

## **ПРАКТИЧНІ АСПЕКТИ ПРОЕКТІВ З ОЦИФРУВАННЯ ДОКУМЕНТІВ ДЕРЖАВНИХ УСТАНОВ УКРАЇНИ**

**Цирульнєв Ю.Б., генеральний директор**

*ПП "Лабораторія інформаційних технологій", м.Київ*

*Науковий керівник – Бобарчук О.А., к.т.н, доцент кафедри КММТ*

**Вступ.** Україна — найбільша і найстаріша країна Європи з населенням 41 588 354 особи, з більш ніж 1 500 річною історією. Україна має 19 (а раніше 55) Міністерств, 25 державних і національних служб, 18 державних і національних агентств, 6 державних інспекцій, 9 інших центральних органів влади, 6 державних комітетів, 365 політичних партій, 75 діючих банків, 19 800 юридичних осіб, які діють на основі лише державної власності і належать до сфери управління відповідного суб'єкта управління; 431 господарських організацій з корпоративними правами держави; більше 1 млн об'єктів державного майна і 18 морських портів. Архівна галузь України включає 9 Центральних державних архівів, 14 галузевих державних архівів, 27 обласних державних архівів, архівні установи органів самоврядування, архівні підрозділи державних наукових установ, музеїв та бібліотек, архівні підрозділи державних органів, органів місцевого самоврядування, державних і комунальних підприємств, установ та організацій, архівні підрозділи об'єднань громадян, релігійних організацій, а також підприємств, установ та організацій, заснованих на приватній формі власності, архівні установи, засновані фізичними особами та/або юридичними особами приватного права, науково-дослідні установи, а також підприємства та організації у сфері архівної справи і діловодства. В державних архівах, архівних підрозділах державних наукових установ НАН України, архівних відділах райдержадміністрацій і міських рад зберігається 55 370 275 одиниць обліку управлінської документації, 996 345 одиниць обліку

особового походження, 1 164 592 одиниць обліку науково-технічної документації, 75 139 одиниць обліку кінодокументів, 1 362 827 одиниць обліку фотодокументів, 108 119 одиниць обліку фоно-документів, 23 655 одиниць обліку відео-документів, що загалом становить 57 531 212 одиниць обліку документів з паперовим носієм інформації та 1 569 740 одиниць обліку аудіовізуальних документів Національного архівного фонду. [\* Дані наведені станом на 1 січня 2021 року]

**Постановка проблеми і обґрунтування її актуальності.** Зберігання документальної спадщини України в паперовому вигляді та на твердих носіях (магнітних стрічках, кіно-стрічках, фотографіях, мікрофішах тощо), призводить до наступних негативних наслідків:

Оцифрування документальної спадщини України — переведення документів, кіно-фото-відео-аудіо матеріалів в електронні формати, надає наступні переваги громадянам України, державі та світовому суспільству:

Особливими обставинами, які роблять задачу оцифрування документальної спадщини України ще більш актуальнішою, є ризики втрати важливих документів, які пов'язані з поточними військовими конфліктами, терористичними актами та війнами, гострі загрози яких в поточний момент мають місце.

Отже, оцифрування документальної спадщини України є важливим, актуальним і невідкладним практичним завданням сучасності.

**Формулювання новизни і наукові здобутки автора.** Дана доповідь вперше в Україні, узагальнює практичний досвід з оцифрування документів державних установ України, на прикладі проектів з оцифрування більш, ніж 50 000 000 сторінок документів державних установ України, які були успішно реалізовані з 2014 по 2021 роки українськими компаніями ПП «Лабораторія інформаційних технологій» та ТОВ «Цифрові документи», що спільно працюють під торговельною маркою *DIGITAL DOCS*.

**Основне завдання дослідження.** Основним завданням дослідження є визначення раціонального складу проектів з оцифрування документів державних установ України, які зберігаються на паперових носіях.

**Виклад розв'язання поставленого завдання.** Виходячи із реальної історії України та фактичного стану її документальної спадщини, можна затверджувати, що більш 80% документів і кіно-фото-відео-аудіо матеріалів, які являють собою цінність для теперішнього та майбутнього і підлягають оцифруванню, знаходяться і зберігаються у ветохому стані. Щодо паперових документів державних установ України, велика їх частина зберігається у вигляді архівних (судових, інвентаризаційних, тощо) справ, в зшитому вигляді, прошиті ниткою. В більшості випадків, частина інформації, яка міститься на паперових документах, скрита під палітуркою і не може бути якісно сканована без

розширення зшитків. До того ж, виходячи з досвіду реалізованих проєктів, сканування зшитих документів призводить до збільшення фактичних матеріальних витрат на їх оцифрування в 3,7-5,2 рази і до збільшення часу в 5,1-6,2 разів. Отже, у всіх випадках, коли є можливість розширення документів, є сенс розширювати їх для сканування, навіть з урахуванням того, що після сканування виникне потреба зшивання їх в похідний вигляд. Першим процесом в проєктах з оцифрування паперових документів має бути процес підготування документів до сканування, який передбачає: розширення документів, видалення артефактів і часткове реставрування (з метою забезпечення належної якості сканування та не пошкодження оригіналів документів під час сканування).

