DOI 10.47315/archives2026.348.054
УДК 025.321:005.591.6]:025.177-028.63:[004.438+004.8

Костянтин КАРТУЗОВ
заступник директора — начальник відділу з питань
цифрового розвитку та архівних інформаційних ресурсів,
Державний архів Миколаївської області
(Миколаїв, Україна)
e-mail: kartuzov@ukr.net
ORCID: https://orcid.org/0000-0002-2054-7236

Наталя МУСТЯЦА
заступник начальника відділу з питань
цифрового розвитку та архівних інформаційних ресурсів,
Державний архів Миколаївської області
(Миколаїв, Україна)
e-mail: mustyatsa-@ukr.net
ORCID: https://orcid.org/0009-0006-2003-7136

АВТОМАТИЗАЦІЯ ОПИСУВАННЯ ЦИФРОВИХ АУДІОВІЗУАЛЬНИХ ДОКУМЕНТІВ ЗАСОБАМИ PYTHON І ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. Мета розвідки — розроблення та апробація підходу до автоматизованого формування технічних метаданих цифрових аудіовізуальних документів для їх подальшого використання під час укладання архівних описів, а також визначення можливостей і меж застосування генеративного штучного інтелекту під час створення відповідного програмного інструментарію. **Методологія дослідження** ґрунтується на методах аналізу, систематизації, порівняння та експериментальної апробації. **Наукова новизна** полягає в розробленні та практичній апробації моделі, що передбачає визначення правил алгоритмізованого формування технічних метаданих цифрових об'єктів та інтеграцію відповідних автоматизованих операцій у процес підготовки архівного опису. Визначено склад технічних метаданих цифрових відеодокументів, формування яких може бути автоматизовано, обґрунтовано підходи до уніфікованого представлення розміру файлу, роздільності, тривалості запису та геш-значення, а також апробовано програмні сценарії для пакетного опрацювання цифрових відеодокументів. Розглянуто генеративний штучний інтелект як допоміжний засіб створення й адаптації програмного коду за умови обов'язкової перевірки результатів фахівцем. **Перспективи подальших студіювань** вбачаються в поширенні запропонованого підходу на інші види цифрових аудіовізуальних документів і дослідженні можливостей автоматизації окремих змістових і контекстуальних елементів архівного опису. **Висновки.**

© К. Картузов, Н. Мустяца, 2026

Автоматизація формалізованих операцій із технічними характеристиками цифрових документів дає змогу скоротити обсяг повторюваної ручної роботи, забезпечити уніфікованість отриманих даних і зменшити ризик механічних помилок, водночас залишаючи за укладачем опису функції перевірки та професійної інтерпретації результатів.

Ключові слова: цифрові аудіовізуальні документи; архівні описи; автоматизація архівів; генеративний штучний інтелект; технічні метадані; Python; ChatGPT.

Повномасштабна збройна агресія Російської Федерації проти України стала одним із чинників, що суттєво вплинули на трансформацію діяльності українських архівних установ. Війна, з одного боку, створила безпосередні загрози фізичній збереженості документів Національного архівного фонду (НАФ), архівній інфраструктурі та накопиченим цифровим інформаційним ресурсам, а з іншого — прискорила впровадження цифрових технологій в практику архівів. Необхідність зберегти архівну інформацію в умовах ризику пошкодження або знищення оригіналів надала додаткового імпульсу оцифруванню документів і створенню резервних цифрових масивів. Одночасно потреба оперативно фіксувати події війни сприяла розвитку практик ініціативного документування та активнішому комплектуванню архівів документами, що від початку створюються в цифровій формі (born-digital), передусім фото-, аудіо- та відеоматеріалами. У такий спосіб воєнні обставини не лише загострили проблему збереження документальної спадщини, але й прискорили перехід від традиційної моделі роботи з переважно аналоговими документами до середовища, в якому цифрові об'єкти стають повноцінною і дедалі вагомішою складовою архівних фондів. Досвід Державного архіву Миколаївської області є показовим у цьому аспекті. Організаційні й технологічні зміни в процесах оцифрування архівних документів, спричинені воєнними загрозами, а також значення цифрових копій і резервних масивів для збереження архівної інформації вже були предметом окремого дослідження Н. Колесник і К. Картузова[1]. Водночас архів розгорнув систематичне ініціативне документування подій шляхом виявлення, фіксації та описування born-digital аудіовізуальних матеріалів[2].

Розвиток роботи з born-digital документами в українських архівах значною мірою пов'язаний з поширенням практики ініціативного документування подій російсько-української війни. На відміну від

[1] Колесник Н., Картузов К. Оцифрування архівних документів у воєнний час : досвід Державного архіву Миколаївської області. *Архіви України.* 2024. № 3(340). С. 22–32. https://doi.org/10.47315/archives2024.340.022

[2] Картузов К., Мустяца Н. Ініціативне документування війни : досвід Державного архіву Миколаївської області. *Архіви України.* 2025. № 3(344). С. 65. https://doi.org/10.47315/archives2025.344.056

традиційної моделі комплектування, за якої архів переважно приймає документи, вже створені та сформовані фондоутворювачами, ініціативне документування передбачає активну участь самої архівної установи у виявленні, фіксації та збереженні документальних свідчень суспільно значущих подій. В умовах війни така діяльність набула особливої актуальності, оскільки значна частина інформації про події формується й поширюється безпосередньо в цифровому середовищі, а строк її доступності може бути нетривалим. Відеозаписи, повідомлення та інші матеріали онлайн-платформ можуть редагуватися або видалятися, внаслідок чого архівування вже не завжди може бути відкладене до традиційного етапу комплектування.

У практиці Держархіву Миколаївської області ініціативне документування подій війни було організовано як систематичне виявлення, збереження та описування born-digital матеріалів, що з'являлися у відкритому інформаційному просторі. Одним із напрямів стало документування офіційних звернень, оперативних повідомлень та інформаційних відеоматеріалів, оприлюднених у верифікованому Telegram-каналі начальника Миколаївської обласної військової адміністрації. Такий вибір дав змогу окреслити тематичні та джерельні межі формованого документального комплексу й одночасно забезпечити систематичне накопичення цифрових свідчень подій у режимі, максимально наближеному до часу їх створення.

Принципово важливо, що за такого підходу архів не обмежується лише пошуком уже створених документів та організацією їх подальшого приймання на зберігання. У процесі ініціативного документування він сам здійснює фіксацію цифрового контенту, формує автентичну копію цифрового об'єкта та створює необхідний комплекс супровідних даних. До відповідних операцій належать отримання born-digital контенту з відкритих платформ, створення технічних і описових метаданих, формування контрольних геш-значень та оформлення документів, необхідних для подальшого включення матеріалів до архівного фонду. Саме ця особливість дає підстави розмежовувати ініціативне документування та ініціативне комплектування: в першому випадку цифровий об'єкт як архівний документ виникає внаслідок активних дій архіву, тоді як у другому архів приймає вже створені стороннім продуцентом документи.

Такий перехід від епізодичної роботи з окремими цифровими файлами до систематичного документування подій в реальному часі закономірно призводить до накопичення значних масивів аудіовізуальних документів. Водночас кожен із них потребує не лише змістового архівного описування, а й фіксації властивостей самого цифрового об'єкта. Для відеодокументів — це, зокрема, назва і формат файлу, його розмір, тривалість, технічні параметри відео- та аудіопотоків, а також контрольне

геш-значення. Чинні методичні підходи до роботи з цифровими аудіовізуальними документами передбачають фіксування криптографічного алгоритму гешування та відповідного геш-значення в описах цифрових відеодокументів постійного зберігання.

