

УДК 004.912

## ТЕХНОЛОГІЯ СТВОРЕННЯ ЕЛЕКТРОННОГО ІЛЮСТРОВАНОВОГО ВИДАННЯ У ФОРМАТІ DJVU

© О. П. Кохановський, к.ф.-м.н., доцент, Ю. В. Сіряк,  
магістрантка, НТУУ «КПІ», Київ, Україна

**Проанализирована проблематика создания книг с иллюстрациями. Разработана технология создания электронных книг с иллюстрированным материалом в формате DJVU.**

**The need to create electronic versions of printed books with illustrations was analyzed. Development of automated technology for processing e-books with illustrated material in DjVu format was proposed.**

### **Постановка проблеми**

В наш час завдяки швидкому удосконаленню і розвитку електроніки та комп'ютерних інформаційних технологій електронні видання набувають все більшого поширення. Процес оцифрування у стислому виді включає в себе процес сканування та наступну обробку отриманих зображень. Постають проблеми вибору скануючого пристрою та режимів сканування, а також збереження складної верстки, вибору кінцевого формату, обробка ілюстративного та текстового матеріалу окремо і т. д.

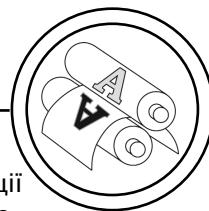
Задача нашого дослідження полягає у спробі розв'язати поставлені проблеми для багатоілюстрованого друкованого видання у вигляді створення доступної технології.

### **Аналіз попередніх досліджень**

Електронні носії інформації та засоби їх відтворення спричинилися до справжньої науко-

во-технічної революції у видавничій справі.

Дані соціологічного дослідження, проведеного компанією GFK Ukraine (за ініціативою Міжнародного фонду «Відродження» та The Fund for Central and European Book Projekt (Amsterdam) і за підтримки програми МАТРА відповідно до проекту «Дослідження книжкового ринку» засвідчують, що українці втрачають інтерес до книжкової продукції, зокрема 58 % із них протягом останніх трьох місяців не придбали жодної книги. Головними аргументами низького попиту на книги є брак грошей (19 %) і висока вартість необхідних видань (11 %). Такі проблеми можуть усунути електронні книги, оскільки вони характеризуються низькою вартістю (а багато навіть містяться в мережі Інтернет у вільному доступі). Крім того електронні книги підвищують інтерес до читання, особливо серед дітей (дослідження видавничої ком-

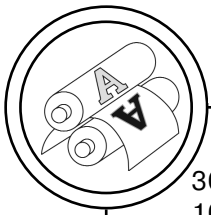


панії Scholastic and Harrison [10]), тому електронні книги починають набувати популярності і серед навчальної літератури. У багатьох випадках частковий або повний перехід на електронне забезпечення навчальними матеріалами цілком виправданий. Ефективність такої заміни тим вище, чим більше розмаїття навчальних посібників і нижче їх тиражі [2, С. 26–27].

Слід також зазначити, що розвиток електронного книговидання — це одна з провідних світових тенденцій. Значна кількість провідних бібліотек світу реалізують проекти зі створення електронних копій своїх фондів. У зарубіжних країнах активно проводиться робота із оцифрування фондів, особливо значні успіхи мають європейські країни. Свідченням є офіційне відкриття Європейської цифрової бібліотеки — Europeana [11], яка надаватиме вільний та швидкий доступ до європейської культурної спадщини та шедеврів в єдиній багатомовній віртуальній бібліотеці. Україна тільки починає робити певні кроки в цьому напрямку. Зокрема держава приєдналася до Хартії ЮНЕСКО про збереження цифрової спадщини, а постановою [1] № 956 від 17 серпня 2011 року Уряд затвердив Державну цільову національно-культурну програму створення єдиної інформаційної бібліотечної системи «Бібліотека — XXI». Все це свідчить про актуальність створення електронних книг та бібліотек на сьогоднішній день. І, звичайно, наявність сучасних потужних видавничих систем і різних за-