Процес сканування документів має виконуватися з наступними вимогами:

- застосування устаткування для сканування (сканерів), які обладнані механізмами запобігання пошкодженню оригіналів документів під час їх сканування;
- використання спеціалізованого програмного забезпечення, яке дозволяє забезпечувати відповідну якість та швидкість сканування та не має ліцензійних обмежень, щодо використання в комерційних цілях: ліцензійні угоди, щодо використання стандартного програмного забезпечення, яким комплектуються сканери, що використовується для сканування (*End User Licensed Agreement — EULA*) як правило, забороняють використання його в комерційних проєктах, та/або в інтересах третіх осіб;
- забезпечення сканування з розподільчою якістю не менш 300 *ppi* та динамічним діапазоном, який наближається до  $D=4$  (з метою забезпечення розпізнання тексту засобами *OCR*, та/або якісного друку в масштабі 1:1) з глибиною кольору не менш 8 *bit* у випадку сканування у відтінках сірого та 24 *bit* у випадках сканування документів в повному кольорі.

Сканування паперових документів не забезпечує можливість їх читання, аналізу, редагування електронними засобами у вигляді текстів, а також не надає можливості пошуку фрагментів тексту по тексту документів. Результатом сканування є електронні файли, які містять графічну копію паперового документу у вигляді зображення.

Для забезпечення можливості повнотекстового пошуку по документам, текст документів має бути розпізнаним з растрового графічного зображення, яке є результатом сканування, в *ANSI/ASCII* символи, які можуть оброблятися комп'ютерним засобом, як текст (*Optical Character Recognition — OCR*), а електронні копії мають бути асоційовані з метаданими. Результатом *OCR* є електронні файли, які містять тексти документів в *ANSI/ASCII* символах.

Для забезпечення можливості швидкого пошуку електронних копій документів по ключовим (індексним) полям, електронні копії документів мають бути проіндексовані, тобто кожній електронній копії документу має бути надана відповідність потрібній кількості ключових слів (цифр). Результатом індексування є електронні (індексні) файли, в яких наведена інформація про відповідність електронних копій індексним полям.

Для забезпечення можливості отримання інформації про документи, їх електронні копії мають бути асоційовані з метаданими (даними про документи). Результатом індексування є електронні файли, в яких наведена інформація про відповідність електронних копій метаданим про документи.

Зшивання документів після сканування не є доцільним. Документи слід зберігати в архівних коробах в розшитому вигляді. Зберігання документів в зшитому вигляді підвищує ризики їх втрати. Результатом кінцевого комплектування документів можуть бути документи, які укомплектовані в архівні коробки, або зшитки, якщо зшивання є обов'язковою вимогою.

**Висновки і основні отримані результати.** Проекти з оцифрування документів повинні передбачати (містити) наступні ключові процеси:

- підготування документів до сканування;
- сканування документів;
- обробка електронних графічних копій документів;
- розпізнання текстів в *ANSI/ASCII* символи;
- індексування електронних копій;
- введення метаданих;
- комплектація документів в архівні коробки, або зшивання документів в зшитки (архівні справи).

Оцифрування великих обсягів документів, які утворювалися в попередні періоди (Залпове оцифрування) має сенс здійснювати ресурсами компаній, які мають спеціалізовані призначені для оцифрування великих обсягів документів технології, обладнання, програмне забезпечення та персонал.

Оцифрування документів, які утворюються в повсякденній діяльності (Поточне сканування) має сенс здійснювати власними ресурсами організацій з використанням автоматизованих робочих місць і ділянок оцифрування документів, що створюються компаніями, які мають відповідні компетенції.

Для Залпового оцифрування потрібне спеціалізоване устаткування та програмне забезпечення, яке надає ряд важливих можливостей із забезпечення якості та швидкості сканування. Для Поточного оцифрування можливо використання стандартного устаткування та програмного забезпечення, яким комплектуються сканери.

Зшивання документів після сканування не доцільно. Якщо зшивання є обов'язковою вимогою, його слід доручати компаніям, які мають відповідну компетенцію.