

Лабораторная работа №27.

Инструмент для поиска «битых ссылок» XENU Link Sleuth

1. Цель работы

Изучить принципы работы инструмента для поиска неработающих ссылок на сайте XENU Link Sleuth. Научиться анализировать построенные отчеты.

2. Общие сведения и ход работы

Неработающая ("Битая") ссылка – это ссылка на несуществующую страницу, документ, изображение. Переходя по битой ссылке пользователь, в лучшем случае, видит сообщение о том что страница не найдена. В худшем случае он видит служебную страницу сервера.

Почему возникают битые ссылки?

Основные причины возникновения битых ссылок:

- **Удаление старых страниц и документов.** Удаляя что-либо на сайте не стоит забывать, что в других разделах Вашего сайта могут быть ссылки на удаляемый документ.
- **Изменения на внешних сайтах.** Вы разместили ссылку на сторонний сайт. Через неделю этот сайт сменил структуру, переехал на другое имя или просто удалил тот документ, на который Вы ссылались.
- **Автоматическое обновление данных.** Если на сайте настроено автоматическое обновление данных из внешнего источника (к примеру, синхронизация товаров в интернет-магазине со складским ПО), задача поддержания актуальности ссылок становится как нельзя более важной - малейший сбой в алгоритме обновления может привести к массовым негативным последствиям и владелец сайта должен быть первым, кто об этом узнает.
- **Редизайн и переработка сайта.** Переработка сайта – это кропотливая работа, требующая особой внимательности и терпения. Иногда можно просто не заметить ссылку на уже несуществующую страницу.
- *Некоторые ссылки на сайте доступны только при вводе пароля. Если программа поиска битых ссылок не может вводить пароли для доступа, то она может посчитать такие ссылки неработающими.*

Как найти битые ссылки?

Существует несколько автоматизированных решений:

- **Инструментарий вебмастера Яндекс и Google**

- <http://webmaster.yandex.ru/wmconsole/>
- [Инструменты для веб-мастеров от Google](#)

Действительно, ведущие поисковики предоставляют отчет по найденным битым ссылкам. Как правило, поисковик полностью перепроверяет сайт раз в несколько недель. Это означает что о новых битых ссылках Вы будете узнавать далеко не сразу. И уж тем более не сможете получить оперативной картины – все ли в порядке на Вашем сайте именно сейчас, а не месяц назад.

- **Специализированное программное обеспечение**

Существует целый ряд разного рода программ, так или иначе проверяющих сайты на битые ссылки. Большинство из них платные (например, Semonitor) и весьма не дешевые. Но имеется прекрасная бесплатная программа для проверки ссылок на вашем сайте - Xenu's Link Sleuth. Недостатком использования такой программы является только то, что проверка ведется с Вашего компьютера, а, значит, используется Ваш интернет-канал. В лучшем случае это приведет к замедлению работы компьютера, в худшем Вы получите весьма ощутимый счет за израсходованный трафик.

- **Интернет-сервисы по проверке битых ссылок**

Преимущества: проверка осуществляется с сервера сервиса и не зависит ни от Вашего компьютера ни от Вашего интернет-канала. Вам не нужно ничего скачивать и устанавливать, не нужно разбираться с новыми программами.

Недостатки: некоторые бесплатные сервисы обычно проверяют ограниченное число ссылок, а полную проверку предлагают осуществить за некоторую плату. Но этот недостаток присущ не всем сервисам, он отсутствует, например, у сервисов, которые вы найдете по двум первым ссылкам в приведенных ниже примерах. **Другой недостаток:** вам придется довольно долго ждать, пока сервис завершит опрос сайта. При этом нельзя разрывать соединение с сервисом, чтобы не потерять результаты.

Примеры Интернет-сервисов:

- <http://validator.w3.org/checklink> Если вы хотите только найти "битые" ссылки, то проставьте галочку "Hide redirect". Чтобы просматривался весь сайт, а не только заглавная страница, задайте глубину просмотра, отличную от 0. Ждать окончания проверки придется достаточно долго (если сайт не маленький). В результате получите список проверенных страниц с указанием "битых" ссылок для каждой страницы.