На нашу думку, ініціативне документування породжує не лише новий спосіб комплектування архівів цифровими документами, а й нове технологічне навантаження на архівіста. Якщо при роботі з поодинокими файлами отримання їхніх технічних характеристик вручну є прийнятним, то систематичне опрацювання сотень і тисяч born-digital аудіовізуальних документів перетворює цю процедуру на значний за обсягом комплекс повторюваних операцій. Виникає потреба відокремити змістову частину архівного описування, яка потребує професійного аналізу й інтерпретації, від формалізованих технічних операцій, результат яких може бути отриманий безпосередньо з цифрового файлу програмними засобами.

У сучасних українських архівознавчих дослідженнях дедалі більше уваги приділяється автоматизації технологічних процесів роботи з документами в електронній формі та управлінню їхнім життєвим циклом. Окремим напрямом цієї проблематики є автоматизоване опрацювання метаданих цифрових об'єктів. Проблему автоматизації роботи з електронними документами безпосередньо порушили Г. Охріменко та О. Федорук[3]. Дослідниці розглянули автоматизоване управління життєвим циклом електронних документів як засіб оптимізації роботи сучасних архівних систем, проаналізували етапи створення, модифікації, упорядкування та зберігання цифрових об'єктів і наголосили на доцільності впровадження автоматизованих технологій в діяльність архівних установ.

Найближче до проблематики застосування сучасних інтелектуальних технологій в архівному описуванні підійшли А. Хромов, Л. Ковальська та Г. Ковальський[4]. Досліджуючи інформаційно-аналітичне забезпечення цифрової конверсії архівних документів, автори розглянули удосконалення описів як комплексний процес, що поєднує методологічну, аналітичну та технологічну складові. Особливо важливо, що серед перспектив подальших студіювань ними безпосередньо визначено використання штучного інтелекту для автоматизованого описування архівних документів відповідно до міжнародних стандартів.

Водночас програмне отримання технічних характеристик документів в електронній формі має в українській архівній практиці триваліший досвід. Т. Кручініна, узагальнюючи практику Центрального державного

---

[3] Охріменко Г., Федорук О. Автоматизація управління життєвим циклом електронних документів у сучасних архівних системах. *Архіви України.* 2023. № 4(337). С. 58–70. https://doi.org/10.47315/archives2023.337.058

[4] Хромов А., Ковальська Л., Ковальський Г. Інформаційно-аналітичне забезпечення цифрової конверсії архівних документів. *Архіви України.* 2026. № 1(346). С. 27–42. https://doi.org/10.47315/archives2026.346.027

електронного архіву України (ЦДЕА України), зазначала, що описова стаття одиниці зберігання містила «ідентифікаційні, контекстні та технічні характеристики». Для визначення технічних характеристик файлів використовувалися як можливості операційної системи, так і спеціалізоване програмне забезпечення MediaInfo, зокрема для отримання відомостей про відео- й аудіодані. Окремою складовою технічної перевірки було обчислення геш-значень файлів для контролю їх цілісності. Таким чином, використання програмних засобів для отримання й перевірки технічних характеристик цифрових документів уже було складовою практики ЦДЕА України[5].

Подібні підходи застосовуються і в зарубіжній практиці роботи з цифровими документами. Зокрема, в системі цифрового збереження Archivematica окремим технологічним процесом є характеристика (characterization) цифрового об'єкта, тобто формування та вилучення його технічних метаданих. Залежно від типу файлу для цього використовуються спеціалізовані програмні засоби, серед яких FFprobe та MediaInfo, а отримані відомості можуть включатися до структури метаданих пакета цифрового збереження[6]. Досвід практичного застосування програмних засобів для опрацювання цифрових архівних документів представлено, зокрема в дослідженні грецького науковця І. Клапсопулоса (Ioannis Clapsopoulos) і його співавторів на матеріалах Історичного архіву Університету Фессалії. Автори апробували інтегроване використання систем Archivematica, AtoM і VuFind для збереження, формування метаданих, архівного описування та забезпечення доступу до цифрових архівних колекцій[7].

Отже, в сучасному архівознавстві вже сформовано значний науковий доробок із проблем цифрової трансформації архівів, електронного архівознавства, комплектування та зберігання електронних і аудіовізуальних документів, ініціативного документування, управління життєвим циклом цифрових об'єктів та автоматизації окремих архівних процесів. Водночас в українській архівознавчій літературі недостатньо представлено дослідження, предметом спеціального розгляду яких є

---

[5] Кручініна Т. Г. Особливості комплектування документами особового походження в електронній формі : досвід Центрального державного електронного архіву України. *Студії з архів. справи та документознавства.* Київ, 2017. Т. 24–25. С. 248–252. URL: https://nbuv.gov.ua/j-pdf/sasd_2017_24-25_28.pdf

[6] Archivematica. Preservation Planning. Characterization. URL: https://www.archivematica.org/en/docs/archivematica-1.18/user-manual/preservation/preservation-planning/

[7] Clapsopoulos I., Laliotou I., Doropoulos S., Fanakis A., Bourdakis V., Stoyannidis Y., Pentazou I., Kalaouzis G., Selitsaniotis T. Managing digital archival collections by integrating Archivematica, AtoM and VuFind software : the University of Thessaly Historical Archive case. *Journal of Integrated Information Management.* 2022. Vol. 7. № 1. P. 7–18. https://doi.org/10.26265/jiim.v7i1.4506

автоматизоване формування технічних метаданих цифрових аудіовізуальних документів безпосередньо для укладання архівних описів із використанням доступного програмного забезпечення з відкритим кодом. Мало дослідженим залишається і практичний потенціал генеративного штучного інтелекту як допоміжного інструмента, за допомогою якого архівіст, не маючи професійної підготовки програміста, може розробляти та адаптувати спеціалізовані сценарії автоматизації архівних технологічних операцій. Саме ця лакуна визначає проблемне поле нашого дослідження.

Попри значний інтерес дослідників до цифрової трансформації архівної справи, комплектування та зберігання електронних і аудіовізуальних документів, питання автоматизації окремих операцій під час їх архівного описування залишається недостатньо розробленим. Особливо це стосується технічних метаданих, значна частина яких не потребує змістової інтерпретації документа і може бути безпосередньо отримана з цифрового файлу програмними засобами. Визначення розміру файлу, тривалості аудіо- чи відеозапису, роздільності, формату та інших технічних характеристик, а також обчислення контрольних геш-значень є формалізованими й повторюваними операціями. За умов опрацювання значних масивів born-digital документів, їх виконання вручну потребує значних витрат робочого часу та створює додатковий ризик механічних помилок.

Проблема полягає не лише в можливості автоматичного отримання відповідних характеристик, а й у включенні такого процесу до усталеної технології архівного описування. Необхідно визначити, які відомості можуть формуватися автоматично без втрати достовірності, якими програмними засобами доцільно їх отримувати, в якому вигляді накопичувати результати та яким чином забезпечити їх подальше використання при укладанні описів цифрових аудіовізуальних документів. Окремим питанням є можливість створення такого інструментарію силами самої архівної установи з використанням програмного забезпечення з відкритим кодом і генеративного штучного інтелекту як допоміжного засобу розроблення програмного коду.

У статті розроблено та апробовано модель автоматизованого формування технічних метаданих цифрових аудіовізуальних документів для їх подальшого використання під час укладання архівних описів із застосуванням програмного забезпечення з відкритим кодом, сценаріїв мовою Python та інструментів генеративного штучного інтелекту. Для цього визначено склад технічних характеристик цифрового аудіовізуального документа, які доцільно отримувати автоматизовано, проаналізовано доступні програмні засоби для їх визначення, розроблено й апробовано алгоритм пакетного опрацювання файлів, перевірено достовірність отриманих результатів і можливості їх використання під час формування

архівного опису. Окрему увагу приділено функціям і межам застосування генеративного штучного інтелекту в процесі створення такого програмного інструментарію.

