

В. О. Рахуба, к. т. н.

ЗАДАЧИ АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ ПОИСКОВОЙ ОПТИМИЗАЦИЕЙ ИНТЕРНЕТ-РЕСУРСОВ

Статья посвящена исследованию поисковой оптимизации интернет-ресурсов как объекта управления. Формализована задача оптимизации и выделены ее составляющие. Приведены результаты эксперимента по влиянию основных факторов на позиции сайтов в поисковой системе Яндекс.

Ключевые слова: поисковая оптимизация web-сайтов (SEO), автоматизированное управление, математическая модель.

Постановка проблемы. На сегодняшний день эффективность работы любого предприятия в значительной степени зависит от способности находить новые рынки сбыта своей продукции. В условиях высокой конкуренции за клиента, которая характерна для экономических отношений в нашей стране и во многих странах мира, одним из важнейших факторов успешной работы предприятия является оптимизация его интернет-ресурса по позиции в выдаче поисковых систем. Среди специалистов, работающих над данным вопросом, эта задача получила устоявшееся название SEO (search engine optimization).

Анализ достижений. Анализ работ, посвященных данной проблеме [1, 2, 3], которых сегодня немало, позволяет сделать вывод, что, несмотря на попытки автоматизации SEO, подавляющее большинство специалистов по оптимизации сайтов выполняет свою работу «в ручном режиме», опираясь на эмпирические результаты собственных экспериментов. Однако следует сразу отметить, что процесс SEO непременно имеет творческую составляющую, которая практически не может быть автоматизирована; следовательно, актуальным является вопрос разработки алгоритмов автоматизированного управления теми этапами SEO, которые могут быть формализованы, и создание на основе этих алгоритмов программного обеспечения для управления процессом поисковой оптимизации сайтов.

Постановка задачи. Целью данного исследования является: 1) определение составляющих SEO, которые могут быть автоматизированы; 2) постановка задач автоматизированного управления; 3) определение проблем автоматизации SEO и поиск путей их решения.

Основной материал и исследования. Процесс поисковой оптимизации web-ресурса как объект автоматизации можно представить схемой (см. рис. 1).

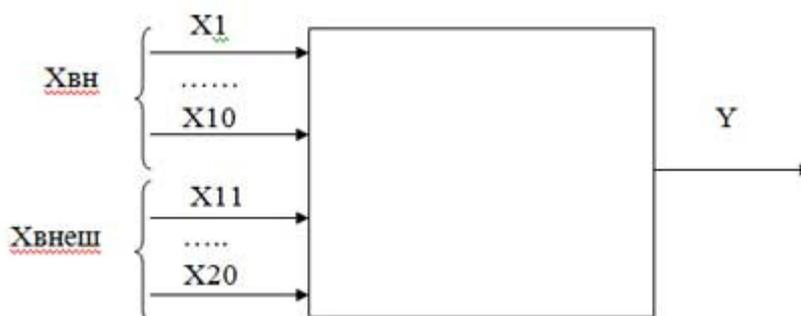


Рис. 1. SEO как объект управления

Откликом Y в данной модели является позиция, которую занимает сайт в выдаче определенной поисковой системы, для которой осуществляется оптимизация.

Все факторы можно разделить на две группы: внутренние $X_{вн}$, характеризующие внутренние свойства web-страницы, продвигаемой в поисковой выдаче, и внешние $X_{внеш}$, характеризующие конкурентность поискового запроса, по которому продвигается страница.

К $X_{вн}$ можно отнести [1 – 5]:

X_1 – возраст домена; X_2 – ТИЦ (или PR) сайта; X_3 – количество и качество внешних ссылок на страницу; X_4 – количество ссылок, исходящих со страницы (в частности, так называемая «внутренняя перелинковка»); X_5 – точное вхождение поискового запроса в теги «Title» и «Description»; X_6 – наличие сайта в каталоге данной поисковой системы; X_7 – наличие файла robots.txt; X_8 – наличие файла sitemap.xml; X_9 – количество точных вхождений поискового запроса в текст страницы; X_{10} – качество представления информации на странице (содержательность, наличие нумерованных списков, подзаголовки, изображения и т. д.).

