

Етапи автоматизованого отримання індексу цитованості наукових електронних документів

Дмитро Тарасов¹, Олег Гарасим²

Кафедра соціальних комунікацій та інформаційної діяльності, Національний університет "Львівська політехніка", УКРАЇНА, м.Львів, вул.С.Бандери, 12,
E-mail: 1. dtarasov@lp.edu.ua, 2. garasym-oleg@rambler.ru

In the article substantiates the need for citation indexes for open electronic archives. Raised the question analysis of the entire array of data archives. Stages of receiving citation index and proposed model automating this process..

Ключові слова – електронний архів, наукові документи, дані, індекс цитованості, спеціалізована система.

I. Вступ

Електронний науковий архів (ЕНА) представляє сукупність документів наукової діяльності, що відкриває можливості дослідження наукового внеску зареєстрованих дослідників, розвиток окремої галузі науки, наукового видання, публікації, в також загальний науковий рівень університету. Сьогодні в архіві зберігаються понад 10000 наукових праць.

Для аналізу архіву постає необхідність отримання відомостей про документи, які проіндексовані спеціалізованою пошуковою системою, а також отримання показників цитованості.

II. Етапи отримання індексів цитування з Google Scholar

2 грудня 2009 року у науково-технічній бібліотеці Львівської політехніки відкритий науковий електронний архів, що базується на платформі DSpace компанії Hewlett-Packard, основними перевагами якого є швидкий пошук, завантаження, перегляд електронних наукових документів. У електронний архів наукових публікацій НТБ увійшли науково-технічні збірники, журнали, матеріали конференцій, публікації та електронні документи працівників НТБ, бібліографічні видання та багато іншого. Сьогодні в архіві зберігаються понад 10000 наукових праць, з яких 1930 проіндексовані Google Scholar.

Обсяг електронних наукових документів архіву Національного університету «Львівська політехніка» щодня поповнюється новими. Його науковий потенціал можна аналізувати за допомогою індексів цитованості. Враховуючи регулярність досліджень та обсяги архіву постає необхідність в автоматизації пошуку індексів цитованості в спеціалізованій системі Google Scholar. Для цього необхідно розробити модуль.

Основними елементами розробленого модуля, який забезпечує пошук та отримання індексу цитування наукових документів ЕНА від Google Scholar є:

- «Опрацювання даних» - метою якого є відправлення запитів до ЕНА для отримання необхідних даних та адаптувати їх для можливості подальшої

роботи з ними. Вибирає тільки необхідні дані, містить ряд правил для коректного пошуку, фільтрації даних;

- «Відправлення на пошук» – задає правила пошуку, містить налаштування для більш точного пошуку (пошук у визначеному домені, вибір галузі науки, можливості простого та розширеного пошуку). З'єднується з базою даних Google Scholar та відправляє http-запит;

- «Google Scholar» – система визначена як така, що найкраще задовольняє вимоги пошуку індексу цитованості наукових документів українських дослідників. Може використовуватися інша система;

- «Web-сторінка результатів» – містить результати пошуку, передається у вигляді html-коду;

- «Аналіз Web-сторінки» – проводиться розбір результатів та їх формалізація;

- «Перевірення відповідності результатів до запиту» – містить правила перевірення на відповідність отриманих результатів науковому документу. Результати записуються до ІС.

Кожен з визначених етапів має вплив на вдосконалення методів отримання кількісних наукових показників і є важливими при практичній реалізації.

ВИСНОВОК

Обґрунтована необхідність автоматизації процесу пошуку та отримання індексів цитованості в спеціалізованих системах. Визначені етапи отримання індексу цитованості наукових електронних документів та їх функціональний опис, що дозволяє реалізувати модуль для автоматизованого отримання індексів цитованості відкритих електронних архівів. Розбиття мети, аналізу наукових електронних документів, на етапи, розкриває потенційні проблеми роботи модуля, можливість передбачити результати аналізу, вдосконалювати методику отримання наукових показників з спеціалізованих систем, виділити правила пошуку, обмеження та настройки, щоб підвищити результати пошуку.

Література

1. Тарасов, Д.О. Технологічні особливості опрацювання документів у електронній формі у бібліотеках / Д.О. Тарасов // Інформаційні системи та мережі. – Львів : Видавництво Національного університету "Львівська політехніка", 2008. – С. 229-232.
2. Висоцька, В.А. Алгоритми та засоби опрацювання інформаційних ресурсів в системах електронної контент-комерції / В.А. Висоцька // Інформаційні системи та мережі. – Львів : 2008. - С. 78-96.
3. Гарасим О.Р. Аналіз засобів управління корпоративною конфіденційною інформацією / Гарасим О.Р., Чирун Л.Б. // Інформаційні системи та мережі: Вісник Національного університету «Львівська політехніка», №689. – Львів, 2010