Для визначення складу технічних метаданих, необхідних під час описування цифрових аудіовізуальних документів, застосовано методи аналізу і систематизації. Розроблені програмні сценарії апробовано на реальному комплексі цифрових відеодокументів Держархіву Миколаївської області, а коректність автоматизованого оброблення перевірено шляхом зіставлення отриманих результатів із характеристиками вихідних файлів. Практичний характер дослідження дав змогу простежити весь технологічний процес: від постановки архівістом завдання до отримання структурованих даних, придатних для включення до архівного опису.

Запропонована й практично апробована модель ґрунтується на розмежуванні відомостей архівного опису за способом їх формування: технічних метаданих цифрового об'єкта, які можуть бути однозначно отримані або обчислені програмними засобами за визначеним алгоритмом, та змістових і контекстуальних відомостей, що потребують професійної інтерпретації архівіста. Новизна запропонованого підходу полягає у визначенні правил алгоритмізованого формування технічних метаданих та інтеграції відповідних автоматизованих операцій у процес підготовки архівного опису. Практичну апробацію моделі в межах цього дослідження здійснено на масиві цифрових відеодокументів.

У межах цієї моделі генеративний штучний інтелект розглядається не як джерело технічних метаданих, а як допоміжний інструмент створення та адаптації програмних рішень архівістом, який не є професійним розробником програмного забезпечення.

Описування цифрових відеодокументів у практиці державних архівів України регламентується комплексом нормативно-правових і методичних документів, які визначають як форму відповідного облікового документа, так і вимоги до відбору, приймання, забезпечення автентичності та фіксації технічних характеристик цифрових об'єктів. Важливо розмежовувати ці функції, оскільки форма опису цифрових відеодокументів визначається нормативно-правовим актом, тоді як окремі технологічні процедури роботи з цифровими файлами деталізовано в галузевих методичних рекомендаціях.

Безпосередньою нормативною основою для укладання опису цифрових відеодокументів є Порядок передавання кіно-, відео-, фото-, фонодокументів на постійне зберігання, затверджений наказом Міністерства юстиції України від 19 липня 2016 р. № 1962/5[8]. Форму «Опис цифрових відеодокументів постійного зберігання» передбачено додатком 11 до

[8] Про затвердження Порядку передавання кіно-, відео-, фото-, фонодокументів на постійне зберігання : наказ Міністерства юстиції України від 19 липня 2016 № 1962/5. URL: https://zakon.rada.gov.ua/laws/show/z1000-16#Text

цього Порядку, яким його було доповнено згідно з наказом Міністерства юстиції України від 23 квітня 2021 р. № 1493/5. Застосування цієї форми також рекомендовано Методичними рекомендаціями «Ініціативне документування подій широкомасштабної збройної агресії Російської Федерації проти України», розробленими Українським науково-дослідним інститутом архівної справи та документознавства і затвердженими наказом Державної архівної служби України від 29 жовтня 2024 р. № 179-НОД. У практиці Держархіву Миколаївської області ця форма застосовується під час укладання описів цифрових відеодокументів постійного зберігання.

Форма опису поєднує відомості різної природи. Частина з них потребує інтелектуальної роботи архівіста та встановлюється на основі аналізу змісту і контексту створення документа. Зокрема, поле «Назва і вид документа» формується відповідно до вимог Правил організації діловодства та архівного зберігання документів у державних органах, органах місцевого самоврядування, на підприємствах, в установах і організаціях[9]. До цієї ж групи належать визначення автора документа та дат, пов'язаних із його створенням і побутуванням. При цьому дату створення документа і дату його оприлюднення слід розмежовувати як різні характеристики. Дата створення може формуватися автоматизовано лише за умови, що відповідне значення однозначно міститься в метаданих цифрового об'єкта або в структурованих вихідних даних і може бути достовірно пов'язане з конкретним документом. Дата оприлюднення характеризує момент введення документа в публічний інформаційний простір і, якщо вона не зафіксована у структурованому джерелі, потребує встановлення на підставі аналізу контексту. У такому випадку її визначення належить до змістових і контекстуальних відомостей, що формуються архівістом.

Іншу групу становлять характеристики, притаманні самому цифровому об'єкту і такі, що можуть бути отримані безпосередньо з файлу програмними засобами. Для цифрового відеодокумента до них належать, зокрема, роздільність зображення, розмір файлу, тривалість запису та криптографічне геш-значення. Саме ця група реквізитів становить особливий інтерес із погляду автоматизації, оскільки їх значення не повинні залежати від суб'єктивної оцінки архівіста і можуть визначатися алгоритмічно.

Показово, що необхідність поєднання змістової та технологічної роботи з аудіовізуальними документами була закладена в українських методичних підходах ще до масового поширення born-digital ресурсів. У Методичних рекомендаціях «Відбір на постійне зберігання аудіовізуальних документів» (2010 р.) до організації ініціативного докумен-

[9] Про затвердження Правил організації діловодства та архівного зберігання документів у державних органах, органах місцевого самоврядування, на підприємствах, в установах і організаціях : наказ Міністерства юстиції України від 18 червня 2015 № 1000/5. URL: https://zakon.rada.gov.ua/laws/show/z0736-15#Text

тування, поряд із вибором об'єкта та проведенням запису чи знімання, віднесено «виробничо-технологічні операції; створення необхідної супровідної текстової документації; підготовку висновку спеціаліста і опису документів»[10]. Отже, технологічне опрацювання аудіовізуального документа є не сторонньою щодо архівного описування технічною процедурою, а складовою цілісного процесу його підготовки до постійного зберігання. У цифровому середовищі частина таких технологічних операцій набуває формалізованого характеру і може виконуватися програмними засобами, що створює передумови для їх автоматизації.

Для документів, створюваних у процесі ініціативного документування, ця нормативно-методична система може мати додаткову специфіку, якщо архів не лише приймає готовий цифровий документ, а сам здійснює захоплення цифрового контенту, формує первинні метадані, фіксує технічні характеристики, встановлює контрольну цілісність за допомогою криптографічного гешування та вносить створений цифровий об'єкт до опису. Відповідно, саме на етапі між отриманням цифрового файлу та внесенням відомостей про нього до опису виникає значна кількість формалізованих технологічних операцій, потенційно придатних для автоматизації.

Практичною базою подальшого дослідження обрано опис № 3 цифрових відеодокументів постійного зберігання за 2022 р., укладений Держархівом Миколаївської області за результатами ініціативного документування матеріалів верифікованого Telegram-каналу начальника Миколаївської обласної військової адміністрації В. Кіма[11]. До опису внесено 341 одиницю обліку цифрових відеодокументів. Він побудований за хронологічним принципом і містить як змістові, так і технічні характеристики кожної одиниці обліку. Зокрема, форма передбачає зазначення назви файлу і номера одиниці обліку, назви і виду документа, виробника[12] та дати, роздільної здатності[13], розміру файлу, тривалості запису, алгоритму

[10] Відбір на постійне зберігання аудіовізуальних документів : методичні рекомендації / Держкомархів України, УНДІАСД, ЦДКФФА України ім. Г. С. Пшеничного; уклад.: К. Т. Селіверстова, А. М. Шелест, О. Г. Саприкіна та ін. Київ, 2010. С. 21.

[11] Опис № 3, цифрових відеодокументів постійного зберігання за 2022 рік. *Державний архів Миколаївської області.* URL: https://archive.mk.ua/nda/avd/d003/ (дата звернення: 12.08.2026).

[12] Використання терміна «виробник» у межах аналізу конкретної форми опису не має на меті його поширення як універсального позначення автора або створювача документа. Для інших видів документів відповідна характеристика може визначатися згідно з термінологією нормативних і методичних документів, що регламентують їх описування.

