

ФУНКЦІОНАЛЬНІ МОЖЛИВОСТІ СУЧАСНИХ СИСТЕМ ЕЛЕКТРОННИХ БІБЛІОТЕК

В.А. Резніченко, Г.Ю. Проскудіна, О.М. Овдій

Інститут програмних систем НАН України,
03187, Київ, проспект Академіка Глушкова, 40.
Тел.: (044)526 6033; e-mail: gupros@isofts.kiev.ua

У роботі йде мова про функціональні можливості сучасних систем електронних бібліотек. Наведені приклади функціональних можливостей найбільш поширених систем – EPrints, DSpace, Fedora, CDSware.

The paper is about functionality of the modern digital library system. Examples of the functional possibilities of the most popular systems such as EPrints, DSpace, Fedora and CDSware are presented.

Вступ

Електронні бібліотеки (ЕБ) це керовані колекції інформації, пов'язані з ними спільноти та сервіси, де інформація зберігається у цифровому форматі і доступна через мережу [1].

У розробках колекцій інформаційних ресурсів електронних бібліотек, забезпеченні їхньої підтримки й доступу до них виявився затребуваним весь спектр ключових технологій керування інформацією, що використовуються у сучасних інформаційних системах – технології баз даних, технології текстового пошуку та веб-технології [1]. Тому не випадково, що тенденції, які склалися та формуються за останні роки, розвитку зазначених технологій істотно впливають на функціональні можливості електронних бібліотек.

Потреба пристосувати особливості сучасних технологій ускладнює розуміння основних понять і функціональних можливостей ЕБ, таким чином створюючи перешкоди й високу вартість створення нових систем ЕБ. Проектувальниками таких систем як правило виступає технічний штат бібліотеки з недостатнім досвідом у розробці такого роду програмного забезпечення (ПЗ) або програмісти, яким бракує знань з області організації, функціонування та побудови ЕБ з використанням веб-технологій. Отже, системи ЕБ часто створюються “з нуля”, використовуючи наявну архітектуру ПЗ, що не враховує переваги ЕБ і наявний досвід побудови такого роду систем. Крім того, часто не враховується той факт, що на сьогодні з'явилося багато доступних інструментів ПЗ для побудови ЕБ або систем електронних бібліотек. Як наслідок це приводить до даремних витрат та нездатності до взаємодії, ізоляції й неминучих витрат у майбутньому.

Дана робота дещо узагальнює, розширює та доповнює роботу [2], де був зроблений порівняльний аналіз функціональних можливостей та особливостей DSpace [3] та EPrints [4] як найбільш популярних вільно розповсюджуваних систем побудови ЕБ, фіксує власний досвід побудови таких ЕБ на їх основі, розглядає функціональні можливості деяких інших відомих систем.

Мета даної роботи – спроба зафіксувати та систематизувати функціональні аспекти ЕБ, підсумовуючи одержаний досвід, та практичні й теоретичні знання за побудовою сучасних систем ЕБ на основі наявного ПЗ.

Робота написана наступним чином. У розділі 1 йде мова про термінологічні та функціональні аспекти ЕБ. У розділі 2 описані функції пошуку і перегляду. Розділ 3 містить визначення поняття користувача електронної бібліотеки, та розгляд функцій безпосередньо пов'язаних з користувачем. Розділ 4 розповідає про функції внесення контенту. Адміністративний аспект ЕБ розглядається у розділі 5.

1. Термінологічні та функціональні аспекти електронних бібліотек

Електронні бібліотеки це – керовані колекції інформації, пов'язані з ними спільноти та функції (сервіси), де інформація зберігається у цифровому форматі і доступна через мережу [1]. Електронну бібліотеку можна розглядати як можливо віртуальну організацію мережевого типу, що збирає, керує та довготерміново зберігає різноманітний *цифровий контент* та пропонує своїм *користувачам* (або спільнотам користувачів) спеціалізовані *функціональні можливості* щодо контенту.

Репозиторії (repositories) це мережеве сховище призначене для зібрання та зберігання інформаційних колекцій, та надання до них доступу [5]; *архіви* це сховища, що спеціально організовані для довготермінового зберігання інформації. Іноді ці терміни: репозиторії, архіви, сховища вживаються як синоніми до терміну електронна бібліотека.

Системи електронних бібліотек (СЕБ) це програмне забезпечення, що базується на визначеній, можливо розподіленій, архітектурі і забезпечує всю функціональність, що потребує окрема електронна бібліотека [5]. Користувачі взаємодіють з ЕБ через відповідну СЕБ.

У літературі під ЕБ часто розуміють абстрактну систему, що складається з двох компонент: фізичної та віртуальної. Під СЕБ розуміють конкретну систему ПЗ. Для кожної ЕБ існує єдина діюча СЕБ (що можливо складатися з декількох взаємодіючих СЕБ).

