

УДК 004.042

В.В. Копина, Г.В. Шимчук

Тернопільський національний технічний університет імені Івана Пулюя

Пошукові системи та загальні принципи їх роботи

V. Kopyna, G. Shymchuk

SEARCH SYSTEMS AND GENERAL PRINCIPLES OF THEIR WORKS

В теперішній час просто неможливо уявити життя без суспільства без виходу в мережу Інтернет. Якщо на початках Інтернет був засовом зв'язку для простих користувачів, то на даний час він став засобом комунікації структур державного рівня.

З розвитком веб-технологій створення сайтів стало більш доступнішим для аудиторії його користувачів і за рахунок цього в Інтернеті почали з'являтися все більше сайтів, тематика яких була абсолютно різною – від сайтів корпорацій, до сайтів маленьких фірм, що пропонують відвідати їх офіси в межах одного міста та замовити в них пакет певних послуг.

На сьогоднішній день пошукові системи це більше ніж просто нелінійні алгоритми. Це дає їм можливість максимально точно оцінити веб-сайт з більше ніж двохсот критеріїв. Тому зараз як ніколи стало актуальним завдання створення не просто якісного веб-сайту, але й підтримка його рівня довіри в пошукових системах, для отримання хорошої відвідуваності та конверсії від потенційних клієнтів.

Майже всі власники сайтів докладають чималих зусиль для їх розкручування та піару і хотіли б бачити сторінки свого сайта на найвищих позиціях в видачі пошукових систем. Досягти дійсно істотного ефекту в цьому напрямі можна тільки за допомогою пошукової оптимізації – комплексу заходів по проведенню SEO.

Актуальність SEO обумовлена великим рівнем конкуренції в мережі Інтернет та можливістю з мінімальними фінансовими затратами потрапити в топ пошукової видачі. Також важливим є той факт, що самі пошукові системи дають рекомендації веб-майстрам як проводити ті чи інші заходи пошукової оптимізації веб-сайту та враховують їх при оцінці релевантності контенту і сайту в цілому. Таким чином сайт належно проведеною SEO-оптимізацією стає більш конкурентно спроможним та приносить більше конверсії для його власника.

При розробці методики просування сайту важливим є питання вивчення механізмів роботи пошукових алгоритмів, оскільки на їх основі складаються закони пошукової оптимізації (так звана SEO-оптимізація) і просування сайтів.

Враховуючи це, весь комплекс заходів з SEO-оптимізації спрямований, власне, на просування українських сайтів саме в цих двох системах – Google і Яндекс. Саме топ-позиції у видачі цих пошукових систем дають реально додатковий трафік на сайт.

Якщо розглядати процес пошуку інформації в мережі, його можна розбити на наступні етапи:

- збір інформації зі сторінок сайтів в мережі Інтернет;
- індексація сайтів;
- пошук за запитом;
- ранжування результатів.

Незважаючи на те, що кожна пошукова система має свої особливості в алгоритмах пошуку та ранжування результатів, принципи роботи всіх пошукових систем загальні.

Збір даних здійснюється спеціальною програмою, яка є у кожній пошукової системи – пошуковий робот. Він призначений для перебору сторінок Інтернету з метою занесення інформації в базу даних пошукової системи. Як тільки ви запустили свій сайт

і дали роботу якої-небудь пошукової системи зрозуміти, що з'явився новий ресурс (з допомогою зовнішніх посилань на ваш сайт, додавання його сервіси додавання add url (англ., додати сайт) або використовуючи інші методи), робот приходить до вас, починає ходити по сторінках і збирати з них дані (це можуть бути текстовий вміст, зображення, відео та інші файли). Цей процес називається збором даних (англ. crawling) і він може відбуватися не тільки при запуску сайту. Робот становить для сайту розклад, коли він має на нього зайти в наступний раз, перевірити стару інформацію і додати нові сторінки, якщо такі є.

Керувати доступом на різні ресурси для пошукового робота можна з допомогою файлу robots.txt, в якому містяться правила і обмеження доступу до вмісту сайту на http-сервері. Карта сайту sitemap.xml (XML-файл з інформацією для пошукових систем про сторінки веб-сайту, які підлягають індексації) також може допомогти роботі, якщо з якихось причин навігація по сайту йому утруднена.

Робот може ходити по вашому сайту довгий час, однак це не означає, що він відразу з'явиться в пошуковій видачі. Сторінок сайту необхідно пройти такий етап, як індексація – складання для кожної сторінки зворотного (інвертованого) файлу індексу. Індекс – всі сторінки, які відомі пошуковій системі, тобто які пройшли індексацію. Індекс слугує для того, щоб швидко з нього виробляти пошук і складається зазвичай із списку слів з тексту та інформації про них (позиції в тексті, вага та ін).

Після того, як пройшла індексація сайту або окремих сторінок, вони з'являються в основній видачі пошукової системи і їх можна знайти за ключовими словами, присутнім у тексті. Процес індексації зазвичай відбувається досить швидко після того, як робот завантажить інформацію з вашого сайту.

При пошуку, насамперед, аналізується запит, введений користувачем (відбувається препроцесинг запиту), в результаті якого обчислюються ваги для кожного з слів. Вагою слова називається відношення частоти використання цього слова до загальної кількості слів, виражене у відсотках.

Далі, пошук здійснюється за інвертованим індексом, знаходяться всі документи в колекції (базі даних пошукової системи), які найбільш підходять під цей запит.

Документи, найбільш схожі з запитом, потрапляють в результати пошуку.

Після того, як найбільш схожі документи були відібрані з основної колекції, вони повинні ранжируватися, щоб у верхніх результати відображені найбільш корисні для користувача ресурси. Для цього використовується спеціальна формула ранжирування, яка для різних пошукових систем має різний вигляд, однак для всіх з них основними чинниками ранжирування є:

- вага сторінки (ТІЦ, PageRank);
- авторитетність домену;
- релевантність тексту запитом;
- релевантність текстів зовнішніх посилань запитом;
- а також безліч інших чинників ранжирування.

Література:

1. Алгоритми пошукових систем [Електронний ресурс]: [сайт]: російська версія / Школа Інтернет маркетингу Empro. – текстові дані. – К., 2014. – Режим доступу: <http://empro.pro/seo-blog>.

2. Мара А. Пошукова оптимізація сайтів / Мара А. – К.: Школа Інтернет маркетингу Empro, 2013. – 129 с.

3. Ситник В.Ф., Писаревська Т.А., Єрьоміна Н.В., Краєва О.С. Основи інформаційних систем. Навчальний посібник. Київ: КНЕУ, 2001, 420