

УДК 004.77; 004.733

Дубина Т. – ст. гр. СНм-51

Тернопільський національний технічний університет імені Івана Пулюя

МЕТОДИКИ ФОРМУВАННЯ РЕЙТИНГУ ВЕБ-САЙТІВ

Науковий керівник: ст. викладач Маєвський О.В.

Dubyna T.

Ternopil Ivan Pul'uy National Technical University

RATING METHODOLOGY OF FORMATION OF WEB SITES

Supervisor: Majevskiy A.

Ключові слова: методика, рейтинг, веб-сайт, ранжування

Keywords: method, rating, website, rankings

Фактично всі методики формування рейтингів є закритими. Для цього є цілий ряд причин. По-перше пошукові машини, інші компанії, для яких рейтинг є бізнес-інструментом, не хочуть розкривати перед конкурентами свої технології, на які витрачають мільйони. По-друге, розкриття особливостей побудови рейтингу дозволить веб-майстрам легко обманювати рейтинг і штучним способом виводити сайти на перші позиції.

Виключенням є лише рейтинги на кшталт Webometrics, які прямим чином зацікавлені в тому, щоб сайти університетів адаптовувалися до їх вимог, і, відповідно, ставали більш прогресивними.

Особливості побудови інших рейтингів зазвичай залишаються до кінця невідомими. Разом з тим практикуючі спеціалісти в сфері оптимізації з часом виявляють закономірності, що існують при формуванні того, чи іншого рейтингу та формують рекомендації, якими наповнений Інтернет.

Варто навести спрощену загальну формулу, яку застосовують більшість пошукових систем при ранжуванні сайтів:

$$\Delta_n = K_1\Delta_1 + K_2\Delta_2 + K_3\Delta_3 \dots + K_m\Delta_m, \quad (1)$$

де: Δ_n – сумарний бал «важливості» сайту, чи ключового слова сайту;

$\Delta_1, \Delta_2, \Delta_3 \dots \Delta_m$ – важливість окремого критерію, за яким оцінюється значення сайту, чи ключового слова сайту, наприклад:

Δ_1 – кількість ключових слів на сайті;

Δ_2 – число ключових слів у веб-сторінці;

Δ_3 – число ключових слів в тегу "title";

Δ_4 – число ключових слів в мета-тегу "description";

Δ_5 – число ключових слів в мета-тегу "keywords";

Δ_6 – кількість ключових слів у заміщуваному тексті;

Δ_7 – місце розташування ключового слова на веб-сторінці;

Δ_8 – глибина веб-сторінки, на якій розташовано ключове слово сайту;

Δ_9 – стиль оформлення ключового слова;

Δ_{10} – кількість посилань на веб-сторінку, на якій розташовано ключове слово.

Δ_m – інші критерії.

$K_1, K_2, K_3 \dots K_m$ – поправочні коефіцієнти, що часто визначаються емпіричним шляхом.

Для окремих критеріїв поправочні коефіцієнти К можуть розраховуватися за спеціальними математичними залежностями. Поправочні коефіцієнти К можуть враховувати, наприклад, щільність ключових слів, розмір шрифту, тип шрифту тощо.

Існує два основних шляхи включення сайту до рейтингу:

- 1) Подача відповідної заявки укладачу рейтингу;
- 2) Автоматичне включення до рейтингу скан-ботом.

Перший спосіб є простим, проте і не надто зручним, адже необхідно затратити час на заповнення та подачу заявки. На опрацювання заявки та прийняття рішення теж піде час, інколи немалий.

Другий спосіб в більшості випадків означає, що бот, скануючи інший сайт за посиланням заходить на Ваш, індексує його та включає до рейтингу. Але для виконання такої процедури необхідно мати зовнішні посилання з уже відомих, і бажано авторитетних, веб-ресурсів на Ваш сайт.

Для цього достатньо зареєструватися у відомих каталогах, таких як DMOZ, або Yahoo. Проте реєстрація на таких ресурсах, в основному, здійснюється шляхом подачі заявок.

Отже можна підсумувати, що шлях до другого способу лежить через перший.

Більшість рейтингів веб-сайтів використовують скан-боти на етапі збору інформації. Важливим нюансом є те, що меню створені на Flash, Java та JavaScript часто блокують роботу сканера (їх опрацювання вимагає забагато ресурсів, тому ігнорується). При верстці сторінок не бажано застосовувати ідентифікатори сеансів cookie, оскільки це ускладнює роботу сканера (адже з кожним документом зв'язується кілька URL-посилань).

Перевагу слід віддавати простим HTML-посиланнями, проте, якщо одними засобами HTML не обійтись, то можна скористатися і Flash, чи будь-якою іншою технологією. В такому випадку важливо правильно оформити файл Robots.txt і мапу сайту. Завдяки цьому можна досягти повної індексації ресурсу.

До речі, інколи JavaScript-посилання можуть бути навіть корисними: якщо посилання, які ведуть на інші сайти виконуються за цією технологією, роботи їх не опрацьовують, що сприяє збільшенню таких показників, як PageRank.

Того ж ефекту можна досягнути помістивши всі вихідні посилання, виконані на HTML на окрему сторінку і заборонивши в Robots.txt індексацію цієї сторінки.

УДК 004.02

Карвацький Р. – ст. гр. СКм-51

Тернопільський національний технічний університет імені Івана Пулюя

ТЕХНОЛОГІЇ КОНКУРЕНТНОЇ РОЗВІДКИ

Науковий керівник: асистент Дмитроца Л.П

Карвацький Р.

Ternopil Ivan Pul'uy National Technical University

TECHNOLOGY COMPETITIVE INTELLIGENCE

Supervisor: Dmytrotsa L.

Ключові слова: конкурентна розвідка, інформація

Keywords: competitive intelligence, information