- ["http://www.graffity.biz/products/link_checker/"](http://www.graffity.biz/products/link_checker/) Отличный и притом бесплатный сервис. К тому же после анализа сайта выдает файл Sitemap.xml (смотри раздел ["Карта сайта и файл Sitemap.xml"](#)). Недостаток заключается в том, что для сайтов с большим числом страниц непосредственно на экране найти битые ссылки очень трудно, массив выдаваемой информации просто огромен (например, это имеет место для сайта <http://rus-linux.net>, у которого больше 6 тысяч страниц).
- [Badlinks.ru](http://badlinks.ru) - сервис автоматического поиска битых ссылок. Все что нужно, это указать адрес сайта и тип проверки. По завершении процесса отчет будет отправлен Вам по электронной почте. Если Ваш сайт большой и часто обновляемый, Вы можете заказать регулярную проверку - с указанной периодичностью робот выполнит проверку и уведомит Вас о результатах. Стоимость тестовой проверки (максимально проверяется 20 страниц) бесплатна, полная проверка - 30 рублей, регулярные проверки от 350 рублей в год.
- [Build your Site Map online](#) - сервис, позволяющий получить бесплатно список битых ссылок и файл Sitemap.xml (но бесплатно обрабатывается не более 500 страниц).

Ход работы:

1. Установить программу Xenu's Link Sleuth.
2. Для того, чтобы запустить проверку "битых ссылок", нужно выбрать в меню **File** пункт **Check URL**. Откроется диалоговое окно.
3. В строке ввода в верхней части этого окна нужно ввести адрес сайта, который вы хотите проверить.
4. **Установка настроек:**
 - a. **Проверка внешних ссылок.** Для того, чтобы в ходе анализа сайта проверялись и внешние ссылки (то есть ссылки, указывающие на внешние сайты), проставьте флажок в чекбоксе "Check external links".
 - b. Экранная кнопка **Local file** используется в тех случаях, когда вы хотите проверить ссылки в локальном файле.
 - c. В блоке **Expand/Restrict** можно либо добавить какие-то типы ссылок к анализу, либо, наоборот, исключить.
 - d. По нажатию на кнопку **More options** откроется диалоговое окно с двумя вкладками: **"Basic"** и **"Advanced"**, с помощью которых нужно сразу задать некоторые параметры анализа (рисунок 36).

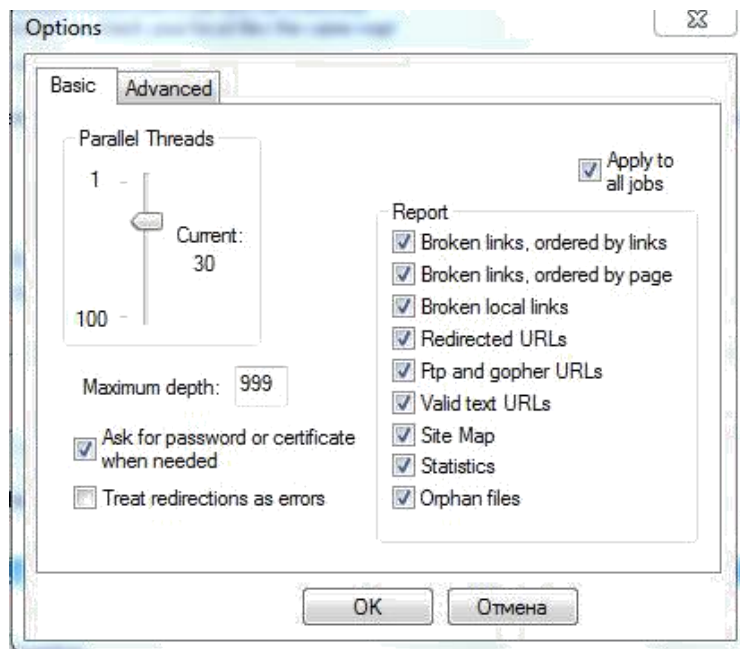


Рисунок 36 – More options в Xenu's Link Sleuth

Ползунком "Parallel Threds" на вкладке "Basic" определяется число одновременно работающих потоков программы, то есть число одновременно проверяемых ссылок. Максимально можно задать 100 потоков. Выбор здесь зависит, очевидно, от объема оперативной памяти компьютера и "толщины" канала связи до провайдера.

