

Способи отримання доступу до метаданих статей через API-інтерфейси пакету програмного забезпечення DSpace

Вінницький національний технічний університет

Анотація

Розглянуто різні способи роботи з метаданими опублікованих наукових матеріалів через програмні інтерфейси пакету програмного забезпечення DSpace. Запропоновано спосіб розширеного доступу до метаданих опублікованих матеріалів з використанням програмного інтерфейсу для модулів DSpace.

Ключові слова: інституційний репозиторій, програмні інтерфейси, веб-додатки, OAI-PMH, DSpace, Google Scholar, Dublin Core, бази даних.

Abstract

Different methods of working with metadata of published science materials through the core application program interface of DSpace software package. Introduced a new method of advanced accessing of published materials using the core application program interface of DSpace.

Keywords: institutional repository, API, web applications, OAI-PMH, DSpace, Google Scholar, Dublin Core, databases

Вступ

Впровадження інституційних репозиторіїв в процес онлайн-публікації наукових статей українських освітніх та наукових установ спричинив значні труднощі в адаптації до нових підходів вже існуючих систем публікацій наукових статей. Найбільш складним процес інтеграції репозиторіїв прийшовся установам, які до цього використовували застарілі програмні рішення, більшість з яких працювали на базі стандартів ГОСТ, що регулюють бібліотечну та видавничу справу. В свою чергу, обов'язковість підтримки власного репозиторію наукових праць призвела до різкого збільшення інтенсивності його наповнення, що призвело до масового ручного перенесення матеріалів з існуючих електронних сховищ електронних матеріалів. Іншою причиною, яка стимулює розвиток системи інституційних репозиторіїв є проект Національного Репозитарію академічних текстів [1], що покликаний створити єдину всеукраїнську базу наукових робіт, яка зможе на більш якісному рівні виявляти плагіат.

Досить велика частина раніше існуючих рішень для збереження електронних версій публікацій не мала жодної підтримки для експорту метаданих у більш сучасні формати збереження метаданих. Основними перешкодами для створення інтерфейсів конвертації метаданих є ліцензійні обмеження раніше використовуваних програмних продуктів та архітектурні несумісності, які полягають в складності підключення динамічних бібліотек 16-бітних додатків, що були написані для ранніх версій ОС Windows. Єдиним рішенням подібних проблем є розробка додатків, які зможуть виконувати роль програмного інтерфейсу у зв'язці з драйвером бази даних, що підтримує старі формати сховищ баз даних відповідних рушіїв баз даних [2].

Результати дослідження

Одним із методів переносу матеріалів та їх метаданих в інституційний репозиторій, що працює на основі DSpace, є так званий OAI харвестинг. OAI харвестинг передбачає збір частини або усіх доступних матеріалів на OAI-джерелі. Протокол OAI-PMH - це протокол передачі метаданих у вигляді XML з використанням просторів імен Dublin Core. Він використовує протокол HTTP для передачі даних, при чому веб-сервер завжди виконує роль джерела даних [3]. OAI харвестинг застосовується для проведення індексації наукових праць, розміщених в репозиторіях. Модуль

сервера OAI-PMH наявний у багатьох пакетах програмного забезпечення репозиторіїв, таких як EPrints, Fedora, а власне в DSpace. Модуль клієнта OAI-PMH (рис. 1) реалізований в такому програмному забезпеченні, як Open Harvester Systems, а також у подібних сервісах збору матеріалів як Google Scholar, ROAR, DOAR та Національній Репозитарій академічних текстів [4]. OAI-харвестер може мати в базі значно меншу кількість полів, ніж він може отримати. Зменшення кількості полів дозволяє оптимізувати швидкодію веб-додатку. У випадку, якщо потрібно зберігати всі метадані, отримані в результаті OAI-запиту, можна зберігати у базі даних весь запит. Найбільш простим прикладом OAI-харвестера є простий цикл на bash, з використанням wget.

Крім використання функції серверу OAI-PMH, DSpace також може виступати в ролі OAI-харвестера (клієнта). Дана функція є досить корисною у випадку, коли відбувається заміна програмного забезпечення, що використовується. Проте функціональність класу OAIHarvester модуля імпорту є досить обмеженою в різноманітності полів, які підтримуються [5]. Це може викликати значні проблеми при імпорті статей з Microsoft SharePoint та порталів, що використовують програмне забезпечення проекту РКР. В останній версії (6.0) DSpace підтримується лише "правильний" Dublin Core, що описаний в специфікаціях OpenAIRE.