Інформація в ЕБ декларується в термінах інформаційних об'єктів або ресурсів, які можуть містити текстову або мультимедійну інформацію (зображення, відео, аудіо), анотації та метадані. Це часто називають контентом. Тобто під контентом ми розуміємо ту інформацію, що обробляє ЕБ і надає своїм користувачам [6]. Контент складається з множини інформаційних ресурсів, організованих в колекції. Термін контент поєднує всі форми інформаційних об'єктів, присутніх у ЕБ. Хоча розходження між даними й метаданими часто залежить від контексту, метадані звичайно структуровані й покривають різні категорії інформації про інформаційний ресурс (об'єкт). Найбільш загальний вид метаданих – описові метадані, що присутні в каталогах та індексах, та містять стислу інформацію використовувану для опису інформаційного об'єкта в ЕБ, що має велике значення для синтаксичної, семантичної та контекстної інтерпретації.

Під функціональними можливостями ми розуміємо сервіси, що пропонує ЕБ своїм користувачам, їх групам чи окремим користувачам. Добре, коли ЕБ має широкий спектр функціональних можливостей і сервісів, але існує і необхідний мінімум функцій, що забезпечується у ЕБ: реєстрація нового інформаційного об'єкта, індексування, пошук та перегляд. Ці функції можуть бути пристосовані до специфічних потреб спільнот користувачів ЕБ та окремих вимог, що висуваються специфікою інформаційних ресурсів, які вона містить.

Як правило стандартна дистрибуція подібних систем ЕБ включає наступні функціональні аспекти [7]:

– *Перегляд*, де передбачені можливості щодо перегляду розділів (підрозділів) бібліотеки, їхніх колекцій, елементів (наприклад, документів) та файлів, а також візуалізацію окремих елементів та здійснення пошуку.

– *Користувальницький аспект*, де передбачено функції щодо входу користувачів у систему (відповідно – виходу з системи), реєстрації нових користувачів, функцій відновлення втрачених аутентифікаційних даних, керування профілем користувача, зміна пароллю.

– *Внесення*, тут реалізовані функції, що відповідають за внесення нового об'єкта (наприклад, документа) у систему, визначення процесу внесення або робочого процесу (workflow process), і нарешті, поглинання нових об'єктів безпосередньо в самому репозиторії (ЕБ).

– *Адміністративний аспект*, де включені функції керування самою системою, наприклад, створення, модифікація, вилучення розділів (підрозділів), колекцій, користувачів, груп користувачів, облікових записів користувачів, функцій авторизації.

Далі розглянемо ці аспекти більш детально демонструючи функціональну реалізацію на прикладах на сьогодні відомих нам систем для створення ЕБ. Функціональність розглядається на основі вивчення та застосування ПЗ DSpace, EPrints, Greenstone, що докладно розглянуті у [8–10], нині досліджуються також системи Fedora [11] та CDSware [12].

Fedora (Flexible Extensible Digital Object Repository Architecture) заснована на сервісах система з відкритим кодом для керування репозиторіями цифрових об'єктів. Вона є базовою архітектурою для електронних репозиторіїв скоріше ніж готовим рішенням для ЕБ і на відміну від систем “під ключ” з фіксованим інтерфейсом таких як DSpace та EPrints дозволяє створювати різноманітні системи для різних галузей. Система надає два види доступу: клієнт для внесення, підтримки та експорту електронних об'єктів та веб-сервіси доступу через інтерфейси SOAP або REST (використовує стандартні HTTP методи для запиту та маніпулювання веб-ресурсами). Об'єкт у системі може мати декілька представлень (наприклад, декілька форматів), також за допомогою заснованих на RDF моделі зв'язків об'єкти можуть бути пов'язані між собою та з зовнішніми ресурсами. Центральний сервіс репозиторію, на якому заснована архітектура системи, надає чотири основних API – керування, доступу, пошуку та OAI сервіс для харверстину метаданих [11].

CDSware (CERN Document Server Software – ПЗ сервера документів європейського центру ядерних досліджень) – пакет прикладних програм, який забезпечує структуру та інструментальні засоби для побудови та керування сервером ЕБ. В галузі фізики елементарних частин необхідність разом використовувати інформацію між науковими організаціями світу привела до народження всесвітньої павутини, сьогодні на часі – рішення для побудови великомасштабних систем відкритого доступу. CDSware – результат більш ніж десятилітнього розвитку, націленого на злиття кращих традицій бібліотечних систем та кращих сучасних технологій інформаційного пошуку. ПЗ CDSware призначене перш за все для керування великими колекціями (більш ніж 1 млн. записів у CERN) та має продвинуті бібліотечні функціональні можливості. Головна особливість архітектури CDSware полягає у її модульній логіці, де кожний модуль втілює специфічні, визначені функціональні можливості системи ЕБ. Модулі взаємодіють між собою, базою даних та інтерфейсними шарами. Система є такою, що добре налаштовується та розширюється [13].

