

СУЧАСНА АРХІВАЦІЯ ВИДАНЬ: ПРАКТИЧНИЙ АСПЕКТ

Наша епоха характеризується переходом від суспільства індустріального до суспільства інформаційного. Адже саме *інформаційна революція*¹, загалом четверта², мала своїм наслідком те, що наприкінці ХХ століття людська цивілізація опинилася на шляху від індустріальної фази свого розвитку до інформаційної.

В *інформаційному суспільстві*³ діяльність як окремих людей, так і колективів усе в більшій мірі буде залежати від їхньої поінформованості та спроможності ефективно використовувати наявну інформацію. Матеріально-технічною основою інформаційного суспільства стануть різноманітні системи на базі комп'ютерної техніки та комп'ютерних мереж, інформаційної технології, телекомунікаційного зв'язку. Вже зараз у світі накопичено величезний інформаційний потенціал, яким люди не можуть повноцінно скористуватися в наслідок обмеженості своїх можливостей. Така ситуація, що отримала назву *інформаційної кризи*, поставила суспільство перед необхідністю пошуку виходу з цього стану. Впровадження сучасних засобів переробки та передачі інформації в різноманітні сфери діяльності стало початком переходу від індустріального до інформаційного суспільства. Цей процес називається *інформатизацією*⁴. Роль цього процесу до певної міри аналогічна процесу «індустріалізації», який означав перехід від суспільства аграрного до індустріального.

Інформатизація суспільства є однією із закономірних ознак сучасного соціального прогресу. Сьогодні термін «інформатизація» рішуче витісняє термін «комп'ютеризація», який широко використовувався до недавнього часу. Незважаючи на зовнішню схожість ці терміни мають суттєву відмінність.

При *комп'ютеризації суспільства* основна увага приділяється впровадженню та розвитку технічної бази – комп'ютерів, які забезпечують оперативне отримання результатів переробки інформації та її накопичення.

При *інформатизації суспільства* основна увага приділяється комплексу заходів, спрямованих на забезпечення повного використання достовірного, вичерпного та оперативного знання в усіх видах людської діяльності.

Таким чином, інформатизація суспільства є ширшим поняттям, ніж комп'ютеризація. Акцент у ньому робиться не стільки на технічні засоби, скільки на сутність і цілі соціально-технічного прогресу в цілому. Комп'ютери – це лише базова технічна складова процесу інформатизації суспільства [1, 9-15].

Інформатизація не обходить і широкі кола науковців. Комп'ютер стає робочим інструментом не тільки фахівців з природничих наук, але й займає почесне місце і на столах «гуманітаріїв» [3, 48-49].

Науково-дослідницька діяльність своєю невід'ємною складовою має пошук, зберігання, переробку, виробництво, обмін та продаж інформації. Будь-який дослідник накопичує чималий особистий науковий архів за час своєї діяльності [3, 52-53]. Розглянемо, як комп'ютер може допомогти вченому на етапі зберігання інформації, яким чином можливо здійснювати *ефективну архівацію видань*⁵ на персональному комп'ютері.

У першу чергу розглянемо процес сучасної архівації видань на паперових носіях на персональному комп'ютері. Як правило, за допомогою сканера здійснюється оцифрування інформації на паперових носіях. Результат сканування однієї сторінки – комп'ютерний файл в одному з графічних форматів. Недолік такого підходу – досить великий розмір комп'ютерного файлу. Одна сторінка формату А4 займатиме від одного мегабайту (бітовий формат) до 10 мегабайтів (відтінки сірого) і набагато більше (у випадку кольорового зображення) дискового простору. Використання «стискуючих» графічних форматів зменшує розмір файлів, але призводить до можливої втрати якості комп'ютерної копії. Виходом із такого становища буде використання прогресивного формату «DjVu» [2].

Для практичного використання цього формату на персональний комп'ютер потрібно встановити⁶ одну з програм-редакторів⁷ (для створення архівних файлів у форматі «DjVu») та програму-переглядач⁸ (для комфортного перегляду архіву). Ми практично розглянемо процес

використання програми «DjVu Editor 4.1.0», яку можна безкоштовно «скачати» з Інтернету. Наш вибір пояснюється тим, що доступна російськомовна версія програми.

