

## СЕКЦИЯ 4 АКТУАЛЬНЫЕ ПРОБЛЕМЫ ИНФОРМАТИКИ

УДК 338.27+338.22.021.1

### ПОИСКОВЫЙ МОДУЛЬ ДЛЯ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ

В.А. Зеленко<sup>a</sup>, А.Ю. Сикиржицкий<sup>b</sup>

#### Аннотация

Статья посвящена идее улучшения качества поиска книг в электронной библиотеке. Предложена модель поиска электронных книг по реквизитам из личных карточек, хранящихся в базе данных. Представлен алгоритм, реализующий метод анализа поискового запроса пользователя для нахождения атрибутов, являющихся отличительными параметрами искомых объектов. На основе алгоритма разработан поисковый модуль, который может быть подключен к любой поисковой системе в данной области.

**Ключевые слова:** модуль, поиск, модель поиска, поисковый запрос, библиотека, книга.

### THE SEARCH MODULE FOR E-REPOSITORY

V. Zelianko<sup>a</sup>, A. Sikirzhytski<sup>b</sup>

#### Abstract

The article is dedicated to the idea of improving the quality of search for e-Books. A model of search for e-Books through the information of users' personal cards stored in the database was proposed. There is the algorithm which implements the method of analysis of user's search query to find attributes that are distinctive parameters of the desired objects. On the basis of a search algorithm a search module was developed which can be connected to any search engine.

**Keywords:** module, search, model search, search request, the library, the book.

#### Введение

В настоящее время информация является одним из ценнейших ресурсов общества. Но для того чтобы использовать информацию, нужно для начала ею владеть. Существует огромное множество информационных массивов, и очень важно иметь возможность быстро находить нужную информацию. Для решения этой задачи и предназначены информационно-поисковые системы.

Информационно-поисковая система – это система, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска [1]. Таким образом, основа реализации поискового модуля – это базы данных, в которых хранится систематизированная информация, относящаяся к применяемой сфере. Систематизированная информация подразумевает присвоение дополнительных характеристик объектам поиска.

<sup>a</sup> Зеленко Виталий Анатольевич,  
Минский университет управления,  
магистрант  
Zelianko Vitali,  
Minsk University of Management,  
Master's degree student  
prisoner08071991@yandex.ru

<sup>b</sup> Сикиржицкий Артём Юрьевич,  
Минский университет управления,  
магистрант  
Sikirzhytski Artsiom,  
Minsk University of Management,  
Master's degree student  
mastak2@tut.by

Для книг это может быть дата издания, год издания, жанр, авторы и т.д. Для картинок можно выделить такие параметры как разрешение, формат, тематика и т.д.

Любая поисковая система оперирует структурами данных и исполняет алгоритм. Поисковый алгоритм можно представить себе как последовательность поисковых действий, в которой вначале осуществляется поиск наиболее популярных совпадений, затем, при неудаче, выполняется подробный анализ поисковой строки путем разделения ее на составляющие. Таким образом система пытается удовлетворить часть запроса, после чего выдает конечный результат. Поисковый алгоритм ищет совпадения по ключевым словам запроса с информацией в базе данных и находит релевантные результаты. На сегодняшний день существует мало поисковых алгоритмов, поэтому необходима разработка новых методов поиска, обеспечивающих эффективное и быстрое нахождение релевантной информации.

Когда речь заходит о повышении качества поиска, всё растущем объеме информации и увеличивающемся потоке пользовательских запросов, необходимо оперировать теоретическим аппаратом – так называемой моделью поиска. Модель поиска – это сочетание следующих составляющих [2]:

- способ представления документов;
- способ представления поисковых запросов;
- вид критерия релевантности документов.

В настоящей статье представлены модель поиска электронных книг по реквизитам из личных карточек, хранящихся в базе данных, и алгоритм поиска релевантной информации с использованием полученной модели.

