

СУЧАСНІ ТЕХНОЛОГІЇ ФОРМУВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ

Надано огляд технологічних підходів для формування ресурсної бази електронної бібліотеки, зауважено на відсутності типових технологічних рішень для формування повнотекстових ресурсів електронних бібліотек. Наведено досвід інтеграції технології електронної бібліотеки і автоматизованої бібліотечної технології у єдиний технологічний процес.

Ключові слова: інформаційні технології, електронні ресурси, оцифровка, інформаційний пошук, інтеграція технологій.

Представлен обзор технологических подходов для формирования ресурсной базы электронной библиотеки, акцентировано внимание на отсутствии типовых технологических решений для формирования полнотекстовых ресурсов электронных библиотек. Приведен опыт интеграции технологий электронной библиотеки и автоматизированной библиотечной технологии в единый технологический процесс.

Ключевые слова: информационные технологии, электронные ресурсы, оцифровка, информационный поиск, интеграция технологий.

Створення електронних бібліотек пов'язано з формуванням цифрових колекцій, створенням розподілених ресурсів і організацією широкого спектру онлайн-інформаційних сервісів, які зводяться до двох основних типів послуг: *пошук* інформації та *першоджерел* і *доступ* до цифрових об'єктів. Тому однією з головних технологічних задач у процесі створення електронної бібліотеки є формування її інформаційних ресурсів.

Розвиток системотехнічних рішень і програмного забезпечення забезпечують сприятливі умови для розбудови електронних бібліотек, надають широкий спектр інструментальних засобів. Але розробка певного технологічного рішення щодо формування інформаційних ресурсів електронної бібліотеки безпосередньо залежить від запланованого функціоналу створюваної системи, моделі керування та тієї інформаційної інфраструктури, на базі якої планується розбудова електронної бібліотеки.

Ресурси електронної бібліотеки (як й традиційної бібліотеки) можуть бути представлені двома рівнями:

рівень даних – контент, – це текстові документи, зображення, звукові, відео- документи та їх поєднання – мультимедійні ресурси;

рівень метаданих – масиви вторинних даних про першоджерела – бібліографічні, реферативні ресурси тощо (для традиційної бібліотеки – це каталоги та картотеки).

Можна виділити основні технологічні напрямки формування

повнотекстових інформаційних ресурсів електронних бібліотек.

Перший технологічним напрямком є *оцифровка* — *переведення в цифрові (електронні) формати* інформаційних джерел, уже існуючих на різних носіях, насамперед, паперових. Оцифровку друкованих документів доцільно здійснювати у ході ретроспективних робіт, таких як переведення в електронну форму стародруків, рукописів, рідкісних видань, архівних документів.

Технологіям оцифровки різних видів документів присвячено багато досліджень, публікацій, міжнародних форумів, європейських, світових та національних проєктів. Прикладом є серія Міжнародних конференцій під загальною назвою “Electronic Imaging & Visual Arts” (www.eva-conferences.com). Заслугує уваги досвід оцифровки у тому числі з посимвольним розпізнаванням текстів, викладений у публікаціях Р. Buonora [1], К.В. Вигурского [2]. Комерційні підприємства пропонують технологічні рішення і послуги аусорсінга для оцифровки різних видів документів.

Слід констатувати, що незважаючи на розвиток програмно-технічних засобів для оцифровки, цей метод створення ресурсів електронних бібліотек є досить дорогим, потребує великої кількості усіх виробничих ресурсів. Це є головною причиною того, що у вітчизняних установах технології оцифровки практично не застосовуються. При цьому, у більшості випадків результатом стають графічні образи текстових документів. Розпізнаванням для цифрового представлення текстів практично не займаються. Технології оптичного розпізнавання текстів (OCR) поки що не задовольняють вимогам масового виробництва, тому переведення текстів у посимвольні цифрові формати супроводжується великою кількістю операцій верифікації та контролю якості. Інколи розпізнавання здійснюється сумісно з клавіатурним введенням даних.

