

УПРОВАДЖЕННЯ ЕЛЕКТРОННОГО АРХІВУ НАУКОВИХ ПУБЛІКАЦІЙ У НАУКОВО-ТЕХНІЧНІЙ БІБЛІОТЕЦІ НА ОСНОВІ ПРОГРАМНОЇ ПЛАТФОРМИ DSPACE

© Андрухів А., Тарасов Д., 2010

Розглянуто основні моменти впровадження електронного архіву наукових публікацій у Науково-технічній бібліотеці на основі програмної платформи Dspace.

Ключові слова: електронний архів, логічна структура електронного архіву, Dspace, електронні документи, електронні бібліотеки.

The article discussed the main points of the introduction of electronic archive of publications in scientific and technical library-based software platform Dspace.

Keywords: digital archive, logical structure of electronic archive, Dspace, digital documents, digital libraries.

Постановка проблеми

У Науково-технічній бібліотеці (НТБ) Національного університету "Львівська політехніка" створено відкритий електронний архів наукових публікацій. В усьому світі університети та дослідницькі інститути активно пропагують свої наукові розробки за допомогою Інтернету. Web-доступ забезпечує нові можливості для своєчасного поширення наукової інформації [1]. З метою ефективного поширення електронних ресурсів, а також підвищення доступності обміну науковою інформацією ці проекти розробляються з ініціативи «Відкриті архіви» (OAI, Open Archives Initiative) [2].

Аналіз останніх досліджень

Сьогодні на ринку лідерами є три безкоштовні системи побудови електронних архівів – Eprints, Dspace, Greenstone [3]. Системи мають підтримку обміну метаданими OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting), що забезпечує глобальні послуги доступу та пошуку. Для побудови електронного наукового архіву у Науково-технічній бібліотеці вибрано систему Dspace.

Dspace – це вільно поширюване програмне забезпечення (ПЗ) під ліцензією GNU, що застосовується для формування та управління Відкритими Архівами [5]. ПЗ Dspace може використовуватись для створення архівів робіт наукових досліджень, зображень, даних та інших видів цифрової інформації. Сьогодні в Україні створено ряд електронних архівів з використанням Dspace, зокрема, такі архіви мають Наукова електронна бібліотека періодичних видань НАН України, Central and Eastern European Marine Repository, Севастопольський інститут біології, Львівський національний університет ім. І. Франка, Київ-Могилянська академія, Харківський національний університет імені В.Н. Каразіна тощо.

Формулювання цілей статті

Розглянуто ключові моменти впровадження електронного архіву наукових публікацій у Науково-технічній бібліотеці на базі програмної платформи Dspace. Правильно організована структура електронного архіву та схема внесення та обробки електронних документів ЕА забезпечить ефективний менеджмент електронних документів. Завдяки цьому істотно спрощується

завдання адміністрування, наповнення та супровід ЕА, пошук потрібної інформації, що забезпечує збільшення кількості відвідувачів та активних учасників.

Долучившись до цього проекту, автори отримають такі вигоди:

- поширення індивідуальних досліджень (у разі окремого депозитора) та досліджень, створених інститутом, працівниками, студентами наукового підрозділу;
- зростання індексу цитування – деякі дослідження показують, що цитування наукових робіт, які є у відкритому доступі, зростають у декілька разів порівняно з тими, до яких немає доступу;
- довготерміновий доступ та зберігання матеріалів.

Проект сприятиме розвитку досліджень в українській та світовій науковій спільноті та спонукатиме до активних дій та кооперації у напрямі вільного поширення та доступу до наукових інформаційних ресурсів університетів як основних наукових установ, задля соціальної трансформації ролі науки у сучасному суспільстві, забезпечить середовище, що дасть змогу науковим підрозділам університету, працівникам та студентам легко розміщувати наукові дослідження в електронній формі у надійний та добре організований архів.

Мета та завдання проекту

Створення електронного архіву (ЕА) дасть змогу безперешкодно отримати доступ до наукового потенціалу у вигляді електронних матеріалів, що стимулюватиме читачку ініціативу, розвиватиме дослідницьку діяльність, а також отримати наукові публікації в електронному вигляді, автоматизувати процес пошуку наукових публікацій, організувати механізм співпраці учасників проекту (бібліотекарів та вчених) з метою поширення наукової інформації, популяризувати науково-дослідницьку роботу в університеті.