Почнемо з процесу інсталяції. Після скачування відповідного запакованого файлу з Інтернету слід розпакувати його в папку⁹ на робочому столі. Далі потрібно пошукати в цій папці файл на зразок *read.me*, ознайомитись з його змістом та розпочати процедуру інсталяції відповідно до отриманих з цього файлу вказівок. Загалом ця процедура полягає у запуску програми *setup.exe* або *install.exe*. У випадку, коли такі файли відсутні, слід застосовувати інші процедури¹⁰. Після закінчення стандартної процедури інсталяції на робочому столі з'явиться відповідний ярлик, за допомогою якого можна буде запускати програму. Також можна буде запускати програму і звичайним чином (використовуючи, наприклад, кнопку «Пуск»).

Програма «DjVu Editor 4.1.0» має стандартний «віконний» інтерфейс операційної системи «Windows» (див. рис. 1).

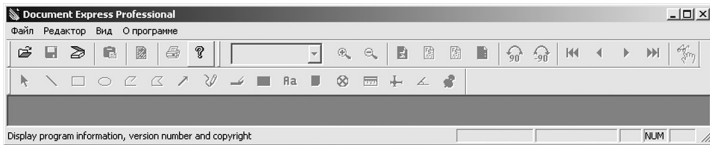


Рис. 1

Під титульним рядком (на ньому напис «Document Express Professional») знаходиться рядок меню («Файл», «Редактор», «Вид» та «О программе»). Наступний рядок займають дві панелі: «Главная панель» та «Панель DjVu». Рядком нижче розташовується «Панель аннотаций». Ще нижче знаходиться місце для документів (на малюнку темно-сірий колір). Завершує «вікно» програми «Панель состояния». Зауважимо, що російськомовна версія програми не містить OCR¹¹, довідки та слайд-шоу. Отже, меню «О программе» та відповідна кнопка («Знак запитання») головної панелі фактично зайві. За допомогою меню «Вид» можливо сховати всі панелі (або деякі з них). При наведенні курсору миші на будь-яку кнопку тієї чи іншої панелі з'являється допоміжна інформація щодо функції кнопки під курсором миші.

Розглянемо основні меню «Файл» та «Редактор». На початку роботи з програмою меню файлів містить такі позиції: «Открыть...»,

«Сканировать...», «Выбрать TWAIN...», «Загрузить отсканированные файлы» (ця позиція не активна) та «Выход». Позиція «Открыть...» має звичайне клавіатурне скорочення «Ctrl+O». На головній панелі їй відповідає окрема кнопка (перша ліворуч) із стандартною позначкою. Ця позиція відкриває стандартне діалогове вікно вибору файлів для завантаження у програму.

Позиція «Сканировать...» відкриває діалогове вікно, яке дозволяє розпочати процес сканування паперових оригіналів для подальшого завантаження у програму з наступним перетворенням у формат DjVu. На головній панелі їй відповідає окрема кнопка (третя ліворуч) із стандартною позначкою (стилізоване зображення сканера).

Позиція «Выбрать TWAIN...» відкриває діалогове вікно, яке дозволяє вибрати «джерело» для процесу сканування паперових оригіналів. Інакше – можна вибрати, яким сканером потрібно користуватися.

Останні дві позиції меню файлів дозволяють завантажити у програму вже відскановані файли (коли процес сканування завершений) та здійснити вихід із програми.

Меню редактора містить позиції «Вставить как новый документ...» та «Настройки». Позиція «Вставить как новый документ...» (на початку роботи з програмою дезактивована) має звичайне клавіатурне скорочення «Ctrl+V». Позиція «Настройки» дозволяє налаштувати дисплей (коригування кольорів) та друк на принтері. За умовчанням рекомендується активувати контроль «Как на мониторе».

Слід зауважити, що як зміст рядка меню, так і самих меню залежить від того, в якому стані знаходиться програма. Рисунок 1 відповідає випадку, коли у програму ще не завантажено жодного файлу та самою програмою ще не користувалися. Якщо у програму завантажено будь-який файл, то рядок меню збільшується (див. рис. 2). При цьому збільшується і кількість позицій у самих меню.

