

ОТЧЕТ

По предмету: «Программирование на PHP»

На тему: «Search Engine»

Выполнил: студент 4 курса, спец. ВТиПО - 442 Баярысов Жанали

Проверил: доцент, к.э.н Астаубаева Г. Н.

Алматы, 2019 г.

**Содержание**

Введение …………………………………………………………….

База Данных Проекта………………………………………………

Методы Объектно-Ориентированного Программирования …….

Заключение ………………………………………………………

Список литературы ……………………………………………

**Введение**

**О проекте:**

Цель проекта:

Целью проекта является разработка поисковика Zhanali Search Engine с использованием HTML, CSS, JavaScript, jQuery и PHP.

Задачи проекта:

Задачей проекта является создать поисковую систему со следующим функционалом:

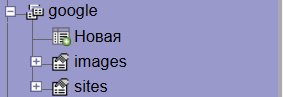
1. Поиск сайтов по ключевым словам;
2. Поиск изображений по ключевым словам;
3. Внедрение системы пагинации (pagination system);
4. Предварительный просмотр изображения при нажатии на него;
5. Обновление базы данных сайтов и изображений.

Актуальность:

Search Engine или поисковая система является необходимым средством для работы, автоматизации систем, а также продвижения бизнеса. В связи с тем, что популярные поисковые системы, вроде Google, Bing и Яндекс начали подозреваться в использовании специальных алгоритмов, которые скрывают некоторые результаты поиска, что вызвало волну недоверия к этим компаниям. Предоставление пользователям новой поисковой системы является отличной возможностью получить определенную долю этого огромного рынка.

**База Данных Проекта**

Как правило, в качестве хранилища данных используются базы данных. PHP позволяет использовать различные системы управления базами данных, но наиболее популярной на сегодняшний день в связке с PHP является MySQL. MySQL представляет бесплатное программное обеспечение, позволяющее взаимодействовать с базами данных с помощью команд языка SQL [1]

База Данных состоит из двух таблиц. В таблице Images хранятся ссылка на сайт, ссылка на картинку, описание картинки, название картинки, количество щелчков на картинку и «испорченная» ли ссылка на картинку. В Таблице sites хранится ссылка на сайт, оглавление сайта, описание сайта, ключевые слова и количество щелков мыши по ссылке. 



В таблице sites содержатся такие столбцы как: “id”, “url”, “title”, “description”, “keywords”, “clicks”. В нём хранятся ссылки на сайт, оглавление сайта, описание сайта, ключевые слова и количество щелчков мыши по ссылке для определения релевантности веб-сайта, которая поможет сервису выводить часто посещаемые пользователями сайты на первой странице в целях достижения удобства.

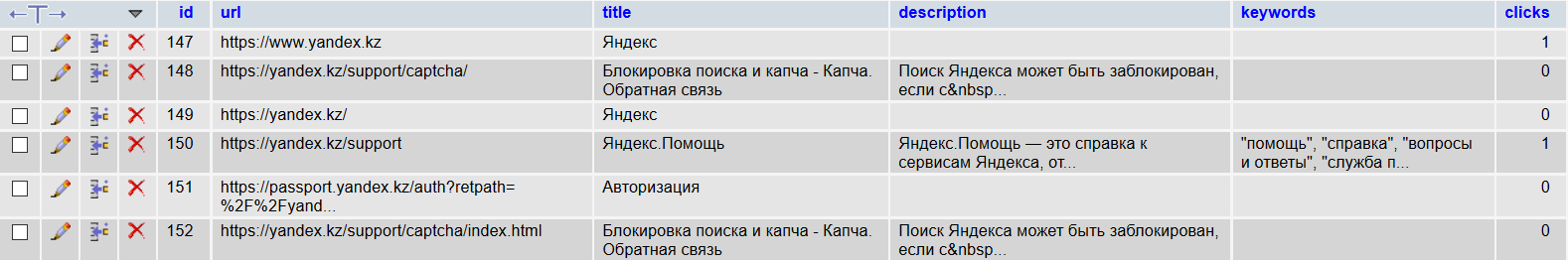
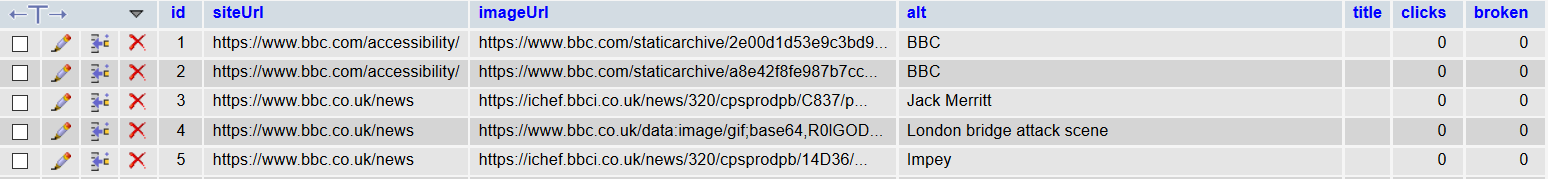