К $X_{внеш}$ относят [1 – 5]:

X_1 – общее количество ответов по данному запросу; X_2 – количество главных страниц в ТОПе поисковой выдачи; X_3 – среднее количество и качество ссылок на сайты в выдаче; X_4 – средний возраст сайтов в выдаче; X_5 – количество точных вхождений поискового запроса в теги «Title» и «Description» сайтов в выдаче; X_6 – наличие блоков контекстной рекламы в выдаче поисковой системы; X_7 – наличие страниц Википедии и Ютуба по данному запросу; X_8 – наличие географической карты в результатах выдачи; X_9 – среднее значение ТИЦ сайтов в выдаче; X_{10} – другие внутренние характеристики сайтов в поисковой выдаче.

Итак, задачу оптимизации поискового продвижения web-ресурса целесообразно представить в виде

$$Y = f(X_1, \dots, X_{22})$$

$$Y \rightarrow \min$$

При чем ограничения на процесс определяют исходя из бюджета, выделяемого на данный проект.

В целом в процессе поисковой оптимизации можно выделить следующие составляющие:

1. Определение ограничений для задачи поисковой оптимизации данного информационного ресурса:

а) анализ предметной области, в которой будут продвигать сайт с целью исследования частотности и конкурентности поисковых запросов;

б) разработка вариантов продвижения сайта и расчет вариантов бюджета для их реализации;

с) выбор варианта поискового продвижения сайта.

2. Синтез семантического ядра информационного ресурса:

а) формирование базы потенциальных поисковых запросов, по которым может продвигаться данный ресурс;

б) отбор оптимальных запросов в соответствии с критерием управления;

с) построение двух- или трехуровневой структуры семантического ядра по частотности запросов.

3. Собственно оптимизация страниц сайта по выбранным запросам путем управления внутренними факторами в приведенной выше модели:

а) создание рационального дизайна сайта;

б) написание текстов с качественным представлением информации и оптимальным количеством вхождений запросов;

с) рациональное заполнение тегов страниц и главной страницы сайта;

д) создание файлов robots.txt и sitemap.xml, регистрация в каталогах;

е) создание рациональной внутренней перелинковки;

f) закупка рекламных ссылок с других сайтов в соответствии с выбранным вариантом бюджета.

Для задач автоматизированного управления наибольший интерес представляют пункты 1b, 2b и 2c, поскольку они требуют разработки моделей продвижения web-ресурсов в поисковых системах, учитывающих влияние приведенных выше многочисленных факторов на позицию в выдаче, а также требуют синтеза критериев отбора запросов, которые обеспечат решение поставленной задачи оптимизации. Также частично можно автоматизировать пункты 1a, 2a, 3e и 3f, тогда как большинство составляющих пункта 3, требующих творческого труда людей, не могут быть формализованы. На этом этапе работой системы управления оптимизацией web-ресурсов должно быть предоставление персоналу массива поисковых запросов, сформированного в виде иерархической структуры, необходимого количества точных вхождений соответствующих запросов на соответствующих страницах и схемы перелинковки страниц с указанием нужных анкоров. Пункт 1c предусматривает согласование с заказчиком проекта. Также актуальна разработка подсистемы прогнозирования трафика на сайт в зависимости от выбранного варианта оптимизации его поискового продвижения.

Для установления зависимостей, составляющих основу алгоритмов управления, необходимо провести исследования выдачи поисковых систем. Очевидно, что для предварительного исследования влияния указанных выше факторов на отклик может быть выбран только метод пассивного эксперимента. При чем целесообразно провести несколько различных экспериментов, выбирая для каждого определенный набор внутренних или внешних факторов. Для того чтобы результаты обработки пассивного эксперимента имели смысл, целесообразно выбирать не более, чем четыре фактора для одной серии исследований [6].

Для такого эксперимента выбрана образовательная тематика, в частности предметная область – профориентация. Результаты такого исследования могут быть интересны вузам и другим организациям, которые предоставляют образовательные услуги.

Проведены две группы по три серии исследований ТОП-10 выдачи поисковой системы Яндекс. Первая группа позволила учесть влияние внутренних факторов в каждой серии, эксперименты проводили по запросам Z1 – «Как выбрать профессию», Z2 – «Какую профессию выбрать», Z3 – «Тест какую профессию выбрать», имеющие частотность соответственно 1542, 637 и 213 показов в месяц. В качестве факторов выбраны X5, X9, X3/10, X2/10. Исследования второй группы учитывали влияние внешних факторов, эксперименты проводили для сайтов моеобразование.ru, kartaznaniy.ru и moya-ramyat.ru; факторами выступали X1/млн, X9/10, X3/10, X5.