Далі опишемо методику роботи з програмою «DjVu Editor 4.1.0» зі створення електронної версії паперових оригіналів. Перш за все домовимось відносно методики роботи щодо переведу інформації з паперових носіїв у формат DjVu. Зрозуміло, що ця робота складається з двох основних етапів: перший – оцифрування паперового оригіналу; другий – переклад цифрового зображення паперового оригіналу у формат DjVu. Перший етап виконується за допомогою сканера та супроводжуючих його програм¹². На другому етапі

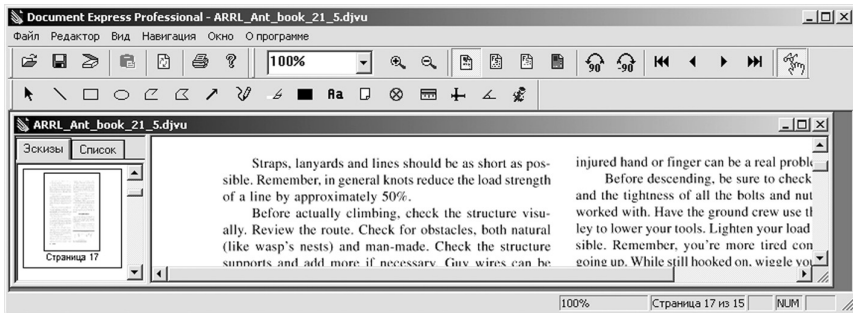


Рис. 2

застосовується «DjVu Editor 4.1.0». Аналіз інтерфейса користувача програми показує, що обидва етапи можливо виконувати безпосередньо в самій програмі, не виходячи з неї. Мається на увазі використання позиції «Сканировать...» меню файлів або відповідної кнопки головної панелі. Але чи буде доцільною така організація роботи з програмою?

На наш погляд, потрібно брати до уваги такий чинник, як використання ресурсів комп'ютера. Адже саме робота з графічними об'єктами вважається найбільш витратною стосовно комп'ютерних ресурсів. Крім того, доволі часто на етапі сканування виникає потреба скоригувати¹³ цифрове зображення паперового оригіналу. Виконувати таку корекцію доводиться за допомогою вельми складних програм – редакторів растрової графіки. Оскільки ресурси комп'ютера обмежені, виникатимуть перевантаження системи, що може мати наслідком «відмови» в його роботі аж до повного «зависання».

Доцільнішим буде розділити перший етап на дві частини – *сканування* та *коригування*. На стадії сканування паперовий оригінал сканується безпосередньо у графічний файл. Якщо дозволяє програмне забезпечення сканера, на цій стадії можна виконувати певне редагування зображення (видалення деяких артефактів, поворот тощо). На стадії коригування здійснюють остаточне редагування цифрових зображень, коли в цьому виникає потреба. На другому етапі працюють із отриманими комп'ютерними файлами-копіями паперових оригіналів.

Такий підхід до роботи з «DjVu Editor 4.1.0» дозволяє значно зменшити навантаження на комп'ютерну систему й уникнути небажаних наслідків

усіляких «аварійних» ситуацій. До речі, подібний підхід цілком виправданий і при використанні різноманітних OCR-програм¹⁴.

Будемо вважати, що перший етап уже виконано: паперові оригінали проскановано, відкориговано і відповідні комп'ютерні файли зображень-копій знаходяться у папці проекту на робочому столі. Розглянемо виключно другий етап. Проілюструємо роботу з програмою «DjVu Editor 4.1.0» на прикладі конкретного проекту з архівації електронного видання класичного типу¹⁵. На робочому столі папка проекту має назву «ARRL_Ant_book_21». Подамо декілька роз'яснень відносно змісту цієї папки. У ній знаходиться 31 файл (00.pdf, 01.pdf, ... 27.pdf, 28.pdf, appendix.pdf та index.pdf), які являють собою електронну версію¹⁶ книги технічної тематики обсягом близько тисячі сторінок формату А4. Загальний обсяг цього електронного видання приблизно двісті мегабайтів. Кожен файл відповідає окремій главі паперового варіанту видання.

