МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Физико-технический факультет**

**Кафедра теоретической физики и компьютерных технологий**

**КУРСОВОЙ ПРОЕКТ**

**ИССЛЕДОВАНИЕ И СОЗДАНИЕ ЗЕРКАЛА САЙТА**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Гайдар Николай Александрович

Курс 2

Направление 09.03.02 Информационные системы и технологии

Научный руководитель

канд. физ.-мат. наук, преподаватель\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_О.М. Жаркова

Нормоконтролер инженер \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Г. Д. Цой

Краснодар 2017

|  |  |
| --- | --- |
| **СОДЕРЖАНИЕ** |  |
| Введение . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  | 3 |
| 1 Алгоритм создания зеркала сайта . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  | 5 |
| 1.1 Понятие зеркала сайта . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  | 5 |
| 1.1.1 Определение главного зеркала для Яндекс . . . . . . . . . . . . .  | 7 |
| 1.1.2 Определение главного зеркала для Google . . . . . . . . . . . . . .  | 7 |
| 2 Способы создания зеркал сайта . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  | 9 |
|  2.1 Программы для создания зеркала сайта . . . . . . . . . . . . . . . . . . . . . . .  | 9 |
|  2.2 Проблемы, связанные с созданием зеркал . . . . . . . . . . . . . . . . . . . . .  |  36 |
| Заключение . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  |  38 |
| Список использованных источников . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  |  41 |

Введение

Актуальность выбранной для исследования проблемы может быть обозначена следующим образом. Было ли у вас такое пытается зайти на сайт, а у вас никак не получается. Пробовали через анонимайзер, он позволяет обходить блокировки - не получается, перезагружали компьютер, ну мало ли, всё может быть - тоже никак. Жаль, что у этого сайта не оказалось зеркала.

Зеркало - это точная копия данных одного сервера на другом. В интернете зеркалом сайта называют точную копию другого сайта. Наиболее часто зеркала используется для предоставления нескольких источников одной и той же информации.

Для чего же это делается. Во-первых, это отличный способ подстраховаться, если с сайтом что-то случится, его собираются закрыть или изменить доменное имя, а так как процесс этот не быстрый пользователей автоматически переправляют на другую страничку с почти идентичным содержанием информация, которая нам нужна.

Во-вторых, зеркалирование применяется если доступ к сайту блокирован или закрыт. Зачастую, это происходит с информационными и оппозиционными сайтами. Как, к примеру, произошло в Таджикистане, когда был блокирован доступ к трём ресурсам: tjknews.com и сайты информагентств tojnews.tj и avesta.tj. Пользователи последнего могли попасть на зеркало портала, так как его сервер находится в Скандинавии, поэтому физически закрыть avesta.tj невозможно. Сайт tojnews.tj был полностью выключен. По словам пользователей, сервер этого сайта размещается в Таджикистане и, как результат, был просто обесточен.

Также зеркальные сайты приходят на помощь, когда нужно скачать много больших по размеру файлом, выложенные на файлообменник.

Доменные зеркала тоже часто называют зеркалами. В основном их используют на сайтах с одинаковым звучанием и написанием для того чтобы пользователи не путались. Например, vkontakte.ru и vk.com – обе эти ссылки ведут на один и тот же сайт.

Экономия и скорость, куда удобнее пользоваться внутренним, а не внешним трафиком сейчас таких зеркал сотни. Попробуйте в адресе сайта, которым вы часто пользуетесь, поменять доменную зону, например, KZ на RU и наоборот. Ну и напоследок, вместо спасительного зеркала вам помогут подсунуть фейковую страницу идентичного внешнего вида и содержания, но создают её мошенники, чтобы выкрасть ваши пароли или выдать себя за другого человека, так что будьте бдительны.

Объект исследования – зеркало сайта.

Предмет исследования – свойства и особенности создания зеркала сайта.

Целью работы зеркало сайта и его создание.

Задачи исследования:

* рассмотреть понятие зеркала сайта;
* выявить плюсы и минусы зеркал сайтов.
* проанализировать процесс создания зеркала сайта.
* выявить проблемы во время создания зеркала сайта и предложить меры по устранению выявленных недостатков.

Последовательное решение задач курсовой работы обусловливает ее структуру, которая включает введение, две главы, разделенные на параграфы, заключение, а также список использованных источников и литературы.

**1 Алгоритм создания зеркала сайта**

**1.1 Понятие зеркала сайта и зачем оно нужно**

Эта тема посвящена такому важному моменту, как определение главного зеркала сайта. Давайте определимся какие сайты называются зеркалами.

Главное зеркало – основная версия сайта. Зеркалом именуют полную или практически полную копию сайта, размещённую по другому интернет-адресу. Она повторяет контент на 80% и выше. Для поисковых роботов, наличие приставки www перед адресом делает сайт совершенно другим, то есть www.SITE.ru и SITE.ru – понимаются, как физически отличающиеся адреса.

Также выделяют такое понятие, как доменные зеркала, которые тоже называют просто «зеркало». Если звучание при написании таких имен похоже, то нередко создаются такие типы зеркал. Например, есть сайт, который называется vk.com. A ещё есть ресурс, называемый vkontakte.com. Это тот же самый vk.com, но написание несколько другое. Зеркала являются очень важной частью интернета и польза в них просто неоценима.

Яндекс и Google индексируют сайты по-разному: первый может считать главным сайтом www.сайт.ru, а второй – сайт.ru. Или наоборот. Именно поэтому важно настроить их на главный и зеркальный сайты.