Для улучшения поисковой системы электронного учебно-методического комплекса (ЭУМК) решено внедрить автоматический модуль разбора строки. ЭУМК позволяют собрать в единый комплекс практически все информационные материалы, требуемые для изучения той или иной дисциплины, вследствие чего пользователь может испытывать затруднение в поиске нужной книги. Модуль будет основан на нахождении в поисковой строке реквизитов из карточек электронных книг, что не только упростит, но и улучшит качество поиска нужной информации. ЭУМК обеспечивает необходимую сегодня интерактивность, наглядность, мобильность, компактность и низкую стоимость тиражирования, многовариантность, многоуровневость и разнообразие проверочных заданий и тестов. К достоинствам современных ЭУМК относятся прежде всего возможность эффективной организации самостоятельной работы и активизация роли обучаемого в процессе обучения.

При использовании поискового модуля в информационно-поисковых системах можно выделить плюсы и минусы данного подхода.

Плюсы:

- пользователю не нужно будет обращаться к расширенному поиску информационно-поисковой системы;
- увеличение точности нахождения релевантной информации;
- возможность подключения к другим поисковым системам данной информационной сферы;
- исключение нахождения ложной информации, найденной при помощи поиска по тексту.

Минусы:

- неэффективен при маленьких и отсутствии необходимых атрибутов в пользовательских запросах;
- тратится дополнительное время на анализ пользовательского запроса.

Для точного ответа на поисковый запрос достаточно будет вычислить следующие реквизиты электронных книг:

- авторы;
- жанр;

- издательство;
- год публикации;
- количество страниц;
- название.

Ядром модуля является процесс анализа поискового запроса с целью нахождения в нем атрибутов, которые позволят находить совпадения вне зависимости от склонения слов. Модель поиска представлена в виде входных данных, поискового модуля, в котором при помощи словарей и данных происходит поиск атрибутов и формирование XML ответа (рисунок 1).

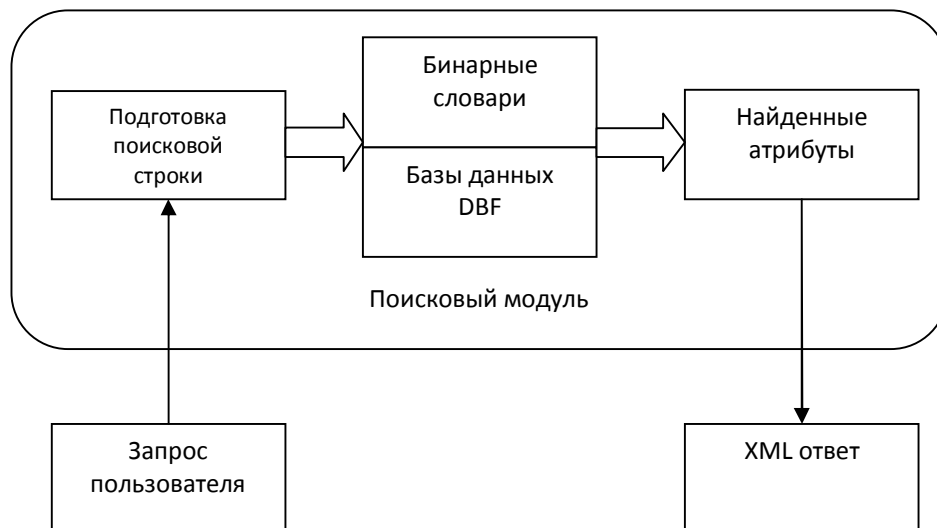


Рисунок 1 – Модель поиска модуля

Последовательность нахождения реквизитов:

- название;
- авторы;
- издательство;
- жанр;
- год публикации;
- количество страниц.

Входными данными поискового модуля являются три строковые переменные:

- поисковая строка, введенная пользователем;
- путь к базам данных, в которых хранится систематизированная информация, относящаяся к применяемой сфере;
- путь к бинарным словарям, которые позволяют реализовать технологию приведения слов к каноническому виду.

На рисунке 2 представлен алгоритм работы поискового модуля.

