

РАЗРАБОТКА МЕТАПОИСКОВОЙ СИСТЕМЫ ПО ЭЛЕКТРОННЫМ ИНФОРМАЦИОННЫМ БИБЛИОТЕЧНЫМ РЕСУРСАМ

Спирина А.В.

Научный руководитель – старший преподаватель Бархатов А.В.

Сибирский федеральный университет

Научная библиотека Сибирского федерального университета является одной из крупнейших вузовских библиотек в России. Она имеет доступ к онлайн базам данных крупнейших российских и зарубежных производителей информационных продуктов, содержащим: электронные полнотекстовые версии научных журналов, газет и книг, материалов конференций, патентную, библиографическую и наукометрическую информацию.

На сегодняшний день пользователи вынуждены производить один и тот же поиск по многочисленным электронным информационным ресурсам Научной библиотеки Сибирского федерального университета. Данные ресурсы отличаются друг от друга поисковыми интерфейсами, навигацией, правилами запросов. Данная проблема характерна не только для Научной библиотеки Сибирского федерального университета, но и для других вузов и научных учреждений, имеющих доступ к различным электронным информационным ресурсам.

В связи с этой проблемой возникла идея разработать метапоисковую систему по электронным информационным библиотечным ресурсам, чтобы оптимизировать работу пользователей с электронными каталогами и онлайн базами данных различных электронных библиотек.

Метапоиск или метапоисковая система – это поисковая система, которая использует для выдачи результатов данные нескольких поисковых систем. То есть метапоисковая система не имеет собственной базы данных и поисковой машины (поисковых роботов, которые индексируют и оценивают сайты), а обладает только интерфейсом и механизмом запроса к базам других поисковых систем.

Разработанная метапоисковая система выступает посредником между пользователем и поисковыми системами электронных каталогов и онлайн баз данных, рассылая введенный пользователем запрос в поисковые системы электронных информационных ресурсов и обрабатывая полученную информацию, затем выдает обработанную информацию пользователю.

В Научной библиотеке Сибирского федерального университета один из электронных каталогов работает в АБИС Ирбис, которая позволяет осуществлять метапоиск по электронным каталогам библиотек Красноярска, работающих на АБИС Ирбис. Этот метапоиск осуществляется по протоколу Z39.50.

Также достаточно известен совместный информационный проект Научной библиотеки МГУ и компании "Библиотечная компьютерная сеть" – СИГЛА. Это портал межбиблиотечной информации, который помогает сформировать запрос и найти документы в электронных каталогах крупнейших библиотек России и зарубежных стран, поддерживающих протокол Z39.50.

Для разрабатываемой метапоисковой системы был выбран механизм взаимодействия с поисковыми системами онлайн баз данных по протоколу HTTP, так как доступа к другим, более эффективным, механизмам, таким как: OAI PMH, Z39.50 – нет. Для передачи данных запроса в различных поисковых системах используется

HTTP-метод POST или GET. Поэтому разрабатываемая метапоисковая система использует оба метода передачи данных запроса. Многие электронные информационные ресурсы работают с пользователем с использованием cookies. Вследствие чего разрабатываемая система устроена таким образом, что вначале она получает от данного электронного информационного ресурса cookies, а затем пересылает эти данные веб-серверу данного ресурса в HTTP-запросе при попытке открыть страницу соответствующего сайта.

Так как все электронные информационные ресурсы имеют различные поисковые формы (с различными полями для поиска), то для поисковой формы разрабатываемой системы были выбраны те поля, которые встречаются во всех электронных ресурсах. Все электронные информационные ресурсы относятся к различным отраслям наук. Для оптимизации поиска по электронным информационным библиотечным ресурсам, была проведена классификация этих ресурсов по отраслям науки, таким образом, пользователь может выбрать поиск только по тем электронным информационным ресурсам, которые удовлетворяют выбранной отрасли науки.

Для разработки метапоисковой системы был выбран язык программирования PHP. Для метапоиска необходимо было реализовать “модуль мультизапроса”, который позволил бы одновременно сделать несколько запросов к сайтам электронных ресурсов. Данный модуль был реализован с использованием библиотеки cURL (реализация псевдо многопоточности на PHP). cURL – это свободная, кроссплатформенная служебная программа командной строки для передачи данных по различным протоколам с синтаксисом URL.

Принципы работы метапоисковой системы представлены на рисунке 1.