У електронний архів наукових публікацій НТБ увійшли науково-технічні збірники, журнали, матеріали конференцій, публікації та електронні документи працівників Бібліотеки, бібліографічні видання та багато іншого. Особливістю цього проекту можна вважати самостійне розміщення наукових праць ("самоархівування").

Виклад основного матеріалу

Під час розроблення логічної структури електронного архіву (ЕА) ми дотримувались таких вимог [1]:

1. *Горизонтальний ріст дерева розділів не повинен призводити до необхідності реструктуризації (переміщення) документів.*

При появі нової гілки (розділу) структури ЕА, документ, який належав до однієї гілки, повинен у ній і зберігатись. Переміщення документа може відбуватись за рахунок уточнення назви.

2. *Принципи структуризації повинні забезпечувати можливість автоматизованого об'єднання чи поділу гілок.*

За наявності різних зібрань однієї тематики їх доцільно об'єднувати, а сильно нарощені розділи ділити для ефективнішого пошуку, причому цей процес повинен бути автоматизований.

3. *Структура повинна корелювати зі структурою вищого навчального закладу та акцентувати увагу на тематиці основних джерел інформації (кафедри, бібліотеки, видавництва тощо).*

При побудові ЕА необхідно враховувати джерела надходження інформації та структуру ВНЗ для спрощення процесу адміністрування, пошуку, ведення статистики.

4. *Назва фондів повинна відображати вміст документів фонду.*

5. *Узгодження логічної структури повинно відбуватись за участі науковців та професорсько-викладацького складу ВНЗ.*

Потрібно враховувати побажання професорсько-викладацького складу ВНЗ, щодо структури, оскільки саме вони становитимуть основний контингент відвідувачів та активних учасників (редакторів) ЕА.

6. *Доцільно об'єднати документи у розділи за такими критеріями: тематична ознака, джерело надходжень, ліцензійні вимоги до використання, вид документа.*

7. *Структура відкритого ЕА повинна містити розділ для вільного розміщення авторських праць.*

У архіві передбачено розділ для вільного розміщення наукових праць усіма, хто бажає. Наявність такого розділу відповідатиме принципу відкритості електронного архіву. Розділ для

вільного розміщення авторських праць може мати обмеження за тематикою (наприклад, тематичний технічний відкритий ЕА), за статусом автора (наприклад, лише працівники та студенти ВНЗ), за статусом документа (наприклад, лише попередньо опубліковані праці у фахових виданнях або лише навчально-методичні видання тощо).

Програмне забезпечення Dspace уможливило створення структури архіву за допомогою фондів, підфондів та зібрань документів.

Логічна структура створеного електронного архіву наукових публікацій [6] наведена на рис. 1.



Рис. 1. Логічна структура наукового електронного архіву НТБ

Деталізована схема структури фонду «Електронний науковий архів» подана на рис. 2. Фонд «Електронний архів НТБ» має структуру, зображену на рис. 2.



Рис.2. Структура фонду «Електронний архів НТБ»

Фонд «Електронний архів НТБ» містить матеріали, отримані від Видавництва Національного університету «Львівська політехніка», фізичних та юридичних осіб.

У нього входить ряд підфондів та зібрань, зокрема:

«Вісники та науково-технічні збірники» — міститиме матеріали вісників та науково-технічних збірників, виданих у Львівській політехніці.

«Публікації та електронні документи НТБ, бібліографічні видання» – наукові публікації, опубліковані працівниками НТБ, бібліографічні покажчики, інші матеріали, пов’язані із професійною діяльністю НТБ.

«Автореферати дисертацій» – міститиме автореферати дисертаційних робіт, захист яких відбувався у Львівській політехніці, та додатково надані для НТБ електронні варіанти авторефератів.

«Матеріали конференцій» – матеріали конференцій, що проводились у Національному університеті «Львівська політехніка» та за його межами.

«Навчально-методичні матеріали» – міститиме методичні інструкції до виконання лабораторних, курсових та інших робіт, навчальні посібники, збірники задач тощо.

«Інформаційні та рекламні видання Національного університету «Львівська політехніка» – сюди увійдуть інформаційні та рекламні видання Університету.