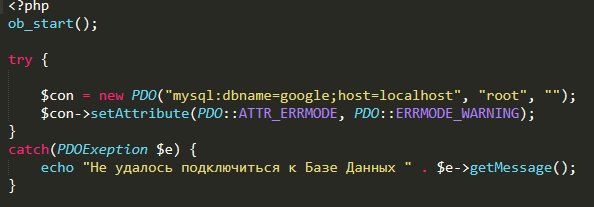
Таблица images содержит следующие столбцы: “id”; “siteUrl” - ссылка на сайт; “imageUrl” - ссылка на картинку; “alt”, в котором записано описание картинки; “title”; “clicks”, определяющее количество щелчков по картинке; “broken”, для проверки не является ли ссылка повреждённой.



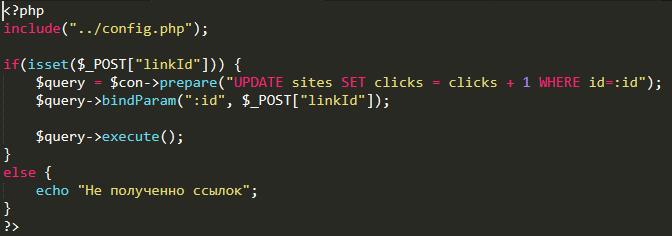
Выполнение запросов:

За выполнение запросов отвечают функции mysqli\_query(), mysqli\_real\_query() и mysqli\_multi\_query(). Чаще всего применяется функция mysqli\_query(), так как она выполняет сразу две задачи: выполняет запрос и буферизует на клиенте результат этого запроса (если он есть). Вызов mysqli\_query() идентичен последовательному вызову функций mysqli\_real\_query() и mysqli\_store\_result().

Ниже приведенный код назначает конфигурацию, то есть определяет где находится база данных, авторизуется в систему как "root" и в случае появления исключения выводит сообщение об ошибке, а также назначает дополнительные атрибуты.



Теперь, когда у нас есть база данных, у нас появляется необходимость с этой базой выполнять различные манипуляций. Например:



Каждый раз, когда пользователь переходит по ссылке или открывает картинку, то в базе необходимо обновить значение clicks, для того чтобы в следующий раз этот результат выводился первым на экране. Происходит это, потому что программа сортирует результаты по количеству переходов по ссылке или нажатий на картинку.

**Объектно-Ориентированное Программирование**

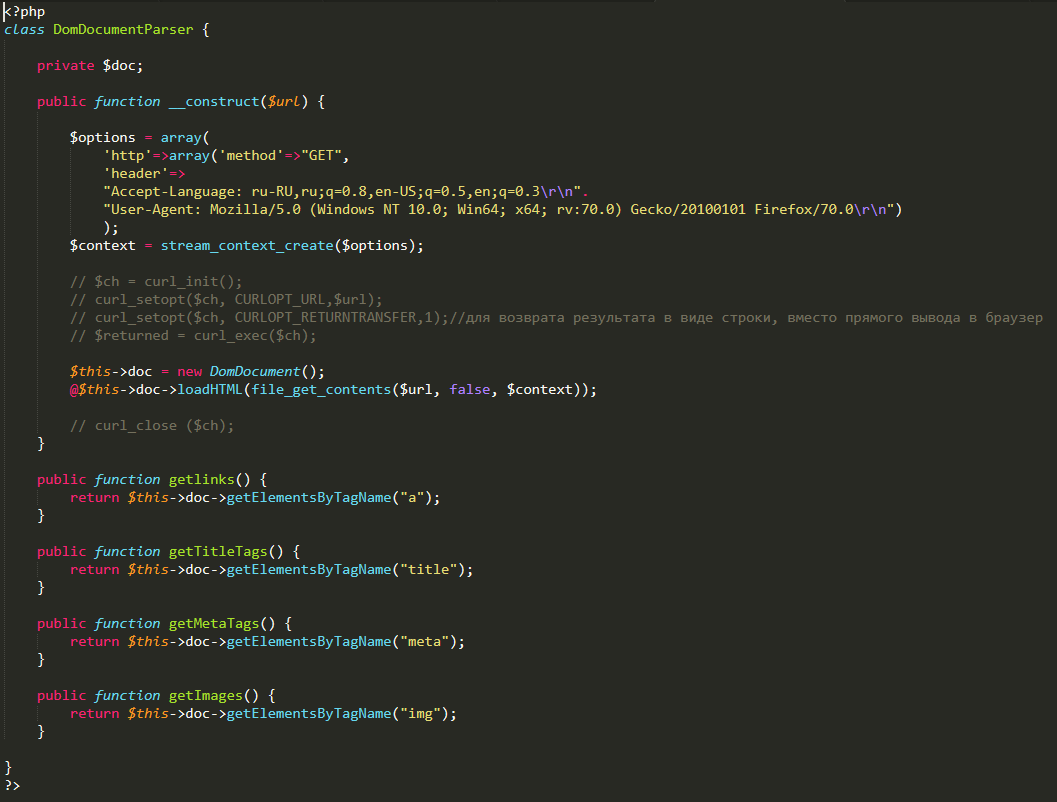
Для удобства и большей читабельности кода, были использованы методы Объектно-Ориентированного Программирования. Все часто используемые куски кода были перенесены в отдельные файлы.