Размещение зеркала сайта может преследовать различные цели, например, в большинстве случаев зеркала сайтов создаются для предоставления интернет пользователям одной и той же информации на разных сайтах, соответственно и на разных серверах, чтобы при возникновении неполадок на одном из них, другой продолжал функционировать и предоставлять информацию с данного ресурса. Создаётся своего рода резерв. При возникновении технических неполадок на [сервере](http://statiami.com/hosting-eto.php), пользователь будет автоматически перенаправлен на зеркало этого ресурса, если таковое существует. Также часто зеркалами пользуются крупные ресурсы, которые содержат различные, пользующиеся спросом у посетителей, файлы, для увеличения скорости загрузки и распределения нагрузки между серверами.

Из всех причин создания зеркал сайта можно выделить несколько основных:

- защита от повреждения информации (чаще всего жёстких дисков), резервирование данных;

- создание копии сайта на время технических работ на основном сервере;

- доступ через альтернативный путь к заблокированной и недоступной по каким-либо причинам, информации;

- снижение нагрузки на главный сайт;

- если внутренний трафик в вашей сети значительно дешевле, чем внешний, то было бы целесообразнее создать зеркало в собственной Интернет зоне.

Поскольку зеркала – это копии сайтов, то поисковые системы для того, чтобы снизить нагрузки на серверы, не индексируют их содержание и выводят в результатах ссылки только на основной ресурс. Эта мера позволяет избежать и выдачи идентичных страниц.

Те алгоритмы, которые выявляют дубликаты этого рода, постоянно меняются. Нередко это порождает ситуации, в которых в качестве главного сайта поисковая система определяется зеркало и игнорирует основной ресурс. Это негативно влияет и на ссылочные показатели, и на индекс цитирования.

Например робот системы [Яндекс](http://www.seowizard.ru/publications/article/yandeks/) в случае обнаружения зеркал выбирает из всех них только один сайт, которому присваивает статус основного. Остальные сайты робот просто удаляет индекса. Если не контролировать этот процесс, то система самостоятельно выберет то зеркало, которое ей понравилось, а это может быть вовсе не запланированный владельцем сайта ресурс.

Для того, чтобы избежать таких случаев, веб-мастера нередко запрещают доступ робота к зеркалам. Это можно сделать до индексации при помощи файла [robots.txt](http://www.seowizard.ru/publications/article/robotstxt/). Кроме того, существует способов указать поисковых системам на тот сайт, который они должны считать основным. Для Яндекса можно указать в robots.txt директиву Host, содержащую имя основного зеркала.

Итак, существуют специальный функционал в панелях для веб-мастеров у Яндекса и Google для определения главного зеркала сайта.

* + 1. **Определение главного зеркала для Яндекса**

Для этого нам потребуется внести изменения в файл robots.txt Если нам необходимо, чтобы главным зеркалом Яндекс считал адрес вида **www.site.ru**, то прописываем следующее:

User-agent: Yandex
Disallow:
Host: www.site.ru

и соответственно если мы хотим основное зеркало сайта вида **site.ru**

User-agent: Yandex
Disallow:
Host: site.ru

**1.1.2 Определение главного зеркала для Google**

Для этого необходимо зарегистрировать сайт в сервисе Google, под названием «[Инструмент для веб-мастеров](http://www.google.com/webmasters/tools)».

После этого нужно подтвердить права на управление сайтом.

Затем переходим по ссылке «Настройки» — «Основной домен», ставим галочку, на то, что нам необходимо, и сохраняем, после этого склейка может растянуться на 2-3 месяца, а может склеится и в течение недели.

В заключении, хочется сказать, что с помощью зеркал можно решить следующие задачи: повысить надежность сайта, увеличить его посещаемость и ускорить раскрутку, уменьшить нагрузку на основной сайт. Но также возникают проблемы во время зеркалирования, а именно, этот процесс занимает длительное время (особенно для Яндекса) и большая вероятность попадания под фильтр.

Плюсы: повышается надежность сайта, увеличивается посещаемость, уменьшается нагрузка на сайт.

Минусы:процесс зеркалирования занимает длительное время (особенно для Яндекса), большая вероятность попадания под фильтр.

**2 Способы создания зеркал сайта**

Любая поисковая система расценивает каждое зеркало исключительно как дубликат сайта. Индексацию таких ресурсов она проводит следующим образом: «склеивает» зеркала вместе и допускает к участию в поиске только один из ресурсов. Он и становится основным зеркалом. Здесь при продвижении сайта важно проследить за тем, чтобы был выбран именно нужный сайт, потому что в противном случае поисковый робот все выполнит на основании своего алгоритма. Веб-мастеру нужно сообщить поисковой системе адрес нужного ресурса, который будет выбран в качестве основного.

Сделать это можно одним из следующих способов:

- с использованием директивы Host;

- с использованием инструмента «Главное зеркало», предоставляемого сервисом Яндекс Вебмастер;

- с использованием серверного редиректа;

- с использованием файла robots.txt.

**2.1 Программы для создания зеркала сайта**

Для процесса зеркалирования сейчас создано большое количество специальных программ для скачивания сайтов, чтобы можно было просматривать его локально на своем компьютере без выхода в интернет.

Веб-разработчикам интересно скачивания целого файла, чтобы посмотреть как реализована верстка сайта, как оптимизированы изображения и многое другое.

Наиболее известны такие программы как:

- Offline Explorer и Offline Explorer Pro

- Teleport Pro

- Wget

- HTTrack Website Copier

А теперь подробнее о каждой программе.

[Offline Explorer](http://www.metaproducts.com/mp/mpProducts_Detail.asp?id=1) — один из лучших офлайн-браузеров. Удобен он как для начинающих, так и для профессионалов, как для частного, так и для корпоративного пользователя. Offline Explorer выпускается в разных версиях: для частного пользователя — Offline Explorer, для профессионального пользователя — Offline Explorer Pro и для корпоративного пользователя — Offline Explorer Enterprise.

