

Модель суммарной оценки сайта в сети Интернет на основе факторов ранжирования

П.В. Пестерев, А.Г. Янишевская

Омский государственный технический университет, Омск

Аннотация: В статье описаны модель идеального сайта с точки зрения поисковой системы на основании поискового запроса. Данная модель учитывает различные факторы ранжирования, к которым относятся поведенческие факторы ранжирования, внешние факторы ранжирования и внутренние факторы ранжирования. Описана модель суммарной оценки сайта в сети Интернет в зависимости от факторов ранжирования.

Ключевые слова: фактор ранжирования, сортировка сайтов, оценка сайта, поисковая система, оптимизация сайта, Интернет.

Введение

В настоящее время существует большое количество поисковых систем в сети Интернет, наиболее популярными из которых в России являются Yandex, Google, поиск Mail и т.д. [1]. Данные ресурсы обеспечивают поиск необходимой информации в зависимости от запросов пользователей. При этом необходимо, чтобы поисковой ресурс выводил в начало списка сайты, информация на которых более точно соответствует запросам. К таким сайтам, исходя из исследований [2], относятся сайты, на которых учтено больше групп факторов (особенно более значимых из них) и проведена качественная поисковая оптимизация [3-5]. Сортировка сайтов осуществляется за счет факторов ранжирования: внутренние факторы ранжирования, внешние факторы ранжирования и поведенческие факторы ранжирования [6]. Использование тех или иных факторов имеют определенные преимущества и недостатки. Существует возможность накручивания данных факторов разработчиками сайтов с целью улучшения положения разработанного ресурса в поисковой выдаче. Однако это может привести к санкциям со стороны поисковых систем. Также в последнее время поисковые системы начали учитывать и социальные сети. Исследования

социальных сетей проводится достаточно давно [7, 8], но совсем недавно замечено их влияние на позиции сайтов в поисковых системах.

Тем не менее, такой сайт, по крайней мере, какое-то время будет ухудшать поиск. Поэтому необходимо совершенствовать алгоритмы и математические модели, на основе которых работают поисковые системы. Наименее подвержены накручиванию поведенческие факторы ранжирования, т.е. факторы, учитывающие предпочтения реальных посетителей сайтов. Анализ таких факторов, включающий анализ внешней среды, отслеживание ведущих рыночных трендов и рыночной конъюнктуры [9], намного более объективен, т.к. работает с предпочтениями большой целевой аудитории, являясь неотъемлемым элементом эффективного функционирования любого сайта. Поэтому становится актуальным создание алгоритмов, в которых данные факторы занимают основное положение. В то же время происходит постепенный отказ от внешних факторов, которые чаще всего искусственно завышают. Например, одним из интересных нововведений Яндекса должен стать алгоритм ранжирования без учета ссылок, проанонсированный руководителем поисковых систем Яндекса 5 декабря 2014 г. Александром Садовским на конференции IBS Russia [10]. В данной статье описывается предложенная модель суммарной оценки сайта в сети Интернет на основе факторов ранжирования. Данная модель является основой для дальнейшего совершенствования и перехода к модели, в которой планируется отдать основное предпочтение поведенческим факторам ранжирования и уменьшить влияние внешних факторов.

Модель идеального сайта

С точки зрения поисковых систем идеальным сайтом можно назвать сайт, занимающий первую позицию в результатах выдачи в зависимости от

поискового запроса. Однако сайта, идеально соответствующего всем поисковым запросам, не существует.

Для построения идеальной модели сайта был проведен подробный анализ факторов ранжирования первых десяти сайтов в результатах поисковой выдачи Yandex и Google по определенным поисковым запросам и выбраны максимальные значения по каждому из факторов.

$$Y_i = f(\alpha_i)_{max} \quad (1)$$

где i – порядковый номер фактора;

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

Данный подход предполагает то, что сайт, имеющий первую позицию в поисковой системе, не является идеальным, т.к. значения показателей его факторов ранжирования не будут максимальными.

Определение для идеального сайта значений факторов ранжирования необходимо с целью построения модели суммарной оценки сайта, находящегося в зависимости от запросов пользователей на любой позиции в результатах поисковой выдачи.