2. Пошук та перегляд

Основу ЕБ складають інформаційні функції, що забезпечують навігацію за інформаційними ресурсами. Навігація за інформаційними ресурсами забезпечується функціями пошуку та перегляду.

Пошук інформаційних ресурсів. Пошук – істотний компонент виявлення об'єктів у будь-якій бібліотечній системі. З огляду на високі користувальницькі вимоги до пошукових машин, надання як можна більшого числа пошукових можливостей – важлива особливість більшості сучасних бібліотечних систем. Пошук має надавати

можливість знаходити в системі інформаційні ресурси, їх властивості та/або зміст. Пошукова функція ЕБ задовольняє наступним вимогам.

Інтерфейс пошукових засобів побудований так, щоб він був інтуїтивно сприятливим для користувачів. Як правило використовується механізм побудови пошукових запитів. У найпростішому випадку це фіксована множина пошукових запитів, в які заноситься значення параметрів, що задається користувачем. Там, де це можливо, значення пошукових термінів вибираються з заданих списків (наприклад, мова документа, формат документа, перелік використовуваних метаданих, дата створення документа тощо).

Як правило системи ЕБ надають два типа пошуку: *стандартний* або *простий* та *розширений*. Користувач має можливість самостійно вибирати той чи інший пошук.

Стандартний пошук надає деякі мінімальні пошукові можливості, якими можуть легко оволодіти більшість користувачів і, які задовольняють достатньою повнотою. Пошук здійснюється в усіх індексах, які побудовані у системі, тобто в усіх описових полях та в усіх текстах.

Розширений пошук надає користувачам усі можливості пошуку:

- повнотекстовий пошук;
- пошук усіх описових метаданих;
- пошук вибраних полів метаданих (набір полів, як правило 3-4 поля, визначається користувачем, поєднуючи їх бульовими операторами).

Звуження простору пошуку. У сучасних системах ЕБ можна *звужити простір пошуку*, а також здійснювати *багатокроковий пошук*. Звуження простору пошуку передбачає можливість відокремити підмножину інформаційних ресурсів, що шукаються. Критеріями такого відокремлення можуть бути наступні:

- вибір ресурсів певного виду (книги, журнали, дисертації тощо);
- вибір ресурсів, що відповідають певним значенням метаданих (дата створення, автор, організація, мова документа, формат представлення документа тощо);
- використання ієрархічних словників пошукових термінів, предметних рубрикаторів, тезаурусів або класифікаторів.

Багатокроковий пошук передбачає те, що пошук ведеться серед тих ресурсів, які були знайдені в результаті попереднього пошуку. В результаті такого пошуку користувачеві надається можливість послідовно знаходити більш релевантні ресурси.

Результати пошуку можуть бути відсортовані за:

- автором;
- назвою;
- датою випуску;
- ступенем релевантності;
- іншими характеристиками.

Мова пошуку у СЕБ має необхідну для пошуку релевантних документів повноту, що забезпечує достатню гнучкість, виразність та наочність. Для цього вона задовольняє наступним можливостям:

- пошук документів за їх повними текстами та/або за описовими даними (метаданими);
- пошукові терміни мають складатися з окремих пошукових слів або фраз;
- відсікання закінчення або початку слів у пошукових термінах;
- використання групових символів (що замінюють один або декілька символів: знак питання (?) може представляти будь-який одиничний символ; зірочка (*) використовується для представлення будь-якого символу чи групи);
- використання бульової логіки (як правило, логічних зв'язок ТА-АБО-НІ);
- пошуку за словоформами та синонімами пошукових термінів, а також морфологією мови (пошук слова в усіх його морфологічних формах), на якій формулюються пошукові терміни;
- чутливість або нечутливість до регістру символів;
- пошук за близькістю розміщення слів у тексті;
- пошук за фонемним звучанням пошукових термінів.

Пошукові засоби СЕБ надають також наступні функціональні можливості:

- налаштування параметрів пошуку;
- збереження результатів пошуку для наступного використання,
- збереження текстів запитів та їх повторне використання самостійно або у складі інших запитів;
- представлення результатів пошуку у різних форматах;
- допомога користувачам за використанням пошукових засобів та підвищенням ефективності пошуку.

Послідовний перегляд інформаційних ресурсів. Інший важливий механізм виявлення матеріалів у системах ЕБ – перегляд, процес, де користувач переглядає конкретний індекс, наприклад, індекс назви і в процесі пошуку елементів, що цікавлять, проводить за ним навігацію. Цей вид пошуку надає можливість зробити послідовні кроки, що приводять до знаходження необхідного документа. Ця послідовність кроків може бути наступною:

- вибір типу документа (книга, стаття з журналу, звіт і т.д.);
- із предметного покажчика (класифікатора, рубрикатора) вибір теми документа;

– подальше уточнення необхідного документа. Наприклад, для журналів: вказівка назви журналу, року, випуску, зміст, стаття.

На кожному кроці послідовного перегляду має надаватися можливість організації пошуку в межах зазначеного контексту.

