

ПАРСИНГ САЙТОВ. «SITEANALYZER» КАК ИНСТРУМЕНТ ДЛЯ ПРОВЕДЕНИЯ ТЕХНИЧЕСКОГО SEO-АУДИТА САЙТОВ

Ситник А.В.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Файзрахманов Ф.М. – ст. преп.

Веб-сайт – эффективный многоцелевой инструмент любой компании. В настоящее время многие белорусские организации, осознав важность регулярного аудита собственных веб-ресурсов, сталкиваются с проблемой трудоемкости проведения самостоятельных мероприятий по оптимизации сайтов, связанной с временными, материальными и энергозатратами. Использование инструментов, автоматизирующих данный процесс, позволяет выявить несовершенства системы и принять решения по их устранению, не прикладывая при этом огромных усилий.

Парсинг сайтов (англ. web scraping) – это процесс программного обхода и загрузки информации с веб-сайтов, извлечения неструктурированных или слабо структурированных данных в упорядоченный формат [1]. Инструмент, с помощью которого происходит выполнение указанного процесса, называется парсером.

Существует множество целей проведения парсинга сайтов, т.к. это универсальный процесс сбора информации с веб-страниц, позволяющий получить данные для выполнения коммерческих, технических, личных и других задач.

Среди основных целей парсинга веб-страниц можно выделить следующие:

- технический SEO-аудит сайта;
- мониторинг цен и ассортимента продукции конкурентов;
- сбор контактной информации;
- поиск целевой аудитории компании;
- проверка актуальности информации, размещенной на сайте.

Следует отметить, что парсинг сайтов является правомерной деятельностью. Согласно статье 16 Закона Республики Беларусь «Об информации, информационных технологиях и защите информации», к общедоступной информации относится информация, доступ к которой, распространение и (или) предоставление которой не ограничены [2]. Исходя из этого следует, что парсинг веб-ресурсов является законным действием при условии, что полученная информация размещалась в открытом доступе и не является коммерческой, профессиональной и иной, охраняемой законом тайной.

Выбор парсеров достаточно велик. Они отличаются друг от друга функционалом, удобством использования и ценой, можно использовать уже созданные парсеры либо разработать собственную программу. Из этого следует, что выбор парсера зависит от конкретных целей, стоящих перед компанией.

Для технического SEO-аудита сайтов предназначены такие парсеры, как «Screaming Frog», «Netpeak Spider», «WildShark SEO Spider» и «SiteAnalyzer». Первые два инструмента являются платными, а последние – их бесплатными аналогами. Их сравнение было произведено в рамках проводившегося нами исследования.

«Screaming Frog» – один из самых распространенных парсеров. Является мощным и гибким инструментом, способным эффективно обрабатывать как небольшие, так и многостраничные сайты. Парсер используется для поиска битых ссылок, проверки заголовков и мета-данных, анализа ключевых слов и множества других функций [3].

«Netpeak Spider» является комплексным инструментом для глубокого аудита внутренней оптимизации сайта [4]. Программа имеет интерактивный и удобный интерфейс. Среди особенностей можно выделить возможность экспортирования отдельных отчетов, парсинг сайтов, требующих авторизации, возможность останавливать и возобновлять процесс анализа веб-страниц в любое время.

Парсер «WildShark SEO Spider» представляет собой простой удобный инструмент для быстрой оценки технического состояния сайта [5]. Для детального аудита веб-ресурсов следует выбрать вышеперечисленные программы, т.к. функционал данного парсера достаточно узкий.

«SiteAnalyzer» – программа, предназначенная для анализа сайта и выявления технических ошибок, а также ошибок и недоработок в SEO-оптимизации [6]. Данный инструмент позволяет производить сканирование всех страниц и изображений сайта, поиск битых ссылок, анализ контента и других параметров. В общей сложности анализу подвергается около 60 таких параметров. По окончании технического аудита можно экспортировать отчеты с результатами работы программы в CSV, MS Excel и в PDF-формат.

С целью анализа функциональных возможностей перечисленных парсеров сайтов были разработаны соответствующие шкалы, представленные в таблице 1.