Рекомендуется в рамках лабораторной работы указать до 20 потоков.

Можно задавать глубину просмотра ссылок с помощью параметра **Maximum depth**.

В блоке **Reports** указывается детальностью получаемого отчета (рисунок 37).

- а. На форме добавления URL для проверки имеется кнопка **e-mail**. С помощью этой кнопки можно настроить отправку электронных писем, извещающих об окончании процесса сканирования сайта. При выборе опции извещения об окончании процесса сканирования программа Xenu завершится, а на указанный почтовый адрес придет только отчет.

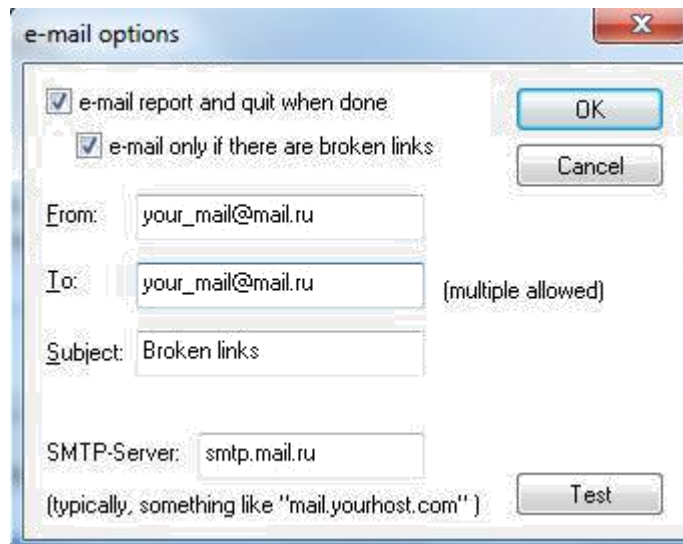


Рисунок 37 – Настройка отправки отчёта на e-mail

- После завершения установки требуемых настроек можно запустить процесс сканирования, нажав на кнопку ОК (рисунок 38):