Модель суммарной оценки сайта в сети Интернет

Суммарную оценку сайта предлагается производить из расчета показателей факторов ранжирования идеального сайта (см. формула 1) и оценки каждого из них по десятибалльной шкале.

Затем с целью оценки факторов ранжирования предполагается произведение сравнения значений факторов анализируемого сайта с соответствующими значениями факторов идеального сайта.

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

где $f(\alpha_i)$ – значение фактора анализируемого сайта;

Y_i – значение фактора идеального сайта;

10 – предложенный множитель.

Значения показателей для каждого из факторов либо берется из сервисов Яндекс.Метрика и Google Analytics, либо рассчитывается по соответствующим формулам.

Значение фактора скорость загрузки сайта производится по формуле:

$$f(\alpha) = t_a + t_h + t_c + t_j \quad (3)$$

где t_a – время ответа сервера;

t_h – время загрузки HTML документа;

t_c – время загрузки таблицы стилей;

t_j – время загрузки JavaScript.

Определение значения фактора «Google PageRank» производится по формуле:

$$f(\alpha) = (1 - d) + d \left(\frac{PR(\tau_1)}{c(\tau_1)} + \dots + \frac{PR(\tau_n)}{c(\tau_n)} \right) \quad (4)$$

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

n – количество страниц с установленными обратными ссылками;

C – общее количество установленных на странице внешних ссылок;

T – номера ссылающихся страниц, $1 \leq T \leq n$.

$\frac{f(\alpha_i)}{n} = 1$ при условии, если значение фактора анализируемого сайта

больше значения фактора идеального сайта.

Значение фактора анализируемого сайта в большинстве случаев не превышает значения фактора идеального сайта, но возможны исключения.

Например, анализ факторов ранжирования первых десяти сайтов в поисковых системах Яндекс по поисковому запросу «Вузы Омска» показывал максимальное значение показателя одного из факторов «ТИЦ» (тематический индекс цитирования) меньше значения показателя «ТИЦ» сайта ОмГТУ, не входящего в список первых десяти сайтов в поисковой выдаче по заданному поисковому запросу. В этом случае оценка фактора «ТИЦ» для сайта ОмГТУ при составлении суммарной оценки составляет десять баллов по десятибалльной шкале.

Расчет суммарной оценки сайта производится на основе расчета суммы произведений коэффициентов значимости факторов ранжирования на оценку факторов по десятибалльной шкале:

$$P_{sq} = \sum_{i=1}^N (\sigma_i * f(\varphi_i)) \quad (5)$$

где i – порядковый номер фактора;

σ_i – коэффициент значимости фактора;

$f(\varphi_i)$ – оценка фактора по десятибалльной шкале.

Вычисление суммарной оценки сайта необходимо для анализа сайтов по заданным поисковым запросам, а также с целью прогнозирования позиций анализируемого сайта. Сравнение суммарной оценки анализируемого сайта с суммарной оценкой сайта, расположенного на десятой позиции в поисковой

система, возможно, даст оценку вероятности попадания сайта в первые десять в поисковой выдаче. На основании данных результатов можно создать математическую модель системы автоматического ранжирования сайтов в поисковых системах в зависимости от запросов пользователей, более точно учитывающую именно поведенческие факторы ранжирования и снижая значимость внешних факторов.

Заключение

Анализ страниц с результатами поиска по определенному поисковому запросу показал, что при построении поисковыми системами Яндекс и Google поисковой выдачи используется огромное количество факторов, которые возможно оценить.

С точки зрения поисковых систем идеальным сайтом можно назвать сайт, расположенный на первой позиции в результатах поисковой выдачи. Значения факторов ранжирования для таких сайтов по выбранному поисковому запросу выше. На основании данной информации была предоставлена модель идеального сайта.

Увеличение значений показателей факторов ранжирования анализируемого сайта до значений показателей идеального сайта будет приводить к его высоким позициям в результатах выдачи в поисковой системе. Также необходимо учитывать тот факт, что значения показателей факторов ранжирования сайта, находящегося на первой позиции в результатах поисковой выдачи, не всегда являются максимальными.