Другим сталим технологічним підходом є *комп'ютерний набір інформації*. Такі технології є характерними для виробників інформаційних продуктів та послуг, що співпрацюють із крупними видавництвами та базами даних. Основними недоліками цього технологічного напрямку є великі витрати виробничих ресурсів і часу, що оправдано лише для прибуткових інформаційних виробництв (комерційних інформаційних продуктів). Для більшості вітчизняних бібліотек, навпаки, клавіатурний набір залишається основним способом створення електронних ресурсів – метаданих.

Третій напрямок формування інформаційно-ресурсних складових електронних бібліотек, що був актуальним на початку розвитку електронних бібліотек — це *використання загальнодоступних Internet-ресурсів*. Актуальність цього напрямку пов'язана з необхідністю інтегрування в електронні бібліотеки велику кількість тих електронних ресурсів, зокрема, електронних публікацій, які розташовані на багатьох серверах та сторінках національної складової глобальної мережі. Є окремі приклади успішної реалізації бібліотечної системи автоматизованого збору та опрацювання періодичних

видань, що розміщуються в глобальній мережі з подальшим формуванням як бібліографічних, так і повнотекстових електронних ресурсів [3].

Але на сьогодні розвиток пошукового інструментарію Інтернет-систем та спеціалізованих систем пошуку у ресурсах глобальній мережі [4] відкидають необхідність "переробляти" ресурси Інтернету, проте можуть надати бібліотекам, у тому числі електронним, інструментарій для додаткові сервісів з бібліотечно-інформаційного обслуговування.

Найбільш привабливим для формування ресурсів електронних бібліотек є четвертий технологічний напрям, пов'язаний з *використанням інформації, яка відразу виробляється в електронній формі*: різноманітні електронні документи, електронні видання, електронні публікації, електронні інформаційні матеріали, що формуються у процесі комп'ютерної верстки друкованих видань, записи бібліографічних і реферативних баз даних, які створюються в інформаційних органах тощо.

Трапилися значні зміни у видавничих технологіях. Як прогнозував В.М. Глушков, автоматизація інформаційних технологій (виробничих процесів, де основною продукцією, що обробляється, є інформація) спричиняє до витискування паперових носіїв, доводячи їх роль до розумних границь [5]. В умовах бурхливого техніко-технологічного розвитку змінюється парадигма видавничої справи, перспектива належить виготовленню цифрових книжок та електронних видань. Інформацію в електронному вигляді, треба максимально використати в усіх технологічних процесах від видавництва друкованої продукції до агрегації електронних "першоджерел" в інформаційні продукти та електронні бібліотеки. Отже пріоритети в розробці інформаційних технологій зміщуються в *напрямку технологій створення, збору, обробки та використання електронної інформації*. Цей напрямок набуває стратегічного значення з погляду на формування інформаційних ресурсів електронних бібліотек.

Вибір того чи іншого із зазначених шляхів залежить від різних факторів, але доцільною є інтеграція до єдиної технології усіх методів і технологій.

У бібліотечній справі технологічні зміни останніх десятиріч уплинули на розвиток програмно-апаратних засобів, які застосовуються у технологічних бібліотечних процесах. Сьогодні у бібліотеках впроваджені типові автоматизовані бібліотечні інформаційні системи (АБІС), які забезпечують автоматизацію формування метаданих та систематизацію документів.

Проте, незважаючи на десятилітню історію активного дослідження електронних бібліотек як інформаційних систем та практичних розробок в цій сфері (включаючи як високовартісні продукти, так і вільне програмне забезпечення), вітчизняна бібліографічно-інформаційна сфера досі не має базових технологічних рішень та типових інструментальних засобів для забезпечення процесу створення повнотекстових ресурсів для електронних бібліотек.