[13] У затвердженій формі опису цифрових відеодокументів використано назву графи «Роздільна здатність». У цій статті для позначення відповідної характеристики самого цифрового зображення застосовується термін «роздільність»; назву графи нормативно встановленої форми наведено без змін.

криптографічного гешування і геш-значення файлу, його первинної назви з розширенням і приміток. Саме структура цього опису використовується в дослідженні як вихідна модель для визначення тих реквізитів, формування яких може бути автоматизовано програмними засобами.

У межах цього дослідження виходимо із ситуації, коли архівна установа вже має сформований масив цифрових відеодокументів, відібраних для подальшого описування. До безпосереднього укладання опису необхідно здійснити попередню перевірку файлового масиву та, за потреби, вирішити питання уніфікації найменувань файлів.

Методичні рекомендації «Цифровий фонд користування документами Національного архівного фонду: створення, зберігання, облік та доступ до нього» визначають найменування файлу одним із елементів його ідентифікації. Зокрема, після створення цифрових копій рекомендується «присвоювати унікальні найменування кожному файлу для їхньої швидкої ідентифікації». Запропонована схема передбачає включення до найменування ідентифікатора країни походження та архівної установи, типу шифру, типу і характеристики цифрової копії, номера фонду, опису, справи або одиниці зберігання та, за необхідності, інших складових. Така модель забезпечує однозначну ідентифікацію цифрового об'єкта без звернення до додаткових облікових відомостей і цілком може застосовуватися архівними установами під час організації відповідних цифрових масивів[14].

У практиці Держархіву Миколаївської області було враховано інший спосіб організації цифрового сховища. Відеодокументи зберігаються на файловому сервері в структурованому каталозі \Audiovisual\digital-003\, тому повторне включення до найменування кожного файлу відомостей, які вже визначаються його місцем у файловій структурі, було визнано надлишковим. Додатковим аргументом є специфіка організації аудіовізуальних документів. Чинні Правила роботи архівних установ України передбачають можливість нефондової організації аудіовізуальних документів і відповідну систему їх обліку. Зокрема, до основних облікових документів віднесено «аркуш обліку кіно-, фото-, фоно-, відеодокументів», який застосовується за умови їх нефондової організації»[15]. Отже, для аудіовізуальних документів нефондової організації номер архівного фонду не є обов'язковим елементом їх ідентифікації, що також зменшує доцільність використання для них розгорнутого файлового шифру, орієнтованого на структуру «фонд/опис/справа». У будь-якому разі система

[14] Цифровий фонд користування документами Національного архівного фонду : створення, зберігання, облік та доступ до нього : методичні рекомендації / Держ. арх. служба України, Укр. наук.-дослід. ін-т арх. справи та документознавства; уклад.: Л. В. Дідух, Н. В. Залєток, Т. М. Ковтанюк. Київ, 2018. С. 52–55.

[15] Про затвердження Правил роботи архівних установ України : наказ Міністерства юстиції України від 08 квітня 2013 № 656/5. URL: https://zakon.rada.gov.ua/laws/show/z0584-13#Text

найменування відеофайлів має забезпечувати можливість їх автоматичного сортування в тій самій послідовності, в якій відповідні одиниці обліку розміщуватимуться в архівному описі.

Наступним етапом підготовки опису є визначення того, які його поля можуть формуватися програмними засобами, а які потребують безпосередньої участі укладача. Вихідним критерієм такого розмежування, на нашу думку, має бути не складність виконання окремої операції, а можливість однозначно отримати відповідне значення з властивостей цифрового об'єкта або обчислити його за наперед визначеним алгоритмом. Якщо результат не залежить від змістового тлумачення документа та за однакових вихідних даних є відтворюваним, виконання такої операції доцільно делегувати програмному забезпеченню.

У межах цього дослідження під технічними метаданими (*англ.* technical metadata) цифрового документа розуміємо структуровані відомості про технічні властивості цифрового об'єкта, які можуть бути безпосередньо отримані з файлової системи або структури файлу чи обчислені за визначеним алгоритмом без необхідності змістової інтерпретації документа. До таких відомостей можуть належати, зокрема, розмір і формат файлу, характеристики відео- та аудіопотоків, тривалість запису, роздільність відео та геш-значення.

З огляду на це поля опису цифрових відеодокументів можна умовно поділити на три групи. До першої групи належать поля, що можуть формуватися автоматично за наперед визначеним алгоритмом. Для описування цифрових відеодокументів до таких полів можуть належати роздільність відео, розмір файлу, тривалість запису, геш-значення за алгоритмом SHA-256[16] (рекомендовано в Методичних рекомендаціях)[17] і первинна назва файлу з розширенням. Значення перелічених характеристик безпосередньо містяться в структурі цифрового файлу, визначаються файловою системою або можуть бути обчислені програмними засобами без необхідності інтерпретації змісту документа.

Другу групу становлять поля, автоматизація яких можлива лише після встановлення укладачем певних правил або за наявності попередньо структурованих вихідних даних. Зокрема, програмне забезпечення може автоматично присвоювати послідовні номери одиницям

[16] SHA-256 (Secure Hash Algorithm 256) — алгоритм криптографічного гешування, результатом застосування якого є 256-бітне геш-значення. В архівній практиці геш-значення може фіксуватися під час приймання цифрового документа та надалі використовуватися для контролю незмінності його файлового вмісту під час зберігання шляхом повторного обчислення і порівняння з первинно зафіксованим значенням.

[17] Забезпечення збереженості цифрових аудіовізуальних документів Національного архівного фонду : метод. рекомендації / Держ. арх. служба України, Укр. наук.-дослід. ін-т арх. справи та документознавства; уклад.: Л. В. Дідух, В. В. Міхалко, Є. М. Чорноморець. Київ, 2024. 182 с.

обліку, однак порядок їх розташування попередньо має бути визначений укладачем опису. Автоматизованим може бути й заповнення відомостей про автора або дату документа, якщо такі дані вже наявні в структурованому вигляді та можуть бути однозначно пов'язані з відповідним відеофайлом. У цьому випадку програмне забезпечення виконує не самостійне визначення архівної характеристики, а перенесення та нормалізацію вже встановлених відомостей.

До третьої групи належать поля, формування яких передбачає інтелектуальне описування документа. Насамперед це — назва і вид документа, а також змістові примітки. Їх визначення потребує встановлення змісту відеозапису, контексту його створення, належності до певної події та інших характеристик, які не можуть бути однозначно виведені лише з технічних параметрів файлу. Програмні засоби, в т. ч. засоби генеративного штучного інтелекту, потенційно можуть використовуватися для допоміжного опрацювання таких відомостей, однак остаточне формулювання відповідних елементів опису залишається функцією його укладача.

У межах запропонованої моделі поля архівного опису розмежовуються за способом формування відповідних відомостей. До першої групи належать технічні метадані цифрового об'єкта, значення яких можуть бути однозначно отримані з файлової системи або структури файлу чи обчислені за визначеним алгоритмом. Операції з формування таких відомостей є формалізованими, відтворюваними та придатними для автоматизації. Другу групу становлять змістові та контекстуальні відомості, визначення яких потребує аналізу змісту документа, його походження та обставин створення і тому залишається сферою професійної інтерпретації архівіста. Отже, автоматизація не замінює інтелектуальної роботи укладача опису, а охоплює ті його елементи, формування яких може бути алгоритмізоване.

У межах цієї статті предмет дослідження свідомо обмежено тими полями опису цифрових відеодокументів, заповнення яких може бути формалізоване у вигляді програмного алгоритму та виконуватися автоматизовано. Питання формування змістових і контекстуальних елементів опису, що потребують інтелектуального опрацювання укладачем, а також можливості використання для цього засобів генеративного штучного інтелекту виходять за межі цієї роботи й можуть стати предметом окремих наукових розвідок.

