

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АВТОМАТИЧНОЇ АГРЕГАЦІЇ МЕТАДАНИХ НА БАЗІ OAI-PMH

С. О. Петров, асистент; С. О. Глушко, студент,
Сумський державний університет
sergpet@gmail.com

Значне зростання кількості інформації за останні роки викликало широкий попит на створення інформаційних ресурсів (репозиторіїв зберігання інформації різного призначення), та зараз постає завдання забезпечення їх взаємної інтеграції як на технологічному, так і на технічному рівні. Тобто використання спільного інструменту для обміну інформацією про обсяг даних, які зберігаються в окремому репозиторії, а також отримання даних про інформацію яка зберігається в інших репозиторіях. Прикладами таких репозиторіїв можуть бути: репозиторій журнальних статей «DOAJ», бібліотека «ЛОІППО», репозиторій ePrints та інші.

Тому є необхідність у розробленні деякого інструменту (технології) з можливістю отримання доступу до колекцій, створення нових сервісів, які будуть забезпечувати як програмний, так і користувацький інтерфейс взаємодії з репозиторіями. Лідером у даній сфері є організація Open Archives Initiative (OAI), яка розробила протокол OAI-PMH (OAI Protocol for Metadata Harvesting), призначений для створення архівів публікацій, обміну метаданими і пошуку інформації [1]. Протокол OAI-PMH визначає механізм збору записів, які зберігають метадані із репозиторіїв, та надає можливість зробити їх метадані доступними для сервісів, що базуються на стандартах HTTP (Hypertext Transport Protocol) і XML (Extensible Markup Language). У рамках OAI-PMH описують дві основні складові: постачальники даних (data providers), які роблять доступними метадані і є серверами OAI [2], та сервіс-провайдери (harvesters), вони використовують протокол для збору даних, їх обробки і підвищення значущості їх метаданих (рис. 1).

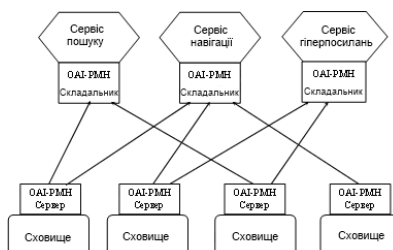


Рис. 1. Топологія постачальника даних і сервіс-провайдера

Для реалізації поставленого завдання, необхідно розробити компонент, який би при установленні в існуючу систему реалізовував подану вище схему взаємодії, тобто дозволяв створювати репозиторій даних на будь-якій існуючій системі. Розроблюваний компонент повинен поділитися на дві частини. Перша частина за допомогою існуючих API функцій проводить збір даних з існуючих репозиторіїв, а друга частина за визначеними правилами проводить генерацію, даних доступних по протоколу HTTP, які потрібні іншим репозиторіям для отримання інформації про обсяг інформації поточного репозиторія, для їх взаємної інтеграції. Цю схему роботи можна зобразити графічно (рис. 2).

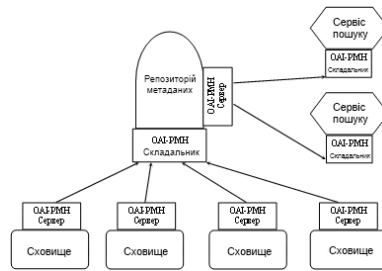


Рис. 2. Репозиторій метаданих на базі OAI-PMH

Для реалізації запропонованої моделі в рамках поставленого завдання, була вибрана мова програмування PHP у поєднанні з CMS Joomla. Таким чином, розроблено web-сайт, який має уніфікований, дружній користувачеві та функціонально-програмний інтерфейс, який на рівні модуля може бути інтегрований у діючу систему, та забезпечує створення узагальнюючого депозитарію метаданих. Пошук оновлених даних відбувається автоматично по всіх репозиторіях, з яких було отримано дані, і виводяться на одній сторінці.

Таким чином, процедура пошуку інформації функціонально розширилась, а технологічно спростилась. Ця технологія була випробувана на базі існуючого Web-сайта «Журнал нано- та електронної фізики» і показала позитивні результати.

1. The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14. <http://www.openarchives.org/OAI/>

2. Barrueco J.M., Subirats I. Open Archives Initiative Protocol for Metadata Harvesting : description, functiones and applications. El Profesional de la Informacion 12, 2, 2003, 99-106 pp.