Для более эффективной работы модуля поиск атрибутов происходит в строгом порядке, начиная с важных, которые помогают точнее определить искомую информацию, и кончая второстепенными. Во избежание определения одного атрибута за другой (номер за дату или наоборот) найденная информация удаляется из поискового запроса. Для нахождения цифровых и буквенных атрибутов используются регулярные выражения в связке с базами данных. Регулярные выражения предоставляют мощный, гибкий и эффективный метод обработки текста [3].

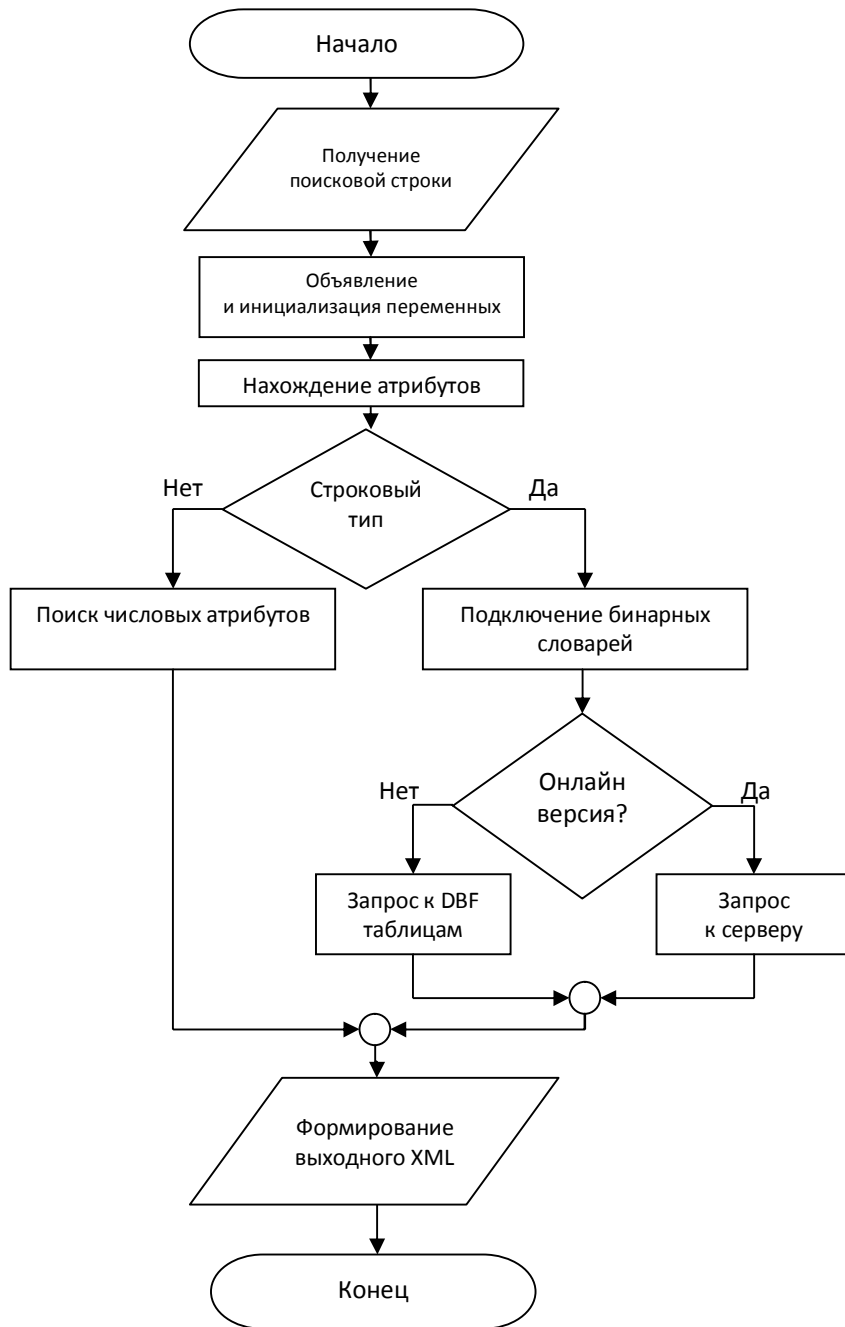


Рисунок 2 – Алгоритм поиска

На рисунке 3 представлена формируемая после выполнения всех функций результирующая XML строка.

