

Тема выпускной квалификационной работы:

«Исследование алгоритмов поисковых систем,
разработка системы прогнозирования позиций
сайтов в результатах поисковой выдачи»

Выполнил:

Дьяконов Дмитрий Валерьевич

Научный руководитель:

д.т.н., профессор

Янишевская Анна Генриховна

АКТУАЛЬНОСТЬ

Ежедневно в интернете появляются тысячи новых веб-сайтов. Чтобы занять и удерживать лидирующие позиции, в условиях постоянного изменения алгоритмов поисковых систем, а так же привлекать больше целевых посетителей, чем у сайта-конкурента, необходимо серьезно подойти к вопросам поисковой оптимизации и продвижения сайта именно в поисковых системах.



НАУЧНАЯ НОВИЗНА

Научной новизной является алгоритм прогнозирования позиций сайтов в результатах поисковой выдачи и методика продвижения веб-ресурсов в современных поисковых системах.



ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Провести анализ алгоритмов ранжирования современных поисковых систем;
2. Построить математическую модель системы ранжирования сайтов;
3. Разработать систему анализа сайтов и прогнозирования их позиций в результатах поисковой выдачи;
4. Разработать методику продвижения сайтов в современных поисковых системах.



ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ

Объект исследования – поисковые системы Google, Yandex, и веб-ресурсы.

Предметом исследования являются алгоритмы ранжирования современных поисковых систем.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

P_{sq} – суммарная оценка сайта s в поисковой системе по поисковому запросу q

$$P_{sq} = \prod_{i=1}^n (\sigma_i \cdot f(\varphi_i)) \quad (1)$$

где i – порядковый номер фактора,
 σ_i – коэффициент значимости фактора,
 $f(\varphi_i)$ – оценка фактора по 10-балльной шкале.

$$f(\varphi_i) = \frac{f(\alpha_i)}{\gamma_i} * 10 \quad (2)$$

где $f(\alpha_i)$ – значение фактора анализируемого сайта,
 γ_i – значение фактора идеального сайта,
 10 – предложенный множитель.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

$$\text{if } f(\alpha_i) \geq \gamma_i, \text{ then } \frac{f(\alpha_i)}{\gamma_i} = 1$$

$$\gamma_i = f(\alpha_i)_{max} \quad (3)$$

где $f(\alpha_i)_{max}$ – максимальное значение фактора среди факторов каждого из 10-ти первых сайтов по выбранному поисковому запросу.

Значение $f(\alpha_i)$ на примере фактора “скорость загрузки сайта”

$$f(\alpha) = t_a + t_h + t_c + t_j \quad (4)$$

где t_a – время ответа сервера, t_h – время загрузки html, t_c – время загрузки таблицы стилей, t_j – время загрузки javascript.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Значение $f(\alpha_i)$ на примере фактора "PageRank"

$$f(\alpha) = (1 - d) + d\left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)}\right) \quad (5)$$