Для частного пользователя интерес представляют классические функции офлайн-браузеров: загрузка сайтов в режиме максимальной загрузки канала и просмотр сайтов в офлайн-режиме. Механизм загрузки, используемый компанией MetaProducts, вероятно, один из лучших среди созданных механизмов. Но интерес представляет не только он, но и сервис вокруг этого механизма, который позволяет сделать работу даже неопытного пользователя комфортной и быстрой.

В окне Offline Explorer имеются девять основных разделов, но по функциональности всех их можно разбить на четыре условные группы (рисунок 1).

Проводник проектов. Проект — это самостоятельная задача загрузки. В проводнике мы будем создавать эти проекты, а также удалять, изменять, перемещать или копировать, и даже вставлять из буфера обмена. Проекты можно удобно рассортировать по тематическим папкам и каждой папке задать свой каталог для хранения на диске. Проводник имеет три закладки, на первой из которых отображается список проектов, на второй — структура сохраняемого сайта, а на третьей — очередь файлов, ожидающих загрузки.

Статус загрузки. В нижней части экрана в реальном времени можно наблюдать за процессом загрузки файлов — выводятся их размер, процент загрузки, адрес и другие параметры.

Окно просмотра. Служит для отображения загруженных сайтов во внутреннем браузере, который основан на Internet Explorer. Забегая вперед, отметим, что страницы для просмотра можно открыть и во внешнем браузере.

Управление. Меню программы и дублирующие его панели управления, адресная строка и кнопка быстрой загрузки — все это служит для настройки и управления всеми процессами офлайнового браузинга.

|  |
| --- |
|  |
|  |
| Рисунок 1 - Главное окно Offline Explorer |

Создание проекта. Проект для загрузки легче всего создать с помощью встроенного мастера, который позволяет пошагово задать все параметры проекта, причем выбор каждого параметра доходчиво комментируется прямо в диалоговых окнах. Процесс создания нового проекта упрощает использование шаблонов для загрузки. Достаточно сделать несколько таких шаблонов с типовыми установками, и можно моментально создавать практически любой проект, слегка изменяя его свойства. Несколько готовых шаблонов имеются в составе Offline Explorer. Добавление проекта в шаблон по умолчанию возможно двумя способами: вводом URL в поле адреса проекта или перетаскиванием ссылки из браузера в дерево проектов Offline Explorer.

Параметры проекта. При настройке проекта можно редактировать следующие его свойства (рисунок 2).

Проект. Название, URL и глубина загрузки. Параметр "Уровень" определяет глубину закачки (количество ссылок от первой страницы до самой последней). Параметр "Не загружать существующие файлы" позволяет при последующем обновлении загруженного проекта скачивать только отсутствующие файлы.

Фильтры файлов. Фильтрация по типам файлов. Для каждого типа файлов следует выбрать "Загружать только с начального сервера" и ограничить размеры файлов приемлемыми для вас цифрами, скажем 300 Кб.

|  |
| --- |
|  |
|  |
| Рисунок 2 - Окно свойств проекта |

Фильтры URL. Настройки этого блока применяются к типам файлов, для которых в поле "Расположение" включено "Загружать согласно фильтрам URL". Все четыре фильтра работают как единое целое и позволяют исключить загрузку проекта по протоколу, имени сервера, пути и даже имени файла. Каждый из фильтров имеет три режима: глобальный — позволяющий загружать все; начальный — разрешающий загружать только начальный протокол, файлы с начального сервера или домена, с начального каталога и ниже и с начальным именем; режим произвольной конфигурации — включающий или исключающий загрузку файлов по ключевым словам для фильтрации URL.

Дополнительно. В этом блоке можно включить опцию сохранения старых файлов, и Offline Explorer будет делать их копии (до 10 тысяч) вместо перезаписи. Здесь же устанавливается расписание для загрузки проектов и выбирается режим перевода ссылок: "Трансляция для локального просмотра", "Онлайн-трансляция" или "Без изменения ссылок". Оптимальным выбором является второй режим, при котором только ссылки, разрешенные для загрузки, изменяются на локальные.

Загрузка проектов. При загрузке проектов программа позволяет загружать созданные проекты в разных режимах и в любом количестве в зависимости от толщины канала доступа в Интернет. При использовании бесплатного доступа в ночные часы стоит воспользоваться функцией "Управление расписанием". Проекты для загрузки просто перетаскиваются мышью в специальное окно, выстраиваются в той очередности, в которой их хочется загрузить, и для всей этой очереди устанавливается расписание на загрузку.

Offline Explorer может загружать до 500 файлов одновременно, но рекомендуется использовать не более 10-20 одновременных соединений, хотя программа позволяет использовать до 100 соединений. Меньшее их количество может замедлить загрузку проекта, а большее — сделать ее неустойчивой из-за большого времени ожидания ответа сервера (тайм-аута).

Во время загрузки проекта можно изменять скорость, делать паузу и совсем останавливать проект. Последнее нужно делать только в том случае, если вы совсем передумаете загружать какой-то проект. А если намерены возобновить его загрузку позже, например после перезапуска компьютера, лучше воспользоваться уникальной возможностью Offline Explorer — "Приостановить в файл". В этом случае состояние загрузки проекта сохраняется в отдельном файле, а по команде "Восстановить из файла" загрузка моментально продолжается с места останова. (После полной остановки проекта Offline Explorer долго бы сличал загруженное на диск с содержимым сайта.) Кстати, приостановить в файл можно не только отдельный проект, но и целую запущенную очередь.