собів, обробки і конвертації електронних документів, дозволяє переводити документи в різноманітні формати (такі як PDF, PostScript, Tex, Html). Проте існує одна велика проблема. Існує величезна кількість видань, публікацій, документів, які були випущені в паперовому вигляді ще до засилля комп'ютерних технологій, серед яких багато таких, які будуть безповоротно втрачені для людства, якщо їх не перевести в цифровий формат (документи у численних історичних архівах, рідкісні видання, видання, які були випущені малими тиражами і стали біографічною рідкістю). А сподіватися найближчим часом на їх перевидання не доводиться. Поправити цю ситуацію намагаються деякі інтернет-проекти — Vivos Voco, Наука і Техніка та ін. Але це крапля в морі, і основною причиною цього, в першу чергу, є складність переведення паперових документів в цифровий формат, зручний як для електронних бібліотек, так і читачам. Крім того, створення електронних бібліотек вимагає дорогого апаратного та програмного забезпечення, що стримує їх розвиток.

Дослідники з лабораторій AT&T Labs пропонують своє рішення — нову технологію стиснення зображень DjVu, яка дозволяє легко розміщувати в Інтернеті високоякісні версії відсканованих документів. За оцінками розробників пропонується технологія дозволяє досягти коефіцієнта стиснення 1000:1 (відсканована при 300 dpi кольорова сторінка займає —



30–80 Кб, чорно-біла — 10–30 Кб). Потенційною сферою застосування технології розробники називають обробку відсканованих книг, журналів, каталогів, інструкцій, історичних і рідкісних документів та розміщення їх цифрових копій в Інтернеті.

### **Мета роботи**

Мета роботи полягає в розробленні сучасного технологічного процесу створення електронного ілюстрованого видання в форматі DjVu.

### **Результати проведених досліджень**

Першим етапом в оцифруванні книг є вибір сканера та режимів сканування. Враховуючи формат видання та характер ілюстрацій, які необхідно передати з високою якістю, було обрано сканери, а також за допомогою порівняння шляхом аналізу пелюсткових діаграм було обрано найбільш продуктивний сканер — Canon image Formula DR X10C — для оцифрування друкованих сторінок видання.

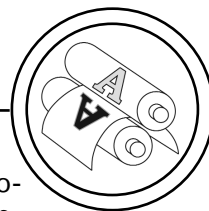
Для оброблення відсканованих ілюстрацій було обрано програмний пакет IrfanView 4.1 [9], оскільки це — вільне програмне забезпечення, крім того дана програма є компактною і в той же час потужною для перегляду і конвертації графічних файлів з багатьма можливостями, а також IrfanView підтримує безліч форматів графіки, містить опції редагування. Враховуючи багатоілюстративність видання та необхідність передати ілюстрації з високою якістю

було встановлено у вказаному вище програмному пакеті оптимальні параметри сканування: роздільна здатність — 600 dpi, серія зображень (пакетний режим). Важливим є вибір формату вихідного файлу. Відскановані сторінки видання було збережено в форматі Uncompressed (Нестиснутий) TIFF, оскільки він має перевагу над іншими форматами, які «розмивають» скан-зображення сторінок книги (наприклад, JPG).

Крім того, відскановані зображення були оброблені за допомогою плагіну Descreen фірми Sattva на предмет видалення поліграфічного растру та муару. Висловлюємо вдячність фірмі Sattva за безкоштовне надання вказаного програмного продукту кафедрі репрографії з навчальною метою.

На сьогоднішній день зусиллями багатьох людей створено достатньо великий набір програм для відновлення паперових книг в електронному виді. В основному функції цих програм полягають у наступному: виправлення орієнтації сторінок, поділ розворотів, компенсація нахилу (поворот зображення кожної сторінки, щоб текст став горизонтальним), відрізання непотрібних темних смуг на краях, створення сторінок однакового розміру, очищення «забруднення» на сторінках (включаючи помарки від руки, штампи та ін.).

