

Реферат

Магистерской аттестационной работы на тему:

«Средства многокритериальной оценки качества веб-сайтов и поиск
поддельных сайтов»

Рублёва Дмитрия Сергеевича

Актуальность работы

Мы живем в век стремительного развития информационных технологий. Они проникают во все сферы жизни современного человека. Возрастает ценность информации, переосмысливается ее роль в жизни человека. Со стремительным развитием сети Интернет, становится все труднее среди информационного «шума» найти нужную нам информацию. Сейчас, для создания Интернет-сайта не требуется знать HTML и языки программирования, поскольку существует большое количество программных инструментов для создания веб-страниц, где вся работа по разработке страницы сводится к перетаскиванию графических элементов и ввода текстов. Во многих случаях, о качестве сайтов, созданных таким образом лучше молчать - при отсутствии понимания основных принципов построения веб-сайтов, веб-стандартов и иногда даже обычных правил грамматики, то о каком качестве может идти речь? Существование бесплатных хостингов позволяет достаточно быстро заполнять виртуальное пространство сайтами, информативность и полезность которых стремится к нулю, и которые осложняют поиск полезной информации в глобальной информационной сети.

На существовании «мусорных» сайтов сейчас строится целый бизнес - это так называемая «черная» поисковая оптимизация и «манимейкинг». Оба явления пользуются тем, что количество в определенное время перерастает в качество - создав большое количество дорвеев и сателлитов, можно на некоторое время увеличить посещаемость выбранного Интернет-ресурса. При создании «мусорных» сайтов, вопросы их содержания часто стоит не на первом месте, и наполнение сайта создается за счет заимствования текстов с других

сайтов, переписыванием текстов с заменой слов синонимами, а иногда даже генерацией текста из обрывков фраз, выбранных случайным образом. Во всех случаях, от существования таких сайтов страдают пользователи Интернета, а в первом - еще и авторы заимствованных текстов.

Но заимствование текстов иногда случается не только в среде черной поисковой оптимизации, но и среди крупных сайтов известных компаний. Существуют случаи заимствования текстов довольно крупными корпорациями, которые приводили к открытию судебных дел и наказания виновных.

Особняком стоит категория сайтов, полностью дублируют внешний вид какого-то определенного сайта. Основная доля таких - сайты в мошеннических целях. Сделав точную копию страницы авторизации в популярной онлайн-банковской системе, и разослав спам-сообщения, с просьбой зайти на созданную страницу и авторизоваться под любым предлогом. Если такое письмо попадет к невнимательному пользователю банка, страницу авторизации которого скопировали - он (пользователь) фактически сам отдаст вора свои пароли.

В связи с такой ситуацией, вопрос оценки качества веб-сайтов и выявления поддельных сайтов встает довольно остро.

Цель работы

Целью данной работы является исследование возможности оценки качества сайтов, определения критериев, на него влияющих, или дающих возможность сравнивать сайты, и использования оценивания сайтов для обнаружения подделок и плагиата.

Достигнутые результаты

В работе рассмотрена возможность оценивать качество сайтов. При отсутствии четкого определения качества веб-сайтов, избран основные критерии, которые можно толковать, как показатели качества, сделан обзор

инструментов для оценки этих критериев и описаны пути повышения качественных характеристик сайтов.

Среди критериев, по которым можно определять качество веб-сайтов выделено 4 главных:

1. Информационное наполнение сайта;
2. Цитируемость информации;
3. Техническая реализация;
4. Соответствие сайта веб-стандартам;

Среди существующих инструментов оценки веб-сайтов были рассмотрены как онлайн-сервисы, так и прикладные программы.

Научная новизна

Впервые были рассмотрены подходы к:

1. Оценке ценности информации для поисковой системы;
2. Поиску плагиата с определением источника оригинальных текстов;
3. Уменьшения негативных последствий от заимствования информации чужими сайтами.

Практическая ценность работы

В настоящее время система имеет лишь базовую функциональность, на основе которой можно разработать более мощные инструменты для определения заимствования информации между Интернет-сайтами.

Выводы

Из-за невозможности оценки качества, как критерия, который в настоящее время не может быть сформулирован и использован для оценки сайтов, был рассмотрен подход к оценке Интернет-сайтов по критериям, которые используются поисковыми системами для ранжирования информации.

В результате проведенных опытов, было определено, что наилучшим путем для создания и последующего развития эффективных сайтов является

использование CMS с канонизацией URL, клиентской оптимизацией содержимого страниц и автоматическим информированием поисковых систем об обновлении информации. Этот подход позволяет улучшить индексацию и уменьшить негативное влияние от заимствования информации с сайтов.

Работа содержит 98 стр., 21 рис., 1 табл., 26 источников

Ключевые слова: CMS, W3C, PHISHING, DOORWAY, ПЕРВИЧНОСТЬ ИНФОРМАЦИИ, ПОИСКОВЫЙ СПАМ, ИНДЕКСАЦИЯ ИНТЕРНЕТ-САЙТОВ, КЛИЕНТСКАЯ ОПТИМИЗАЦИЯ.