

УДК 519.95

ТЕХНИЧЕСКИЙ АНАЛИЗ ВЕБ-РЕСУРСОВ С ИСПОЛЬЗОВАНИЕМ АППАРАТА ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

© Д.И. Панин

Ключевые слова: математическое моделирование; искусственные нейронные сети; технический анализ веб-ресурсов.

Разработана автоматизация технического анализа сайта для создания более качественных веб-ресурсов.

Технический анализ сайта (технический аудит сайта) необходим для того, чтобы определить техническую готовность сайта к продвижению в поисковых системах, выявить недостатки в эргономике, функциональности ресурса и подготовить рекомендации по устранению этих недостатков. Технический анализ сайта является неотъемлемой частью работ по поисковому продвижению сайта.

В данный момент времени технический анализ плохо автоматизирован, большинство факторов собирается и анализируется человеком, не исключены ошибки, которые человек заметить просто не в силах. Автоматизация должна подтолкнуть пользователей и разработчиков создавать более качественные сайты, приятные пользователю и дружественные для поисковых роботов.

Для достижения указанной цели необходимо решить следующие задачи.

1. Анализ проблематики данной области.
2. Разработка модульной системы комплексного сбора информации о сайте.
3. Разработка статичных алгоритмов, предназначенных для количественного анализа факторов, влияющих на рейтинг и качество веб-ресурса.
4. Разработка обучающихся адаптивных нейронных сетей с целью выявления факторов, которые не могут быть проанализированы стандартными методами в силу уникальности каждого объекта.
5. Разработка механизма составления рекомендаций для перечисления найденных недоработок и методов их разрешения.
6. Разработка системы статистического анализа собранных наблюдений с целью выявления значений коэффициентов каждого фактора в отношении к позиции сайта в ранжировании поисковой системой.

Сбор информации об исследуемом ресурсе – это первый этап технического анализа сайта. Он включает в себя полный или частичный перебор всех страниц ресурса, выделение в них необходимой информации для анализа, а также структурирование полученной информации для последующей работы. Часть информации собирается и анализируется человеком в силу своей абстрактности и сложности для выполнения машиной. Для исключения человека из этой системы необходима реализация искусственной нейронной сети,

которая частично или полностью заменит человека на данном этапе анализа.

Собранные данные анализируются в двух направлениях:

- 1) статичный анализ – данные могут быть проанализированы жесткими правилами и алгоритмами;
- 2) динамический анализ – данные анализируются адаптивной системой, производящей поиск по ключевым значениям имеющихся в базе знаний.

Пример факторов, подвергающихся анализу.

Статичные (могут быть заданы стандартными алгоритмами) факторы.

- Нет ошибочных ссылок на сайте.
- Несуществующие страницы возвращают 404 заголовков.
- Присутствует карта сайта (HTML/XML).
- Присутствует верный тег используемой кодировки, он совпадает с заданным в заголовке.
- HTML разметка проходит валидацию по используемому стандарту.
- Все страницы, изменившие адрес, должны иметь перенаправление с кодом 301 со старой на новую.

Динамичные (целесообразно использование нейронных сетей) факторы.

- Поиск уязвимостей на сайте (XSS атака, SQL- Injection).
- Симуляция примитивных действий пользователя (серфинг по сайту, соответствие выдаваемой информации к запрошенной).
- Проверка работоспособности форм (заказа, обратной связи).
- Проверка цветов фона, текста и изображений на контрастность и восприятие человеком.
- Определение категории сайта (магазин, информационное хранилище, сайт-визитка и т. п.) для применения в последующих стадиях анализа.
- Определение смысловых дубликатов страниц.

Составление рекомендаций является важной частью технического анализа, т. к. жизненно необходимо подробно описать проблему и пути ее решения. Отчет должен формироваться по критичности найденных недоработок и содержать оптимальное количество необходимых эксперту данных.

В результате множественного анализа различных веб-ресурсов можно вычислить приблизительную закономерность влияния факторов состояния сайта на его рейтинг в поисковых системах. Это отношение можно представить в форме матрицы:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

Количественные показатели факторов a_{mn} , где n – порядковый номер показателя, m – порядковый номер просканированного ресурса, высчитывается путем анализа сайта. Значение b_m равняется текущей позицией сайта в ранжировании исследуемой поисковой систе-

мы. X_n – неизвестный коэффициент фактора анализа сайта, характеризующий отношение количественного показателя к рейтингу.

Итоговая точность эксперимента будет зависеть от количества используемых факторов, а также от количества экспериментов.

Такая схема может применяться для любых поисковых систем и выявления ключевых факторов, влияющих на позиции сайта.

Поступила в редакцию 12 ноября 2010 г.

Panin D.I. Technical analysis of web-resources with use of apparatus of artificial neuron networks

The automatizations of technical cite analysis for creation of better web-resources is developed.

Key words: mathematical models; artificial neuron networks; technical analysis of web-resources.