Для вибору оптимального програмного пакету для вище описаних цілей, було проведено порівняльну характеристику найбільш популярних та функціональних програм для



обробки відсканованих зображень — ScanTailor [3], ScanKromsator [4], BookRestorer [5].

В результаті аналізу було обрано Scan Tailor [3], оскільки він безкоштовний і повністю дозволяє розв'язати поставлені нами задачі до оцифрування фотоальбому. Зазначимо, що ця програма, зручна для використання, має розвинену інтерактивність та засоби для автоматизації процесу.

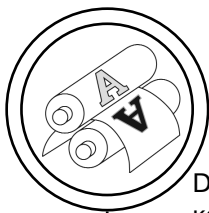
Корисну площу сканованих зображень слід розділити на текстову та ілюстраційну частини, оскільки до них висуваються різні задачі обробки. У текстовій частині слід вирівняти рядки, які спотворились після сканування, згладити контури букв, відрегулювати жирність, ліквідувати розриви тощо. Для ілюстраційної частини слід провести операції видалення растру та муару, масштабування, ретушування, провести кольорокорекцію і т. д. Тому наступним кроком у створенні електронного ілюстрованого видання було розділення текстової та ілюстративної інформації. Для виконання цієї операції було використано програму STSeparator [6].

Обробка ілюстративної інформації фотоальбому «Советское кино» здійснювалась в програмному пакеті Adobe Photoshop CS5, яка в даному випадку дозволяє здійснювати такі основні операції з ілюстраціям: очищення від шуму; виконання вибіркової корекції та ретуші (очищення від плям, затемнених та засвітлених ділянок, які виникли в результаті сканування); затемнення та освітлення деяких ділянок, в тих

місцях, де після сканування кольори втратили свій першочерговий вигляд; підсилення різкості; налаштування адекватної кольоропередачі; здійснення ефективної кольорокорекції тощо.

Одним із найважливіших етапів при створенні електронного видання є вибір кінцевого формату. В наш час існує декілька форматів, в яких зберігаються електронні книги (PDF, DjVu і PostScript). Для вибору найбільш оптимального була проведена порівняльна характеристика найбільш поширених форматів. В результаті порівняння було обрано графічний формат DjVu, оскільки він є оптимальним з точки зору ступеня стиснення і якості картинки, має унікальну особливість — надзвичайна компактність при зберіганні зображень з високою роздільною здатністю. Серед переваг даного формату також є: спеціалізованість для електронізації паперових текстів (DjVu завжди використовує 3-шарову сегментацію контенту, яка забезпечує оптимальне стиснення); малий розмір файлу; збереження точної відповідності верстки електронної копії оригіналу в результаті мінімізації розміру файлу, крім того, в DjVu є орієнтованість на електронний вид документа (можливість перегляду файлу, який ще не завантажився).

Завершальним етапом створення ілюстративного електронного видання є об'єднання вже відредагованих ілюстрацій та тексту в єдину електронну книгу. Для цього було використано програми DjVuSmall [7] та



DjVuImager [8]. DjVu Small — це компактний і зручний кодувальник з/в DjVu. Дана програма кодує як окремі файли, так і групи графічних файлів, декодує одно- і багатосторінкові DjVu-файли (в пакетному режимі) в графічний формат TIFF [7]. DjVuImager — програма, яка призначена для кодування DjVu сканів книг, що містять ілюстрації (напівтонові і/або багатокольорові) [8]. Отже, спочатку за допомогою DjVu Small кодуються файли з розширенням .tif папки \txt\ в DjVu-файл txt.djvu. Далі, за допомогою DjVuImager кодується в профілі photo вміст папки \pic\ і вставляється результат в чорно-білий txt.djvu.