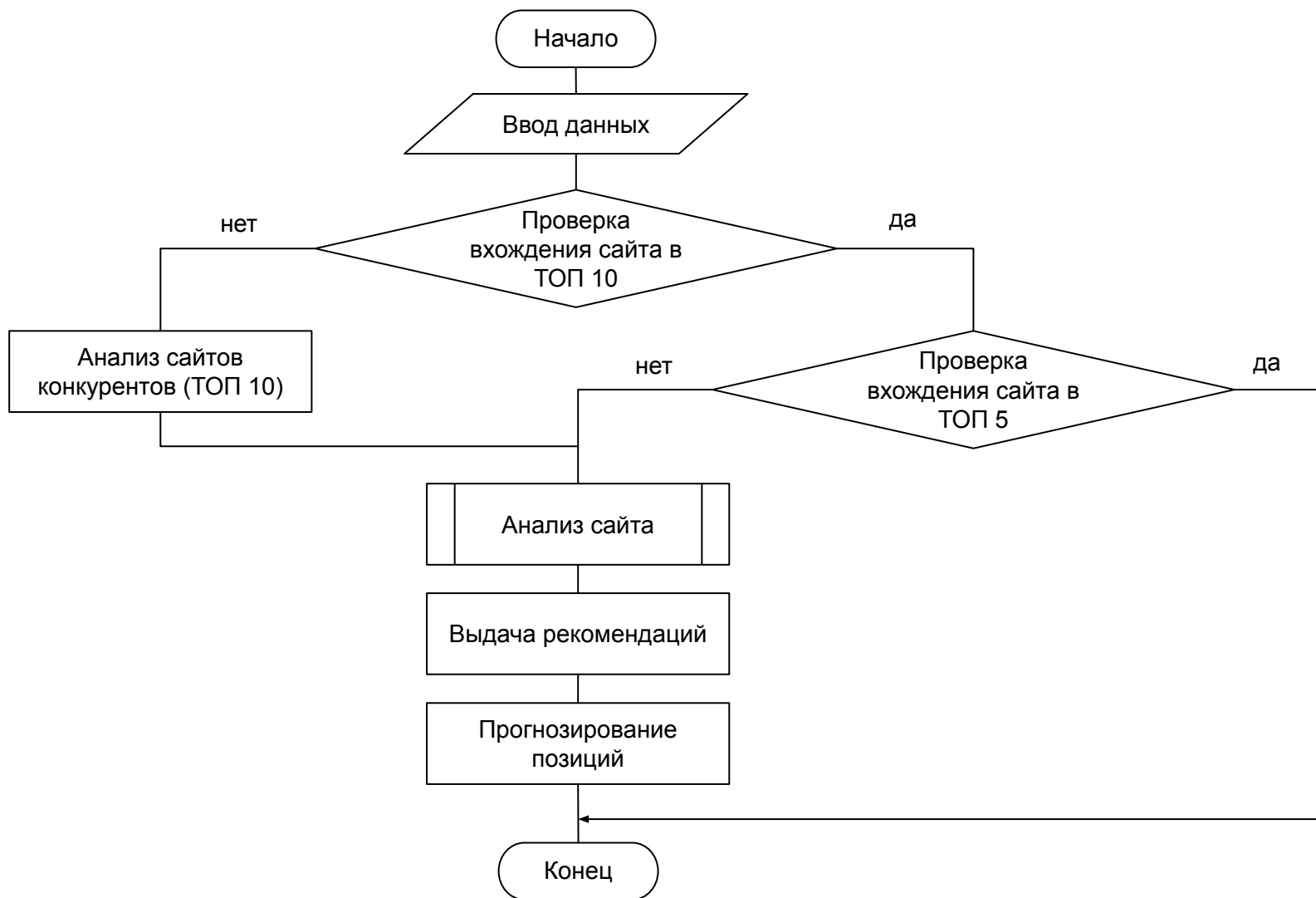
где d – демпфирующий коэффициент, отображающий «количество авторитетности»,

n – количество страниц, на которых установлены обратные ссылки,

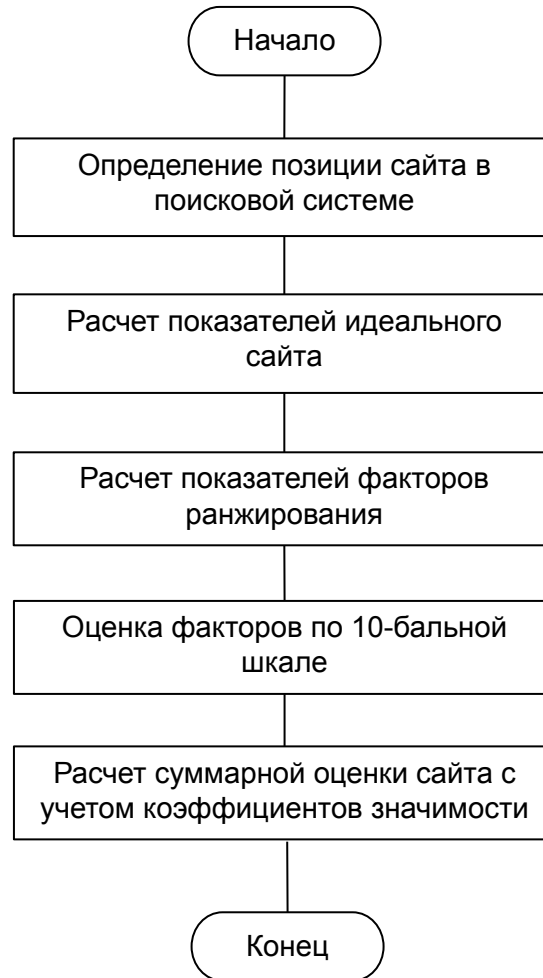
C – общее количество внешних ссылок, установленных на странице,