Просмотр и экспорт проектов. Начинать просматривать сайт автономно можно, не дожидаясь его окончательной загрузки. Статичные проекты, которые не требуется обновлять (например, электронные книги), после загрузки лучше перенести на вечное хранение в отдельный каталог. Более того, для компактности хранения эти проекты лучше экспортировать в ZIP-файлы, MHT-архивы или файлы HTML Help (.CHM). Два последних формата удобны тем, что все экспортируемые сайты сохраняются в единый файл, для просмотра которого уже не требуется Offline Explorer. HTML Help выгоднее тем, что его файл занимает меньше места на диске из-за применения сжатия. MHT-архив не использует сжатия, зато его файлы могут просматриваться без использования MS IE. Еще одна интересная разновидность экспорта в MS IE cache — все экспортируемые сайты становятся доступными в режиме "Автономная работа".

В особенности Offline Explorer Pro входит загрузка "хитрых" сайтов. Далеко не все сайты отдают свои страницы различным программам-качалкам, в том числе и офлайновым браузерам. С Offline Explorer такие сайты скачать можно, если в настройках соединения программы выполнить следующее:

- прикинуться "безобидным" браузером (установить переключатель в положение "Идентифицировать как" и выбрать Microsoft Internet Explorer из списка);

- установить "Количество соединений" равным 1;

- установить "Задержку между загрузками" в несколько секунд, скажем 5;

- отключить загрузку картинок.

"Жадный" сервер будет считать, что это вы вручную по одной странице открываете его страницы в IE, и проект загрузится без проблем. Кстати, если задать случайную задержку между загрузками (ввести в это поле интервал, скажем 3-10), имитация ручного просмотра сайта станет еще более правдоподобной. После этого останется убрать паузы между загрузками, включить загрузку изображений (и, наоборот, отключить загрузку страничек), установить любимое количество соединений и просто обновить проект.

Загрузка сайтов с паролями. Сайты, использующие формы и cookies для идентификации пользователя (пример — www.itknowledge.com), труднодоступны для загрузки офлайновыми браузерами. В Offline Explorer для их получения нужно всего лишь один раз в дополнительных свойствах проекта ввести свое имя и пароль, а в тех же настройках соединения активизировать "Иcпoльзoвaть MS IE cookies" — и все, запароленный сайт можно закачивать обычным образом. Кстати, любую заполненную веб-форму в Offline Explorer можно сохранить как отдельный проект и потом просто ею пользоваться (рисунок 3).

Макрокоманды — отдельная фишка программы. Они очень полезны для планировки автоматических загрузок в случае, когда каждый раз меняется URL проекта. Например, ежедневную картинку с погодного сайта можно получать по ссылке www.weather.com/image{:day}.png, где {:day} текущий день (1-31). Макрокоманд таких множество, и их применение подробно расписано в справке к программе.



Рисунок 3 - Offline Explorer очень удобная система поиск

Достоинства:

- очень быстрый, простой и надёжный возврат файлов из Web (включая FTP, HTTP, HTTPS, MMS, PNM и RTSP-сайты);

- поддержка HTTP, SOCKS4, SOCKS5 и User@Site Proxy. Автоматическое определение настроек прокси. Определение настроек прокси из конфигурации URL;

- загрузка до 500 файлов одновременно. Каждый Проект загружается параллельно с другими. Можно настроить количество соединений для загрузки каждого Проекта и пауз между загрузками Проекта;

- возможность ограничения повторного получения для HTTP, HTTPS, FTP и RTSP загрузок;

- загруженные web-страницы могут быть распечатаны из встроенного браузера. Также возможна печать полностью целого сайта;

- загруженные web-страницы могут быть опубликованы на Web или экспортированы в другой каталог;

- встроенный редактор с подсветкой синтаксиса для HTML, CSS и javascript-файлов;

- поддержка Drag-and-Drop (перетаскивания) ссылок из браузеров;

- отслеживание содержимого Буфера обмена для ручного или автоматического создания новых Проектов;

- полностью настраиваемый пользовательский интерфейс (включая панель инструментов) с удобным Мастером создания Нового Проекта;

- возможность фонового выполнения загрузок;

- обработка javascript для поиска большего количества ссылок;

- единственный офлайн-браузер, полностью поддерживающий все известные XML-технологии, включая XSL и DTD файлы;

- поддержка использования Microsoft Internet Explorer cookies при загрузке;

- многоразовые Шаблоны делают создание нового Проекта простым;

- встроенное Резервное копирование/Восстановление Проектов;

- поиск в загруженных файлах по ключевым словам в именах файлов или в тексте HTML. Найденные ключевые слова подсвечиваются в общем списке результатов поиска и в найденных web-страницах во время их просмотра;

- операции с множеством Проектов быстро выполняются в дереве Проектов;

- встроенный дозвон с возможностями отключения связи и автоматического выключения компьютера по завершении загрузки;

- гибкие Фильтры URL с мощной поддержкой ключевых слов для тонкой настройки Ваших загрузок;

- управление скоростью загрузки;

- загрузка возможна по расписанию, очереди или запущена из командной строки;

Недостатки:

- слишком простые неэффективные фильтры закачивают весь сайт вместе со всей рекламой, баннерами и прочим мусором;

- загрузка более 500 файлов требует платной версии программы.