Парсер

**Парсер** - это программное обеспечение для сбора данных и преобразования их в структурированный формат, чаще всего работа с текстовым типом информации. Предположим, вам нужно разместить тысячу карточек товаров в вашем интернет-магазине. Это занятие не на один вечер. Вам нужно собрать много информации, обработать, переписать и заполнить карточки.

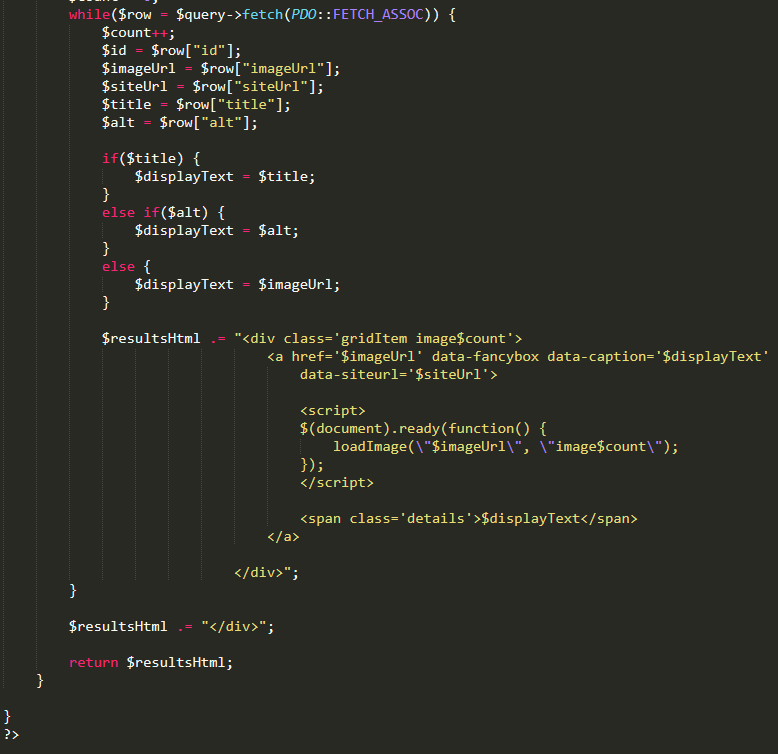
Для решения подобных задач, появились программы, сканирующие наполнение веб-страниц на просторах сети Интернет, результатов выдачи поисковых систем, и копирующие однотипные (текст или картинки) либо универсальные (текст и картинки) сведения. Они позволяют распознавать огромные объемы непрерывно обновляемых значений [2].

* **DomDocumentParser.php** – отвечает за подключение к сайту и скачивания его HTML-кода.



* **ImageResultsProvider.php** – отвечает за запрос и вывод всех картинок из Базы Данных.