У межах визначених вище обмежень до полів опису цифрових відеодокументів, заповнення яких доцільно делегувати програмному забезпеченню, відносимо назву файлу документа та номер одиниці обліку, первинну назву файлу з розширенням, розмір файлу, роздільність відео, тривалість запису та геш-значення файлу за алгоритмом SHA-256. Частина цих відомостей безпосередньо зчитується з файлової системи або структури мультимедійного файлу, інша може бути отримана шляхом однозначних математичних обчислень. Водночас автоматичне

отримання значення ще не означає відсутності методичних проблем: для кожного з перелічених полів необхідно попередньо визначити, яку саме характеристику цифрового об'єкта слід фіксувати та в якому уніфікованому вигляді вона має відображатися в описі.

Для поля «Назва файлу документа/№ од. обліку» принциповою є відповідність між послідовністю файлів у цифровому масиві та порядком одиниць обліку в описі. Якщо прийнята система найменування передбачає використання послідовної нумерації, програмний алгоритм може встановлювати або перевіряти цю відповідність. Водночас первинна назва файлу з розширенням становить окрему характеристику, яку доцільно зберігати незалежно від подальшого перейменування документа. Вона фіксує назву цифрового об'єкта в тому вигляді, в якому він був отриманий архівною установою, і дає змогу зберегти зв'язок між первинним та архівним найменуваннями.

На перший погляд, визначення розміру цифрового файлу є однозначною операцією, однак на практиці вона потребує попереднього встановлення одиниці вимірювання та способу представлення отриманого значення. На цю проблему звернули увагу К. Картузов і Н. Мустяца, аналізуючи досвід укладання описів цифрових відеодокументів у різних операційних середовищах. Автори встановили, що Windows і macOS можуть по-різному відображати розмір одного й того самого файлу в мегабайтах або гігабайтах через застосування різних способів перерахунку, тоді як «розмір файлів у байтах збігається в обох системах»[18]. З огляду на це ними було запропоновано використовувати саме байти як однозначну величину, що не залежить від способу представлення більших одиниць інформації. Такий підхід узгоджується і з Методичними рекомендаціями щодо цифрового фонду користування, де серед відомостей, що підлягають фіксації в облікових документах, передбачено «розмір файлу в байтах». Водночас для зручнішого візуального сприйняття значення в описі подається в мегабайтах відповідно до передбаченої форми опису. У програмному сценарії таке представлення формується з точного значення розміру файлу в байтах шляхом ділення на 1024^2 (1 048 576) та округлення отриманого результату до двох знаків після коми. Отже, під час розроблення програмного алгоритму необхідно чітко розмежовувати фактичний розмір файлу в байтах, його представлення в похідних одиницях та обсяг дискового простору, який він займає на конкретному носії.

З огляду на це доцільно розмежовувати точне значення розміру цифрового об'єкта та спосіб його представлення в облікових документах або інформаційній системі. Первинним значенням, яке має зберігатися без округлення, є розмір файлу в байтах. Саме воно може надалі використовуватися для автоматизованого контролю наявності цифрових документів, планування обсягів систем зберігання та інших

[18] Картузов К., Мустяца Н. Зазнач. твір. С. 65. https://doi.org/10.47315/archives2025.344.056

технологічних операцій. Подання цього значення в мегабайтах, гігабайтах або терабайтах слід розглядати як похідне представлення, призначене для зручності користувача. Якщо відповідна форма облікового документа визначає конкретну одиницю вимірювання, значення подається відповідно до встановленої форми; в інших випадках спосіб його відображення може визначатися прийнятими в архівній установі правилами ведення обліку або налаштуваннями інформаційної системи. Перерахунок у похідні одиниці та округлення не повинні призводити до втрати первинного точного значення в байтах.

У програмній реалізації, апробованій в межах цього дослідження, до CSV-файлу виводиться розмір у мегабайтах, округлений до двох знаків після коми відповідно до форми опису. Водночас під час подальшої інтеграції такого алгоритму в архівну інформаційну систему доцільно поряд із похідним представленням зберігати точне значення розміру цифрового об'єкта в байтах.

Роздільність і тривалість відеозапису також належать до характеристик, які можуть бути визначені автоматично, хоча й потребують уніфікованого способу представлення в описі. У запропонованій програмній реалізації значення роздільності та тривалості визначаються за першим відеопотоком мультимедійного файлу, що дає змогу застосовувати єдине правило до всього опрацьовуваного масиву документів. Для файлів, що містять кілька відеопотоків, вибір потоку, характеристики якого мають бути відображені в описі, потребує окремого методичного визначення. Тривалість відеофайлу може бути визначена з точністю до дробової частини секунди, тоді як в описі її доцільно подавати цілим числом секунд із округленням до найближчої секунди. Просте відкидання дробової частини, яке іноді використовується програмним забезпеченням при визначенні тривалості відео, може дати розбіжність в 1 секунду.

Окремо слід розмежовувати первинну назву цифрового файлу та його найменування після опрацювання архівною установою. У полі «Первинна назва файлу документа з розширенням» доцільно фіксувати назву саме в тому вигляді та регістрі, в якому вона була отримана від програмного забезпечення або інформаційної системи, що слугувала джерелом документа. Зокрема, для відеофайлів, експортованих із Telegram, зберігається оригінальне написання назви й розширення, сформоване відповідним програмним забезпеченням. Якщо ж отримання відеодокументів здійснюється за процедурою, під час якої архівна установа може наперед визначити параметри збереження файлу, наприклад під час завантаження матеріалів із YouTube, доцільно заздалегідь установити єдині правила формування назв, використання регістру та розширень. Водночас файл, що надалі зберігається в архівному файловому масиві, за

потреби може бути перейменований відповідно до прийнятої в установі системи ідентифікації та впорядкування.

Після перейменування файлу його первинна назва зберігається як окремий елемент опису в спеціально передбаченому для цього полі «Первинна назва файлу документа з розширенням». Це забезпечує можливість встановити початкове найменування цифрового об'єкта незалежно від того, яке архівне ім'я було присвоєно йому під час упорядкування. Сформоване архівне найменування, своєю чергою, має забезпечувати можливість автоматичного сортування файлів у тій самій послідовності, в якій відповідні одиниці обліку розміщуються в архівному описі.

Нарешті, геш-значення SHA-256 принципово відрізняється від більшості перелічених характеристик тим, що воно не зчитується з наявних метаданих файлу, а обчислюється на основі послідовності його байтів. Зафіксоване під час описування геш-значення надалі дає можливість перевірити незмінність файлового вмісту шляхом повторного обчислення та порівняння отриманого результату з первинно зафіксованим. У згадуваному вже дослідженні К. Картузова та Н. Мустяци використання SHA-256 пов'язувалося із забезпеченням автентичності та цілісності цифрових документів. Таке трактування загалом відображає роль криптографічного гешування в комплексі процедур забезпечення збереженості цифрових документів, однак потребує певного уточнення. Саме по собі геш-значення не встановлює походження документа й не може розглядатися як самодостатній доказ його архівної автентичності. Його безпосереднє призначення полягає насамперед у контролі незмінності (fixity) цифрового об'єкта після моменту фіксації контрольного значення. Водночас у поєднанні з документованими процедурами отримання, обліку, зберігання та контролю цифрового документа гешування може бути одним із технічних засобів підтримання доказів його цілісності та автентичності.

Важливим є також визначення моменту обчислення геш-значення файлу. Просте перейменування цифрового об'єкта не змінює його байтового вмісту, а отже, не впливає на значення SHA-256. Натомість будь-яке втручання у файловий вміст, зокрема перекодування, зміна мультимедійного контейнера (remux), обрізання або інша модифікація призводить до формування іншої послідовності байтів і, відповідно, нового геш-значення. З огляду на це вважаємо за доцільне утримуватися від перекодування, конвертації або інших операцій, що змінюють файловий вміст, обмежуючись упорядкуванням і перейменуванням файлів та зберігаючи їхню первинну файлову структуру й технічні характеристики.