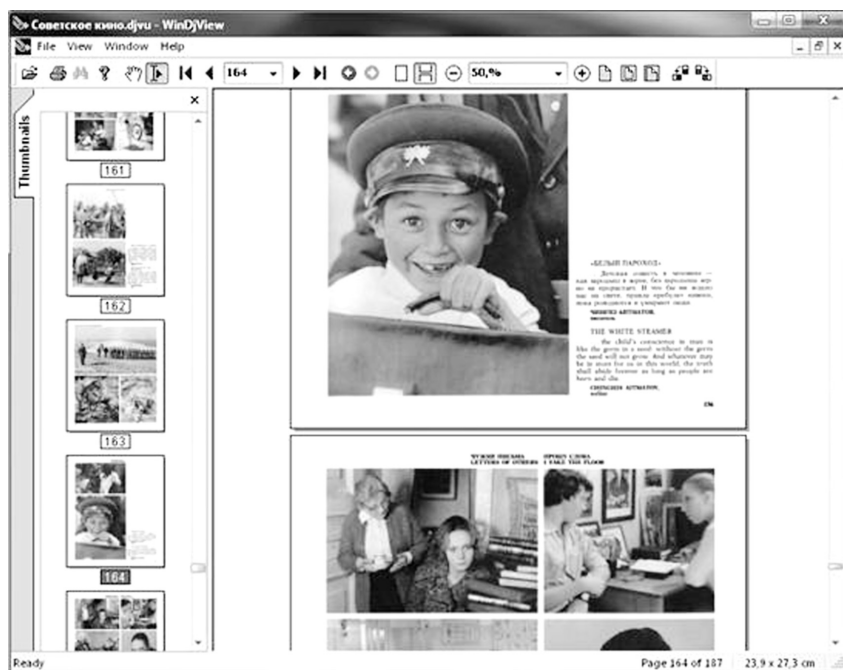
Для тестування запропонованої технології було використане друковане видання «Советское кино». Таким чином, як ре-

зультат, було отримано електронне ілюстроване видання «Советское кино» в форматі DjVu з обробленими ілюстраціями та відредагованим текстом. Реальний вигляд даного електронного видання у форматі DjVu зображено на рис.

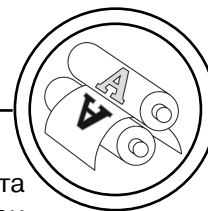
### Висновки

1. Запропонована нами технологія випробовувалася та тестувалася на прикладі багатоілюстрованого видання — фотоальбому «Советское кино» (Москва: «Планета», 1979). Досвід впровадження можна вважати успішним. На основі результатів була захищена бакалаврська робота «Фотоальбом Советское кино» з детальною розробкою електронного видання.

2. Запропонований процес створення електронної версії



Перегляд електронного видання «Советское кино» у форматі DjVu за допомогою WinDjView



дозволяє створювати електронні версії багатоілюстрованих видань. Він забезпечує реальне створення електронних бібліотек місцевого значення та може використовуватися при розв'язанні відповідних задач в поліграфічній галузі.