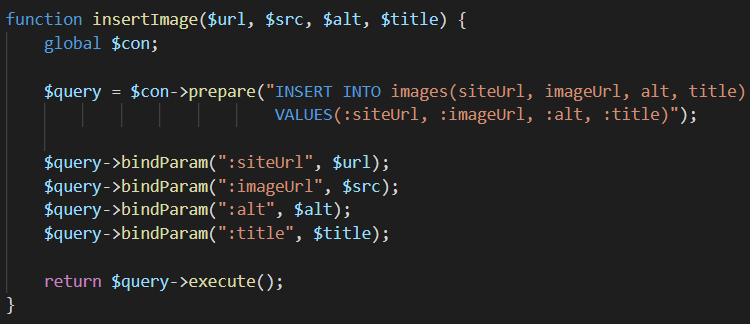


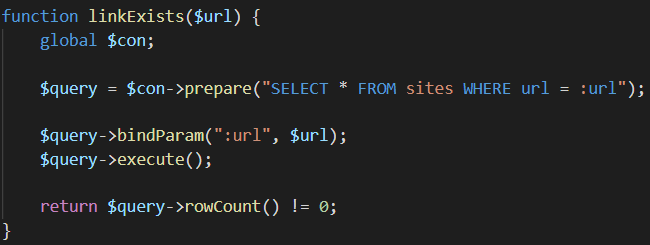
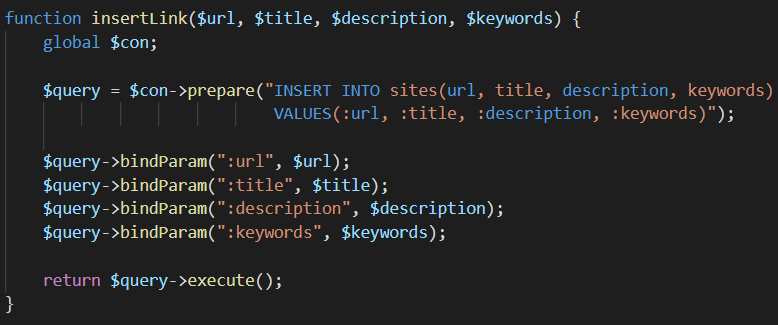
* **SiteResultsProvider.php** – отвечает за запрос и вывод всех сайтов из Базы Данных.





* **Crawl.php -** отвечает за проверку существует ли уже данная ссылка в БД, вставляет изображения и сайты, создаёт ссылки, получает из БД подробную информацию об изображениях и сайтах.



