

Цурін О.П., Іванченко Р.М.
 ННК "ІПСА" НТУУ "КПІ"

Автоматизовані засоби підняття світового Webometrics-рейтингу Web-сайтів наукових установ

Згідно з методикою, розробленою Webometrics [1], рейтинг наукової установи формується на основі 4 показників: 1. Розміру домену наукової установи, наприклад, університету (S); 2. "Видимості" домену (V), де V це кількість унікальних зовнішніх посилань на сторінки наукової установи; 3. Об'єму файлів, розміщених в домені (R); 4. Кількості робіт, що містять посилання на домен та проіндексовані системою Google Scholar (Sc).

Саме тому, за вимогами п.1, всередині домену слід створювати максимальну кількість сторінок. Всі сайти наукового закладу повинні містити якнайбільшу кількість публікацій: рефератів, звітів, тез, анотацій, наукових статей. Зважаючи на те, що українська мова, як і російська, для InterNet є мовою "національних меншин", наявність англійських сторінок є обов'язковою. Всі сторінки у HTML-тілі повинні містити відповідні метадані (скорочений зміст, ключові слова). Після створення та наповнення контентом InterNet-сторінок всі вони повинні бути проіндексовані в системах Yahoo, Open Directory Project, Live Search, Exalead, Google та Google Scholar [2].

В доповіді розглядаються два етапи автоматизації підняття Webometrics-рейтингу наукової установи. Перший етап виконується одразу після створення, або корінної модернізації Web-сторінок. Саме тоді варто подбати про індексацію сторінок всіма пошуковиками. В цьому може допомогти програмний комплекс, котрий позбавить розробників від рутинної праці та проіндексує сайти автоматично. Існують як платні, так і безкоштовні системи, здатні впоратися з таким завданням. Серед перших – Semonitor, CS Yazzle, AllSubmitter, до других належать такі програми, як Site-Auditor. Хоча використати готовий продукт простіше, проте деякі каталоги, як от Open Directory Project, забороняють автоматичну індексацію і "поведінку" існуючих програм впізнають одразу. Щоб обійти цей момент необхідно використовувати або власний продукт з унікальною "поведінкою", або проводити індексацію на таких ресурсах "вручну".

Другий етап, на якому може бути здійснена автоматизація – це стадія контролю за станом та наповненням вже існуючих сторінок. На цьому етапі необхідно перевіряти кількість сторінок в домені, працездатність сторінок, кількість "викладених" в домені публікацій, їх "видимість" для пошуковиків. Використання власного комплексу, хоч і тягне за собою додаткові організаційні та матеріальні витрати, однак дає можливість отримати більш повну картину стану наукової установи в Webometrics-рейтингу. Більше того, побудова власного комплексу дає можливість автоматично вирахувати приблизний Webometrics-рейтинг InterNet-представництва кожного підрозділу наукової установи (Університету). Це, в свою чергу, дозволяє періодично, визначати лідерів та тих, хто відстає, вчасно вживати організаційних заходів.

Література

1. Методика складання рейтингу наукових установ Cybermetrics Lab
<http://www.webometrics.info/methodology.html>
2. Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Science Department, Stanford University, Stanford, CA 94305, USA, 2000.