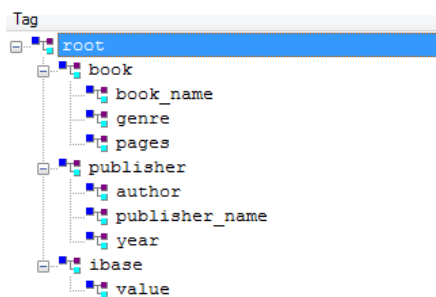


Рисунок 3 – Дерево XML ответа

### Заключение

Существует большое количество методов по разработке поисковых систем, но каждый из них не является совершенным. Качество информационного поиска современных систем неуклонно растет. Это обусловлено растущими вычислительными мощностями и объемами памяти современных компьютеров, что позволяет реализовывать все более сложные и ресурсоемкие алгоритмы.

Многие модели, особенно связанные со сложным анализом текста, таким как синтаксический анализ, пока просто невозможно реализовать в полном объеме, поэтому трудно оценить их влияние на качество поиска. Другие пока еще недостаточно изучены с точки зрения их влияния на качество поиска и требуют дополнительных исследований.

Разработка поискового модуля является хорошим способом улучшения поисковой системы. При наличии обширной базы данных и систематизированных данных данное направление дает значительные преимущества для нахождения релевантной информации. Неоспоримый плюс поискового модуля заключается именно в его модульности, что позволяет расширять и модернизировать модуль вне зависимости от изменения в поисковой системе. Еще одним немаловажным преимуществом является возможность удобного подключения к любой поисковой системе, базирующейся в данной информационной сфере.

Поисковый модуль удовлетворяет всем требованиям для усовершенствования поиска и является новейшей технологией по улучшению качества работы информационно-поисковых систем. Модуль является универсальным и позволяет находить точную и достоверную информацию в применяемой сфере благодаря технологии регулярных выражений. ЭУМК позволяют собрать в единый комплекс практически все информационные материалы, требуемые для изучения той или иной дисциплины. В результате накопления большого массива информации пользователь может испытывать сложности в поиске нужного ему материала. Благодаря поисковому модулю данная проблема исчезает, так как он позволяет отсеять лишнее и предоставить пользователю желаемый результат поиска.

### Литература / References

1. Карманова, А.П. Структура и классификация автоматизированных информационных систем / А.П. Карманова [Электрон. ресурс]. – 2013. – Режим доступа : [http://ru.science.wikia.com/wiki/Информационно-поисковые\\_системы](http://ru.science.wikia.com/wiki/Информационно-поисковые_системы). – Дата доступа : 12.10.2014.  
Karmanova, A.P. Struktura i klassifikatsiya avtomatizirovannykh informatsionnykh sistem / A.P. Karmanova [Electronic resource]. – 2013. – Mode of access : [http://ru.science.wikia.com/wiki/Informatsionno-poiskovyie\\_sistemy](http://ru.science.wikia.com/wiki/Informatsionno-poiskovyie_sistemy). – Date of access : 12.10.2014.
2. Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский. – М. : ДМК Пресс, 2003. – 288 с.  
Kogalovskiy, M.R. Perspektivnyye tekhnologii informatsionnykh sistem / M.R. Kogalovskiy. – М. : DMK Press, 2003. – 288 s.
3. Фридл, Дж. Регулярные выражения / Дж. Фридл. – М. : Символ-Плюс, 2008. – 160 с.  
Fridl, Dzh. Regulyarnyye vyrazheniya / Dzh. Fridl. – М. : Simvol-Plyus, 2008. – 160 s.