Teleport Pro. При первом запуске программы Teleport Pro откроется окно Мастера Новых Проектов (New Project Wizard). Если Вы запускаете программу не в первый раз, это окно можно открыть двумя способами:

- в меню File выбрать команду New Project Wizard;

- на панели инструментов нажать кнопку ;

Чтобы создать определенный проект, нужно будет выбирать один из возможных вариантов выбора (рисунок 4):



Рисунок 4 – Первый этап создания копии сайта

- сreate a browsable copy of a website on my hard drive (создать копию сайта на винчестере для просмотра) - все файлы сайта будут записаны в папку проекта, причем вложенных папок там не будет;

- duplicate a website, including directory structure (создать точную копию сайта, включая структуру каталогов) - то же, что и пункт 1, только в папке проекта будут находится подкаталоги, соответствующие каталогам скачиваемого сайта. Для просмотра скачанного сайта можно выбрать любой из этих двух пунктов;

- search a website for files of a certain type (искать на сайте файлы определенного типа) - сайт не скачивается, но скачиваются определенные файлы (графические, звуковые и т.д.), содержащиеся на этом сайте;

- explore every site linked from a central site (исследовать все сайты, связанные с заданным сайтом) - файлы сайта не скачиваются, а создается только их полный список. Работает очень быстро. Полезно использовать, если Вы хотите проверить объем сайта и узнать сколько времени Вам придется потом его скачивать: 15 минут или трое суток;

- retrieve one or more files at known addresses (скачать один или несколько файлов с известными адресами)- сайт не скачивается, скачиваются только файлы с указанными адресами. Эти адреса в случае выбора данного пункта нужно ввести в форму, которая появится на экране после нажатия кнопки "Далее" ("Next");

- search a website for keywords (искать ключевые слова на сайте) - аналогично пункту 1, только скачиваются не все файлы сайта, а только те, которые содержат данные ключевые слова.

После выбора переходим к [следующему шагу](http://telpro.narod.ru/nachalo/hag2.htm) (рисунок 5).



Рисунок 5 – Ввод адреса проекта

Здесь необходимо ввести стартовый адрес проекта (в верхнем поле) и глубину поиска (в нижнем поле).

Что такое глубина поиска можно объяснить на примере. Если введен адрес <http://www.proba.ru>, а глубину поиска 3, то файл http://www.proba.ru/proba1/proba2/index3.html скачиваться будет, а файл http://www.proba.ru/proba1/proba2/proba3/proba4/index5.html - не будет, т.к. глубина файла index5.html равна 5 (больше 3).

После переходим к следующему шагу (рисунок 6).



Рисунок 6 – Выбор типов скачиваемых файлов

Типы скачиваемых файлов:

- только текст;

- текст и графика;

- текст, графика и музыка;

- все файлы.

При необходимости можно указать логин (Account) и пароль (Password), под которыми можно зайти на сайт.

Для перехода к [следующему шагу](http://telpro.narod.ru/nachalo/hag4.htm) нажимаем кнопку "Далее" ("Next") .

Проект создан. Осталось нажать кнопку "Готово" для сохранения файла проекта на винчестере (рисунок 7).



Рискнок 7 – Финал создания копии сайта

Сохранение проекта (рисунок 8).



Рисунок 8 – Сохранение копии сайта на ПК

Открывается стандартное окно Сохранить Как (Save As), в котором нужно выбрать папку для файла проекта (а при необходимости ее создать), придумать имя файлу проекта, которое соответствует тематике сайта и сохранить.

Выполнение проекта. Итак, на данном этапе у нас есть созданный проект в виде файла на диске, но для того, чтобы скачать сайт из Интернет, этот проект нужно выполнить (рисунок 9).



Рисунок 9 – Выполнение проекта

Чтобы выполнить проект, нужно выбрать в меню Project команду Start или нажать кнопку  на панели инструментов.

Во время выполнения проекта Ваш компьютер должен быть обязательно подключен к Интернет, а до начала выполнения подключение не требуется.

В левой части окна проекта Вы увидите карту проекта, в которой будут отображаться названия страниц, где побывал Teleport Pro.

Правая часть представляет собой список файлов, которые Teleport Pro скачал для каждой из страниц в карте проекта.

Сайт можно скачивать в несколько этапов (т.е. не за один раз). Для управления работой проекта во время его выполнения можно воспользоваться четырьмя кнопками на панели инструментов :

- начать или продолжить выполнение проекта ;

- временно приостановить выполнение проекта (использовать не рекомендую, а в крайнем случае можно использовать на очень непродолжительное время)  ;

- остановить выполнение проекта . После нажатия на эту кнопку Teleport Pro докачает те файлы, которые он уже начал скачивать, после чего остановится;

- прервать выполнение проекта  . После нажатия на эту кнопку Teleport Pro сразу закончит выполнение проекта. Все недокачанные файлы в следующий раз придется скачивать заново (не рекомендую).

В случае разрыва соединения после повторного подключения Вы можете продолжить выполнение проекта, нажав на кнопку .

Просмотр сайта в off-line. Оff-line - это режим отключения от Интернет, т.е. после окончания выполнения проекта можно отключиться от Интернет и не спеша просматривать скачанные странички.

Для этого в правой части окна проекта нужно выбрать любой файл формата html (или htm) и открыть его (рисунок 10).



Рисунок 10 - Просмотр сайта в off-line

Главная же страница сайта чаще всего имеет имя index.html или index.htm.

Все скачанные файлы хранятся в папке, которая имеет такое же имя, что и файл проекта и находится в том же каталоге (рисунок 11).



Рисунок 11 – Скаченные файлы

А это значит, что для просмотра скачанного сайта теперь не нужен Teleport Pro. С помощью любой стандартной программы Windows для работы с файлами, например, с помощью "Мой компьютер" или "Проводник" можно найти папку со всеми скачанными файлами, где находится файл  с расширением \*.html (или \*.htm), который можно открыть.

Достоинства:

- многопоточная закачка, способная намного повысить скорость закачки файлов;

- поддержка нескольких режимов скачивания:

 а) полная копия («[зеркало](https://ru.wikipedia.org/wiki/%D0%97%D0%B5%D1%80%D0%BA%D0%B0%D0%BB%D0%BE_%28%D0%B2_%D1%81%D0%B5%D1%82%D0%B8%29)») сайта, с сохранением структуры каталогов;

 б) копия сайта для просмотра офлайн;

 в) скачивание отдельных файлов.