Для анализа собранных данных использовали методику обработки результатов пассивных экспериментов, представленная в [6]. Согласно данной методике, принимаем предположение о линейном виде регрессионной модели; составляем матрицу центрированных факторов с дополнительным фиктивным столбцом; затем определяем коэффициенты регрессии путём решения матричного уравнения, составленного из преобразованных матриц исходных данных. По результатам экспериментов, получены следующие регрессионные зависимости:

для первой группы

$$Y = 5,5 - 2,69x_1 + 2,038x_2 + 0,003x_3 - 0,081x_4$$

$$Y = 5,5 - 4,519x_1 + 0,94x_2 + 0,009x_3 - 0,004x_4$$

$$Y = 5,5 - 3,544x_1 - 0,226x_2 + 0,075x_3 - 0,072x_4,$$

для второй группы

$$Y = 2,5 - 0,23x_1 - 0,0007x_2 - 0,0004x_3 - 0,548x_4$$

$$Y = 3,78 + 2,47x_1 + 0,087x_2 - 0,119x_3 - 1,781x_4$$

$$Y = 11,1 - 1,2x_1 - 0,117x_2 + 0,077x_3 - 4,482x_4,$$

где x_1, x_2, x_3, x_4 – центрированные значения выбранных факторов соответственно.

Проверить адекватность полученных моделей классическим способом (по критерию Фишера) невозможно ввиду отсутствия параллельных опытов, поэтому для проверки адекватности использовали метод, согласно которому дисперсию воспроизводимости заменяют дисперсией отклика [6]. В результате применения данного метода сделан вывод, что две модели, из полученных в первой серии опытов, можно считать адекватными. Модели, характеризующие влияние внешних факторов, неадекватны.

Таким образом, анализ результатов проведенных исследований позволяет сделать следующие выводы:

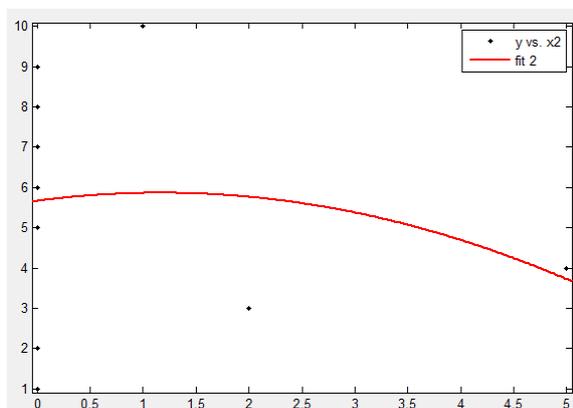
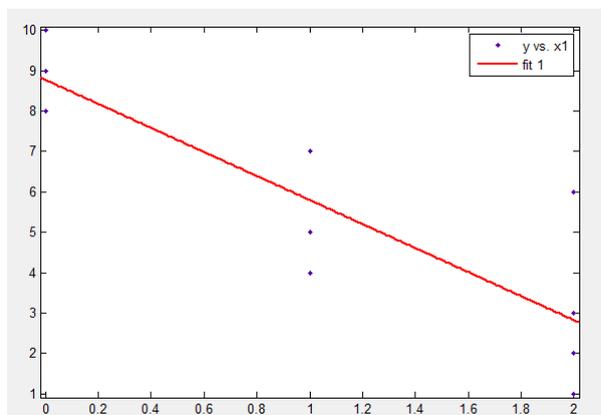
1. Используемый в данной работе метод получения регрессионных моделей влияния внутренних факторов оптимизации ресурса можно применять для предварительных расчётов на этапах 1а, 1б;

2. Наибольший вес при ранжировании ресурса в выдаче имеет фактор точного вхождения поискового запроса в теги «Title» и «Description». На рис. 2 графически показана зависимость позиции сайта в выдаче от каждого из выбранных для эксперимента факторов.

Невозможность получить адекватные зависимости для второй серии экспериментов указывает на то, что в формировании величины, называемой конкурентностью запроса, играет роль большее число факторов и (или) они взаимодействуют в более сложной связи.

Выводы. Проведенные предварительные исследования подтверждают сложность описания ранжирования ресурсов в поисковой выдаче и дают обоснование разработке моделей с помощью современных методов моделирования сложных систем. Принимая во внимание такие характеристики алгоритмов ранжирования, как высокая сложность, секретность, функционирование в области искусственного интеллекта, изменчивость и непрерывная адаптация, наиболее целесообразным считаем использование нейронных сетей [7] как метода моделирования процессов, на базе которых осуществляется оптимизация поискового продвижения web-ресурсов.

