его функциональности, так и удобства его использования самыми широкими слоями населения и в том числе непрофессионалами в юридической области.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объект исследования: информационно-поисковая система правовой информации.

Предмет исследования: программно-технологические решения по организации поиска в информационно-поисковой системе правовой информации.

Цель исследования: совершенствование поиска правовой информации в информационно-поисковой системе, функционирующей в рамках государственной системы правовой информации, посредством выработки единых подходов к организации поиска и разработки программно-технологического комплекса, обеспечивающего реализацию этих подходов.

Программный комплекс должен отвечать современным требованиям, в том числе обеспечивать:

1. Высокую скорость обработки больших объемов информации.
2. Высокую скорость поиска информации в банках данных.
3. Релевантность представления результатов поиска.

В соответствии с поставленной целью были определены следующие **задачи исследования:**

1. Аналитический обзор литературы по вопросам организации поиска в информационных системах.
2. Анализ существующих систем полнотекстового поиска.
3. Разработка программно-технологического комплекса, обеспечивающего реализацию единых подходов к организации поиска.

Основным итогом магистерской работы является разработка технорабочего проекта, включающего разработку (модернизацию) программно-технологического комплекса, обеспечивающего совершенствование и реализацию единых подходов к организации поиска в информационно-поисковой системе правовой информации.

Научная и практическая значимость исследования: создание и внедрение разработанного продукта в рамках требований Национального центра правовой информации.

Апробация и публикации: на основе полученных в ходе исследований данных была опубликована статья «Организация поиска в информационной системе» [1–А].

СОДЕРЖАНИЕ РАБОТЫ

Общий объем магистерской диссертации составляет 74 страницы, включая 28 иллюстраций, 4 таблицы, библиографический список из 30 наименований, 1 приложение.

В **общей характеристике работы** сформулированы цель и задачи исследования, даны сведения об объекте и предмете исследования, приведены апробации и публикации результатов.

Во **введении** дано обоснование актуальности работы, описывается назначение и применение разработанной системы.

В **первой главе** изучены общие вопросы осуществления поиска информации в компьютерной среде и подходы к организации поиска в информационных системах, проведен анализ существующих систем полнотекстового поиска. В первом разделе приведена краткая историческая справка по развитию информационного поиска, описаны основные особенности организации информационного поиска. Во втором разделе рассмотрены основные методы обработки поисковых запросов. В третьем разделе приведено краткое описание методики оценки качества информационного поиска. В четвертом разделе рассмотрены специальные программные инструменты, решающие задачу поиска по тексту и организующие работу со сложными индексами и большими объемами данных.

Во **второй главе** приведено описание теоретической базы исследования. В первом разделе представлена общая схема построения информационно-поисковых систем, приведено описание основных компонентов системы. Во втором разделе проведен анализ существующего поискового механизма.

В **третьей главе** приведено подробное описание разработанной системы. В первом разделе приведена методика оценки, анализа и сравнения результатов поиска, применяемая в ходе разработки системы. Во втором разделе представлено описание работы анализатора входной строки, выполняющего морфологическую обработку данной строки посредством использования информации из лингвистической базы данных, а также выполняющий определение реквизитов документов в поисковой строке. В третьем разделе описан блок управления ранжированием, с помощью функциональных возможностей которого осуществляется управление и настройка факторов ранжирования, использующихся в системе поиска. В четвертом разделе описаны методы расчета основных факторов ранжирования, таких как фактор частоты встречаемости слов поискового запроса в тексте документа, порядка (близости) следования слов и другие. В пятом разделе приведено описание всех факторов ранжирования, которые используются в системе поиска для ранжирования документов в результатах поиска. В шестом разделе приведено краткое описание

блока сбора статистики о работе пользователей, который обеспечивает непрерывное наполнение информации о действиях пользователей при работе с поисковой системой. Данные сведения используются для анализа успешности выполнения поисковых запросов, а также для формирования дополнительных коэффициентов популярности документов и таким образом влиять на ранжирование результатов поиска (поведенческий фактор ранжирования). В седьмом разделе представлена архитектура системы, описаны основные инструменты и программные модули, которые были применены в разработке. В восьмом разделе представлены результаты сравнительного тестирования поиска ИПС "ЭТАЛОН-ONLINE" и интернет-систем сторонних организаций-распространителей правовой информации.

В **заключении** сформулированы основные результаты работы.

ЗАКЛЮЧЕНИЕ

В рамках проведенного исследования были изучены теоретическая и практическая составляющие информационного поиска. В результате работ выявлено, что современные средства, с помощью которых осуществляется информационный поиск, в основном, строятся по схожим принципам и в большинстве своем успешно решают основную задачу информационного поиска – удовлетворяют информационные потребности пользователей. Однако, следует отметить, что список задач информационного поиска постоянно расширяется, и данное обстоятельство предполагает проведение работ по его совершенствованию.

В ходе исследования были подробно рассмотрены различные специальные программные инструменты, решающие задачу поиска по тексту и организующие работу со сложными индексами и большими объемами данных. Выявлено, что по данному направлению работ с целью эффективного поиска, в основном, применяются системы полнотекстового поиска, среди таких некоммерческих систем самыми востребованными являются системы Sphinx, Solr и ElasticSearch.

При проведении исследования, учитывая специфику поиска правовой информации, необходимо отметить, что для эффективной работы поиска недостаточно только программного инструмента рассматриваемых систем поиска. Необходима либо доработка функционала систем поиска под специфику ИПС правовой информации либо создание отдельного программного модуля, взаимодействующего с поисковой машиной и ее настройку обработки получаемой от модуля метайнформации.

В рамках исследования разработан и внедрен в опытную эксплуатацию технорабочий проект, включающий разработку программно-технологического комплекса, обеспечивающего совершенствование и реализацию единых подходов к организации поиска в рамках информационно-поисковых систем правовой информации.

Внедрение результатов исследования способствует:

совершенствованию поиска правовой информации в информационно-поисковых системах, функционирующих в рамках государственной системы правовой информации;

совершенствованию деятельности государственных органов и организаций посредством совершенствования процессов предоставления доступа к актуальной правовой информации.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1–А.] Глацкевич, О. В. Организация поиска в информационной системе / О. В. Глацкевич // Молодой ученый. — 2021. — Май. — № 20 (362). — С. 22—24.

Библиотека БГУИР