Бібліотечні системи в конфігурації за замовчуванням дозволяють користувачам здійснювати перегляд за індексами назв, автором, датою та предметною темою. Цей перелік перегляду може бути змінено при додатковій настройці.

Приклади реалізації функцій пошуку і перегляду у розглянутих нами системах. Модуль індексації й пошуку DSpace [7] має дуже простий API, що надає можливість індексації нового контенту, регенерації індексу й виконання пошуку по всьому архіву, розділу, підрозділу або колекції. За API стоїть вільно розповсюджувана пошукова Java-машина Lucene [14], що надає пошук за полями метаданих, видаленням стоп-слів, виконанням пошуку слова у всіх його морфологічних формах, можливістю інкрементно додавати новий проіндексований контент без регенерації всього індексу. Індекси пошуку можна конфігурувати, що дозволяє налаштувати індексовані поля метаданих. Система надає простий і розширений пошук. Підсистема перегляду в DSpace забезпечує простий API, де надається можливість визначити індекс і його підрозділ. Потім підсистема перегляду розкриває ту частину індексу, що цікавить користувача. Індекси, які можуть бути переглянуті: назва, автор, дата випуску документа й предметні (тематичні) терміни. Додатково, перегляд може обмежуватися до документів усередині визначеної колекції або розділу.

У системі EPrints [8] доступно кілька типів навігації в репозиторії; початкова конфігурація включає перегляд за роком видання, за темою; є можливість включити перегляд документів за будь-яким полем метаданих; можна здійснювати простий пошук та розширений пошук, де для багатокритеріального пошуку представлено безліч полів метаданих; тут також є доступ до повнотекстового пошуку; доступна стрічка RSS (Really Simple Syndication) [15] про останні надходження до архіву.

Fedora [10] надає два типи пошукових індексів. Простий пошук – заснований на простому індексі побудованому на властивостях об'єкта та елементах набору метаданих Дублінського ядра. Та пошук заснований на RDF-індексі ресурсів, який включає більше інформації про об'єкт та зв'язки між об'єктами. Обидва види пошуку доступні за допомогою REST-інтерфейсу. У термінах користувальницького інтерфейсу система надає простий та розширений пошук через веб-форму. Всі запити нечутливі до регістру. Простий пошук дозволяє шукати за словом або фразою, у будь-яких індексованих метаданих об'єкта. Розширений пошук за полями метаданих з будь-якою комбінацією елементів використовуючи оператори порівняння для строкових полів (= та ~), та операторів порівняння для дат (=, >, ≥, <, ≤), крім того, можуть використовуватись групові символи * та ?. У останній версії системи був доданий сервіс GSearch (Generic Search), який дозволяє індексувати текстовий зміст об'єктів та виконувати пошук, дозволяючи вбудовувати різні пошукові машини, на даний момент – Lucene та Zebra.

3. Користувальницький аспект

Користувальницький аспект передбачає перш за все функції щодо входу користувачів у систему (відповідно – виходу з системи), реєстрації нових користувачів, функції відновлення втрачених аутентифікаційних даних, керування профілем користувача та зміна паролю.

Під користувачем ми розуміємо того (людина або машина), хто взаємодіє з ЕБ, яка з'єднує користувача з запитуємою ним інформацією та забезпечує повідомлення про її надходження. З поняттям користувача також пов'язуються такі елементи як права доступу, що користувач має в межах системи, профілі користувачів з їх характеристиками, що показують їх поведінку у системі або представляють цих користувачів у співробітництві з іншими користувачами.

Користувальницький аспект та пов'язаний з ним набір функціональних можливостей, що надається у ЕБ, залежить від ступеню відкритості ЕБ:

– повністю відкрита ЕБ або вільного доступу, наприклад, Вікіпедія [16], де всі матеріали, що містяться у системі є необмеженого, відкритого доступу, а також будь-який користувач може вносити свої матеріали до бібліотеки. Тому у такій системі реєстрація користувачів можлива, але не є обов'язковою;

– частково відкрита ЕБ, де зберігаються матеріали як відкритого, так і обмеженого доступу;

– ЕБ закритого доступу, де матеріали доступні тільки тим користувачам, що мають відповідні права.

Реєстрація, аутентифікація та обмеження прав доступу має виконуватися, в основному, в тому випадку, коли в ЕБ є інформаційні об'єкти і сервіси, доступ до яких обмежується. Це може бути обумовлено наступними факторами:

– економічні фактори – деякі інформаційні ресурси і послуги надаються на платній основі;

– фактори таємності – доступ до деяких ресурсів може бути обмежений у зв'язку з наявністю в них секретної чи конфіденційної інформації;

– нормативні фактори – доступ обмежений у зв'язку, наприклад, з необхідністю захисту авторських прав.