1. Постанова Кабінету міністрів України від 17 серпня 2011 р. № 956, Київ. «Про затвердження Державної цільової національно-культурної програми створення єдиної інформаційної бібліотечної системи «Бібліотека — XXI»». 2. Вуль В. А. Электронные издания : Учебник / В. А. Вуль. — М. — СПб. : Петербургский институт печати, 2001. — 308 с. 3. Інтерактивний інструмент для обробки відсканованих сторінок Scan Tailor. [Електронний ресурс]. — Режим доступу : URL: <http://scantailor.sourceforge.net/> — Вигляд з екрану. 4. Посobie по программе ScanKromsator [Електронний ресурс]. — Режим доступу : URL: <http://www.djvu-soft.narod.ru/kromsator/> — Вигляд з екрану. 5. Устранение кривизны строк и перекоса текста в Book Restorer 4.1 [Електронний ресурс]. — Режим доступу : URL: [http://www.djvu-soft.narod.ru/scan/curved\\_text.htm/](http://www.djvu-soft.narod.ru/scan/curved_text.htm/) — Вигляд з екрану. 6. STSeparator [Електронний ресурс]. — Режим доступу : URL: [http://sourceforge.net/apps/mediawiki/scantailor/index.php?title=ST\\_Separator](http://sourceforge.net/apps/mediawiki/scantailor/index.php?title=ST_Separator). 7. Использование DjVuSmall [Електронний ресурс]. — Режим доступу : URL: [http://www.djvu-soft.narod.ru/scan/curved\\_text.htm/](http://www.djvu-soft.narod.ru/scan/curved_text.htm/) — Вигляд з екрану. 8. Использование DjVu Imager [Електронний ресурс]. — Режим доступу : URL: [http://www.djvu-soft.narod.ru/scan/djvu\\_imager.htm/](http://www.djvu-soft.narod.ru/scan/djvu_imager.htm/) — Вигляд з екрану. 9. IrfanView [Електронний ресурс]. — Режим доступу : URL: <http://www.irfanview.com/> — Вигляд з екрану. 10. Електронний ресурс. — Режим доступу : <http://mediaroom.scholastic.com/node/378>. 11. Електронний ресурс. — URL : Режим доступу : <http://www.europeana.eu/portal/>.

1. Postanova Kabinetu ministriv Ukrainy vid 17 serpnia 2011 r. № 956, Kyiv. «Pro zatverdzhennia Derzhavnoi tsilovoi natsionalno-kulturnoi prohramy stvorennia yedynoi informatsiinoi bibliotechnoi systemy «Biblioteka — XXI»». 2. Vul' V. A. Jelektronnye izdaniija : Uchebnik / V. A. Vul'. — М. — СПб. : Peterburgskij institut pečatati, 2001. — 308 s. 3. Interaktyvnyi instrument dlia obrobky vidskanovanykh storinok Scan Tailor. [Elektronnyi resurs]. — Rezhym dostupu : URL: <http://scantailor.sourceforge.net/> — Vyhliad z ekranu. 4. Posobie po programme ScanKromsator [Elektronnyi resurs]. — Rezhym dostupu : URL: <http://www.djvu-soft.narod.ru/kromsator/> — Vyhliad z ekranu. 5. Ustranenie krivizny strok i perekosa teksta v Book Restorer 4.1 [Elektronnyi resurs]. — Rezhym dostupu : URL: [http://www.djvu-soft.narod.ru/scan/curved\\_text.htm/](http://www.djvu-soft.narod.ru/scan/curved_text.htm/) — Vyhliad z ekranu. 6. STSeparator [Elektronnyi resurs]. — Rezhym dostupu : URL: [http://sourceforge.net/apps/mediawiki/scantailor/index.php?title=ST\\_Separator](http://sourceforge.net/apps/mediawiki/scantailor/index.php?title=ST_Separator). 7. Ispol'zovanie DjVuSmall [Elektronnyi resurs]. — Rezhym dostupu : URL: [http://www.djvu-soft.narod.ru/scan/curved\\_text.htm/](http://www.djvu-soft.narod.ru/scan/curved_text.htm/) — Vyhliad z ekranu. 8. Ispol'zovanie DjVu Imager [Elektronnyi resurs]. — Rezhym dostupu : URL: [http://www.djvu-soft.narod.ru/scan/djvu\\_imager.htm/](http://www.djvu-soft.narod.ru/scan/djvu_imager.htm/) — Vyhliad z ekranu. 9. IrfanView [Elektronnyi resurs]. — Rezhym dostupu : URL: <http://www.irfanview.com/> — Vyhliad z ekranu. 10. Elektronnyi resurs. — Rezhym dostupu : <http://mediaroom.scholastic.com/node/378>. 11. Elektronnyi resurs. — URL : Rezhym dostupu : <http://www.europeana.eu/portal/>.

Рецензент — О. М. Величко,  
д.т.н., професор, НТУУ «КПІ»