Найбільш розповсюдженими підходами до організації повнотекстових

ресурсів в бібліотеках, що забезпечують у кращому випадку лише пошук та перегляд електронних документів, залишаються створення графічних цифрових копій першоджерел та забезпечення засобами АБІС гіпертекстових зв'язків бібліографічних ресурсів з цими копіями, а також з електронними документами що публікуються в Інтернет.

Проте головним завданням будь-якої бібліотеки, і перш за все електронної, є надання користувачу комплексу інформаційних ресурсів і сервісів. Сучасному ж користувачу потрібен інструментарій, який дозволяє не тільки ефективно шукати першоджерела, але й інформацію в них, а також максимально зручно обробляти матеріал для власних цілей (навчання, дослідження тощо).

Треба враховувати той факт, що стратегічні напрями розвитку систем інформаційного пошуку пов'язані з використанням лінгвістичних технологій, перш за все, морфологічного і смислового аналізу текстів (для флективних мов), з подальшим використанням все більших ефективних методів лінгвістичного аналізу в багатомовних текстових ресурсах, а також с використанням інтелектуальних систем обробки мови.

З урахуванням наведених двох факторів в Національній парламентській бібліотеці України (НПБУ) реалізується комплексний підхід до формування ресурсів електронної бібліотеки. У центрі уваги поставлено завдання формування комплексу ресурсів для забезпечення найповнішого функціоналу електронної бібліотеки.

В якості ресурсно-сервісної системи електронної бібліотеки обрана інформаційна система T-Libra, призначена для створення багатофункціональних електронних бібліотек з можливістю гнучкого повнотекстового пошуку різного виду, що параметризується [6].

Електронні ресурси, що створюються, мають бути орієнтовані на реалізацію основних сервісів електронної бібліотеки – пошук інформації та першоджерел, доступ до документів та отримання необхідної інформації з них.

Розробка технологічного забезпечення ресурсної складової електронної бібліотеки спрямована на формування трикомпонентних масивів електронних ресурсів: цифрові представлення першоджерел, їх метадані та текстові данні для інструментарію повнотекстового пошуку, та ведеться за такими напрямками:

- суміщення ІС T-Libra та автоматизованих бібліотечних систем на рівні обміну метаданими (як початковий етап інтеграції інформаційних систем),
- розробка таких режимів (процесів) створення та обробки повнотекстових ресурсів, котрі дозволяють максимально автоматизувати процеси формування метаданих, текстових ресурсів та фондів електронних документів,
- розробка технології підготовки ресурсів для всіх видів публікацій та документів, що є в наявності в бібліотеках та інших інформаційних організаціях, а також для різних видів електронних документів (електронних

аналогів сучасних публікацій, цифрових копій документів з бібліотечних фондів, електронних видань);

- розробка комплексу цифрових представлень документа в для забезпечення багатоформатного представлення першоджерела кожного виду, а також технологій підготовки кожного представлення;

- розробка типових рішень для організації файлових сховищ – технологічних, страхових, архівних та інших видів електронних фондів.

Формування ресурсів електронної бібліотеки з оцифрованого матеріалу здійснюється з врахуванням необхідності багатоаспектної і багатоцільової візуальної репрезентації першоджерела в інтерфейсі користувача, а також для організації зручної навігації и режимів роботи з текстом. Для цього створюється низка цифрових (електронних) представлень кожної книги у різних файлових форматах:

- **архівне представлення:**
 - посторінкове графічне (TIFF);
- **депозитарні представлення:**
 - посторінкове графічне (JPG);
 - повна графічна копія (PDF графический);
 - повна текстова копія (PDF символный);
- **текстове представлення** (на основі RTF);
- **презентаційне представлення** (HTML).

Повні електронні копії та презентаційні представлення забезпечуються навігацією за структурою видання (за змістом). Текстове представлення використовується для організації повнотекстового пошуку у системі та презентування результатів пошуку користувачеві.