T (от 1 до n) – номера ссылающихся страниц.

АЛГОРИТМ РАБОТЫ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ПОЗИЦИЙ САЙТОВ



АЛГОРИТМ РАБОТЫ МОДУЛЯ АНАЛИЗА САЙТА



ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

1	A	B	C		D		E		F	
			Использование при ранжировании		Коэффициент значимости					
2			Google	Yandex	Google	Yandex				
3	Домен	Возраст домена	1	1	0.32	0.41				
4		История домена	1	1	0.11	0.09				
5		Ключевые слова в доменном имени	1	1	0.13	0.14				
6		Субдомен или нет	1	1	0.21	0.44				
7		Размещение домена	1	0	0.17	0				
8		IP домена	0	0	0	0				
9		Размещение IP адреса / сервера	0	1	0	0.21				
10		История санкций домена	1	1	0.39	0.39				
11		Ключевое слово в имени субдомена	1	1	0.40	0.31				
12		Правильность регистрационных данных домена	1	1	0.22	0.27				
13		Alexa Rank	1	1	0.19	0.12				
14		ТИЦ домена	0	1	0	0.7				
15										
16		Сервер	Недоступность сайта	1	1	0.5	0.54			
17	Долгий отклик сайта		1	1	0.28	0.3				
18	Uptime сервера		1	1	0.28	0.29				
19										
20	Структура сайта	Использование Header тегов	1	1	0.22	0.19				
21		Вид URL	1	1	0.48	0.35				
22		Использование внешних CSS/JS файлов	1	1	0.51	0.60				
23		Ключевое слово в имени папки	1	1	0.08	0.11				
24		Ключевое слово в параметрах URL	1	1	0.33	0.31				
25		Ключевое слово в расширении файла страницы	0	1	0	0.09				
26		Динамические URL (= вид URL)	1	1	0.4	0.46				
27		Использование SSL	1	1	0.7 *	0.65 *				
28		Использование Google Search Appliance	1	0	0.57	0				
29		Содержимое страницы	1	1	0.32	0.35				
30		Плотность ключевых слов	1	1	0.04	0.11				
31		Ключевое слово в теге Title	1	1	0.11	0.15				
32		Ключевое слово в Meta Description	0	1	0	0.12				
33		Ключевое слово (H, H ...)	1	1	0.09	0.11				
34		Ключевое слово в содержимом страницы	1	1	0.14	0.14				
35		Новизна контента	1	1	0.64	0.73				
36		Ключевое слово URL	1	1	0.05	0.21				
37		Ключевое слово в имени файлов	1	1	0.01	0.02				
38		CSS валидация	1	1	0.33	0.31				
39		Наличие рекламы	-1	1	0.01	0.01				
40		Наличие AdSense	-1	1	0.07	0.01				
41		Наличие в Google news RSS ленты сайта	1	0	0.22	0				
42		Адаптивность	1	1	0.8 *	0.3				

СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЗИЦИЙ САЙТОВ



[Главная](#) [Инструменты](#) [Аккаунт](#) ▾

Сайтовед Online!

Сервис анализа сайтов и прогнозирования их позиций в результатах поисковой выдачи

[Начать использование](#)



Более 30 инструментов для анализа вашего сайта

[Просмотр инструментов](#)

Вот лишь некоторые из наших инструментов ...