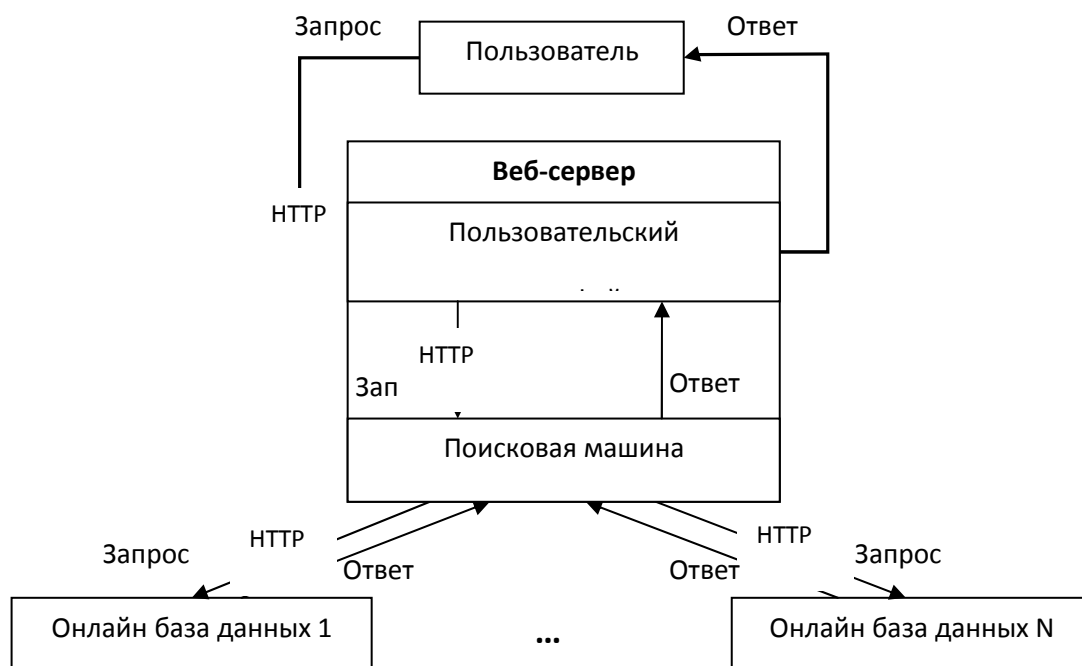


Рисунок 1 – Схема работы метапоисковой системы

Данная система работает следующим образом:

Пользователь заполняет форму запроса и отправляет запрос на веб-сервер.

Этот запрос (адаптированный/трансформированный под конкретную поисковую систему) с веб-сервера отправляется в поисковые системы электронных ресурсов (по протоколу HTTP).

На стороне электронных информационных ресурсов происходит обработка запроса.

Метапоисковая система получает html-коды страниц с результатами поиска, в которых содержатся метаописания найденных документов.

На веб-сервере происходит разбор html-кода полученных страниц для выделения нужной информации.

Обработанная информация предоставляется по мере поступления результатов пользователю с разбивкой по электронным ресурсам.

Разработанная структура модулей представлена на рисунке 2.