Таким образом, представленные в данной работе результаты формализации процесса поисковой оптимизации web-ресурсов могут быть использованы при разработке систем управления их продвижением. Актуальной остаётся проблема разработки моделей ранжирования сайтов в поисковых системах, для чего наиболее целесообразным представляется использование нейронных сетей.



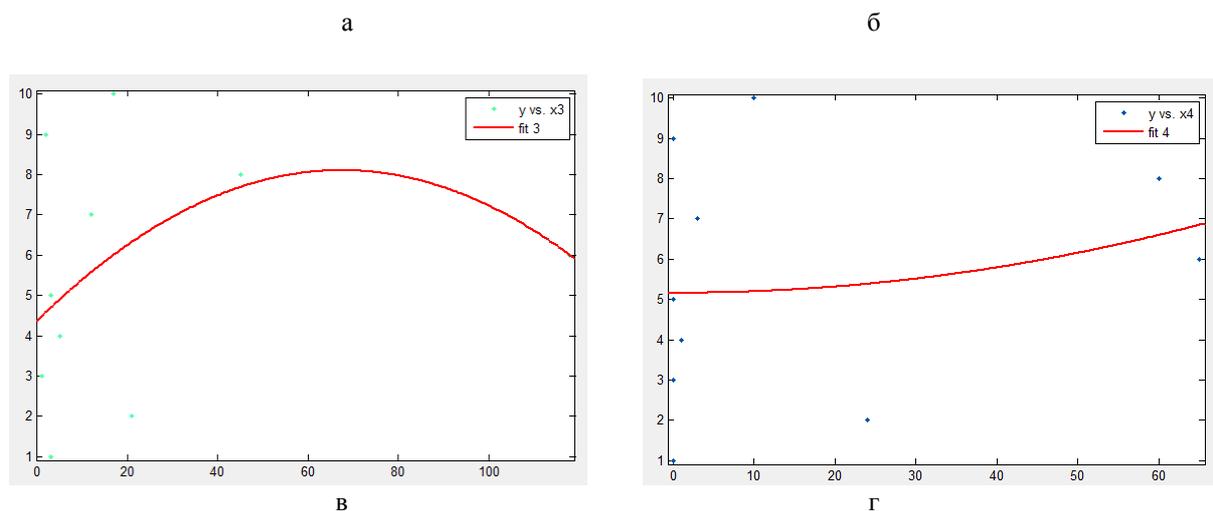


Рис. 2. Влияние внутренних факторов оптимизации X5, X9, X3/10, X2/10 (рис. а, б, в, г соответственно) на позицию сайта в выдаче ПС Яндекс (точками показаны экспериментальные данные, а кривыми – аппроксимированные)

СПИСОК ЛИТЕРАТУРЫ

1. Актуальные факторы ранжирования [Электронный ресурс] / Энциклопедия поисковых систем. – Режим доступа : http://www.searchengines.ru/articles/aktualnye_fakto.html/. – Назва з титул. екрану.
2. Три способа оценить конкурентность поискового запроса [Электронный ресурс] / Е. Камская. – Режим доступа : <http://optimizerha.ru/master-classes/konkurentnost-zaprosa>. – Назва з титул. екрану.
3. Оценка конкуренции по запросу в Яндексе. 1 часть [Электронный ресурс] / И. Токарев. – Режим доступа : <http://whiteseo.ru/content/view/521/>. – Назва з титул. екрану.
4. Как раскрутить сайт? [Электронный ресурс] / В. Беляев. – Режим доступа : <http://trynyty.ru/kak-raskrutit-sajt/>. – Назва з титул. екрану.
5. Алгоритмы поисковых систем [Электронный ресурс] / Вэбэффектор. – Режим доступа : http://www.webeffector.ru/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B_%D0%9F%D0%A1. – Назва з титул. екрану.
6. Костин В. Н. Статистические методы и модели : [учеб. пособие] / В. Н. Костин, Н. А. Тишина. – Оренбург : ГОУОГУ, 2004. – 138 с.
7. Методы и модели искусственного интеллекта [Электронный ресурс] / Jahrbuch fur EcoAnalytic und EcoPatologic. – Режим доступа до журн.: <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content128/Content128.htm>. – Назва з титул. екрану.

Рахуба Виктория Олеговна – к. т. н., доцент кафедры автоматизированного управления технологическими процессами.

Запорожская государственная инженерная академия.