

*Рубльов Д.С. – рецензент Цурін О.П.
ННК “ІПСА” НТУУ “КПІ”, Київ, Україна*

Дослідження засобів для визначення “сміттєвих” інтернет-сайтів

Розвиток глобальної мережі Інтернет, збільшення популярності пошукових систем та доступність засобів для створення інтернет-сайтів стають причиною для виникнення такого явища, як різного роду “сміттєві” сайти.

Такі типи сайтів мають назви – дорвеї (від англ. “doorway” – “прохідник”), сателіти (від англ. “satellite” – “супутник”), що є однією з складових частин “чорної оптимізації” інтернет-сайтів. Існування таких сайтів має лише одну мету – збагачення свого власника, що досягається декількома шляхами (перенаправлення пошукового трафіку на сайти замовників, або підвищення пошукових позицій сайту і т. д.).

При створенні дорвеїв переслідується ціль отримання високих позицій у результатах видачі пошукових систем для отримання високого пошукового трафіку. Отриманий трафік конвертується у різних партнерських програмах, перенаправляючи відвідувачів на інші сайти (на сайти замовників “чорної оптимізації”). Вміст сайту, при цьому, може складатися з великої кількості слабо пов’язаних між собою ключових слів, і не нести корисної для користувачів інформації. Такі сайти зазвичай мають примітивну структуру і дизайн, дуже часто розміщуються на безкоштовних хостингових площадках для зменшення затрат вебмайстрів із-за великої вірогідності бути відфільтрованими пошуковими системами за короткий проміжок часу. Майже всі посилання зі сторінок такого сайту, що не задіяні у навігації його розділам, ведуть на сторінки з переспрямуванням на інший сайт.

Сателіти є складовою частиною як “сірої”, так і “чорної” пошукової оптимізації, і призначені для підвищення позицій інших сайтів у видачі пошукової системи, будучи при цьому абсолютно не спрямованими на надання релевантної інформації для користувачів. Свою роль сателіт виконує, посилаючись на сайт, позиції якого у видачі пошукової системи потрібно підвищувати. Зазвичай, один сателіт посилається на декілька сайтів з кожної зі своїх сторінок. Сателіти можуть мати досить складну структуру, і нормальний дизайн сторінок, і більш свідоме наповнення, хоча при цьому не представляти ніякої корисної інформації для користувача, який зайде на такий сайт з пошукової системи.

Оцінка структури сайтів, якості їх наповнення, зв’язків з іншими сайтами через переспрямування та посилання, хостингу, на якому вони знаходяться та доменного імені дозволяють з великою точністю визначати подібні сміттєві сайти у мережі Інтернет.

Література

1. Поиск неестественных текстов [Електронний ресурс]: (Научные статьи и отчеты – Компания Яндекс) / Евгений Гречников, Глеб Гусев, Андрей Кустарев, Андрей Райгородский // RCDL’2009 – Режим доступа: http://download.yandex.ru/company/A_Kustarev_A_Raigorodsky_poisk_neestestvennih_textov_statia.pdf – Назва з екрана.
2. Combining Anchor Text Categorization and Graph Analysis for Paid Link [Електронний ресурс] / Kirill Nikolaev, Ekaterina Zudina, Andrey Gorshkov // WWW’2009: The 18th International World Wide Web Conference. – Режим доступа: <http://www2009.org/proceedings/pdf/p1105.pdf> – Назва з екрана.
3. Автоматическая классификация web-сайтов [Електронний ресурс]: (Научные статьи и отчеты – Компания Яндекс) / Михаил Маслов, Алексей Пяллинг, Сергей Трифонов // RCDL’2008. – Режим доступа: http://download.yandex.ru/company/experience/rcdl2008/rcdl_sites_autoclassification.pdf – Назва з екрана.