Схема опрацювання даних у електронному архіві

Загалом дані опрацьовуються так:

- підготовка інформації про електронний документ (створення шаблону для вводу, визначення тематики та розділу для розміщення документа);
- внесення та обробка електронного документа (редакторську обробку документа див. табл. 1);
- генерування системою індексних файлів з метою організації повнотекстового пошуку;
- формування системою тематичного списку нових надходжень;
- розсилання системою повідомлень про нові надходження для зацікавлених користувачів (e-mail subscription);
- реєстрація системою подій у log-файлах. Оновлення статистичних даних;
- контроль системою часу закінчення ембарго для доступу до файлів;
- надання доступу до електронного документа з урахуванням попередньо визначених повноважень.

Матеріали, що потрапляють у ЕА, опрацьовуються за схемою, поданою у таблиці. Після остаточного правлення (редагування) метаданих матеріал потрапляє у електронний архів та стає доступним користувачам.

Схема внесення та обробки електронних документів ЕА

		Складові частини матеріалів ЕА	Джерело інформації
Матеріал (Item)	Метадані (Metadata)	Загальна фактографічна інформація	
		Інформація про авторів	Електронна версія публікації. Вноситься редакторами та авторами
		Назва	
		Анотація	
		Повний бібліографічний опис	Електронний каталог та інші бібліотечні системи. Визначається редактором.
		Інша інформація	Електронна версія публікації. Вноситься редакторами та авторами
		Додаткова інформація	
		Ключові слова	Електронна версія публікації. Вноситься редакторами та авторами
		Інформація про спонсора	Електронна версія публікації. Вноситься редакторами та авторами
		Прив’язка до тематичної структури архіву	Метадані. Визначається редакторами. Можливість зміни тематичної належності
	Файли	Системна інформація	Створюється ПЗ архіву
		Інша інформація	Метадані. Визначається редакторами та авторами
		Електронні версії матеріалів (публікації)	
		Повнотекстові файли	Електронна версія публікації та додаткові електронні документи. Вноситься редакторами та авторами
		Електронна презентація доповіді	
		Інформація про авторів (їх контакти, фотографії)	
		Аудіо-, відеоматеріали, що доповнюють електронний документ	
		Інша інформація	

Ключові характеристики ПЗ Dspace

Під час вибору програмної системи для побудови наукового архіву потрібно було б, щоби ця система максимально підходила за своїми можливостями Львівській політехніці. Dspace, передусім, відрізняється від своїх аналогів способом організації даних. Необхідно, щоб вони були організовані так, щоби максимально відобразити структуру організації, у якій би була розгорнута система. Особливістю є те, що кожна сторінка Dspace поділяється на розділи, що можуть містити підрозділи та колекції, колекція складається із елементів [1, 7] і т.д., що є повністю адекватним до структури Львівської політехніки.

Крім того, у Dspace є підтримка трьох видів метаданих, а саме – описові, адміністративні та структурні. Така ієрархія дає змогу об'єднувати подібні системи між собою.

У системі доволі добре розвинена та гнучка система прав доступу. Багато функцій Dspace можуть виконуватись анонімно, але, щоб виконати занесення документа, користувачу потрібно зареєструватись. Така система прав доступу дасть змогу чітко та зрозуміло побудувати політику безпеки із супроводу системи.

Вибране програмне забезпечення Dspace забезпечує:

- створення електронних архівів інформаційних ресурсів;
- швидкий доступ до електронного архіву;
- надання результатів дослідження широкому колу користувачів;
- збереження результатів наукових досліджень;
- обмін інформацією на локальному та глобальному рівнях;
- підтримка різноманітних форматів електронного документа, зокрема – pdf, xml, txt, asc, htm, html, css, doc, docx, ppt, pptx, xlsx, xls, jpeg, jpg, gif, та багато інших.

До головних системних недоліків ПЗ Dspace належать:

- Прив'язаність до деревоподібної структури ЕА.

Не можна побудувати структуру за попередньо сформованим списком ключових слів (модерованим списком значень), списком ключових слів, визначених авторами (хмарками тегів, довільними значеннями) або їх об'єднанням.

- Через інтенсивний ріст структури ЕА виникнення складнощів адміністрування прав доступу через перенасичення групами засобів авторизації та неможливість видалення користувачів, що внесли документ у архів.