Метаданні (каталогізаційні данні, бібліографічний опис) створюються відповідно до стандартів і форматів інформаційно-бібліотечної галузі.

На Рис. 1. наведена схема інтеграції технології створення ресурсів, яка реалізована для створення цифрових копій, а також для обробки текстів документів з метою використання їх у системі повнотекстового пошуку й для автоматизованого формування метаданих.

Технологічний процес переведення в цифрову форму рідкісних книжок є найповнішим (по відношенню до інших видів документів) і складається з таких процесів:

- відбір книжок для оцифрування та формування тематичних колекцій;
- оцифрування;
- створення цифрових страхових копій та електронних представлень книжок для надання користувачам;
- бібліотечне опрацювання електронних копій (створення каталогізаційного запису);
- формування страхового фонду цифрових копій;
- формування електронного депозитарію та електронного каталогу

оцифрованих видань у спеціалізованій інформаційно-пошуковій системі (IC T-Libra);

- створення супровідних інформаційних матеріалів для користувачів.

Технології електронної бібліотеки

- технології створення електронних ресурсів
- технології обробки електронних ресурсів
- технології інформаційного обслуговування на базі електронних ресурсів

Інтеграція технологій електронної та автоматизованої бібліотек



Рис. 1. Схема інтеграції технологій електронної та автоматизованої бібліотек на рівні метаданих

На основі ресурсів, що формуються, в електронній бібліотеці користувачам надаються різні режими роботи з матеріалами:

- пошук електронних документів у каталозі;
- робота з депозитарієм ресурсів за тематичним рубрикатором;
- повнотекстовий пошук (простий і багаторівневий абзацно-орієнтований; абсолютний та відносний частотні запити).

В усіх режимах користувачу надаються повні електронні копії і біографічні описи знайдених першоджерел. У режимі повнотекстового пошуку результатом є авторські абзаци творів з релевантним текстом. Кожний знайдений текстовий абзац супроводжується такими сервісами:

- перегляд тексту знайденого абзацу, збільшення контексту;
- перегляд графічної копії сторінки оригіналу зі знайденим текстом, а також попередньої та наступної сторінок книги;
- створення оцінок і коментарів до знайдених та обраних текстових фрагментів;
- формування файлу результату із знайдених та обраних текстових фрагментів із бібліографічними посиланнями на першоджерела та коментарями користувача, при цьому результати ранжуються відповідно до оцінки користувача;

- термінологічний аналіз твору або групи творів;
- збереження історії запитів користувача.

Комплекс наведених ресурсів і сервісів, що надається інформаційно-пошуковою системою, дозволяє кожному користувачеві обрати найбільш оптимальний для нього режим пошуку першоджерел та/або інформації, а також обрати зручний саме для нього режим роботи з оцифрованою книгою: перегляд графічних копій або роботу з текстом (як наведеними вище засобами системи, також за допомогою функціонала поширених програм для перегляду PDF-файлів), комбінувати роботу з текстом та перегляд сторінок оригіналу. На допомогу користувачеві формується низка допоміжних інформаційних матеріалів: абетковий та тематичний бібліографічні покажчики доступних до використання оцифрованих видань; бібліографічні відомості про авторів творів; стислі описи електронних колекцій, що формуються; інструктивні матеріали.

В електронній бібліотеці НПБУ вже містяться як цифрові копії рідкісної книги з фонду бібліотеки, а також електронні аналоги видань НПБУ, та інформаційні матеріали, що виготовляються лише в електронному вигляді у процесі інформаційного обслуговування користувачів. Мережевий доступ до електронної бібліотеки надається через офіційний веб-сайт НПБУ (<http://tlibra.nplu.org>).

Висновок

На сьогодні при формуванні ресурсів електронних бібліотек основна увага приділена текстовим документам та двомірним зображенням. Причому тексти – як контент, так і текстова інформація рівня метаданих – становляться не тільки цифровою формою представлення та опису першоджерел, але й найважливішою ресурсною базою для технологій контент-аналізу, комп'ютерної лінгвістики та штучного інтелекту.