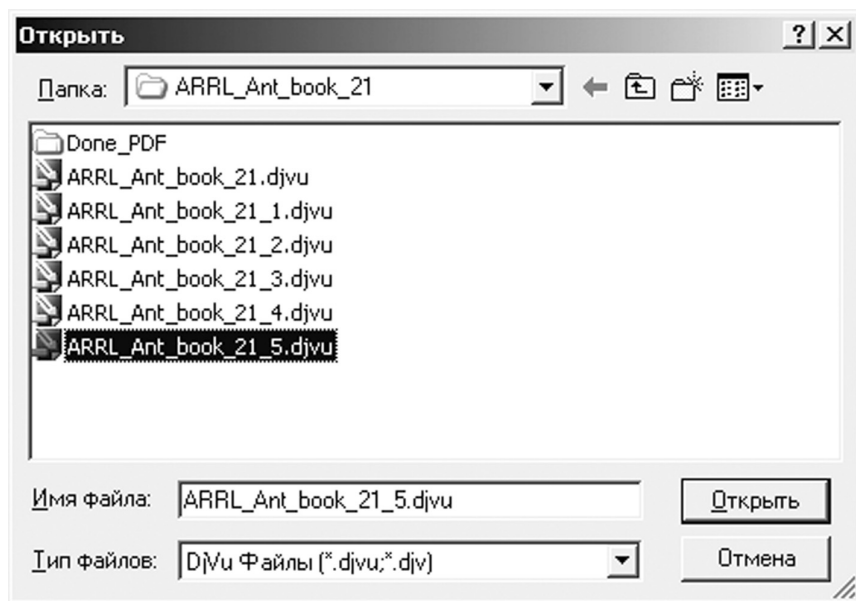


Рис. 3

Кожна глава, починаючи з палітурки та вступу (файл 00.pdf), «сканується»¹⁷ і проскановані сторінки архівуються у файл, спочатку ARRL_Ant_book_21_0.djvu¹⁸, потім ARRL_Ant_book_21_1.djvu і т. д. Тобто, після додавання наступної глави номер файлу збільшується на одиницю і, загалом, відповідає номеру доданої глави. Опрацьовані файли переносяться до папки «Done_PDF» (щоб не заважали).

Після завершення процесу архівації файл ARRL_Ant_book_21_31.djvu буде потрібно перейменувати на ARRL_Ant_book_21.djvu та видалити всі «проміжні» файли у форматі DjVu. Така організація роботи видається дещо складною, проте вона дозволяє надійно «захистити» вже отримані результати у випадку відмови комп'ютера, наприклад, його зависання.

Опишемо процес на прикладі додавання до архівного файлу шостої глави (файл 06.pdf – 52 сторінки). Процес «сканування» цієї глави створить 52 файли (06_Print_v5_01.tif, ... 06_Print_v5_52.tif)¹⁹ у папці проекту.

Запустимо програму «DjVu Editor 4.1.0» і відкриємо документ ARRL_Ant_book_21_5.djvu (до якого вже додано п'яту главу). Для цього клацнемо лівою кнопкою миші на меню «Файл», потім на позицію «Открыть...». З'явиться стандартне вікно «Открыть» операційної системи, в якому вибираємо²⁰ потрібне ім'я файлу (див. рис. 3). Для завершення операції відкриття документу клацнемо на кнопку «Открыть».

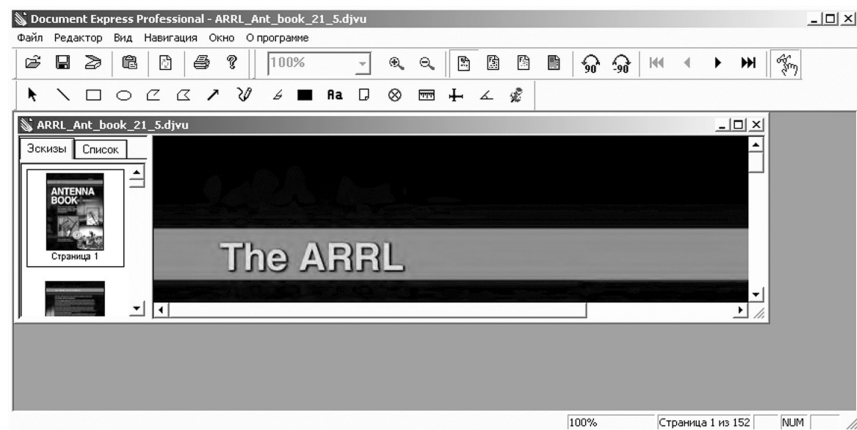


Рис. 4

Вікно програми прийме вигляд, показаний на рис. 4 (зверніть увагу на панель навігації ліворуч у вікні документу).

Далі переходимо до процесу додавання сторінок шостої глави. Для цього скористаємось меню редактора: клацнемо лівою кнопкою миші на меню «Редактор», потім на позицію «Добавить страницы...». Знову з'явиться стандартне вікно «Открыть» операційної системи. На цей раз у полі «Тип файла» потрібно вибрати «tiff Файлы (*.tiff; *.tif)»²¹. З'явиться перелік усіх графічних файлів цього типу, які знаходяться у папці проекту.

