

УДК 004.9

ИССЛЕДОВАНИЕ ЗНАЧИМОСТИ ПРИЗНАКОВ РАНЖИРОВАНИЯ РЕЗУЛЬТАТОВ РАБОТЫ ПОИСКОВЫХ СИСТЕМ

Зосимов В.В.

*Миколаївський національний університет ім. В.О. Сухомлинського,
вул. Никольська, 24, Миколаїв, 54030, Україна*

zosimovvv@bk.ru

У статті описаний експеримент з просування веб-ресурсу у видачі пошукових систем. В ході експериментів було вивчено вплив кількісних показників ознак ранжирування на позиції веб-ресурсів в рейтингу пошукових систем. Ознаки ранжирування для експериментів були відібрані з використанням узагальненого ітераційного алгоритму методу групового обліку аргументів.

Ключові слова: пошукова система, ранжування, МГУА, веб-ресурси, пошук інформації.

The article describes an experiment to promote a web-resource in search results. During the experiments, the effect of quantitative indicators ranking features the position of web resources in the ranking of search engines was studied. Ranking signs for the experiments were selected using a generalized iterative algorithm for the group method of data handling.

Keywords: search engine rankings, GMDH, web resources, information search.

В статье описан эксперимент по продвижению веб-ресурса в выдаче поисковых систем. В ходе экспериментов было изучено влияние количественных показателей признаков ранжирования на позиции веб-ресурсов в рейтинге поисковых систем. Признаки ранжирования для экспериментов были отобраны с использованием обобщенного итерационного алгоритма метода группового учета аргументов.

Ключевые слова: поисковая система, ранжирование, МГУА, веб-ресурсы, поиск информации.

Введение

Сегодня сложно представить себе человека, которых бы хоть раз в день не пользовался поисковыми системами в сети Интернет. По мере роста и развития всемирной паутины росли и требования к поисковым системам. Изменялись алгоритмы их работы. Непрерывно растущее количество веб-ресурсов вынуждало разработчиков поисковых систем искать пути ускорения работы своих алгоритмов. Но скорость обработки данных не единственная и не самая важная проблема, которая возникала на пути разработчиков. Основной задачей является создание оптимальных моделей ранжирования результатов поиска. Они должны отвечать ряду требований:

- удовлетворять информационные потребности всех пользователей;
- обеспечивать высокие значения полноты и точности поиска;
- эффективно отсеивать поисковый спам, сайты-клоны;

- иметь сложную внутреннюю структуру для минимизации возможности внешнего воздействия на результаты выдачи.

Поисковые системы не всегда могут эффективно выполнять свои функции. С одной стороны это обусловлено развитием сети и появлением в ней новых типов электронных информационных ресурсов. С другой же стороны стремительное развитие сферы мобильных приложений, облегчающих жизнь пользователей, делает их более требовательными к качеству предоставляемых им информационных услуг.

1. Построение модели ранжирования

В качестве базы для проведения экспериментов была выбрана поисковая система Google [1], как наиболее востребованная и технически оснащенная. Для получения наиболее точных результатов необходимо построить актуальную модель ранжирования поисковой выдачи. Для этого была использована методика, описанная в [2]. В качестве признаков ранжирования были использованы следующие критерии:

1. Количество ключевых слов на сайте.
2. Количество ключевых слов на странице.
3. Соотношение общего числа слов на сайте к количеству ключевых слов на сайте.
4. Соотношение общего числа слов на странице к количеству ключевых слов на странице.
5. Google PR.
6. Популярность тематики.
7. Число запросов по конкретному ключевому слову за определённый период времени.
8. Общее количество веб-страниц сайта.
9. Объем текста сайта.
10. Объем сайта.
11. Объем текста веб-страницы.
12. Возраст сайта.
13. Наличие ключевого слова в URL сайта (имя домена)
14. Периодичность обновления информации на сайте.
15. Последнее обновление страниц сайта.
16. Число картинок (рисунков) на сайте.
17. Количество мультимедийных файлов.
18. Наличие замещающих надписей на рисунках (картинках).
19. Длину (в количестве символов) замещающих надписей рисунков (картинок).
20. Использование фреймов.

21. Язык сайта.
22. Размер шрифта, которым оформлены ключевые слова.
23. Жирность шрифта ключевых слов.
24. Написаны в разрядку или нет ключевые слова.
25. Написаны или нет заглавными буквами ключевые слова.
26. Как далеко от начала веб-страницы располагаются ключевые слова.
27. Наличие ключевых слов в заголовке.
28. Наличие ключевых слов в мета-тэгах.
29. Наличие файла «robot.txt».
30. Географическое месторасположение сайта.
31. Комментарии внутри программного кода сайта.
32. К какому типу страниц относится каждая страница сайта: html или asp.
33. Наличие в составе сайта flash модулей.
34. Наличие в составе сайта веб-страниц с незначительными отличиями друг от друга.
35. Соответствие ключевых слов сайта тому разделу каталога поисковой машины, в котором зарегистрирован сайт.
36. Наличие «шумовых слов» («стоп слов»).
37. Общее количество гиперссылок сайта.
38. Количество внутренних гиперссылок сайта.
39. Количество внешних гиперссылок сайта.
40. Глубина сайта.
41. Количество внешних ссылок, содержащих в названии ключевое слова.