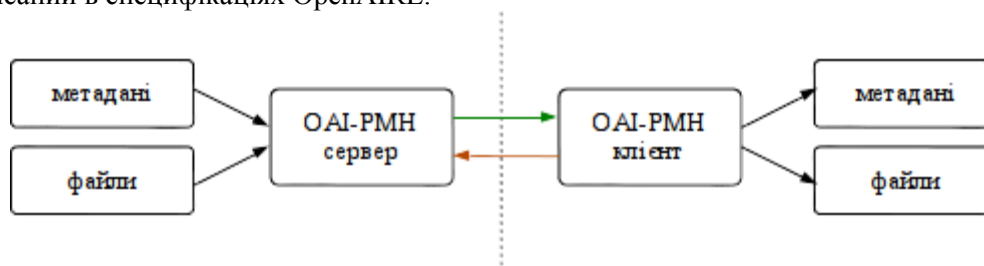


Рисунок 1. Застосування протоколу OAI-PMH для збору статей

В ході дослідження можливостей OAI-PMH було створено програмний модуль на мові програмування Python, що дозволяє переносити метадані наукових праць викладачів ВНТУ в каталог матеріалів персональних сторінок з використанням бібліотеки ruoai. Отримання доступу до додаткових полів реалізовано за допомогою REST API.

Висновки

Отже, при побудові програмних інтерфейсів зовнішніх джерел метаданих використання протоколу OAI-PMH є досить продуктивним рішенням, що дозволяє швидко перенести статті з застосуванням програмних інтерфейсів, які основані на форматі опису метаданих у форматі Dublin Core. Бібліотеки OAI-PMH випускаються під вільними ліцензіями та є кросплатформеними, що дозволяє підключати їх сумісно з відносно старими драйверами застарілих рушіїв баз даних.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Постанова Кабінету Міністрів України "Про затвердження Положення про Національній репозитарій академічних текстів" від 07.12.2016 р. [Електронний ресурс]. – Режим доступу: <http://mon.gov.ua/Новини/2016/12/07/polozhennya-repozit-07-12-16.doc>
2. Бісікало О. В. Концептуальні основи корпоративної бази даних з WEB-технологіями доступу [Текст] / О. В. Бісікало // Прикладні комп'ютерні програми для навчальної, методичної та організаційної роботи у вищих аграрних навчальних закладах III-IV рівнів акредитації : доповіді, виступи та повідомлення семінару, Суми, 1–3 листопада 2000 р. – Київ : Аграрна освіта, 2001. – С. 86–90.
3. Малініч І. П. Особливості дослідження політичної участі Інтернет-спільнот [Електронний ресурс] / І. П. Малініч // Матеріали XLV Науково-технічної конференції ВНТУ, Вінниця, 23-24 березня 2016 р. - Електрон. текст. дані. - 2016. - Режим доступу : <http://conferences.vntu.edu.ua/index.php/all-hum/all-hum-2016/paper/view/717>.
4. Криштафович Л. А. Електронний архів ВНТУ як інформаційний навігатор у світі інтелектуальних ресурсів відкритого доступу [Текст] / Л. А. Криштафович // Електронні інформаційні ресурси: створення, використання, доступ : збірник матеріалів Міжнародної науково-практичної Інтернет-конференції, 1-2 грудня 2015 р. – Вінниця : ВНТУ, 2016. - С. 224-242.
5. Клас OAIHarvester. [Електронний ресурс] – Режим доступу: <https://github.com/DSpace/DSpace/blob/master/dspace-api/src/main/java/org/dspace/harvest/OAIHarvester.java>

Малініч Ілля Павлович, студент групи 1ПЗ-16м, факультет ІТКІ, Вінницький національний технічний університет, м. Вінниця, e-mail: goosyara@malinich.net

Паламарчук Євген Анатолійович, кандидат технічних наук, доцент кафедри автоматичної та інформаційно-вимірювальної техніки, Вінницький національний технічний університет, м. Вінниця

Malinich Illia, group 1PZ-16m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: goosyara@malinich.net

Palamarchuk Yevhen A., PhD, Docent of Automatics and Informatics and Measurement Techniques Department, Vinnytsia National Technical University, Vinnytsia, email: p@vntu.edu.ua