Прогноз обратных
ссылок



Анализ ссылок



Сравнить
PageRank



Детали браузера



Мой IP Address



Обфускация Email



Генератор мета-
тегов



Получить размер
веб-страницы



Whois сервис



Количество
обратных ссылок

Логин

Пароль

[Регистрация](#)

[Вход](#)

СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЗИЦИЙ САЙТОВ

Инструменты для анализа вашего сайта

Добро пожаловать, admin!



Сравнить PageRank



Проверка на адаптивность



Видимость сайта поисковой системой



Генератор ключевых слов



Генератор Sitemap



Обфускация Email



Check Competition



Узнать IP Address сайта



Проверка на наличие фильтров



Узнать Ваш IP адрес



Количество обратных ссылок



Проверка доступности сайта



Проверка скорости сайта



Получить Мета-теги



Информация о браузере



Оценка PageRank



Проверка заголовков



Ссылки с государственных ресурсов



Whois сервис



Генератор Robots.txt



Анализ ссылок

СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЗИЦИЙ САЙТОВ

Критические факторы

✓ Доступность для поисковиков

✓ Доступность файла robots.txt

Статус проверки:

Файл с информацией для поисковых роботов `robots.txt` найден и целевой url разрешен к индексированию

Объяснение:

В `robots.txt` содержится информация для поисковых роботов - какие страницы открыты или закрыты для индексации. Убедитесь, что в `robots.txt`

Если Вы уверены, что Ваш `robots.txt` файл не содержит запрета на индексирование запрошенного URL - проверьте кодировку файла. UTF-8 (согласно спецификации Google).

✗ Использование ключевого слова в тексте

На вашей странице нет ни одного ключевого слова, по которому Вы производите анализ сайта.

Рекомендации:

Применяйте ключевое слово, по крайней мере, один раз в тексте документа на этой странице.

✗ Использование ключевого слова в заголовке страницы

СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЗИЦИЙ САЙТОВ

Другие факторы

✖ Оптимальная длина Заголовка страницы

Статус проверки:

Заголовок страницы: "ОмГТУ: Главная. Омский Государственный технический университет. Официальный сайт."

Длина заголовка: 81 символов

Объяснение:

В выдаче поисковой системы отображается примерно 66 символов текста с заголовков страниц. В случае, если заголовок длиннее, он «обрезается» заголовок не более 66 символов длиной.

✖ Выделение ключевых слов жирным, курсивом, подчеркиванием и т.д.

✖ Использование ключевого слова в Meta описании

✔ Использование META-описаний

✔ Избегайте излишнего количества внутренних ссылок

Статус проверки:

Кол-во внутренних ссылок: 94 [показать ссылки](#)

Объяснение:

Наиболее оптимальное решение - удаление «сквозных» ссылок с сайта. Это позволяет точно организовать структуру сайта. Помимо всего, большое количество ссылок говорит о попытке манипуляции выдачей и может привести к низкому ранжированию сайта по соответствующим запросам. При необходимости размыслите постарайтесь изменить ссылочные якоря.

СПИСОК НАУЧНЫХ ПУБЛИКАЦИЙ

1. Дьяконов Д.В. Проблемы продвижения сайтов в современных поисковых системах / Д.В. Дьяконов // Международная научно-практическая конференция «Наука и образование в XXI веке»: тез. докл. – Тамбов, 2014. – С. 123-125.
2. Дьяконов Д.В. Влияние факторов ранжирования на позиции сайтов в поисковых системах / Д.В. Дьяконов., П.В. Пестерев., А. П. Рудюк., А.Г. Янишевская // Инженерный вестник Дона: тез. докл. – Ростов-на-Дону, 2014. – С. 131-140.
3. Дьяконов Д.В., Пестерев П.В. Сервис автоматической генерации карты сайта. ОФЭРНиО. Свидетельство № 20682 от 16.12.2014.