

*Victoria O. Rakhuba,
ScD, associate professor,
Zaporozhye State Engineering Academy*

Search Engine Optimization as a Controlled Object

Key words: search engine optimization (SEO), automated control, mathematical model.

Annotation: The article deals with research of search engine optimization as a controlled object. The optimization task is formalized. The results of Yandex search engine's top research experiment are presented.

В условиях высокой конкуренции за клиента, которой сегодня характеризуются экономические отношения 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 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 данным процессом.

Целью данного исследования является установление зависимостей для создания теоретических основ алгоритмов управления поисковой оптимизацией сайтов.

Процесс поисковой оптимизации web-ресурса как объект автоматизации можно представить следующим образом:

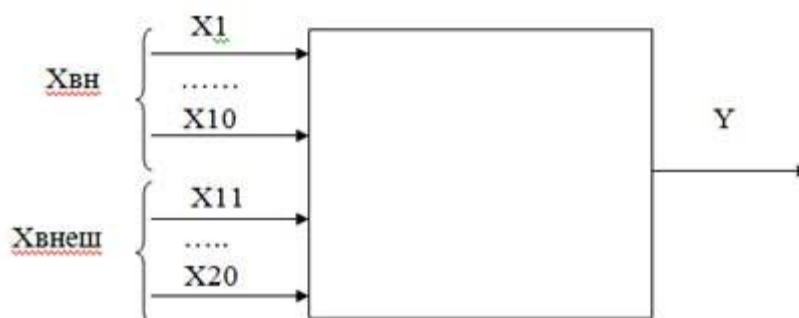


Рис. 1. SEO как объект управления

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 (1-3,4,5):

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; 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.

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

$$Y = f(X_1, \dots, X_{20}), Y \rightarrow \min ,$$

причем ограничения на процесс определяются 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; 2. Синтез семантического ядра; 3. Собственно оптимизация страниц сайта по выбранным запросам путем управления внутренними факторами в приведенной выше модели.

Для установления зависимостей, составляющих основу алгоритмов управления, необходимо провести исследования выдачи поисковых систем. Очевидно, что для предварительного исследования может быть выбран только метод пассивного эксперимента. 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 проводились 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 проводились для сайтов moeobrazovanie.ru, kartaznaniy.ru и moya-ramyat.ru; факторами выступали X1/млн, X9/10, X3/10, X5.

Для анализа собранных данных использовалась методика обработки результатов пассивных экспериментов (6). 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

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

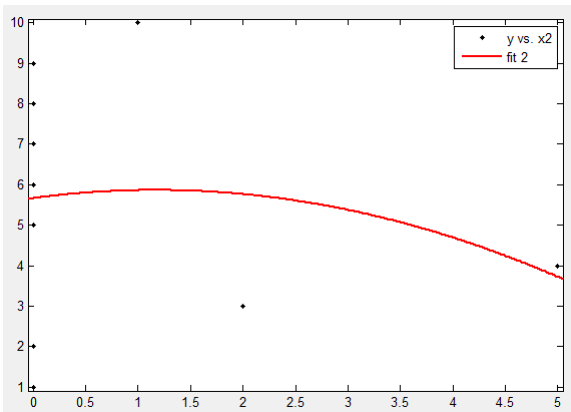
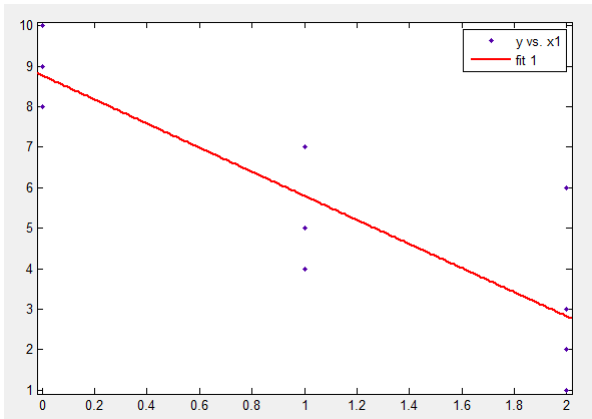
$$Y = 3,78 + 2,47x_1 + 0,087x_2 - 0,119x_3 - 1,781x_4$$