Таблица 1 – Шкалы для оценки функциональных возможностей исследуемых парсеров сайтов

Шкала для оценки интерфейса		Шкала для оценки скорости работы		Шкала для оценки стоимости за год	
Критерий	Оценка	Интервал, мин.	Оценка	Интервал, долл.	Оценка
Очень сложный	1	46-60	1	300-450	1
Сложный	2	31-45	2	150-300	2
Простой	3	16-30	3	1-150	3
Интерактивный и простой	4	1-15	4	Бесплатно	4

Примечания: Источник: собственная разработка.

Каждый знак «+» приравнивается к оценке 1, знак «+/-» – к оценке 0,5, а знак «-» – к оценке 0. В качестве исследуемого веб-ресурса использовался официальный сайт БГУИР [7].

Результаты сравнительной характеристики исследуемых парсеров сайтов представлены в таблице 2.

Таблица 2 – Сравнительная характеристика функциональных возможностей исследуемых парсеров сайтов

Функции	Парсеры сайтов							
	«Screaming Frog»		«Netpeak Spider»		«WildShark SEO Spider»		«SiteAnalyzer»	
Интерфейс	2		4		3		3	
Скорость работы	6:13	4	10:24	4	18:47	3	59:02	1
Стоимость	180,45	2	182,40	2	0	4	0	4
Наличие документации по работе с инструментом	+		+		-		+	
Поиск битых ссылок	+		+		+		+	
Проверка дублей Title, Description и заголовков H	+		+		+/- (только заголовков H)		+	
Обнаружение входящих и исходящих ссылок страницы	+		+		-		+	
Парсинг отдельных URL-адресов	+		+		-		-	
Указание Title, Description, H заголовков, атрибутов Alt и их длины	+		+		+/- (только текст)		+	
Анализ ключевых слов	+		+		+		+	
Обнаружение мета-тега robots	+		+		-		+	
Итого:	16		18		13		15	

Примечания: Источник: собственная разработка с использованием данных [8].

Следует отметить, что характеристика парсеров нами проводилась не по всем параметрам, которыми должна обладать программа для проведения технического SEO-аудита. По нашему мнению, набор исследуемых функций является достаточным для начинающих SEO-специалистов, что позволяет сделать выбор в пользу наиболее подходящего для начального уровня инструмента.

«SiteAnalyzer» является достойным бесплатным аналогом платных инструментов, позволяющим выполнять широкий круг задач в области анализа веб-сайтов. К единственному и главному недостатку этого инструмента относится длительное время парсинга. Несмотря на это, программа имеет широкий функционал, характеризуется простотой и удобством использования, что играет немаловажную роль для начинающих специалистов.

Таким образом, с помощью использования парсера «SiteAnalyzer» можно проводить регулярный технический SEO-аудит сайтов самостоятельно, без привлечения профессионалов, экономя при этом материальные, трудовые и энергоресурсы.

Список использованных источников:

1. Ward, J. Instant PHP Web Scraping / J.Ward. – Birmingham: Packt Publishing, 2013. – 5 p.
2. Национальный правовой Интернет-портал Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://www.pravo.by>. – Дата доступа: 18.03.2020.
3. «Screaming Frog SEO Spider» [Электронный ресурс]. – Режим доступа: <https://www.screamingfrog.co.uk/seo-spider/>. – Дата доступа: 18.03.2020.
4. «Netpeak Spider» [Электронный ресурс]. – Режим доступа: <https://netpeaksoftware.com/ru/spider>. – Дата доступа: 18.03.2020.
5. «WildShark SEO Spider» [Электронный ресурс]. – Режим доступа: <https://wildshark.co.uk/spider-tool/>. – Дата доступа: 18.03.2020.

6. Документация программы «SiteAnalyzer» [Электронный ресурс].– Режим доступа: <https://site-analyzer.ru/documentation/>. – Дата доступа: 18.03.2020.
7. Официальный сайт БГУИР [Электронный ресурс].– Режим доступа: <https://www.bsuir.by/>. – Дата доступа: 18.03.2020.
8. Обзор ТОП-6 парсеров сайтов [Электронный ресурс].– Режим доступа: <https://siteclinic.ru/blog/seo-instrumenty/obzor-parserov/>. – Дата доступа: 18.03.2020.