Также была описана модель суммарной оценки сайта, анализирующая позицию любого сайта в поисковой системе. Позиция сайта в поисковой системе зависит от суммарной оценки сайта, чем она больше, тем выше сайт располагается. На основании данных результатов, как уже было сказано, можно создать математическую модель системы автоматического ранжирования сайтов в поисковых системах в зависимости от запросов

пользователей, более точно учитывающую именно поведенческие факторы ранжирования и снижая значимость внешних факторов, которые больше подвержены искусственному накручиванию.

Литература

1. Блог веб-студии RevolWeb / Самые популярные поисковые системы на 2015 год, 2015. URL: revolweb.ru/prodvizhenie-sajtov/samye-populyarnye-poiskovye-sistemy-na-2015-god.

2. Tober M. Searchmetrics Ranking Factors 2014: Why Quality Content Focuses on Topics, not Keywords // Moz URL: moz.com/blog/searchmetrics-ranking-factors-2014 (date: 2014).

3. Ашманов И.С., Иванов А.А. Оптимизация и продвижение сайтов в поисковых системах. СПб.: Питер, 2011. 464 с.

4. Enge E., Spencer S., Stricchiola J., Fishkin R. The Art of SEO. O'Reilly Media, 2012. 718 p.

5. Этемеэ Е. Поисковая оптимизация от А до Я. М.: Вильямс, 2008. 598 с.

6. Неелова Н. Sembook. Энциклопедия поискового продвижения Ingate. СПб.: Питер, 2014. 520 с.

7. Pinheiro Carlos A.R. Social Network Analysis in Telecommunications. Hoboken: John Wiley & Sons, 2011. 288 p.

8. Розин М.Д., Свечкарев В.П., Конторович С.Д., Литвинов С.В., Носко В.И. Исследование социальных сетей как площадки социальной коммуникации рунета, используемой в целях предвыборной агитации // Инженерный вестник Дона, 2011, №1 URL: ivdon.ru/ru/magazine/archive/n1y2011/397.

9. Берёза Н.В. Современные тенденции развития мирового и российского рынка информационных услуг // Инженерный вестник Дона, 2012, №2 URL: ivdon.ru/magazine/archive/n2y2012/758



10. GM site / Тенденция рынка услуг комплексного интернет-маркетинга в 2014 году, 2014. URL: gmsite.ru/about/news/Market+trends/.

References

1. Blog of webstudio RevolWeb. URL: revolweb.ru/prodvizhenie-sajtov/samyepopulyarnye-poiskovye-sistemy-na-2015-god.

2. Tober M. Searchmetrics Ranking Factors 2014: Why Quality Content Focuses on Topics, not Keywords. URL: moz.com/blog/searchmetrics-ranking-factors-2014 (date: 2014).

3. Ashmanov I.S., Ivanov A.A. Optimizaciya i prodvizhenie saitov v poiskovih sistemah [Optimization and promotion of website in search engine]. SPb.: Piter, 2011. 464 p.

4. Enge E., Spencer S., Stricchiola J., Fishkin R. The Art of SEO. O'Reilly Media, 2012. 718 p.

5. Etemez E. Poiskovaya optimizaciya ot A do Ya [Search engine optimization from A to Z]. M.: Viliams, 2008. 598 p.

6. Neelova N. Sembook. Enciklopediya poiskovogo prodvizheniya Ingate [Sembook. Encyclopedia of search engine promotion Ingate]. SPb.: Piter, 2014. 520 p.

7. Pinheiro Carlos A.R. Social Network Analysis in Telecommunications. Hoboken: John Wiley & Sons, 2011. 288 p.

8. Rozin M.D., Svechkarev V.P., Kontrovich S.D., Litvinov S.V., Nosko V.I. Inzhenernyj vestnik Dona (Rus), 2011, №1 URL: ivdon.ru/ru/magazine/archive/n1y2011/397.

9. Bereza N.V. Inzhenernyj vestnik Dona (Rus), 2012, №2 URL: ivdon.ru/magazine/archive/n2y2012/758.

10. GM site. URL: gmsite.ru/about/news/Market+trends/.