

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ОБЕСПЕЧЕНИЯ ФОРМАЛИЗОВАННЫХ ТРЕБОВАНИЙ К САЙТАМ ВУЗОВ

В.К. Садыхян, магистрант 2 курса, кафедра информационных технологий и систем

*Владивостокский государственный университет экономики и сервиса
г. Владивосток*

Федеральная служба по надзору в сфере образования и науки подготовила список требований к структуре официальных сайтов образовательных учреждений и формату представления информации на них. Согласно нормам, сайт должен быть четко разбит на несколько подразделов, содержать обязательную информацию и документы. В данной работе рассматриваются подходы интеллектуального поиска информации и анализа данных. В ходе научно-исследовательского проектирования будет разработан метод поиска информации и метод контроля наличия информации на сайтах вузов.

Ключевые слова: алгоритм поиска данных, метод сопоставления, анализ результатов, сайт образовательного учреждения, вузы.

INTELLIGENT ANALYSIS ENSURE FORMALIZE REQUIREMENTS FOR UNIVERSITIES SITES

V.K. Sadyhyan, undergraduate of the 2nd year, department of information technologies and systems

*Vladivostok State University of Economics and Service
Russia. Vladivostok*

The Federal Service for Supervision of Education and Science has prepared a list of demands to the structure of the official websites of educational institutions and how to report them. According to the rules, the site should be clearly divided into several subsections contain the mandatory information and documents. This paper examines the intellectual approach of information retrieval and analysis. In the course of scientific research, design will be developed by the method of searching for information and the method of monitoring the availability of information on the websites of universities.

Keywords: algorithm data retrieval, method of comparison, analysis of the results, the site of the educational institution, universities.

В последние десятилетия научное сообщество активно осваивает ресурсы глобальной сети Интернет. Электронные представительства есть практически у каждой организации. Научно-исследовательские институты, научно-образовательные учреждения, библиотеки, научно-производственные предприятия и другие организации научной направленности активно позиционируют себя в Глобальной сети. Интернет предоставляет им колоссальные возможности, стирая временные и пространственные границы [2].

Наиболее ощутимы преимущества использования Интернета в образовательной сфере, особенно в вузах. Теперь стало проще получить самую актуальную и достоверную информацию о любом учебном заведении, не выходя из дома. Абитуриенту нет необходимости покупать дорогостоящие справочники или посещать дни открытых дверей. Студентам предоставлена возможность использовать ресурсы веб-сайта учебного заведения в процессе обучения.

«Под официальным веб-сайтом вуза понимается принадлежащий вузу веб-сайт, предназначенный для всестороннего и достоверного внешних и внутренних посетителей о деятельности вуза, а также представляющий посетителям сайта всю необходимую для обеспечения взаимодействия с вузом, его руководством или его подразделениями справочную информацию» [1].

Настоящая работа посвящена изложению подходов и описанию методов поиска и анализа информации.

Одним из первых опробуемых методов поиска информации по сайтам был – поиск через поисковую систему google.com. Для поиска по сайту были составлены вручную 74 запроса. Поиск осуществлялся по одному сайту образовательной организации. Далее был разработан скрипт, который читал запросы из файла, выполнял их в поисковой системе google.com и разбирал результат выдачи. Однако через несколько десятков запросов google.com распознал, что запросы вводит машина (скрипт) и стал блокировать запросы, компьютерным тестом. На странице поиска отображалась CAPTCHA (Complete-ly Automated Public Turing test to tell Computers and Humans Apart), не давая тем самым обрабатывать скрипту. Данный метод был опробован и для поисковой системы yandex. Результат отказался тем же.

Учитывая данные проблемы, было принято решение перейти к поиску другого подхода поиска информации не используя поисковые системы.

Следующий исследуемый подход к поиску информации заключается в следующем – с помощью скрипта, переходящего по ссылкам на каждом сайте, поиск информации будет осуществляться постепенно на каждом сайте. Программа будет заходить на сайт и, используя html-разметку, искать нужные ссылки, переходить по ним (если они есть) и так далее. Основным минусом такого подхода является то, что такой поиск будет работать намного дольше, чем с использованием поисковых систем.

Описание используемых программных средств.

На первом этапе формируются входные данные. Для данного проекта входными данными является информация для поиска на сайте. Данная информация структурирована и представлена в формализованном виде в методических рекомендациях представления информации об образовательной организации в открытых источниках с учетом соблюдения требований законодательства в сфере образования документе с требованиями от министерства образования.

Для обработки входных данных скриптом их необходимо обработать и перевести в читаемый формат для скриптового языка. Было решено сформировать входные данные в json файл. Для разбора json файла и запуск js-скриптов в порядке очереди используется один из самых распространенных скриптовых языках – PHP. Personal Home Page Tools – скриптовый язык общего назначения, интенсивно применяемый для разработки веб-приложений. В настоящее время поддерживается подавляющим большинством хостинг-провайдеров и является одним из лидеров среди языков, применяющихся для создания динамических веб-сайтов.