- при скачивании можно задать [маску](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB%D1%8B-%D0%B4%D0%B6%D0%BE%D0%BA%D0%B5%D1%80%D1%8B) файлов, которые нужно исключить либо наоборот включить в список (например, все картинки, или все [архивы](https://ru.wikipedia.org/wiki/%D0%90%D1%80%D1%85%D0%B8%D0%B2_%28%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29));

- выбор глубины прохода по сайту, либо ограничения по доменному имени;

- по умолчанию ссылки между страницами сайта заменяются на локальные ссылки;

- скачивание по расписанию;

- ограничение массового скачивания (для предотвращения [бана](https://ru.wikipedia.org/wiki/%D0%91%D0%B0%D0%BD_%28%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82%29%22%20%5Co%20%22%D0%91%D0%B0%D0%BD%20%28%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82%29) на сайте);

Недостатки:

- не определяет правильно ли скачан файл;

- не умеет обходить защиту сайтов, где требуется авторизация;

- не показывает общий объем файлов на сервере.

Программа WGet. Wget — свободная неинтерактивная консольная программа для загрузки файлов и скачивания сайта целиком по сети. Поддерживает протоколы HTTP, FTP и HTTPS, а также поддерживает работу через HTTP прокси-сервер.

Так как Wget является неинтерактивной программой, это означает, что после её запуска пользователь может повлиять на её работу только с помощью средств управления процессами самой операционной системы.

Чтобы получить подсказку по параметрам WGet наберите команду man wget в Linux или wget.exe --help в Windows.

Допустим, мы хотим создать полную копию сайта www.site.com на своем диске. Для этого открываем командную строку (Wget - утилита консольная) и пишем такую команду:

wget.exe -r -l10 -k -p -E -nc http://www.site.com

WGET рекурсивно (параметр -r) обойдет каталоги и подкаталоги на удалённом сервере включая css-стили(-k) с максимальной глубиной рекурсии равной десяти (-l), а затем заменить в загруженных HTML-документах абсолютные ссылки на относительные (-k) и расширения на html(-E) для последующего локального просмотра скачанного сайта. При повторном скачивании не будут лица и перезаписываться повторы(-nc).

Если предполагается загрузка с сайта какого-либо одного каталога (со всеми вложенными в него папками), то логичнее будет включить в командную строку параметр -np. Он не позволит утилите при поиске файлов подниматься по иерархии каталогов выше указанной директории:

wget.exe -r -l10 -k http://www.site.com -np

По умолчанию всё скаченное сохраняется в рабочей директории утилиты. Определить другое месторасположение копируемых файлов поможет параметр -P: wget.exe -r -l10 -k http://www.site.com -P c:\internet\files.

Наконец, если сетевые настройки вашей сети предполагают использование прокси-сервера, то его настройки необходимо сообщить программе:

wget -m -k -nv -np -p --user-agent="Mozilla/5.0 (compatible; Konqueror/3.0.0/10; Linux)" АДРЕС\_САЙТА.

Загрузка всех URL, указанных в файле FILE: wget -i FILE

Скачивание файла в указанный каталог (-P):

wget -P /path/for/save ftp://ftp.example.org/some\_file.iso.

Использование имени пользователя и пароля на FTP/HTTP (вариант 1):

wget ftp://login:password@ftp.example.org/some\_file.iso.

Использование имени пользователя и пароля на FTP/HTTP (вариант 2):

wget --user=login --password=password ftp://ftp.example.org/some\_file.iso.

Скачивание в фоновом режиме (-b): wget -b ftp://ftp.example.org/some\_file.iso.

Скачать страницу с глубиной следования 10, записывая протокол в файл log:

wget -r -l 10 http://example.org/ -o log.

Скачать содержимое каталога http://example.org/~luzer/my-archive/ и всех его подкаталогов, при этом не поднимаясь по иерархии каталогов выше:

wget -r --no-parent http://example.org/~luzer/my-archive/.

Также поддерживается идентификация на сервере:

wget --save-cookies cookies.txt --post-data 'user=foo&password=bar' http://example.org/auth.php

Скопировать весь сайт целиком: wget -r -l0 -k http://example.org/.

Например, не загружать zip-архивы: wget -r -R «\*.zip» http://freeware.ru.

Достоинства:

- неинтерактивная программа, что позволяет запускать ее в фоновом режиме;

- программа Wget может выкачивать файлы "рекурсивно", т.е. автоматически следовать по "ссылкам". Это позволяет, например, скачивать web-страницы вместе с имеющимися в них картинками, или выкачивать по FTP целые директории вместе с содержимым;

- программа Wget умеет "докачивать" файлы после обрыва соединений, что особо ценно при медленной или нестабильной связи.

Недостатки:

- не имеет интерфейса, поэтому рядовому пользователю, незнакомого с командной строкой, будет сложно разобраться;

- разработка Wget идёт медленно;

- могие новые расширения протоколов HTTP, [FTP](https://ru.wikipedia.org/wiki/FTP), сценарии [JavaScript](https://ru.wikipedia.org/wiki/JavaScript%22%20%5Co%20%22JavaScript) и другие функции не поддерживаются.

HTTrack Website Copier. HTTrack — кроссплатформенный [офлайн-браузер](https://ru.wikipedia.org/wiki/%D0%9E%D1%84%D0%BB%D0%B0%D0%B9%D0%BD-%D0%B1%D1%80%D0%B0%D1%83%D0%B7%D0%B5%D1%80). Позволяет загружать веб-сайты из [Интернета](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82) на локальный компьютер. Поддерживает 29 языков мира.

Чтобы скачивать сайт целиком с помощью HTTrack Website Copier нужно установить программу на свой компьютер, создать новый проект, указать название проекта и в какую папку на компьютере сохранить сайт (рисунок 12).