Під час створення нового фонду\підфонду\зібрання створюється нова група прав доступу (при чому назва групи формується автоматично на підставі системного ідентифікатора, а не назви відповідного виду ієрархії, що призводить до неможливості візуальної прив'язки групи до створеного фонду\підфонду\зібрання). Відповідно за інтенсивного росту структури ЕА буде велика кількість груп прав доступу, якими буде дуже важко керувати. Проблемаю також є неможливість видалення користувача, якщо той ввів хоч один документ. Це стосується випадків, коли до наповнення ЕА залучаються «сторонні» особи (студенти, курсанти, працівники тощо) і після введення документів вони з архівом працювати більше не будуть (у базі даних зберігається прив'язка користувач – документ). Щоб видалити користувача, потрібно видалити всі документи, які він ввів.

- Відсутність автоматизованої роботи з підписками про нові надходження.
- Необхідність доступу до Інтернету під час встановлення ПЗ Dspace.

Встановлення ПЗ Dspace потребує надійного та широкого каналу зв'язку з мережею Інтернет, ще на початку його інсталяції. Попередня компіляція системи викачує з різних сайтів різні файли, для того щоб потім можна було зібрати проект. Набагато зручнішим було б створення єдиного дистрибутиву без докачування додаткових файлів, зокрема, це стосується організацій, що не мають виходу в Інтернет або мають дуже вузький канал зв'язку і тоді час розгортання ЕА значно збільшується.

- Малоінформативна система повідомлення про помилки у роботі Dspace.

Система повідомлення про помилки у роботі зведена до надсилання листа на електронну скриньку із вказанням причини виникнення помилки, файла, у якому виникла помилка та безпосередньо зміст помилки. Такий лист є кориснішим для розробників ПЗ Dspace, а не для

адміністратора, який здебільшого не може виправити цієї помилки. В окремих випадках інформація про виникнення помилки не надсилається.

- Відсутність механізмів автоматизованого об'єднання чи поділу гілок логічної структури.

Під час реорганізації архітектури ЕА іноді виникає необхідність перенесення чи дублювання підфонду чи зібрання до іншого фонду. В такому разі доводиться переносити всі документи зібрання окремо.

- Відсутність механізму формування користувацьких статистичних звітів.

Система статистики є доволі абстрактною і не дає якісної статистичної інформації. Наприклад, відсутній механізм виведення кількості введених документів заданим користувачем за заданий період часу, неможливо відстежити динаміку занесення документів на рівні окремого користувача, статистику перегляду чи скачування документа тощо.

- Неможливість генерування статистичних даних на конкретний момент часу. Статистичні звіти формуються тільки за місяць роботи.

- Некоректна робота повнотекстового пошуку (повнотекстовий пошук працює не для всіх текстових документів).

Для організації повнотекстового пошуку потрібно встановлювати додаткове програмне забезпечення, що здійснює конвертування занесених файлів у текстовий формат (наприклад, пакет Xpdf, що виконує екстракцію тексту з файлів формату pdf у формат txt). В результаті такої роботи разом із документом у архіві зберігається ще й текстове представлення цього документа, що, своєю чергою, збільшує об'єм архіву.

- Зберігання файлів із текстом ліцензійної угоди разом із документом.

До кожного документа, що міститься в архіві, додається текстовий файл із ліцензійною угодою. В окремих випадках ліцензійні угоди доступні користувачам разом з матеріалом архіву, причому текст ліцензійної угоди розпочинається із реквізитів особи, що вносила документ (прізвище, ім'я, адреса електронної поштової скриньки). Але автор не завжди хоче, щоб адреса його електронної скриньки була загальнодоступною. Крім цього, масове дублювання текстових файлів із ліцензійною угодою тільки перевантажує архів непотрібною інформацією (доцільніше організувати посилання на стандартну ліцензійну угоду).

- Недосконалий модуль інтернаціоналізації.

Іноді в інтерфейсі Dspace трапляються синтаксичні помилки, коли робота відбувається в українській локалізації. Визначена проблема сортування даних (назви зібрань, підфондів, прізвищ авторів) за літерами «і» та «ї».

Висновки

Протягом останніх кількох років НТБ створює фонд електронних документів (авторефератів дисертацій, статей, книг у електронному вигляді, інших електронних видань). Для ефективного управління (поширення) науковими ресурсами, довготривалого збереження електронних документів створено електронний архів на платформі Dspace.

Dspace – це зручна та легконалаштовувана система з відкритим кодом, що має всі якості, притаманні сучасним програмним продуктам у сфері створення електронних бібліотек. Об'єднання СУБД PostgreSQL, Java та стабільного веб-сервера Apache у системі Dspace є запорукою високої надійності та хорошої швидкодії. Крім цього, система Dspace добре зарекомендувала себе у НТБ як засіб створення відкритого електронного архіву.