Послідовність наступних дій має вирішальне значення для правильного додавання сторінок. *Неухильно дотримуйтесь цих вказівок!*

«Вибираємо» файл останньої сторінки глави (06_Print_v5_52.tif). Його ім'я з'являється у полі «Імя файла».

Потім, утримуючи натиснутою клавішу «Shift», потрібно «вибрати» файл першої сторінки глави (06_Print_v5_01.tif). У полі «Імя файла» з'являється імена файлів-зображень усіх сторінок шостої глави, починаючи з першої (див. рис. 5). У такому порядку і будуть додані сторінки. Завершуємо операцію клацнувши на кнопку «Открыть» діалогового вікна.

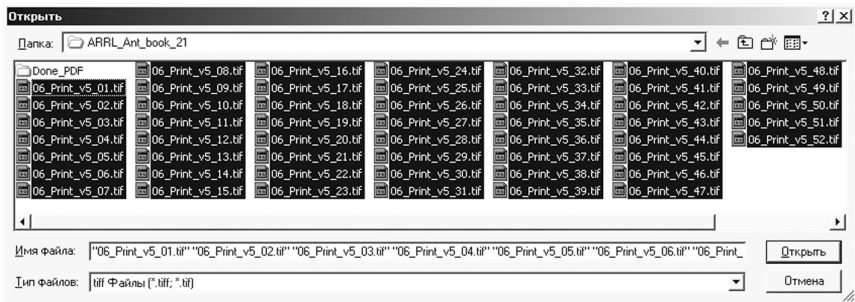


Рис. 5

Комп'ютер розпочне додавати вибрані сторінки згідно порядку файлів, указаних у полі «Імя файла», починаючи з першого. Цей процес триватиме певний час, залежно від потужності комп'ютера та сторінок, що додаються, (кількість, складність тощо). Діалогове вікно «Открыть» може зникнути відразу, а може залишатися на екрані до завершення комп'ютером процесу додавання сторінок. Не треба самому намагатися закрити його – воно у будь-якому разі закриється самостійно при нормальному завершенні процесу.

Пересвідчитись у завершенні додавання сторінок можна глянувши на «Панель состояний» – загальна кількість сторінок збільшиться на число доданих.

Наступний етап – кодування у формат DjVu. Клацнемо лівою кнопкою миші на меню «Файл», потім на позицію «Кодировать в DjVu».

Комп'ютер розпочне тривалий діалог з користувачем. Спочатку з'явиться діалогове вікно «Сохранение...». Погоджуємося з вибором за умовчанням «Один файл», клацнувши лівою кнопкою миші на «ОК».

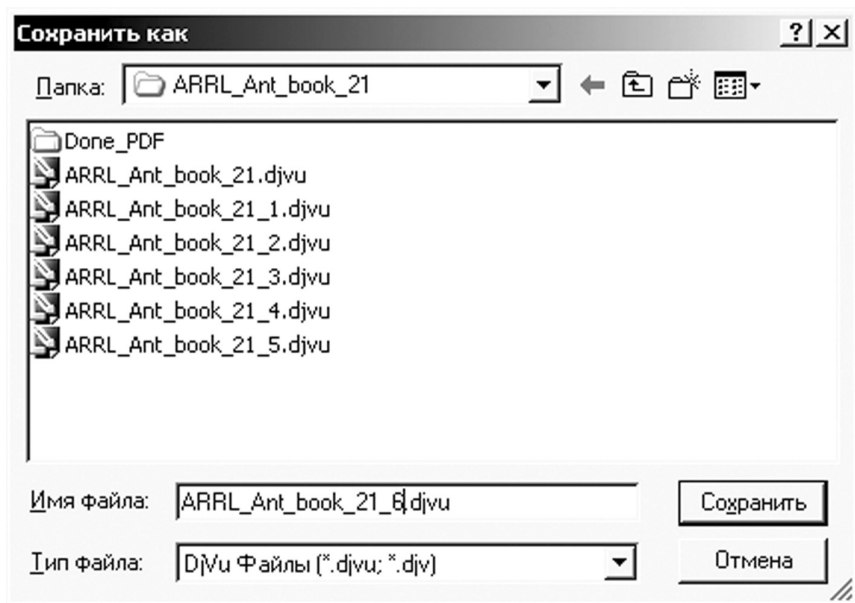


Рис. 6

У наступному діалоговому вікні «Сохранить как» потрібно *виправити* зміст поля «Имя файла» на «ARRL_Ant_book_21_6.djvu» відповідно до номера доданої глави (див. рис. 6). Закриваємо це вікно, клацнувши на кнопку «Сохранить».