Рисунок 12 – Создание проекта

После того как перешли «Далее» в поле «Веб адреса: (URL)» укажите адрес сайта, который надо скопировать и нажмите на кнопку «Задать параметры» (рисунок 13).



Рисунок 13 – Указание адреса

В появившемся окне выставить параметры скачивания.

Если у вас нет прокси оставляем все как есть и переходим к следующей вкладке «Фильтры» (рисунок 14).



Рисунок 14 - Настройка проекта в программе HTTrack Website Copier

Здесь мы избавляемся от мусора и говорим какие файлы нам скачивать не нужно (рисунок 15).



Рисунок 15 - Скачиваем только то что нам нужно

Потом переходим на вкладку «Ограничения» и задаем максимальную глубину загрузки сайта в зависимости от ваших потребностей (рисунок 16).



Рисунок 16 – Свойства проекта

Не забудьте поставить ограничение на скачивание страниц с других сайтов.

Во вкладке «Линки» выставляем флажки как на этом скриншоте, чтобы скачивание сайта происходило максимально быстро (рисунок 17).



Рисунок 17 – Ускорение скачивания страницы сайта

На вкладке «Структура» выбираем тип локальной структуры зеркала (рисунок 18).



Рисунок 18 - Сохраняйте сайты с исходной структурой

После нажатия кнопки «ОК» видно такое окно (рисунок 19).

Здесь вы можете разорвать соединение после скачивания сайта, если у вас интерент не безлимитный. А также поставить флажок напротив пункта «Shutdown PC when finished», если хотите, чтобы HTTrack сам выключил компьютер, когда скачает весь сайт.



Рисунок 19 - Проект готов.

Начался процесс скачивания сайта (рисунок 20).



Рисунок 20 - Окно с подробной информацией по данному проекту

Скачанный сайт можно запустить любым способом, например:

- зайти в папку, в которую сохраняли проект, кликнув по файлу index.html. После этого откроется страница со списком всех проектов. Нужно выбрать нужный;

- или войти в папку с названием проекта, кликнув по файлу index.html, чтобы открылся сайт сразу;

Теперь на компьютере находится точно такая же копия сайта, как и в интернете (рисунок 21).



Рисунок 21 – Скаченный сайт

Достоинства:

- программа HTTrack допускает докачку прерванных загрузок, создавать копий каталогов и отдельных веб-страниц, графических и других файлов;

- при необходимости HTTrack может отыскивать "зеркала" сайта;

- программа рекурсивно создает все директории сайта и воссоздает тем самым оригинальную структуру сайта;

- программа автоматически восстанавливает разорванные соединения, поддерживает функцию обновления уже скачанных сайтов, полностью настраивается и содержит интегрированную справочную систему;

- доступно две графические оболочки: виндовый WinHTTrack и браузерный WebHTTrack, также при желании можно настроить неинтерактивный режим.

Недостатки:

- максимальная скорость закачки 250 [Кб/с](https://ru.wikipedia.org/wiki/%D0%A1%D0%BA%D0%BE%D1%80%D0%BE%D1%81%D1%82%D1%8C_%D0%BF%D0%B5%D1%80%D0%B5%D0%B4%D0%B0%D1%87%D0%B8_%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8) (в старых версиях 50 Кб/с);

- невозможно отключить скачивание внешних сайтов, установка параметра в 0 означает "нет ограничений";

- некоторые сайты скачиваются не корректно.

Сайты начинают считаться самостоятельными и расклеиваются, если поисковики находят существенные различия в их содержимом. Объяснятся это тем, что главной характеристикой зеркала поисковые системы признают его идентичность. В настоящее время все поисковики обзавелись алгоритмом выявления копий всех сайтов одного и того же владельца и удаляют их из индекса, расценивая такие ресурсы как аффилированные, то есть, не дающие пользователю принципиально новой полезной информации по отношению к другим сайтам того же владельца. Ранее же зеркала могли использоваться своими владельцами с целью получения как можно большего количества позиций по запросам.

Все перечисленное выше это лишь общие принципы, поясняющие как создать зеркало сайта. Практическое создание зеркал зависит от конкретного ресурса, сервера на котором он находится и то, что им управляет.

**2.2 Проблемы, связанные с созданием зеркал**

Нужно сказать, что создание зеркала сайта сопровождается появлением нескольких проблем, связанных с работой поисковиков. Поскольку зеркало – это фактически дубликат сайта, то многие поисковые системы (например, Яндекс), чтобы избежать дублирования информации, стараются не индексировать зеркала. Если поисковая система сочла сайт зеркалом, то такой сайт не индексируется, а уже проиндексированные страницы могут быть удалены из базы поисковой системы, поскольку поисковик сочтет материалы одного из зеркал плагиатом. Чтобы избежать подобного, необходимо сообщить поисковой системе, по какому адресу находится основной сайт, а какие адреса являются зеркалами.

Еще одной проблемой является также организация обновлений информации на разных зеркалах. С этой точки зрения зеркала сайта различаются:

* по времени обновления информации: сразу же после изменения информации на основном сайте или спустя короткое время;
* по методу обновления;
* выполняющееся со стороны основного сайта (реплицирование);
* выполняющееся со стороны зеркала;
* вручную, по команде администратора;
* по синхронности обновления;
* синхронное, через определенные промежутки времени;
* асинхронное, например, при изменении содержимого веб-страницы;
* по полноте идентичности: полная копия основного сайта или частичная копия основного сайта (например, копия исходного программного кода без графики).

**ЗАКЛЮЧЕНИЕ**

Основные результаты выпускной (курсовой) работы состоят в следующем:

1 Был проанализирован процесс создания зеркал сайтов.