Отже, перед безпосередньою автоматизацією заповнення технічних полів опису необхідно не лише визначити джерело відповідних даних, а й встановити однозначні правила їх отримання, обчислення,

нормалізації та представлення. Наступним завданням є вибір програмних засобів, здатних реалізувати ці операції для всього масиву цифрових відеодокументів без необхідності ручного опрацювання кожного файлу.

Програмні сценарії, використані в межах цього дослідження, були розроблені із застосуванням генеративного штучного інтелекту. Для генерування та подальшого доопрацювання програмного коду автори використовували платну версію ChatGPT компанії OpenAI. Можна припустити, що аналогічні програмні рішення можуть бути створені й за допомогою інших сучасних великих мовних моделей, однак у межах цього дослідження така можливість експериментально не перевірялася, тому порівняння результатів роботи різних моделей не здійснювалося.

Використання генеративного штучного інтелекту не усуває необхідності фахової участі людини. Постановка завдання, формулювання запитів (промптів), визначення правил опрацювання архівних документів, перевірка згенерованого коду та отриманих результатів, а також упровадження програмного рішення потребують уважності й глибокого розуміння відповідних архівних технологічних процесів. Генеративний штучний інтелект може суттєво знизити технічний бар'єр під час створення спеціалізованих програмних засобів, однак коректність поставленого завдання та відповідність отриманого результату вимогам архівної справи залишаються сферою відповідальності фахівця.

При цьому рівень верифікації має відповідати підготовці користувача. Якщо фахівець не має достатніх навичок для повноцінного аналізу програмного коду, основним способом перевірки має бути верифікація результатів роботи створеного сценарію шляхом їх зіставлення з характеристиками вихідних файлів або результатами, отриманими незалежними програмними засобами. У межах цього дослідження такий підхід було практично апробовано: результати автоматизованого опрацювання порівнювалися з характеристиками вихідних цифрових об'єктів, а окремі алгоритми реалізовувалися різними програмними засобами та зіставлялися між собою. Саме така перевірка дала змогу виявити розбіжності у визначенні тривалості відеозаписів і встановити, що вони були зумовлені різними правилами оброблення дробової частини секунди. Таким чином, використання генеративного штучного інтелекту фахівцем, який не є програмістом, не усуває необхідності контролю, але переносить основний акцент із аналізу програмного коду на перевірку відтворюваності й коректності отриманих результатів.

Використання вже створених і перевірених програмних сценаріїв для автоматизованого опрацювання цифрових відеодокументів не обов'язково потребує професійної підготовки програміста. За наявності належної документації та зрозумілих параметрів запуску такі інструменти можуть застосовуватися досвідченим користувачем, який володіє

базовими навичками роботи з файловою системою та програмним забезпеченням. Водночас розроблення нового сценарію, його суттєва модифікація або адаптація до іншої структури даних потребують ретельної перевірки програмної логіки та результатів його роботи.

Під час вибору програмного забезпечення для автоматизації опрацювання цифрових відеодокументів доцільно, на нашу думку, максимально використовувати програмні засоби з відкритим кодом і стандартні бібліотеки мов програмування. Такий підхід дає змогу мінімізувати залежність архівної установи від пропрієтарного програмного забезпечення, необхідності придбання комерційних ліцензій та конкретного постачальника програмних рішень. Перевагу варто надавати поширеним і документованим інструментам із чітко визначеними умовами ліцензування, які можуть вільно застосовуватися в роботі архівних установ і забезпечують можливість відтворення запропонованого технологічного процесу в інших установах без додаткових фінансових витрат на спеціалізоване програмне забезпечення.

Перед розглядом практичного досвіду застосування генеративного штучного інтелекту необхідно окремо застерегти від використання великих мовних моделей як безпосереднього засобу визначення технічних характеристик цифрових файлів або автоматичного заповнення ними готового архівного опису без подальшої перевірки. Природа генеративних моделей допускає формування правдоподібних, але фактично помилкових відповідей, тому значення розміру файлу, роздільності відео, тривалості запису чи геш-значення не повинні генеруватися засобами штучного інтелекту. У запропонованому підході генеративний штучний інтелект використовується для створення програмного алгоритму, тоді як самі характеристики отримуються детермінованими програмними засобами безпосередньо з цифрових об’єктів. Результатом роботи такого алгоритму є структурований шаблон із технічними даними, який може бути зіставлений з відповідними файлами та дає можливість укладачеві здійснити верифікацію кожної позиції перед перенесенням відомостей до архівного опису.

Перший програмний сценарій для автоматизації цієї роботи було створено в середовищі PowerShell[19]. На початковому етапі максимально використовувалися можливості самої операційної системи: розмір файлу визначався засобами файлової системи, а геш-значення SHA-256 — штатними засобами PowerShell. Однак для отримання специфічних характеристик відеодокументів, насамперед роздільності кадру та тривалості запису, можливостей стандартних засобів виявилося недостатньо. Тому до технологічного ланцюжка було залучено додаткове програмне

[19] PowerShell — кросплатформне середовище автоматизації завдань, що поєднує командну оболонку та мову сценаріїв. Сучасні версії PowerShell є відкритим програмним забезпеченням і працюють у Windows, Linux та macOS.

забезпечення ffprobe, що входить до складу проєкту FFmpeg[20] і дає змогу програмно отримувати технічні параметри мультимедійних файлів.

Розроблення первинного програмного сценарію здійснювалося ітеративно: замість формулювання одного комплексного технічного завдання функціональність програми послідовно розширювалася в процесі взаємодії з генеративним штучним інтелектом. Кожний наступний промпт спирався на вже отриманий програмний код і додавав до нього нову операцію або уточнював спосіб представлення результатів. У скороченому й редакційно адаптованому вигляді послідовність основних запитів була такою:

1. У директорії містяться піддиректорії з файлами. Необхідно визначити розмір кожного файлу в мегабайтах і зберегти результати у файл.

2. Додай до визначення розміру файлів обчислення їх геш-значень за алгоритмом SHA-256.

3. Сформуй результати у форматі CSV[21], придатному для подальшого імпорту до Google Таблиць.

4. Для відеофайлів додай визначення роздільності кадру в пікселях.

5. Чи можна звертатися до MediaInfo через програмний інтерфейс бібліотеки, щоб під час послідовного опрацювання файлів не запускати окремий екземпляр зовнішньої програми для кожного файлу?

6. Які програмні засоби можна використати для отримання технічних характеристик відеофайлів як альтернативу MediaInfo?[22]

7. Перепиши програмний сценарій із використанням засобів FFmpeg для отримання технічних характеристик відеофайлів.

У результаті послідовного уточнення вимог було отримано програмний сценарій PowerShell, призначений для пакетного опрацювання сформованого масиву цифрових відеодокументів. Сценарій запускався безпосередньо на робочій станції під керуванням операційної системи Windows у середовищі PowerShell і не потребував створення окремого спеціалізованого програмного застосунку. Вихідний код цієї реалізації розміщено в репозитарії GitHub[23]:

Особливістю такого технологічного процесу є відсутність потреби в постійному або регулярному виконанні програмного сценарію. Архівний опис цифрових відеодокументів потенційно може охоплювати сотні або

[20] FFmpeg — відкритий набір бібліотек і програмних інструментів для опрацювання мультимедійних даних. До його складу входить утиліта ffprobe, призначена для аналізу мультимедійних файлів і отримання відомостей про контейнер, відео- й аудіопотоки та їх технічні характеристики в придатному для автоматизованого опрацювання вигляді.

[21] CSV (Comma-Separated Values) — текстовий формат подання табличних даних, у якому кожен рядок відповідає окремому запису, а значення полів відокремлюються спеціальним роздільником.