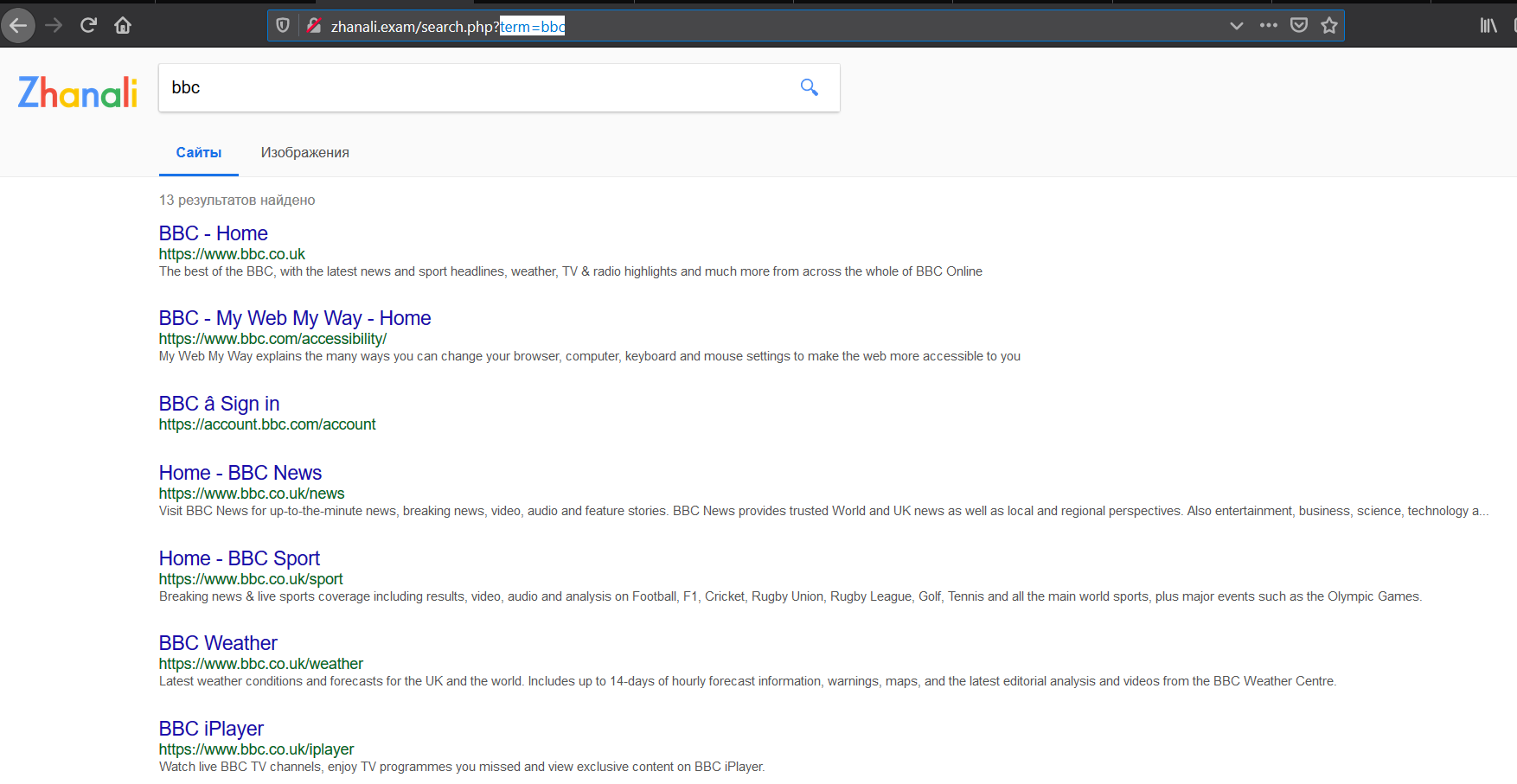


Pagination system

Пагина́ция (от лат. pagina — страница) — многозначный термин, касающийся страниц. В веб-дизайне и дизайне программного обеспечения под *пагинацией* понимают **постраничный вывод** информации, то есть показ ограниченной части информации на одной (веб)-странице (например, 10 результатов поиска или 20 форумных тредов). Она повсеместно используется в веб-приложениях для разбиения большого массива данных на странице и включает в себя навигационный блок для перехода на другие страницы — пагинатор[3].



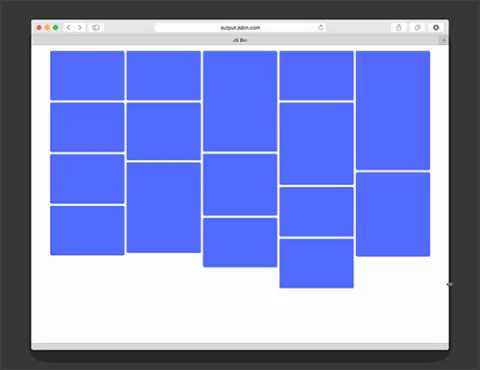
Количество одновременно показываемых сайтов – 20, а изображений 30. Максимальное количество страниц, которые одновременно показывает поисковик – 10.