1. Андрухів А., Тарасов Д. Проект створення електронного архіву наукових публікацій у Науково-технічній бібліотеці Національного університету "Львівська політехніка" / А. Андрухів, Д. Тарасов // Комп'ютерні науки та інформаційні технології: матеріали 4-ї Міжнар. наук.-техн. конф. CSIT' 2009, 15–17 жовт. 2009, Львів, Україна / Нац. ун-т "Львів. політехніка", Ін-т комп'ют. наук та інформ. технологій. – Л. – 2009. – С. 245–248. 2. The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14. -URL: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>, 2010. 3. Кудим К.А., Проскудина Г.Ю., Резниченко В.А. Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1. // Электронные библиотеки.. –URL:

http://dspace.nbuv.gov.ua:8080/dspace/bitstream/123456789/300/1/Проскудина_1.pdf, 2010 4. Сравнительный анализ функциональных возможностей систем электронных библиотек / К.А. Кудим, Г.Ю. Проскудина, В.А. Резниченко // Пробл. програмув. – 2007. – № 4. – С. 32–49. 5. DSpace open source software, -URL: <http://www.dspace.org>, 2010. 6. Андрухів А.І., Тарасов Д.О. Методи та засоби побудови електронного архіву у Науково-технічній бібліотеці Національного університету "Львівська політехніка" / А. Андрухів, Д. Тарасов // "Сучасні проблеми діяльності бібліотек в умовах інформаційного суспільства", наук.-практ. конф. (2009; Львів). Матеріали науково-практичної конференції "Сучасні проблеми діяльності бібліотеки в умовах інформаційного суспільства," Львів, 12 лист. 2009 р.: до 165-річчя Нац. ун-ту "Львівська політехніка" / Нац. ун-т "Львівська політехніка", Наук.-техн. б-ка ; редкол.: О.В. Шишка [та ін.]. – Л. : Вид-во Нац. ун-ту "Львівська політехніка". – 2009. – С. 37–48. 7. Тарасов, Д.О. Технологічні особливості опрацювання документів у електронній формі у бібліотеках / Д.О. Тарасов // Інформаційні системи та мережі: (зб. наук. пр.) / відп. ред. В.В. Пасічник. – Л.: Вид-во Нац. ун-ту "Львівська політехніка", 2008. – С. 229–232. – (Вісник Нац. ун-ту "Львівська політехніка"; № 610).

УДК 004.02

В.А. Висоцька, О.Р. Гарасим

Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

МЕТОД ВИБОРУ ОПТИМАЛЬНОГО АЛГОРИТМУ КРИПТОГРАФІЧНОГО ЗАХИСТУ ІНФОРМАЦІЇ

© Висоцька В.А., Гарасим О.Р., 2010

Вирішується завдання оптимізації вибору алгоритму криптографічного захисту інформації для електронного урядування в Україні за допомогою нелінійної згортки критеріїв на основі методу аналізу ієрархій з урахуванням вимог: безпека, швидкість, характеристика алгоритму. В результаті визначено оптимальний криптоалгоритм, який забезпечує цілісність та доступність інформації під час функціонування електронного урядування, автентифікацію користувачів та неможливість заперечення факту відправлення/отримання інформації.

Ключові слова: електронне урядування, криптоалгоритм, оптимізація вибору.

In the paper a task of cryptographic information security algorithm choice optimization is solving for an electronic management in Ukraine by nonlinear criteria convolution on the basis of hierarchy analysis method taking into account such requirements as safety, speed, algorithm description. As a result it was defined an optimal cryptographic algorithm, which provides integrity and availability of information during functioning electronic management, users authentication and impossibility of fact information sending/receipt denial.

Keywords: electronic management , cryptographic algorithm, optimization selection.

Вступ. Загальна постановка проблеми

Під електронним урядуванням («е-урядуванням») розуміють спосіб організації державної влади за допомогою систем локальних інформаційних мереж та сегментів глобальної інформаційної мережі, що забезпечує функціонування органів влади в режимі реального часу та робить максимально простим і доступним щоденне спілкування з ними громадян, юридичних осіб, неурядових організацій. На практиці це означає організацію управління державою та взаємодії з фізичними, юридичними особами та громадськими організаціями з максимальним використанням в органах публічної адміністрації сучасних інформаційних технологій [1]. Тобто е-урядування