Є ще один аспект ЕБ, який також вимагає цих функцій, а саме – наявність персоніфікованих функцій: створення користувачами персональних профілів ЕБ, розсилання повідомлень про нові надходження, використання e-mail і т.п. Наприклад, функція розсилання повідомлень про нові надходження позбавляє

користувачів від необхідності регулярно відвідувати репозиторій для виконання такої перевірки. Система виконує такі розсилання по e-mail через службу підписки (DSpace, EPrints).

Таким чином, дана служба забезпечує такі функції:

- реєстрація користувачів;
- розподіл прав доступу до інформаційних ресурсів і сервісів;
- аутентифікація (перевірка імен і паролів користувачів).

Реєстрація користувачів може проводитися за визначеними принципами. Реєстрація проводиться заповненням персональної реєстраційної форми користувача. Зареєструватися може будь-який користувач. При реєстрації, крім іншого, треба вказати загальні відомості. Якщо ресурс, яким бажає користуватися користувач, має обмежений доступ, то треба додатково вказати логін та пароль. У випадку використання сервісів, які передбачають зворотній зв'язок з користувачем, наприклад, розсилка повідомлень, то треба додатково вказати адресу електронної пошти. Після цього користувач може входити в ЕБ з будь-якого комп'ютера, підключеного до мережі Інтернет. При цьому будуть представлені всі налагодження користувача (персональний підбір журналів та статей, збережені пошукові запити, персональна статистика).

Наприклад, у системі DSpace про кожного користувача у системі зберігається така інформація:

- адреса електронної пошти;
- прізвище та ім'я;
- пароль;
- список колекцій, відносно яких користувач повідомляється про нові надходження;
- ознака самореєстрації, тобто чи було створено обліковий запис користувача автоматично чи, наприклад, в результаті звернення до адміністратора.

Розмежування прав доступу використовується тільки для тих інформаційних ресурсів і сервісів, які мають обмежене використання. З погляду розмежування прав доступу всі послуги та ресурси діляться на дві категорії:

- повні тексти інформаційних ресурсів;
- всі інші ресурси та послуги ЕБ.

Передбачається, що друга категорія, як правило, буде безкоштовною і нею можуть користуватися всі користувачі. В деяких випадках все ж таки необхідно буде виконати формальну процедуру реєстрації в ЕБ для надання персональних послуг. Що стосується першої категорії послуг, то для них передбачаються наступні права доступу:

- вільний доступ, що надається до більшості ресурсів;
- індивідуальний доступ до окремих ресурсів, у цьому випадку передбачається, що користувач попередньо вирішує проблему доступу до ресурсу (чи то оплачує доступ, чи то отримує дозвіл від власника ресурсу) і після цього йому надається право отримати його в електронному вигляді;
- оформлення підписки, коли користувачі оформлюють підписку доступу до статей всіх або деяких ресурсів ЕБ;
- пільговий доступ. В деяких випадках користувачі визначених категорій (члени деяких організацій та асоціацій, учасники спеціальних проектів і т.д.) отримують пільгові умови доступу до ЕБ, включаючи і вільний доступ до всіх ресурсів.

Аутентифікація (authentication) або підтвердження дійсності (автентичності) — функція перевірки відповідності користувача і того, за кого він намагається себе видати, за допомогою унікальної інформації, у найпростішому випадку – за допомогою ім'я й пароля. Цю функцію слід відрізнити від ідентифікації (впізнання суб'єкта інформаційної взаємодії) та авторизації (перевірки прав доступу до ресурсів системи). Одержавши введений користувачем логін і пароль, комп'ютер порівнює їх зі значенням, що зберігається в спеціальній базі даних і, у випадку збігу, пропускає користувача в систему.

Механізм аутентифікації [17] передбачає попереднє введення ім'я та паролю користувача для доступу до обмежених інформаційних ресурсів ЕБ. У DSpace система формування паролів має достатню захищеність. Система підтримки імен користувачів та їх паролів має можливість нагадати користувачу його ім'я та пароль, якщо він їх забув.

Так, наприклад, у системі Dspace аутентифікація відбувається у випадку, коли прикладний сеанс позитивно ідентифікує себе таким, що належить користувачу (групі). Функція здійснюється через так званий механізм Stackable Authentication: конфігурація DSpace визначає “стек” методів аутентифікації. Потім менеджер аутентифікації випробує кожний із цих методів для того, щоб ідентифікувати користувача (групу) із приналежним сеансом.

4. Внесення

Тут реалізовані функції, що відповідають за внесення нового об'єкта (документа) у систему, визначення процесу внесення або робочого процесу (workflow process), і нарешті, поглинання нових елементів безпосередньо в самому архіві ЕБ.