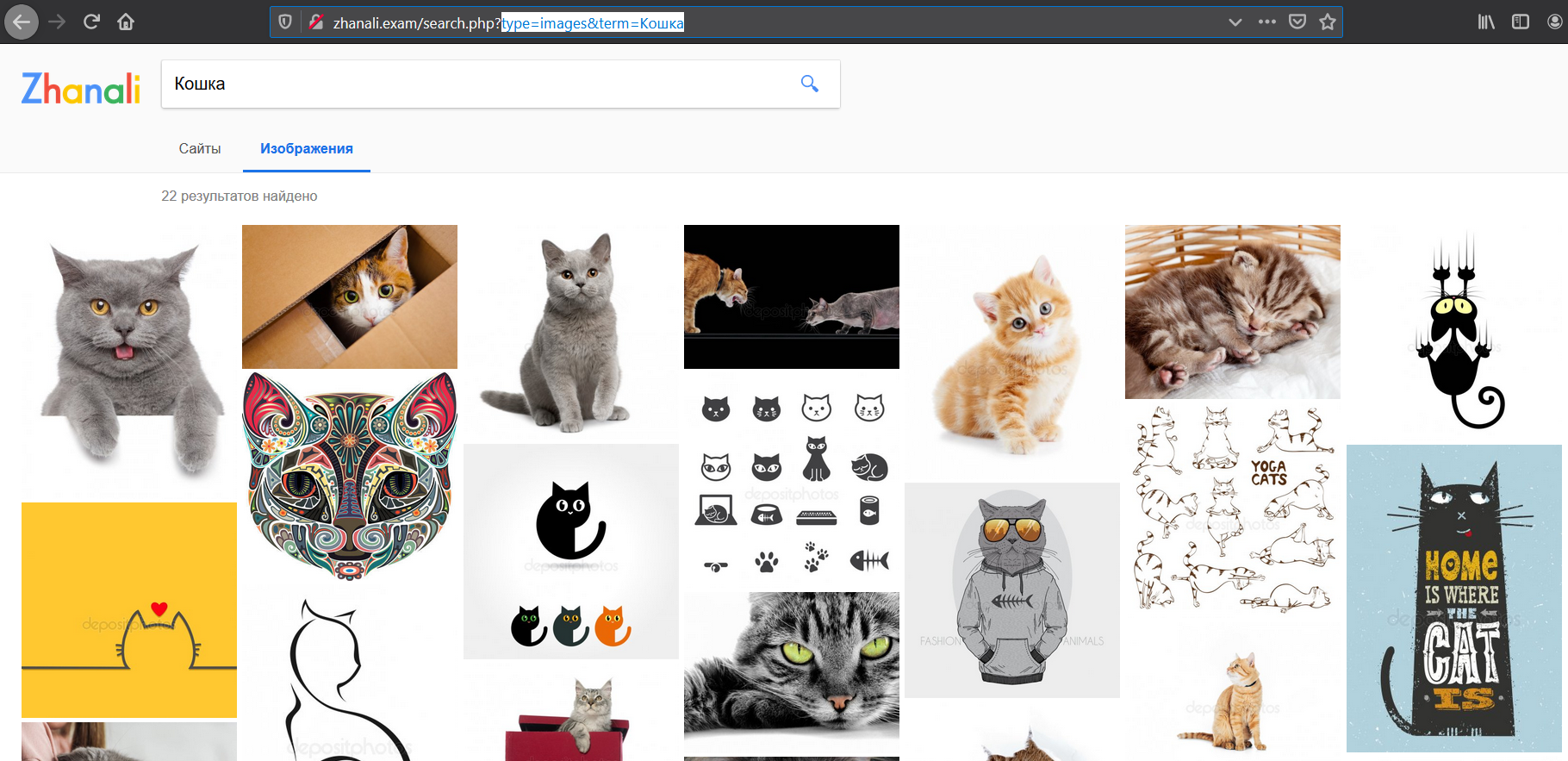


Masonry

Masonry (разработчик David DeSandro) это JQuery-плагин позволяющий быстро и просто организовать динамический layout блоков разного размера практически без потери места, а если подобрать соответствующие размеры блокам — то без пустых мест вовсе.

Под динамическим layout`ом имеется ввиду, что блоки будут располагаться в контейнере в зависимости от его размеров, максимально рационально заполняя его пространство, тем самым экономя место на странице.





**Заключение**

Известно, что интернет пространство сегодня представляет собой источник, содержащий большое количество разнообразной информации. Чтобы ориентироваться в нем и находить нужный объект, необходимы поисковые программы.

В основе каждой поисковой системы стержнем, выполняющим роль навигатора, является машина поиска, которая представляет собой различные программные и аппаратные средства для поиска информации в сети.

Для обеспечения работы нашей поисковой машины мы использовали PHP в связке с MySQL.

В результате проделанной работы, поисковик Zhanali Search Engine решает такие задачи как, поиск сайтов, изображений по ключевым словам; предварительный просмотр изображений, обновление базы данных сайтов и изображений, внедрение системы пагинации.

Список литературы

1.Электронный ресурс. Режим доступа: https://www.php.net/

2.Semantika [Электронный ресурс] Режим доступа: https://semantica.in/blog/chto-takoe-parser.html

3. Электронный ресурс. Режим доступа: https://ru.wikipedia.org/wiki/%D0%9F%D0%B0%D0%B3%D0%B8%D0%BD%D0%B0%D1%86%D0%B8%D1%8F

4. Электронный ресурс. Режим доступа: https://getbootstrap.com/

5. Электронный ресурс. Режим доступа:https://code.tutsplus.com/tutorials/how-to-paginate-data-with-php–net-2928

6. Электронный ресурс. Режим доступа: https://www.myprogrammingtutorials.com/create-pagination-with-php-and-mysql.html

7. Электронный ресурс. Режим доступа: https://habr.com/en/post/322584/

8. Электронный ресурс. Режим доступа: https://canva.com

9. Электронный ресурс. Режим доступа: https://www.festisite.com/logo/google/