2 Было определенно, что зеркало сайта - точная копия одного сайта на другом. Чаще всего зеркала используются для предоставления пользователям одной и той же информации на разных сайтах, на случай, если один из них по каким-либо причинам не работает. В случае технических неполадок на сервере, сбоев, перегрузок, пользователь автоматически будет переноситься на зеркало сайта. Чаще всего располагают на нескольких зеркалах большие и популярные файлы, распределяя нагрузки на серверы и увеличивая скорость загрузки.

3 Выяснено, что причины для создания зеркал могут быть самыми разными. В подавляющем большинстве случаев зеркала создают для того, чтобы иметь возможность предоставлять пользователям один и тот же [контент](http://allforjoomla.ru/info/75-kontent-na-zakaz) на разных веб-сайтах, которые размещаются на разных серверах соответственно.

Если проанализировать все причины создания зеркал, то можно выделить главные:

- резервирование данных, защита информации от повреждения;

- создание работоспособной копии сайта на случай технических работ на основном сервере;

- создание альтернативного доступа к информации, недоступной по каким-либо причинам;

- распределение нагрузки на несколько серверов с целью увеличения быстродействия сайта.

4 Создание зеркала сайта можно осуществить при помощи проверенных программ, способных копировать структуру и html-код интернет проекта:

- HTTrack – это бесплатная полнофункциональная программа, представляющая собой кроссплатформенный оффлайн-браузер.

- Wget является свободной неинтерактивной консольной программой, предназначенной для загрузки файлов по сети. Повлиять на ее работу веб-мастер может только с помощью средств управления операционной системы своего оборудования.

- Teleport Pro представляет собой программу, предназначенную для синхронизации каталогов и файлов с использованием минимального трафика в двух местах. Она осуществляет зеркалирование в каждом направлении одним потоком.

- Offline Explorer – это программное обеспечение, основной задачей которого является распространение и обновление исходных текстов на удаленном сервере.

5 Самым главным критерием зеркала сайта является его полная идентичность основному ресурсу. Оно также, как и основной сайт, может приносить доход. Если оно имеет хорошую посещаемость, можно размещать баннеры и контекстную рекламу.  При обновлении основного сайта, должно также обновляться и зеркало.

6 Зеркала широко используются для продвижения и сохранения показателей сайта, повышения надежности сайта, уменьшения нагрузка на основной сайт. Главными недостатками зеркалирования является то, что этот процесс может происходить более 1-2 месяцев, а также существует вероятность попадания зеркала или основного сайта под фильтр браузера.

7 Для разрешения указанных проблем можно сформулировать следующие рекомендации:

- в целях повышения популярности сайта необходимо определить и настроить главное и дополнительное зеркало сайта и другие механизмы, повышающие посещаемость сайта;

- для принятия оперативных решений и эффективного использования сайта рекомендуется проверить идентичность главного и дополнительного зыкало сайта так, чтобы более 80% их контента совпадали;

- чаще проверять механизм определения зеркала сайта в тех браузерах, в которых вы продвигаете свой сайт.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1 Седерхольм Д. «Пуленепробиваемый Web-дизайн» / Д. Седерхольм. – Питер 2012. –  304 с.

2 Мэтт З. «PHP. Объекты, шаблоны и методики программирования» / З. Мэтт. – Вильямс 2015. – 600 с.

3 Никсон Р. «Создаем динамические веб-сайты с помощью PHP, MySQL, JavaScript, CSS и HTML5» / Р. Никсон. – Питер 2015. –  676 с.

4 Манн И., Иванов М., Фербер М. «HTML5 для веб-дизайнеров» / И. Манн, М. Иванов, М. Фербер. - Миф 2015. - 345 с.

5 Прохоренок Н., Дронов В. «HTML, JavaScript, PHP и MySQL. Джентльменский набор Web-мастера» / Н. Прохоренок, В. Дронов. – БХВ-Петербург 2015. – 766 с.

6 Никсон Р. «Создаем динамические веб-сайты с помощью PHP, MySQL, JavaScript, CSS и HTML5» / Р. Никсон. – Питер 2016. –497 с.

7 Пьюривал С. «Основы разработки веб-приложений» / С. Пьюривал. – Питер 2015. – 272 с.

8 Сидерхолм Д. «CSS3 для веб-дизайнеров» / Д. Сидерхолм. – Миф 2013. – 130 с.

9 Маккоу А. «Веб-приложения на JavaScript» / А. Маккоу. – Питер 2012. – 288 с.

10 Гарднер Л., Григсби Д. «Разработка веб-сайтов для мобильных устройств» / Л. Гарднер, Д. Григсби. – Питер 2013. – 439 с.

11 Томал Р. «Основы Web-Дизайна. Руководство» / Р. Томал. – RafalTomal.com 2015. – 211 с.

12 Закас Н. «JavaScript для профессиональных веб-разработчиков» / Н. Закас. – Питер 2015. – 960 с.

13) Закас Н. «JavaScript. Оптимизация производительности» / Н. Закас. – Символ-Плюс 2012. – 256 с.

14 Роббинс Д. «HTML5, CSS3 и JavaScript. Исчерпывающее руководство» / Д. Роббинс. – Эксмо 2014. – 516 с.

15 Колисниченко Д. «Разработка Web-приложений» / Д. Колисниченко. – БХВ-Петербург 2015. – 543 с.

16 Гоше Д. «HTML5. Для профессионалов» / Д. Гоше. – Питер 2014. – 496 с.

17 Клименко Р. «Веб-мастеринг на 100%» / Р. Клименко. – Питер 2013. – 512 с.

18 Голованов В. «Выразительный Javascript» / В. Голованов. – Питер 2015. – 425 с.

19 Роббинс Д. «HTML5. Карманный справочник» / Д. Роббинс. – Вильямс 2015. – 192 с.

20 Шмитт К., Симпсон К. «HTML5: Рецепты программирования» / К. Шмитт, К. Симпсон. – Питер 2012. – 289 с.