

## **Реферат**

Магістерської атестаційної роботи на тему:

«Засоби багатокритеріальної оцінки якості веб-сайтів та пошук підробних сайтів»

Рубльова Дмитра Сергійовича

### **Актуальність роботи**

Ми живемо в час стрімкого розвитку інформаційних технологій. Вони проникають в усі сфери життя сучасної людини. Зростає цінність інформації, переосмислюється її роль в житті людини. Зі стрімким розвитком мережі Інтернет, стає все важче серед інформаційного «шуму» знайти потрібну нам інформацію. Зараз, для створення Інтернет-сайту не є необхідним знати HTML, та мови програмування, бо існує велика кількість програмних інструментів для створення веб-сторінок, де вся робота з розробки сторінки зводиться до перетягування графічних елементів та введення текстів. У багатьох випадках, про якість сайтів, створених таким чином краще мовчати – при відсутності розуміння основних засад побудови веб-сайтів, веб-стандартів та інколи навіть звичайних правил граматики, то про яку якість може йти мова? Існування безкоштовних хостингів дозволяє досить швидко заповнювати віртуальний простір сайтами, інформативність та корисність яких прямує до нуля, і які ускладнюють пошук корисної інформації у глобальній інформаційній мережі.

На існуванні «сміттєвих» сайтів зараз будується цілий бізнес – це так звана «чорна» пошукова оптимізація та «манімейкінг». Обидва явища користуються тим, що кількість у певний час переростає у якість – створивши велику кількість дорвеїв та сателітів, можна на деякий час збільшити відвідуваність обраного Інтернет-ресурсу. При створенні «сміттєвих» сайтів, питання їхнього вмісту часто стоїть не на першому місці, і наповнення сайту створюється за рахунок запозичення текстів з інших сайтів, переписуванням текстів із заміною слів синонімами, а інколи навіть генерацією тексту з уривків фраз, обраних

випадковим чином. У всіх випадках, від існування таких сайтів страждають користувачі Інтернету, а у першому – ще й автори запозичених текстів.

Але запозичення текстів інколи трапляється не тільки у середовищі чорної пошукової оптимізації, а й серед великих сайтів відомих компаній. Існують випадки запозичення текстів досить великими корпораціями, що призводили до відкриття судових справ та покарання винуватців.

Окремо стоїть категорія сайтів, що повністю дублюють зовнішній вигляд якогось певного сайту. Основна частка таких – сайти у шахрайських цілях. Зробивши точну копію сторінки авторизації у популярній онлайнівій банківській системі, та розіславши спам-повідомлення, з проханням зайти на створену сторінку та авторизуватися під будь-яким приводом. Якщо такий лист потрапить до неуважного користувача банку, сторінку авторизації якого скопіювали – він (користувач) фактично сам віддасть злодіям свої паролі.

У зв'язку з такою ситуацією, питання оцінки якості веб-сайтів та виявлення підробних веб-сайтів постає досить гостро.

### **Ціль роботи**

Метою даної роботи є дослідження можливості оцінювати якість сайтів, визначення критеріїв, що на неї впливають, або дають можливість порівнювати сайти, та застосування оцінювання сайтів для виділення підробок та плагіату.

### **Досягнуті результати**

У роботі розглянуто можливість оцінювати якість сайтів. За відсутності чіткого визначення якості для веб-сайтів, обрано основні критерії, які можна тлумачити, як показники якості, зроблено огляд інструментів для оцінки цих критеріїв та описано шляхи підвищення якісних характеристик веб-сайтів.

Серед критеріїв, за якими можна визначати якість веб-сайтів виділено 4 головних:

1. Інформаційне наповнення сайту;
2. Цитуємість інформації;
3. Технічна реалізація;
4. Відповідність сайту веб-стандартам;

Серед існуючих інструментів для оцінки веб-сайту розглянуто як онлайн-сервіси, так і прикладні програми.

### **Наукова новизна роботи**

Вперше було розглянуто підходи до:

1. оцінки цінності інформації для пошукової системи;
2. пошуку плагіату з визначенням джерела оригінальних текстів;
3. нівелювання негативних наслідків від запозичення текстів чужими сайтами.

### **Практична цінність роботи**

На даний час система має лише базову функціональність, на основі якої можна розробити більш потужні інструменти для визначення запозичення інформації між Інтернет-сайтами.

### **Висновки**

За неможливістю оцінювання якості, як критерію, що на даний час не може бути сформульований та використаний для оцінки сайтів, було розглянуто підхід до оцінки Інтернет-сайтів за критеріями, які використовуються пошуковими системами для ранжування інформації.