Нарешті, на моніторі висвічується останнє діалогове вікно «Установка компресии». Вибираємо радіокнопку «Качественный»²² і закриваємо вікно, клацнувши на кнопку «ОК».

Комп'ютер розпочинає роботу. Вона триватиме певний час. В інформаційному вікні «Encoding in progress...» можна слідкувати²³ за процесом – відображається відсоток виконання. Після завершення кодування просто закриваємо вікно програми «DjVu Editor 4.1.0» – результат вже збережений у файлі ARRL_Ant_book_21_6.djvu разом із щойно доданою шостою главою.

Далі за допомогою програми-переглядача перевіряємо, що все гаразд. Зокрема, що всі сторінки глави були додані у потрібному порядку. Якщо зауважень немає, видаляємо всі допоміжні файли *.tif та *.pdf (при наявності останніх), переносимо файл шостої глави до папки «Done_PDF» і починаємо роботу з сьомою главою...

У випадку порушення порядку доданих сторінок слід видалити файл ARRL_Ant_book_21_6.djvu і повторити процес, використовуючи іншу методику²⁴ вибору сторінок для додавання.

Як бачимо, процес нескладний. Можливо, слід пояснити, як розпочинати роботу з початковою главою. Адже у папці проекту ще немає жодного файлу у форматі DjVu. Просто відкриваємо файл-зображення першої сторінки початкової глави у форматі TIFF. Потім звичайним чином додаємо решту сторінок глави.

Після закінчення робіт обсяг архівного файлу ARRL_Ant_book_21.djvu буде в 11,6 разів менший (16 811 КБ проти 195 169 КБ), ніж обсяг папки «Done_PDF», в якій знаходяться PDF-файли. Якість зображення на моніторі комп'ютера не гірша, а комфортність роботи набагато вища, ніж при використанні відповідного електронного видання класичного типу.

Сподіваємось, що ця стаття «підштовхне» читачів до використання у своїй науковій діяльності формату DjVu.

ПРИМІТКИ

¹ Під поняттям *інформаційна революція* розуміються етапи виникнення засобів та методів обробки інформації, які призвели до кардинальних змін у суспільстві.

² Розрізняють наступні інформаційні революції: *перша* – пов'язана з винаходом писемності; *друга* (середина XVI століття) – пов'язана з винаходом друкарства; *третья* (кінець XIX століття) – пов'язана з винаходом електрики, завдяки якій винайшли телеграф, телефон, радіо (засоби інформаційної комунікації), які дозволяють оперативно передавати і накопичувати інформацію в будь-якому

обсязі; *четверта* (70-ті роки ХХ століття) – пов’язана з винаходом мікропроцесорної технології та появою персонального комп’ютера.

³ *Інформаційне суспільство* – суспільство, в якому більшість працюючих зайняті виробництвом, зберіганням, переробкою, обміном та продажем інформації.

⁴ *Інформатизація* – процес, при якому створюються умови, що задовольняють потреби будь-якої людини в отриманні необхідної інформації.

⁵ Під *ефективною архівацією видань* ми розуміємо створення електронного архіву з мінімальними витратами ресурсів комп’ютера та максимальною комфортністю його використання.

⁶ Нагадаємо, що для встановлення комп’ютерних програм, як правило, потрібно мати права адміністратора.

⁷ Для безкоштовного користування доступні або «DjVu Solo 3.1» або «DjVu Editor 4.1.0». Слід пам’ятати, що створені в другій програмі документи-архіви не відкриваються за допомогою першої. Це не має відношення до програм-переглядачів.

⁸ Можна використовувати або надбудову до одного із стандартних браузерів, або окрему програму [2, 184-185].

⁹ Ця папка створюється програмою-архіватором або автоматично, або її створює власноручно користувач. Якщо виникають труднощі, слід порадитись із кваліфікованим користувачем персональних комп’ютерів, або з фахівцем з комп’ютерних технологій.