[22] MediaInfo — програмне забезпечення з відкритим вихідним кодом для отримання технічних і описових метаданих аудіо- та відеофайлів; не входить до стандартної комплектації Windows і встановлюється окремо.

[23] URL: https://github.com/Kostiantyn-code/archival-video-metadata

тисячі одиниць обліку, однак підготовка відповідного масиву відбувається послідовно: файли відбираються, упорядковуються, перевіряються та формуються в завершений комплекс, призначений для внесення до конкретного опису.

Практична апробація PowerShell-сценарію підтвердила можливість автоматизованого формування необхідних технічних характеристик цифрових відеодокументів. Водночас для подальшого узагальнення запропонованого технологічного підходу постало завдання створити реалізацію, яка якомога менше залежала б від особливостей конкретного робочого середовища та могла бути відтворена в інших установах. Із цією метою було вирішено реалізувати аналогічний алгоритм мовою програмування Python[24].

Вибір Python зумовлений не перевагою цієї мови над PowerShell у виконанні окремої операції, а сукупністю характеристик, важливих для створення відтворюваного програмного рішення: відкритістю, кросплатформністю, широким набором стандартних бібліотек для роботи з файловою системою, криптографічним гешуванням і структурованими даними, а також можливістю підключення спеціалізованих бібліотек для роботи з мультимедійними файлами. У розробленій реалізації операції з файловою системою, формування CSV та обчислення SHA-256 виконуються засобами стандартної бібліотеки Python, тоді як для визначення роздільної здатності й тривалості відеопотоку використано бібліотеку PyAV[25]. Такий підхід усуває необхідність безпосереднього запуску зовнішньої утиліти ffprobe.exe для кожного файлу та дає змогу виконувати аналіз мультимедійного контейнера безпосередньо з програмного сценарію. Водночас PyAV використовує бібліотеки проєкту FFmpeg, тому йдеться не про повну відмову від технологій FFmpeg, а про інший спосіб програмної взаємодії з ними.

Після створення первинного PowerShell-сценарію аналогічний алгоритм було реалізовано мовою Python. Перший варіант Python-сценарію формував той самий набір технічних характеристик, однак порівняння результатів показало, що вихідні дані не є повністю ідентичними. Аналіз розбіжностей засвідчив, що причина полягала не у визначенні роздільної здатності, розміру файлу чи геш-значення SHA-256, а у способі перетворення тривалості відеозапису з дробового значення секунд у формат «хвилини:секунди». Первинний PowerShell-сценарій фактично

[24] Python — високорівнева мова програмування загального призначення, створена нідерландським програмістом Г. ван Россумом (Guido van Rossum) і вперше оприлюднена на початку 1990-х років. Назва мови походить від британської комедійної групи Monty Python.

[25] PyAV — бібліотека Python, що надає програмний інтерфейс до мультимедійних бібліотек проєкту FFmpeg і дає змогу безпосередньо з Python працювати з мультимедійними контейнерами, аудіо- та відеопотоками й отримувати їхні технічні характеристики.

відкидав дробову частину секунди, тоді як перша Python-реалізація застосовувала округлення, унаслідок чого, наприклад, значення 69,500 с могло бути представлене відповідно як 01:09 та 01:10.

Для контрольної перевірки відтворюваності було створено окрему Python-редакцію, логіка якої максимально наближена до первинного PowerShell-сценарію. У ній відтворено той самий порядок обчислення та спосіб представлення результатів, зокрема правило відкидання дробової частини секунди під час формування значення тривалості. Така реалізація дає можливість безпосередньо зіставляти результати, отримані двома різними мовами програмування, та перевіряти коректність перенесення алгоритму.

Водночас аналіз виявленої розбіжності став підставою для розроблення другої, доопрацьованої Python-реалізації. У ній тривалість відеозапису округлюється до найближчої цілої секунди, що методологічно є точнішим способом представлення вихідного значення, оскільки не створює систематичного заниження тривалості.

Під час порівняння двох алгоритмів визначення тривалості на масиві з 341 цифрового відеодокумента розбіжності у сформованих значеннях було виявлено для 176 файлів, що становить 51,6% від загальної кількості. В усіх таких випадках різниця становила 1 секунду й була зумовлена різним підходом до оброблення дробової частини тривалості: її відкиданням у сумісному з PowerShell алгоритмі та округленням до найближчої цілої секунди у рекомендованому Python-сценарії. Така частота розбіжностей свідчить, що вибір правила округлення не є поодиноким технічним нюансом, а має суттєве методичне значення для забезпечення уніфікованого формування технічних метаданих.

Обидві редакції Python-сценарію оприлюднено в репозитарії GitHub разом із первинним PowerShell-сценарієм.

Виконання розроблених сценаріїв не потребує використання спеціалізованого середовища програмування: після встановлення інтерпретатора Python вони можуть запускатися безпосередньо засобами операційної системи, зокрема з командного рядка. Водночас під час розроблення, тестування та апробації сценаріїв автори використовували середовище PyCharm Community Edition[26], яке забезпечувало зручніше візуальне представлення програмного коду, спостереження за процесом його виконання та виявлення й аналіз повідомлень про помилки. Використання цього середовища мало допоміжний характер і не є обов'язковою умовою для практичного застосування розроблених програмних сценаріїв.

Результатом виконання програмного сценарію є файл у форматі CSV, в якому для кожного опрацьованого відеодокумента окремим рядком

[26] PyCharm Community Edition — безкоштовне інтегроване середовище розробки (IDE) для мови Python, призначене для написання, запуску та налагодження програмного коду.

формуються п’ять значень: назва файлу, роздільність відео, тривалість запису, розмір файлу та геш-значення SHA-256. Як роздільник полів використовується крапка з комою. Такий формат не потребує спеціалізованого програмного забезпечення для подальшого опрацювання: отриманий файл може бути відкритий або імпортований до Google Таблиць, Microsoft Excel, OpenOffice Calc та інших табличних процесорів. Шлях до директорії з вихідними відеофайлами та шлях для збереження сформованого CSV-файлу задаються безпосередньо в програмному коді й за потреби можуть бути змінені користувачем відповідно до організації файлової системи конкретної архівної установи.

Зростання обсягів born-digital аудіовізуальних документів, зокрема створюваних і накопичуваних архівними установами в процесі ініціативного документування подій російсько-української війни, актуалізує потребу в автоматизації повторюваних технологічних операцій, пов’язаних з їх описуванням. Проведене дослідження показало доцільність розмежування змістових і контекстуальних елементів архівного опису, які потребують професійного аналізу укладача, та формалізованих технічних характеристик цифрового об’єкта, значення яких можуть бути однозначно отримані або обчислені програмними засобами.

До полів опису, заповнення яких доцільно автоматизувати, належать назва файлу та номер одиниці обліку, первинна назва файлу з розширенням, його розмір, роздільність відео, тривалість запису та геш-значення SHA-256. Водночас автоматичне отримання цих характеристик саме по собі не усуває потреби в попередньому методичному визначенні правил їх формування та представлення. Зокрема, необхідно встановити одиниці вимірювання розміру файлу, джерело визначення характеристик мультимедійного потоку, порядок округлення тривалості, правила найменування цифрових об’єктів і момент обчислення контрольного геш-значення.

Практична апробація програмних сценаріїв засвідчила можливість пакетного формування такого набору технічних метаданих без ручного опрацювання кожного цифрового файлу. Послідовний перехід від первинного сценарію до альтернативної реалізації іншою мовою програмування дав змогу не лише відтворити алгоритм, а й виявити методичну проблему, пов’язану з різними способами представлення тривалості відеозапису. Це підтверджує, що коректність автоматизації має оцінюватися не лише за технічною працездатністю програмного коду, а й за відповідністю отриманих результатів наперед установленим архівним правилам.