В результаті проведених дослідів, було визначено, що найкращім шляхом для створення та подальшого розвитку ефективних веб-сайтів є використання CMS з канонізацією URL, клієнтською оптимізацією вмісту сторінок та автоматичним інформуванням пошукових систем про оновлення інформації. Цей підхід дозволяє покращити індексацію та зменшити негативний вплив від запозичення інформації з сайтів.

Робота містить 98 стор., 21 рис., 1 табл., 26 джерел.

**Ключові слова:** CMS, W3C, PHISHING, DOORWAY, ПЕРВИННІСТЬ ІНФОРМАЦІЇ, ПОШУКОВИЙ СПАМ, ІНДЕКСАЦІЯ ІНТЕРНЕТ-САЙТІВ, КЛІЄНТСЬКА ОПТИМІЗАЦІЯ

## **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:

5. Информационное наполнение сайта;
6. Цитируемость информации;
7. Техническая реализация;
8. 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:

4. Оценке ценности информации для поисковой системы;
5. Поиску плагиата с определением источника оригинальных текстов;
6. 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



## **Abstract**

Master's attestation work on:

“Multi-criteria evaluation tools for websites and fake websites detection”

Rublyov Dmytro

## **Relevance**

We live in a time of rapid development of information technology. They penetrate into all spheres of life of modern man. Increasing the value of information, rethinking its role in human life. With the rapid development of the Internet is becoming more informative among the “noise” to find the necessary information to us. Now, to create a website is not necessary to know HTML, and programming languages, because there are many software tools for creating Web pages, where all the work on the development page is reduced to dragging graphic elements and type. In many cases, the quality of sites created so better keep quiet - the absence of understanding the fundamentals of building web sites, web standards, and sometimes even the ordinary rules of grammar, what kind of quality can we speak? The existence of free webhosts can quickly fill a virtual web site are informative and useful of which tends to zero, and which complicate the search for useful information in global information network.

The existence of “garbage” sites now built a business - so-called “black” search engine optimization and “moneymaking”. Both events use that number in time develops into quality – creating many doorway pages and satellites can be for some time to increase attendance to some Internet resource. When a “garbage” sites are created, the question of their contents is often not the first place, and the site content is created by borrowing texts from other sites, rewriting with replacement of the words verb, and sometimes even generating source of the fragments of phrases, selected at random. In all cases, Internet users suffer from the existences of such sites and in the first - the authors of borrowed texts.

But borrowing texts sometimes occurs not only among black search engine optimization, but also among large sites of major companies. There are cases borrowing texts very large corporations, which led to the discovery of cases and punish the perpetrators.

Unusual stands a category of sites that fully duplicate the appearance of any particular site. The main part of these - sites in the fraud. Having a replica authorization page in a popular online banking system, and sending spam message, asking to visit this page and log in there under any pretext. If this letter gets to the inattentive user bank, which copied the authorization page - it (user) himself actually give thieves their passwords.

In such a situation, the question of assessing the quality of web sites and identify fraudulent Web sites appears quite serious.

### **Objective**

The aim of this work is to study existing Web sites assessment tools by highlighting the basic criteria that together give an idea of quality sites

### **Results**

The paper describes the main criteria to determine quality of the site, an overview of tools to evaluate these criteria and describes ways to improve their own websites.

Among the criteria that define websites quality was selected Top 4:

1. Content;
2. Link popularity;
3. Technical realization;
4. Web standards matching;

Among existing tools for assessing the quality of the website is considered as online services and applications. Using existing programs with reasonable accuracy can judge the quality of websites.

### **Scientific novelty**

For the first time were considered approaches:

1. to evaluate the information for search engines value;
2. to the method of determining the source of original texts;
3. for minimizing negative consequences from borrowing other people's lyrics sites.

### **Practical achievements**

Currently the system has only the basic functionality on which can developed into more powerful tools to identify borrowed information on the Internet.

### **Conclusion**

Due to impossibility of quality evaluation as a criterion that at present cannot be formulated and used to evaluate sites, was considered the approach to the evaluation of Internet sites by the criteria used by search engines for information ranking.

Results of these experiments, it was determined that the best way for the creation and further development of websites is to use CMS with testing situations access to pages with the incorrect address canonization URL, customer optimization and setting up a web server for caching and compression of downloaded information. This approach allows us to keep a clear site structure, reduce time to open the page the user and reduce load on the equipment and bandwidth.

The work contains 98 pages, 21 figures, 1 table and 26 sources.

**Keywords:** CMS, W3C, PRIMACY OF INFORMATION, SEARCH SPAM, INTERNET SITES QUALITY, USER-SIDE OPTIMIZATION.