Часть скрипта для открытия страниц и поиска информации разработана на языке программирования javascript с использованием утилиты для запуска браузера phantomjs.

JavaScript — прототипно-ориентированный сценарный язык программирования. Является диалектом языка ECMAScript. JavaScript обычно используется как встраиваемый язык для программного доступа к объектам приложений. Наиболее широкое применение находит в браузерах как язык сценариев для придания интерактивности веб-страницам.

Phantomjs представляет собой браузер без пользовательского интерфейса, который дает дополнительные возможности в программной навигации по вебу. Базируется на движке webkit.

WebKit — свободный движок для отображения веб-страниц, разработанный на основе кода библиотек KHTML и KJS, используемых в графической среде KDE.

Тестирование метода.

Для размещения информации на официальном сайте должен быть создан специальный раздел «Сведения об образовательной организации», содержащий следующие подразделы обозначение цифрами от 1 до 11 в таблице 1:

1. Основные сведения;
2. Структура и органы управления образовательной организацией;
3. Документы;
4. Образование;
5. Образовательные стандарты;
6. Руководство. Педагогический (научно-педагогический) состав;
7. Материально-техническое обеспечение и оснащенность образовательного процесса;
8. Стипендии и иные виды материальной поддержки;
9. Платные образовательные услуги;
10. Финансово-хозяйственная деятельность
11. Вакантные места для приема (перевода).

Результаты первого этапа тестирования Таблица 1

№	Список сайтов/Подразделы	Сведения об обр. орг.	1	2	3	4	5	6	7	8	9	10	11
1	http://www.mgimo.ru	+	+	+	+	+	+	-	+	+	+	+	+
2	http://www.vvsu.ru	+	+	+	+	+	+	+	+	+	+	+	+
3	http://www.dvfu.ru	+	+	+	+	+	+	+	+	+	+	+	+
4	http://www.kantiana.ru	+	-	-	-	-	-	+	-	-	-	-	-
5	http://www.kpfu.ru	+	-	+	+	+	+	+	-	+	+	+	+
6	http://www.narfu.ru	+	+	+	+	+	+	+	+	+	+	+	+
7	http://www.sfedu.ru	+	+	+	+	+	+	-	+	+	+	+	+
8	http://www.ssau.ru	+	+	+	+	+	+	+	+	+	+	+	+
9	http://www.ncfu.ru	-	-	-	-	-	-	-	-	-	-	-	-
10	http://www.ifmo.ru	+	-	-	-	+	+	-	+	+	+	-	-
11	http://www.spbstu.ru	+	+	-	+	+	+	-	+	+	+	+	-
12	http://www.miet.ru	+	+	+	+	+	+	+	+	+	+	+	+
13	http://www.ursmu.ru	+	+	+	+	+	+	+	+	+	+	+	+
14	http://www.tpu.ru	+	+	+	+	+	+	-	+	+	+	+	+
15	http://www.ranepa.ru	+	+	+	+	+	+	+	+	+	+	+	+
16	http://www.mpei.ru	+	+	+	+	+	+	-	+	+	+	+	+
17	http://www.rea.ru	+	-	-	-	-	-	-	-	-	-	-	-
18	http://www.eltech.ru	+	+	+	+	+	+	-	+	+	+	+	+
19	http://www.tusur.ru	+	+	+	+	+	+	+	+	+	+	+	+
20	http://www.samsu.ru	+	+	+	+	+	+	+	+	+	+	+	-

В результате тестирования данного метода для поиска основных подразделов можно сказать о его положительном эффекте. Тестирование проводилось на 20 сайтах вузов и образовательных организаций, ранее отобранных вручную, и только на 2 сайтах поиск не дал ожидаемого результата. Сначала на сайте происходит поиска основного раздела «Сведения об образовательной организации» и только в случае его наличия продолжается поиск подразделов в нем. Это заметно по результатам, на сайте www.ncfu.ru не было найдено основного раздела и поиск на данном сайте выполнен не успешно. Так же на сайте www.rea.ru основной раздел был найден как текст, поэтому поиск разделов не выполнен. Если рассматривать результаты для остальных ресурсов, то можно говорить о эффективности метода поиска подразделов на сайтах в виде ссылок или текста. Стоит отметить, что тестируемый подход поиска информации не подходит для сайтов с некорректной версткой.

1. Интеллектуальный поиск – научная статья [Электронный ресурс]. – Режим доступа: <http://msdn.microsoft.com/ru-ru/library/ms167167.aspx>.

2. Основные понятия интеллектуального анализа данных [Электронный ресурс]. – Режим доступа: <http://msdn.microsoft.com/ru-ru/library/ms174949.aspx>.