Застосування програмного забезпечення з відкритим кодом дає змогу створювати відтворювані інструменти автоматизації без залежності від спеціалізованих комерційних програмних продуктів. Результатом роботи розробленого сценарію є структурований CSV-файл із технічними характеристиками відеодокументів, придатний для подальшого опрацювання в поширених табличних редакторах і використання

під час укладання архівного опису. Перспективним напрямом подальшої автоматизації є безпосереднє формування на основі отриманих технічних метаданих облікових документів установленої форми або їх автоматизоване передавання до архівної інформаційної системи. Такий підхід не потребує розгортання окремої серверної інфраструктури або постійно діючого програмного сервісу й може бути адаптований до файлової структури конкретної архівної установи.

Генеративний штучний інтелект у запропонованій моделі виконує допоміжну функцію під час створення, пояснення та доопрацювання програмного коду, але не повинен використовуватися як безпосереднє джерело технічних характеристик цифрових документів. Визначення розміру файлу, роздільності, тривалості чи геш-значення має здійснюватися детермінованими програмними засобами безпосередньо на основі цифрового об'єкта. Постановка завдання, визначення методичних правил, перевірка згенерованого коду та верифікація отриманих результатів залишаються сферою відповідальності фахівця. Водночас використання генеративного штучного інтелекту знижує технічний бар'єр для створення спеціалізованих програмних інструментів фахівцями, які не мають професійної підготовки програміста.

Перспективи подальших досліджень пов'язані з адаптацією запропонованого підходу до інших видів цифрових аудіовізуальних документів, розширенням складу технічних метаданих, формування яких може бути автоматизовано, а також окремим вивченням можливостей застосування генеративного штучного інтелекту для допоміжного опрацювання змістових і контекстуальних елементів архівного опису за умови обов'язкового фахового контролю отриманих результатів.

REFERENCES

1. Clapsopoulos, I., Laliotou, I., Doropoulos, S., Fanakis, A., Bourdakis, V., Stoyannidis, Y., … & Selitsaniotis, T. (2022). Managing digital archival collections by integrating Archivematica, AtoM and VuFind software : The University of Thessaly Historical Archive case. *Journal of Integrated Information Management, 7(1),* 7–18. https://doi.org/10.26265/jiim.v7i1.4506 [in English].

2. Didukh, L. V., Zalietok, N. V., & Kovtaniuk, T. M. (Comps.). (2018). *Tsyfrovyi fond korystuvannia dokumentamy Natsionalnoho arkhivnoho fondu : stvorennia, zberihannia, oblik ta dostup do noho* [Digital fund for using documents of the National Archival Fond : creation, storage, accounting and access to it]. Kyiv [in Ukrainian].

3. Didukh, L. V., Mikhalko, V. V., & Chornomorets, Ye. M. (Comps.). (2024). *Zabezpechennia zberezhenosti tsyfrovykh audiovizualnykh dokumentiv Natsionalnoho arkhivnoho fondu* [Ensuring the preservation of digital audiovisual documents of the National Archival Fund]. Kyiv [in Ukrainian].

4. Kartuzov, K., & Mustiatsa, N. (2025). Initsiatyvne dokumentuvannia viiny : dosvid Derzhavnoho arkhivu Mykolaivskoi oblasti [Initiative documentation of the war : experience of the State Archives of the Mykolaiv Region]. *Arkhivy Ukrainy*, *3(344),* 56–72. https://doi.org/10.47315/archives2025.344.056 [in Ukrainian].

5. Khromov, A., Kovalska, L., & Kovalskyi, H. (2026). Informatsiino-analitychne zabezpechennia tsyfrovoi konversii arkhivnykh dokumentiv [Information and analytical support for digital conversion of archive documents]. *Arkhivy Ukrainy*, *1(346),* 27–48. https://doi.org/10.47315/archives2026.346.027 [in Ukrainian].

6. Kolesnyk, N., & Kartuzov, K. (2024). Otsyfruvannia arkhivnykh dokumentiv u voiennyi chas : dosvid Derzhavnoho arkhivu Mykolaivskoi oblasti [Digitization of archival documents in wartime : experience of the State Archives of the Mykolaiv Region]. *Arkhivy Ukrainy*, *3(340),* 22–32. https://doi.org/10.47315/archives2024.340.022 [in Ukrainian].

7. Kruchinina, T. H. (2017). Osoblyvosti komplektuvannia dokumentamy osobovoho pokhodzhennia v elektronnii formi : dosvid Tsentralnoho derzhavnoho elektronnoho arkhivu Ukrainy [Features of acquisition of personal-origin documents in electronic form : experience of the Central State Electronic Archives of Ukraine]. *Studii z arkhivnoi spravy ta dokumentoznavstva*, *24–25,* 248–252. Retrieved from https://nbuv.gov.ua/j-pdf/sasd_2017_24-25_28.pdf [in Ukrainian].

8. Okhrimenko, H., & Fedoruk, O. (2023). Avtomatyzatsiia upravlinnia zhyttievym tsyklom elektronnykh dokumentiv u suchasnykh arkhivnykh systemakh [Automation of electronic document lifecycle management in modern archival systems]. *Arkhivy Ukrainy*, *4(337),* 58–70. https://doi.org/10.47315/archives2023.337.058 [in Ukrainian].

9. Seliverstova, K. T., Shelest, A. M., Saprykina, O. H., et al. (Comps.). (2010). *Vidbir na postiine zberihannia audiovizualnykh dokumentiv* [Selection of audiovisual documents for permanent storage]. Kyiv [in Ukrainian].

Kostiantyn KARTUZOV
Deputy Director — Head of the Department
for Digital Development and Archival Information Resources,
State Archives of Mykolaiv Region
(Mykolaiv, Ukraine)
e-mail: kartuzov@ukr.net
ORCID: https://orcid.org/0000-0002-2054-7236

Natalia MUSTIATSA
Deputy Head of the Department for Digital Development
and Archival Information Resources,
State Archives of Mykolaiv Region
(Mykolaiv, Ukraine)
e-mail: mustyatsa-@ukr.net
ORCID: https://orcid.org/0009-0006-2003-7136

AUTOMATION OF ARCHIVAL DESCRIPTION OF DIGITAL AUDIOVISUAL RECORDS USING PYTHON AND ARTIFICIAL INTELLIGENCE

Abstract. The purpose of the study is to develop and test an approach to the automated generation of technical metadata for digital audiovisual records for use in compiling archival inventories, as well as to determine the possibilities and

limitations of using generative artificial intelligence in the development of relevant software tools. **The research methodology** is based on the methods of analysis, systematization, comparison, and experimental testing. **The scientific novelty** lies in the development and practical testing of a model that establishes rules for the algorithmic generation of technical metadata for digital objects and integrates the corresponding automated operations into the preparation of archival inventories. The study identifies the technical metadata of digital video records whose generation can be automated, substantiates approaches to the standardized representation of file size, video resolution, recording duration, and hash values, and tests software scripts for batch processing of digital video records. Generative artificial intelligence is considered an auxiliary tool for creating and adapting software code, subject to mandatory verification of the results by a specialist. **Prospects for further research** include extending the proposed approach to other types of digital audiovisual records and investigating the possibilities of automating selected content-related and contextual elements of archival description. **Conclusions.** The automation of formalised operations involving the technical characteristics of digital records can reduce repetitive manual work, improve the consistency of the resulting data, and reduce the risk of clerical errors, while verification and professional interpretation of the results remain the responsibility of the archivist preparing the inventory.

Keywords: digital audiovisual records; archival inventories; archival automation; generative artificial intelligence; technical metadata; Python; ChatGPT.

Дата першого надходження статті до редакції: 13.08.2026
Дата прийняття статті до друку після рецензування: 28.09.2026
Дата публікації (оприлюднення) статті: 02.11.2026