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

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

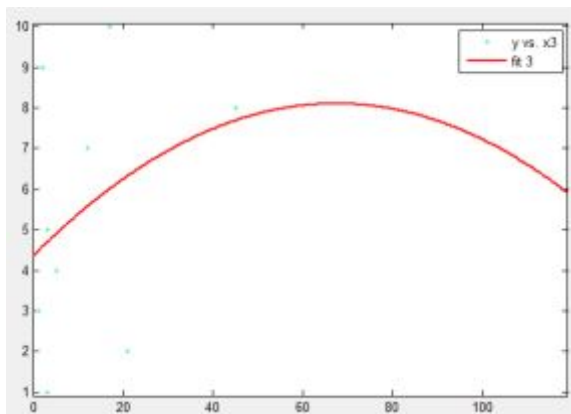
Проверить адекватность полученных моделей классическим способом (по критерию Фишера) невозможно ввиду отсутствия параллельных опытов. Поэтому для проверки адекватности использовался метод, согласно которому дисперсия воспроизводимости заменяется дисперсией отклика (6). В результате применения данного метода сделан вывод, что две модели, из полученных в первой серии опытов, могут считаться 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 может применяться для предварительных расчётов на этапе 1; 2). Наибольший вес при ранжировании ресурса в выдаче имеет фактор точного вхождения поискового запроса в тэги «Title» и «Description». Ниже графически показана зависимость позиции сайта в выдаче от каждого из выбранных для эксперимента факторов (точками показаны экспериментальные данные, а кривыми – аппроксимированные)



а

б



в

г

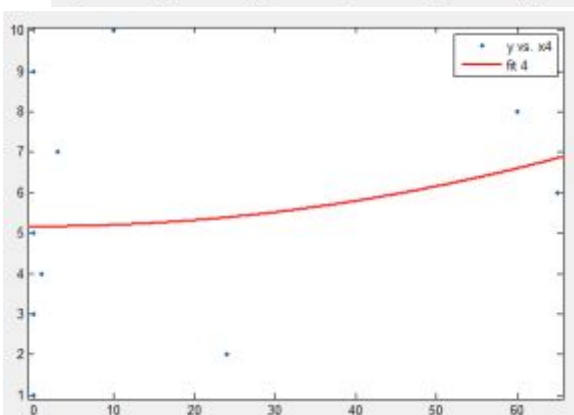


Рис. 2. Влияние внутренних факторов оптимизации X5, X9, X3/10, X2/10 (рис. а,б,в,г соответственно) на позицию сайта в выдаче ПС Яндекс

Невозможность получить адекватные зависимости для второй серии экспериментов указывает на то, в формировании величины, называемой конкурентностью запроса, играет роль большее число факторов и (или) они взаимодействуют в более сложной связи. Но именно оценка конкурентности запроса играет ключевую роль для формирования семантического ядра информационного ресурса на этапе 2. Поэтому разработка методов её моделирования остаётся актуальным вопросом.

Принимая во внимание такие характеристики алгоритмов ранжирования как высокая сложность, функционирование в области искусственного интеллекта, изменчивость и непрерывная адаптация, наиболее целесообразным представляется 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. Актуальным остаётся вопрос разработки моделей ранжирования сайтов в поисковых системах, для чего наиболее целесообразным представляется использование нейронных сетей.

References:

1. *How and why to automate search engine optimization site? [Internet] Available from: http://www.searchengines.ru/articles/why_and_how_to_.html # ixzz2jsYiClYx*
2. *Kama E. Three ways to evaluate the competitiveness of a search query [Internet] Available from: <http://optimizatorsha.ru/master-classes/konkurentnost-zaprosa>*
3. *Tokarev I. Evaluation of the competition on request in Yandex. 1 piece [Internet] Available from: <http://whiteseo.ru/content/view/521/>*
4. *Belyaev V. How to promote my site? [Internet] Available from: <http://trynyty.ru/kak-raskrutit-sajt/>*
5. *The algorithms of search engines [Internet] Available from: 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*
6. *Kostin VN. Statistical Methods and Models: Tutorial: VN. Kostin, NA. Tishina. Orenburg: GOUOGU, 2004; 138.*
7. *Methods and Models of Artificial Intelligence [Internet] Available from: <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content128/Content128.htm>*