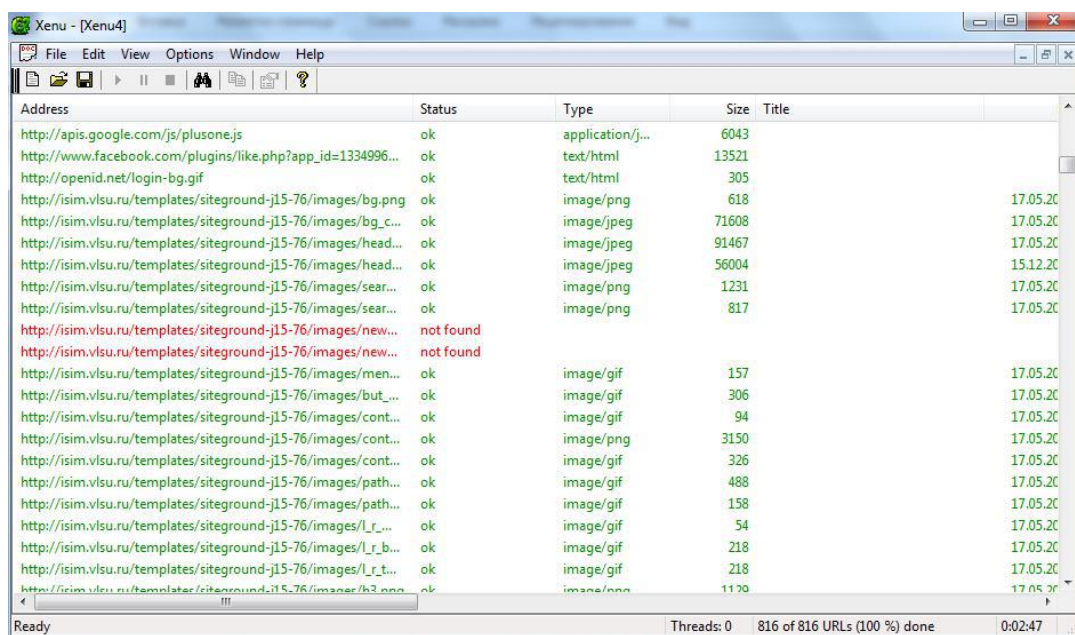


Рисунок 38 – Процесс сканирования Xenu's Link Sleuth

В окне программы в процессе выполнения видно, что рабочие (небитые) ссылки отображаются зеленым цветом, пропущенные в соответствии со сделанными выше настройками внешние ссылки - синим, внутренние ссылки, которые пользователь запретил просматривать - коричневым. Битые ссылки будут выделены красным, а ссылки, просматриваемые в данный момент - серым цветом (на рисунке это не видно). Сам процесс сканирования для

достаточно больших сайтов может продолжаться часами. Когда программа закончит сканирование, будет выдано сообщение.

Можно получить информацию, на какой именно странице искать «битую ссылку». Для этого кликните правой клавишей вашей мышки по соответствующей ссылке в окне программы Xenu's Link Sleuth и выберите из контекстного меню вариант «URL properties». В самом низу дан список адресов, где обнаружена данная битая ссылка.

6. В окне с запросом отчета не торопитесь со щелчком по кнопке "Да". Скажите "Нет". Необходимо проверить, все ли страницы сайта обработаны программой: если имеются строки такие, что в столбце "Status" проставлено значение "pending" или "timeout", сканирование надо повторять. Причем, если сайт у вас содержит достаточно большое число страниц, повторять многократно. Вот когда число известных программе страниц перестанет увеличиваться и все строки будут иметь статус "ok" или "skip", можно уже задать формирование отчета.
7. Отчет можно сформировать в разделе **File – Report**. В окне запроса параметров ftp для удаленного контроля за битыми ссылками (для исправления ссылок из отчета) необходимо нажать Cancel, так как у нас нет реквизитов и прав на доступ к ftp. Сформированный отчет открывается в окне браузера.

Задание:

1. Выбрать сайт для проверки на битые ссылки из предложенного списка:
 - a. <http://isim.vlsu.ru/>
 - b. <http://izi.vlsu.ru/>
 - c. <http://uites.vlsu.ru/>
 - d. <http://ce.vlsu.ru/>
 - e. <http://www.cs.vlsu.ru/>
 - f. <http://ced.vlsu.ru/>
 - g. <http://atf-t.vlsu.ru/>
 - h. <http://ecity.vlsu.ru/>
 - i. <http://iem.vlsu.ru/>
 - j. <http://fiet.vlsu.ru/>
 - k. <http://ui.vlsu.ru/>
 - l. <http://pi.vlsu.ru/>
 - m. <http://phil.vlsu.ru/>
 - n. <http://asf.vlsu.ru/>
 - o. <http://fpmf.vlsu.ru/>

р. <http://fhe.vlsu.ru/>

2. Выполнить анализ сайта на «битые ссылки» со следующими параметрами
 - a. Из анализа исключить хотя бы один раздел сайта.
 - b. Не проверять внешние ссылки.
3. Выполнить анализ сайта по всем разделам, исключая проверку внешних ссылок. Построить отчет по проведенному анализу с использованием инструментария программы Xenu's Link Sleuth, включающий:
 - a. Информацию о битых ссылках сайта
 - b. Информацию о корректных ссылках
 - c. Карту сайта
 - d. Статистику сайта
4. Ответить на вопросы:
 - a. Каково общее число проверенных ссылок?
 - b. Сколько «битых ссылок» обнаружила программа? На каких страницах имеются такие ссылки.
 - c. Какова максимальная глубина вложенности ссылок анализируемого сайта?
 - d. По какой ссылке располагается наибольший объем данных? Почему?
 - e. Как избавиться от обнаруженных «битых ссылок»?
 - f. Проанализировать среднее время отклика ссылок (открытие страниц).

3. Содержание отчета по лабораторной работе

1. Цель работы.
2. Выбранный сайт для анализа.
3. Скриншоты настроек программы для проведения анализа.
4. Скриншоты результатов выполнения (битые ссылки, пропущенные, исключенные).
5. Отчет, построенный программой.
6. Ответы на поставленные вопросы.
7. Выводы по лабораторной работе.