Тому завдання створення повнотекстових ресурсів електронної бібліотеки має розглядатися як комплексне завдання ресурсного забезпечення усього функціоналу електронної бібліотеки, спрямованого на багатоаспектне представлення інформації та надання різноманітних сучасних інформаційних сервісів, перш за все – ефективного пошуку у багатомовних ресурсах.

1. *Buonora P.* Digitization projects in Italian State Archives [Electronic Resource] / P. Buonora. – 2004. – Way of Access: http://www.asrm.archivi.beniculturali.it/CFLR/obbiasco/Atti/Testi/Buonora_en.pdf. – Title from the screen.
2. *Вигурский К.В.* Фундаментальная электронная библиотека "Русская литература и фольклор: опыт создания и использования" / *К.В.Вигурский* // Програма ЮНЕСКО "Інформація для всіх": Всеобщий доступ к информации: Междунар. конф.: Материалы конф., 23-25 июня 2004 г., г. Санкт-Петербург [Электронный ресурс]. – Электрон. дан. – 2004. – 1 электрон. опт. диск (CD-ROM).
3. *Четверов В.В.* Технология создания полнотекстового электронного ресурса с элементами библиографической записи [Электронный ресурс] / *В.В. Четверов* //

Матеріали 11-я Междунар. конф. "Крым 2004": Тр. конф. — М., 2004. — 1 электрон. опт. диск (CD-ROM).

4. Ланде Д.В. Поиск знаний в Internet. — М.: Диалектика, 2005. — 272 с.

5. Глушков В.М. Основы безбумажной информатики / В.М. Глушков. — 2-е изд., испр. — М.: Наука, 1987. — 552 с.

6. Ляпин С.Х. Многоязычный поиск в электронной библиотеке и его реализация в ИС T-Libra 6.x [Электронный ресурс] / Ляпин С.Х., Куковякин А.В. // XI межд. конф. EVA-2008, Москва, 1-5 дек. 2008 г. — Режим доступа: http://conf.cpic.ru/upload/eva2008/reports/doklad_1389.doc. — Загол. с экрана.

Поступила 23.08.2010р.

УДК 621.372

Б.В.Дурняк, Ю.-Ю.М.Коростіль, С.О.Нікулін

ОСОБЛИВОСТІ ОРГАНІЗАЦІЇ ПРОЦЕСУ ЗАХИСТУ ДОСТУПУ СПЕЦІАЛІЗАЦІЇ СИСТЕМИ УПРАВЛІННЯ

Алгоритми реалізації процесу захисту спеціалізованої системи управління представляють собою досить складну структуру в силу наступних причин:

- система алгоритмів складається з алгоритмів окремих функціональних блоків,
- процес функціонування системи захисту як єдиний алгоритм повинен розглядатися як реалізація послідовності певних подій в часі,
- перехід від одного фрагменту алгоритму до іншого може обумовлюватися факторами, які не пов'язані з часом, як базовим параметром системи.

Приймаючи до уваги приведені фактори, перш ніж говорити про формування загального алгоритму функціонування, необхідно розглянути принципи функціонування окремих функціональних блоків системи захисту (SZ) та різні варіанти взаємозв'язку між окремими функціональними блоками.

Блок аналізу загроз (AZG), що входить до складу системи SZ, реалізує задачу виявлення невідомих загроз, які характеризують об'єкт захисту (OZ), і тому його робота ініціюється блоком управління системою захисту (BUSZ), при умові, що з OZ прийшла інформація про успішне завершення однієї з атак. При цьому розпізнавання факту успішного завершення атаки здійснюється відповідним Zh_i на основі даних, які сформувала система управління OZ за участю блока аналізу роботи засобу захисту $ARZh_i$ причиною можливого успішного завершення атаки можуть служити наступні