Коли новий користувач реєструється в системі, йому виділяється окрема робоча область, куди він може завантажувати свої документи. У процесі внесення нового об'єкта користувач має виконати дії, послідовність

яких може залежати від системи та конфігураційних налаштувань, але загалом їх можна перерахувати наступним чином:

- визначити розділ, підрозділ, колекцію, якщо у системі існують ці структурні утворення і до яких користувач має дозвіл вносити свій контент;
- визначити тип документа (книга, глава, стаття, звіт, методичний матеріал і т.п.);
- завантажити файли, при необхідності створити набори файлів;
- внести опис об'єкта, згідно визначеному (на попередніх кроках або у конфігураційному файлі системи) набору метаданих. Деякі описові поля є обов'язковими, що означає неможливість продовження дій внесення у випадку їхнього незаповнення;
- підтвердити ліцензійну угоду, що визначає всі юридичні аспекти збереження та розповсюдження даного об'єкта у системі;
- подати об'єкт на редакторську перевірку.

На будь-якому кроці процес внесення можна перервати без втрати даних, введених на попередніх кроках.

Далі зупинимось на функціях адміністрування щодо внесення контенту.

Функції підтримки кроків внесення. Адміністрування дій внесення дозволяє адміністратору системи визначати число та типи кроків внесення контенту від початкової подачі до остаточного внесення у ЕБ.

Наявність окремої робочої області для подачі документа – забезпечення окремого робочого простору, де зберігаються незавершені або ще не схвалені (наприклад, редактором) документи внесення. Ця функція спрощує процес подачі документа, дозволяючи користувачам зберігати перерване або незавершене з тих чи інших причин внесення, без остаточного розміщення цих документів у репозиторії. У EPrints стан файлів зберігається у БД.

Функціональність підтримки ролей користувачів щодо внесення, яке передбачає конфігуруєми набір функцій перевірки або адміністрування в межах архіву. Наприклад, у CDSware підтримуються ті, хто вносить контент або депоненти; модератори; рецензенти; ті, хто схвалює внесення або редактори, адміністратори. У DSpace можна представляти приблизно ті ж самі ролі [8]. Дещо спрощена підтримка ролей у EPrints, але роль депозитора може бути модифікована та розширена.

Деякі системи застосовують ті ж самі ролі та процеси у всіх колекціях, що зберігаються у ЕБ. Інші ж визначають ці функції на рівні колекції, дозволяючи різним колекціям в рамках однієї системи пропонувати різне представлення та процеси перегляду (DSpace, CDSware).

Функції підтримки внесення (у Fedora не підтримуються ці функції).

E-mail повідомлення депонента – посилає поштового листа користувачу про стан внесеного контенту, наприклад, що документ був схвалений та включений у архів, або був повернутий депоненту на доопрацювання.

E-mail повідомлення адміністратора контенту (рецензента, редактора), коли внесений контент надходить до нього для перегляду (схвалення).

Функція персоніфікованого доступу до системи для зареєстрованих користувачів дозволяє надавати та обробляти інформацію щодо стану внесених документів, наприклад, документ знаходиться у робочій області користувача, документ на розгляді у редактора, документ внесений або знаходиться у архіві та ін.

Перегляд внесеного контенту, дозволяє користувачу-депозитору переглянути свій контент, як прийнятий у репозиторій, так і той що був неповністю внесений, або ще не схвалений. Користувач-адміністратор (редактор) може переглядати контент, що надійшов йому на перевірку.

Функції підтримки ліцензійної угоди. Щоб дозволити організації, якій належить репозиторій адмініструвати та поширювати внесені матеріали, слід укласти з кожним, хто вносить їх до репозиторію, ліцензійну угоду з метою вільного розповсюдження контенту, перетворення його в інші формати довгострокового зберігання та підтримання контенту за можливістю безкінечно довго.

Функція запиту на підтвердження ліцензійної угоди. Підпис ліцензійної угоди є частиною (кроком) процесу внесення контенту. У деяких системах підтримується декілька ліцензійних угод, що можуть різнитися в різних колекціях або для різних користувачів. Інші звертаються до подібних угод за межами ПЗ (CDSware). У Fedora це могла би бути ліцензія на поширення об'єкта або лише простий файл, що зберігається разом з кожним об'єктом.

Збереження ліцензійної угоди разом з контентом. Функція дозволяє зберігати ліцензійну угоду разом з кожним внесенням контенту. У зв'язку з тим, що ліцензійна угода може змінюватися з часом, або за типом контенту, це забезпечує ясність, оскільки кожна угода застосовується до кожного внесення (DSpace, Fedora). Щоб забезпечити цю функціональну можливість EPrints можна конфігурувати.

Підтримка перенесення або зміни формату (format migration) документа – важлива функція репозиторія. Водночас як PDF – стандарт для документів сьогодні, передбачається, що організація підтримки візьме на себе відповідальність за перенесення цих документів у формати, що могли б бути прочитаними програмним забезпеченням у майбутньому.

Приклади реалізації функцій внесення. Процес внесення документів систем EPrints та DSpace докладно розглянуто у [8–9], система CDSware теж має свої особливості. Тут кожна колекція може мати свою політику внесення, що означає можливість виконання:

- безпосереднього внесення;

- контрольованого внесення;
- внесення з простим схваленням;
- внесення з можливістю рецензування та редакційною правкою.