¹⁰ Наприклад, для інсталяції програми «DjVu Editor 4.1.0» з використанням скачаного з Інтернету файлу `djvueditor_pro_rus.zip` (розмір 2 564 636 байтів) досить виконати наступне: 1. Розпакувати цей архів на робочий стіл до запропонованої папки (`djvueditor_pro_rus`). 2. Скопіювати цю папку до підгрупи «LizardTech» групи «Мои программы». Коли на комп’ютері вже встановлено програму-переглядач, ця група вже існує. 3. При потребі, замінити ім’я файлу `DjVuEditorRes_0419.dll` (файл шукаємо у вкладеній папці «bin») на `DjVuEditorRes_0409.dll` (можна з’ясувати номер за ресурсом вже встановленої програми-переглядача). 4. Створити ярлик для файлу `DjVuPro.exe` та скопіювати його на робочий стіл. Зауважимо, що протягом такої «інсталяції» до реєстру комп’ютера не вноситься жодного запису. Отже, для дезінсталяції достатньо просто видалити всю скопійовану папку (`djvueditor_pro_rus`).

¹¹ OCR – оптичне розпізнавання символів (абрєвіатура).

¹² Маються на увазі різноманітні сервісні програми, які поставляються разом із сканером.

¹³ Мається на увазі видалення різноманітних «артефактів», які виникають під час сканування, необхідність повернути зображення на невеличкий кут (формат

DjVu досить чутливий до кутів між прямими лініями і це може призвести до помилок: наприклад, буква «н» може перетворитися на «и») тощо [2, 183].

¹⁴ Маються на увазі програми типу FineReader для робіт, пов'язаних із процесами оптичного розпізнавання символів.

¹⁵ Під електронним виданням класичного типу ми розуміємо файл у форматі PDF.

¹⁶ Нагадаємо, що після завершення підготовки оригінал-макету книжки до друку, відповідне електронне видання класичного типу отримують елементарно – наприклад, шляхом друку на віртуальний принтер Adobe Distiller.

¹⁷ «Сканування» у нашому випадку проводиться дуже просто – досить зберегти файл глави у графічному форматі TIFF за допомогою програми Adobe Acrobat. При цьому утвориться стільки графічних файлів, скільки сторінок у главі. Зауважимо, що ці файли не потребують додаткового коригування на відміну від звичайного сканування за допомогою сканера.

¹⁸ На ілюстраціях цей файл помилково виглядає як ARRL_Ant_book_21.djvu – він був переіменований відповідно до тексту після того, як ілюстрації були вже зроблені.

¹⁹ Такі назви файлів пояснюються тим, що використовувалася п'ята версія Adobe Acrobat, а електронне видання було виготовлене за допомогою набагато новіших версій цієї програми. Отже, попередньо довелось «роздрукувати» главу у файл 06_Print_v5.pdf – інакше було неможливо «відсканувати» її.

²⁰ Нагадаємо, що термін «вибрати» означає клацнути на об'єкт лівою кнопкою миші.

²¹ Для цього досить клацнути на кнопку із «стрілкою вниз» в кінці поля, яка розгортає «свиток» (разворачивающийся список). У розгорнутому списку «вибираємо» потрібний рядок.

²² Цю радіокнопку вибираємо тому, що файл-зображення має високу якість, оскільки його отримано шляхом збереження PDF-файлу в графічному форматі TIFF без застосування сканера. Якщо файл-зображення отримано шляхом сканування або цифрового фотографування паперового оригіналу, слід залишити без змін вибор за умовчанням – «Сканированный».

²³ Зауважимо, що цей процес досить громіздкий – ми не радимо одночасно використовувати комп'ютер для інших справ.

²⁴ Замість клавіши «Shift» застосовують клавішу «Ctrl»: утримуючи її натиснутою послідовно клацають лівою кнопкою миші на файл-зображення передостанньої сторінки і так далі, аж поки не доходять до файлу-зображення першої сторінки. Це вельми марудно, проте сторінки додадуться як слід.

БІБЛІОГРАФІЯ

1. Информатика. 10-11 класс / Под. ред. *Н. В. Макаровой.* – СПб.: Питер, 2002. – 304 с.: ил.
2. *Оскрого В.* Сучасна архівація видань // *Діалог. Медіа-студії: Збірник наукових праць / За заг. ред. Александрова О. В.* – Вип. 6. – Одеса, 2007. – С. 181-186.
3. *Шейко В. М., Кушнарєнко Н. М.* Організація та методика науково-дослідницької діяльності: Підручник. – 3-тє вид., стер. – К.: Знання-прес, 2003. – 295 с.