Для построения модели ранжирования использовался индуктивный подход, а именно обобщенный итерационный метод группового учета аргументов [3].

Модель ранжирования:

$$y = 0,76x_2 + 1,84x_3 + 0,15x_4 + 0,008x_7 - 1,9x_{11} + 0,012x_{22} - 14,8x_{27} - x_{28} - 14,33x_{35} + 2,4x_{40} - 0,3x_{41} - 8,8x_5x_6 + 0,26x_{14}x_{15}^2 + 0,04x_{37}x_{38}x_{39}$$

Из модели видно, что наиболее значимыми для ранжирования являются следующие признаки:

x_2 - количество ключевых слов на странице.

x_3 - соотношение общего числа слов на сайте к количеству ключевых слов на сайте.

x_4 - соотношение общего числа слов на странице к количеству ключевых слов на странице.

- x_5 - Google PR.
- x_6 - Популярность тематики.
- x_7 - Число запросов по конкретному ключевому слову за определённый период времени.
- x_{11} - Объём текста веб-страницы.
- x_{14} - Периодичность обновления информации на сайте.
- x_{15} - Последнее обновление страниц сайта.
- x_{22} - Размер шрифта, которым оформлены ключевые слова.
- x_{27} - Наличие ключевых слов в заголовке.
- x_{28} - Наличие ключевых слов в мета-тэгах.
- x_{35} - Соответствие ключевых слов сайта тому разделу каталога поисковой машины, в котором зарегистрирован сайт.
- x_{37} - Общее количество гиперссылок сайта.
- x_{38} - Количество внутренних гиперссылок сайта.
- x_{39} - Количество внешних гиперссылок сайта.
- x_{40} - Глубина сайта.
- x_{41} - Количество внешних ссылок, содержащих в названии ключевое слово.

2. Экспериментальная проверка важности признаков ранжирования.

Для доказательства важности полученных в модели признаков ранжирования был проведен эксперимент по продвижению веб-ресурса. На каждом этапе эксперимента искусственно завышались показатели определенных параметров модели и оценивались позиции в поисковой выдаче Google.

Веб-ресурс: mebilite.kiev.ua

Ключевые фразы: кухни Киев, мебель на заказ Киев, шкафы купе Киев.

Позиции в поисковой выдаче до начала поисковой оптимизации:

кухни Киев - 46

мебель на заказ Киев - 51

шкафы купе Киев -72

Этап 1. Добавление ключевых слов в заголовки и мета-тэги.

Таблица 1 иллюстрирует график изменения позиций на протяжении шести недель после внесения изменений.

Таблица 1.

График изменения позиций на протяжении шести недель после внесения изменений.

Ключевые фразы	Продолжительность продвижения					
	1 неделя	2 недели	3 недели	4 недели	5 недель	6 недель
кухни Киев	35	32	28	26	25	25
мебель на заказ Киев	46	40	35	27	23	22
шкафы купе Киев	57	50	44	36	36	35

На шестой неделе рост позиций сайта почти прекратился. Для дальнейшего роста позиций необходимо оптимизировать больше параметров.

Этап 2. Размещение внешних ссылок на веб-ресурс

Таблица 2 иллюстрирует график изменения позиций на протяжении шести недель после внесения изменений.

Таблица 2.

График изменения позиций на протяжении шести недель после внесения изменений.

Ключевые фразы	Продолжительность продвижения					
	1 неделя	2 недели	3 недели	4 недели	5 недель	6 недель
кухни Киев	25	20	18	17	15	14
мебель на заказ Киев	22	20	20	16	13	13
шкафы купе Киев	35	30	26	24	21	20

Этап 3. Написание уникального текста с ключевыми словами для главной страницы сайта.

Таблица 3.

График изменения позиций на протяжении шести недель после внесения изменений.

Ключевые фразы	Продолжительность продвижения					
	1 неделя	2 недели	3 недели	4 недели	5 недель	6 недель

Продолжение таблицы 3

кухни Киев	14	11	8	6	6	6
мебель на заказ Киев	13	8	7	7	5	4
шкафы купе Киев	20	20	16	15	10	9

4. Выводы

Проведенные эксперименты по продвижению веб-ресурсов подтверждают достоверность полученной с помощью обобщенного итерационного алгоритма МГУА модели ранжирования поисковой системы Google. Экспериментально доказано, что постепенное наращивание значений основных признаков ранжирования, выделенных на основе модели, позволяет добиться высоких результатов в поисковом продвижении веб-ресурсов.

Литература

1. <http://www.google.com> - главная страница поисковой системы Google
2. Zosimov V., Stepashko V., Bulgakova O. Inductive building of search results ranking models to enhance the relevance of the text information retrieval. – Proc. of the 26th Intern. Workshop “Database and Expert Systems Applications, 1-4 Sept., Valencia, Spain / Ed. by Markus Spies at al. – Los Alamitos: IEEE Computer Society, 2015. – 316 p. / – P. 291-295. – ISSN: 1529-4188.
3. Stepashko V., Bulgakova O., Zosimov V. Performance of Hybrid Multilayered GMDH Algorithm. – Proceedings of the III International Workshop on Inductive Modelling IWIM-2011, 5-9 July 2011, Kyiv-Zhukyn, Ukraine. – Kyiv: IRTC ITS NANU, pp 109-113, 2011.