Кожне внесення має власний процес керування за допомогою інтерфейсу адміністратора, визначення екранів внесення та визначення застосовуваних дій. У системі підтримується режим пакетного внесення [12].

Для внесення даних Fedora надає адміністративного java-клієнта та API, які використовуються розроблюваними для системи фронт-ендами (Fez, Muradora, FABULOUS) для реалізації більш зручного для користувачів, та орієнтованого на конкретну сферу інтерфейсу для внесення даних. Крім того система надає такі корисні утиліти як утиліта міграції, що дозволяє масовий експорт та поглинання об'єктів. Також є утиліта для масового створення та зміни об'єктів репозиторія. Електронний об'єкт може бути представлений у різних xml-форматах. Підтримка системи каталогів базовою системою не надається [11].

5. Адміністративний аспект

Адміністративний аспект включає функції керування самою системою, наприклад, створення, модифікація, вилучення розділів (підрозділів), якщо вони передбачені в моделі даних ЕБ, колекцій, користувачів, груп користувачів, облікових записів користувачів, функцій авторизації.

Функції підтримки колекцій. Системи (наприклад, DSpace) підтримують визначення множини колекцій та/або груп користувачів в межах однієї встановленої системи. Колекції можуть визначатися у різний спосіб: наприклад, за темою, за типом контенту, за метою або аудиторією (наприклад, серія робочих статей або матеріали для навчального процесу). Користувальницькі групи за мірою реалізації потреб установи можуть бути представлені академічними відділами, школами, науково-дослідними інститутами, адміністративними одиницями (наприклад, музеї, лікарні і т.д.).

Наприклад, CDSWare [12] підтримує ієрархію колекцій, а також віртуальні колекції або “горизонтальні представлення”.

Кожна колекція може мати своє власне визначення запису, що включає: поля метаданих (обов'язкові, необов'язкові або факультативні, керовані на час внесення); формати повних текстів; перевірені або доопрацьовані версії. Встановлення різних параметрів внесення для кожної колекції дозволяє адміністратору встановлювати різний контент внесення та параметри перевірки та схвалення (в разі потреби) для кожної з колекцій та/або користувальницьких груп, що визначені у ЕБ (CDSware). У Fedora це здійснюється через застосування шаблонів METS [18]. У EPrints ця функція не підтримується.

Можливість створення домашньої сторінки для кожної колекції. Ця функція підтримується у DSpace, Fedora, у CDSware домашню сторінку колекції можна конфігурувати, у EPrints – не підтримується.

Авторизація незважаючи на те, що великі зусилля додаються для полегшення доступу до електронних матеріалів установи, проте відкривати повний доступ до вмісту репозиторія не завжди доцільно. Крім того, такі функції як депонування й редакторська перевірка мають бути прив'язані до відповідних користувачів і обмежуватися ними. Тому система має функцію авторизації, що заснована на прив'язці дій (в деяких системах такі прив'язки називаються *політики ресурсів*) до об'єктів і списків користувачів (або груп), які можуть ці дії виконувати.

Як правило у бібліотечних системах представлені наступні групи користувачів, які відрізняються правами доступу:

- мінімальний користувач може переглядати вміст ЕБ, підписуватися на списки розсилання, створювати збережені пошуки;
- депонент має права мінімального користувача, та додатково за ним закріплена робоча область, куди він може завантажувати свої об'єкти (документи), і може подавати їх редактору на розгляд;
- редактор має права депонента, і може приймати, відхиляти або вилучати об'єкти, подані користувачами для розміщення в ЕБ;
- адміністратор має права редактора, а також може керувати обліковими записами користувачів та виконувати будь-які дії на сайті.

Користувач може бути асоційований з декількома групами одночасно. Для більш ефективного керування наданням привілеій, адміністратори можуть використовувати групи також і в якості *ролей*. Щоб користувач міг виконати дію з об'єктом системи, він повинен мати дозвіл, що задається явно. Відсутність явно заданого дозволу за замовчуванням приводить до політики заборони.

Якщо в ЕБ реалізована гнучка система прав доступу (DSpace), це дозволяє обмежувати доступ до різних частин архіву. Кожному розділу архіву можна призначити групу користувачів, яким дозволяється доступ до даного розділу. Кожній колекції призначається безліч із окремих користувачів і груп, які будуть депонентами для цієї колекції, будуть мати доступ до вмісту, відігравати роль редакторів, або, нарешті, адмініструвати колекцію.

Керування версіями документа (історія). Керування версіями (version control) – функція для полегшення роботи з інформацією, що часто змінюється. Система керування версіями дозволяє зберігати декілька версій одного і того самого документа, і при необхідності, повертатися до більш ранніх версій, визначати, хто й коли зробив ту чи іншу зміну та інше.