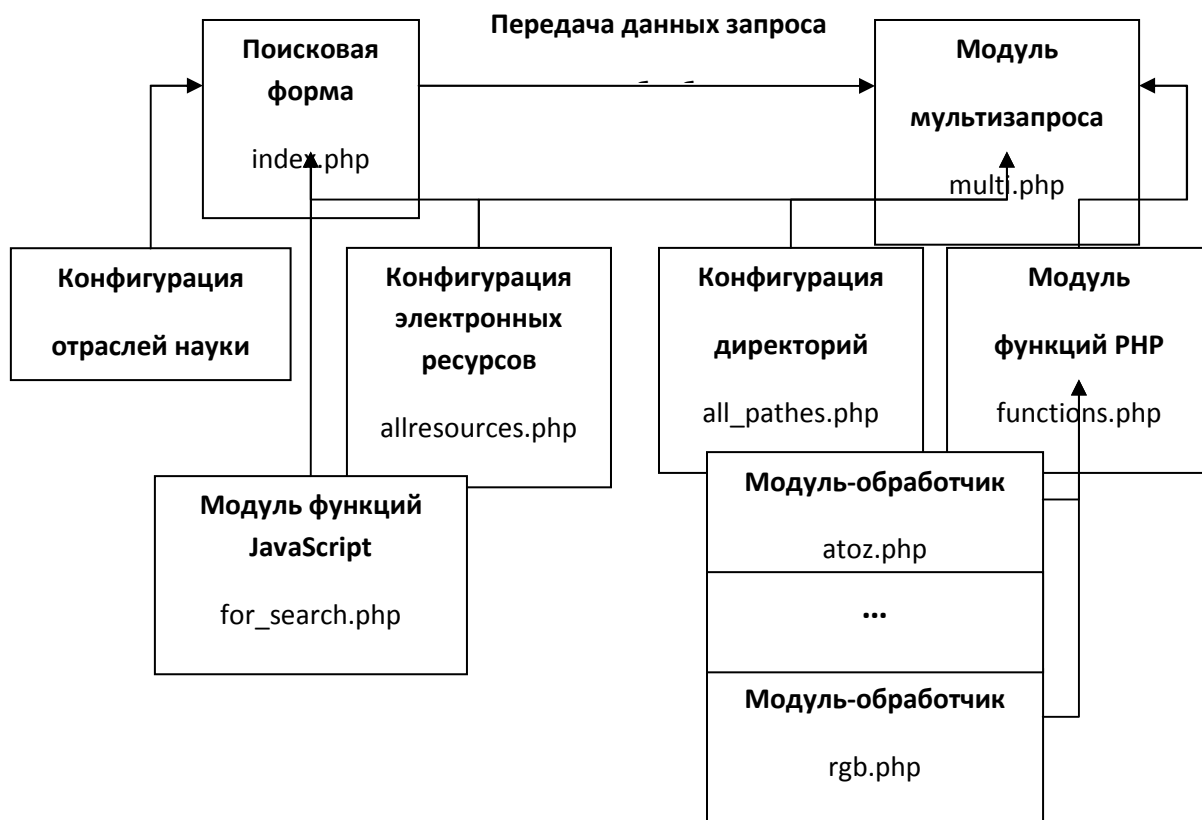


Рисунок 2 - Разработанная структура модулей метапоисковой системы

Метапоисковая система состоит из:

Модуль index.php. Здесь пользователь вводит свой запрос. Данная страница представлена на рисунке 3 и рисунке 4.

The screenshot shows the search interface. At the top, there is a search bar with the text 'естественные науки' (natural sciences) and a dropdown arrow. Below it is a link for 'Расширенный поиск' (Advanced search). The form includes input fields for 'Ключевые слова' (Keywords), 'Заглавие' (Title), and 'Автор' (Author). A section for 'Ресурсы' (Resources) contains a list of checked items: 'Выбрать все ресурсы' (Select all resources), 'Электронный каталог НЕ СФУ: литература по естественным и гуманитарным наукам' (Electronic catalog of NISFU: literature on natural and humanities), 'Архив электронных ресурсов СФУ' (Archive of electronic resources of SFU), 'Видео СФУ' (Video SFU), 'ЭБ ИздДома /Гребенников/' (EB Publishing House /Grebennikov/), 'ЭБД РГБ (БД диссертаций)' (EBD RAS (DB of dissertations)), 'A-to-Z (каталог журналов)' (A-to-Z (journal catalog)), 'American Chemical Society', and 'American Institute of Physics'. At the bottom, there are buttons for 'Поиск' (Search) and 'Очистить форму' (Clear form).

Рисунок 3 - Главная страница метапоисковой системы (поля для простого поиска)

Поиск

[Расширенный поиск](#)

естественные науки

Ключевые слова

Заглавие

Автор

Ресурсы

- Выбрать все ресурсы
- Электронный каталог НБ СФУ: литература по естественным и гуманитарным наукам
- Архив электронных ресурсов СФУ
- Видео СФУ
- ЭБ ИздДома Гребенников/
- ЭБД РГБ (БД диссертаций)
- А-то-З (каталог журналов)
- American Chemical Society
- American Institute of Physics

Год издания

Год издания с

Тематика

Поиск Очистить форму

Рисунок 4 - Главная страница метапоисковой системы (поля для расширенного поиска)

Модуль `multi.php`. Данный модуль организует мультизапрос по полученным данным запроса из модуля `index.php` и предоставляет полученные результаты пользователю.

Модуль `field_conf.php`. Модуль содержит список отраслей науки и их идентификаторы.

Модуль `allresources.php`. Модуль содержит данные об электронных информационных ресурсах в виде ассоциативного массива.

Модуль `all_pathes.php`. Модуль содержит пути ко всем необходимым файлам.

Модуль `functions.php`. В модуле описаны необходимые функции PHP.

Модули – обработчики. Модули обрабатывают html-коды полученных страниц с результатами поиска.

Модуль `for_search.js`. Здесь содержатся необходимые функции JavaScript.

Разработанная система достаточно легко расширяется в случае подключения доступа к новым онлайн базам данных. Необходимо лишь внести необходимую информацию об электронном информационном библиотечном ресурсе в ассоциативный массив в модуль `allresources.php` и добавить модуль-обработчик html-кода.

Время ожидания пользователем результатов поиска зависит от скорости интернет-канала к выбранным ресурсам. Так, время ожидания результатов поиска от российских электронных информационных ресурсов, как правило, меньше, чем от зарубежных ресурсов.

Данная система поддерживает поиск документов на разных языках, при вводе поискового запроса на нужном языке; предоставляет возможность пользователю перейти на страницу электронного информационного ресурса с результатами поиска; поддерживает поиск только по выбранным пользователем электронным информационным ресурсам.

Разработанная метапоисковая система не поддерживает русскую морфологию, дополнительные технологии поиска, такие как кавычки, в которые заключаются фразы, или булевы операторы, так как некоторые поисковые системы электронных информационных ресурсов не поддерживают данные технологии. Так же эта система не поддерживает аутентификацию по логину/паролю, если какие-то электронные библиотеки (электронные информационные ресурсы) требуют регистрации.

По окончании разработки системы планируется ее внедрение на сайт Сибирского федерального университета (по адресу `libsearch.sfu-kras.ru`).