У бібліотечних системах, де ведеться робота з великою кількістю електронних документів, що безупинно змінюються, програмне забезпечення веде історію змін для всіх її статей, використовуючи методи, аналогічні тим, які застосовуються в системах керування версіями.

Функція керування версіями стає все більше й більше важливою, оскільки репозиторії (архіви) електронних ресурсів старішають, і контент поступово мігрує до нових форматів та технологій. Версії можуть використовуватися для підтримки не тільки міграції, але також і для виправлення й технічної модифікації істотно еквівалентного семантичного контенту. У деяких системах, версії також використовуються для семантично різного контенту, наприклад версій статей до публікації та після.

Система генерації користувальницької статистики та звітів. Функції генерації користувальницької статистики та звітів дозволяють адміністраторам репозиторію відстежувати використання та сприйняття репозиторію, це полегшує планування місткості системи та підтримувати розміщення внутрішніх ресурсів та бюджету (DSpace). Fedora генерує логфайли використання та поведінки системи, і хоча ці логфайли представлені в XML-форматі, та могли б бути проаналізовані інструментами, що генерують звіти, проте такі інструменти не є вбудованими в систему. У CDSware використовується інструментальні засоби сторонніх компаній, типу Webalizer.

Висновки

За мірою розвитку ЕБ та зростанням до них вимог [19] змінюються також і їх функціональні можливості. У даній роботі ми спробували зафіксувати та систематизувати функціональність розглянутих та використовуваних систем з метою їх подальшого застосування та розширення.

1. *Коголовский М.Р.* Тенденции развития технологий управления информационными ресурсами в электронных библиотеках // Тр. VIII Всероссийской научн. конф. Электронные библиотеки: перспективные методы и технологи. – Суздаль, Россия. – 2006. – С. 46 – 55.
2. *Кудим К.А., Проскудина Г.Ю., Резниченко В.А.* Сравнительный анализ функциональных возможностей систем электронных библиотек // Проблемы программирования. – 2007. – № 4. – С. 32-48.
<http://eprints.isofts.kiev.ua/320/>
3. *DSpace Federation* Web site <http://dspace.org/>
4. *EPrints Software* <http://software.eprints.org/>
5. *Kahn Robert, Wilensky Robert* A Framework for Distributed Digital Object Services, May 1995.
<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>
6. *Arms William Y.* Key Concepts in the Architecture of the Digital Library, D-Lib Magazine, July 1995.
<http://www.dlib.org/dlib/July95/07arms.html>
7. *Scott Phillips, Cody Green, Alexey Maslov, Adam Mikeal, John Leggett* Manakin: A New Face for Dspace // D-Lib Magazine. – November / December 2007 – Vol. 13. N 11/12.
<http://www.dlib.org/dlib/november07/phillips/11phillips.html>
8. *Кудим К.А., Проскудина Г.Ю., Резниченко В.А.* Создание научных электронных библиотек с помощью системы DSpace // Проблемы программирования. – 2007. – № 3. – С. 49–60.
<http://eprints.isofts.kiev.ua/233/>
9. *Новицкий А.В., Кудим К.А., Резниченко В.А., Проскудина Г.Ю.* Создание научных архивов с помощью системы EPrints // Проблемы программирования. – 2007. – № 1. – С. 46–60.
<http://eprints.isofts.kiev.ua/157/>
10. *Резниченко В.А., Проскудина Г.Ю., Овдий О.М.* Создание цифровой библиотеки коллекций периодических изданий на основе Greenstone. Электронные библиотеки. – 2005. – 8. Вып. 6.
<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>.
11. *Fedora* - Flexible Extensible Digital Object Repository Architecture. <http://www.fedora.info/>
12. *CERN Document Server Software (CDSware).* <http://doc.cern.ch/EDS/current/guide/english/>
13. *CERN Document Server Software: the integrated digital library.* A. Pepe, T. Baron, M. Gracco, J.-Y. Le Meur, N. Robinson, T. Simko, M. Vesely // ELPUB 2005 conference, Heverlee (Belgium), 8-10 June 2005. – <http://cdsware.cern.ch/invenio/doc/elpub2005.pdf>
14. <http://lucene.apache.org/>
15. <http://ru.wikipedia.org/wiki/RSS>
16. *Jeremy Frumkin* The Wiki and the digital library // *OCLC Systems & Services.* – 2005. – Vol. 21. N 1. – pp. 18–22.
<http://www.emeraldinsight.com/10.1108/10650750510578109>
17. *Ричард Э. Смит* Аутентификация: от паролей до открытых ключей = Authentication: From Passwords to Public Keys First Edition. – М.: Вильямс, 2002. — С. 432.
18. *Metadata Encoding and Transmission Standard (METS).* <http://www.loc.gov/standards/mets/>
19. *Dagobert Soergel* A Framework for Digital Library Research // D-Lib Magazine. – November – December 2002. – Vol. 8. – N 11/12.
<http://www.dlib.org/dlib/december02/soergel/12soergel.html#9>