

- obiednan ta ukrainiky*. Available at: <https://tsdahou.archives.gov.ua/doradchi-organy/polozhennya-pro-ekspertno-perevirnu-komisiyu-czentralnogo-derzhavnogo-arhivu-gromadskykh-obyednan-ukrainy/>.
6. Pro zatverdzhennia Metodyky vidnesennia dokumentiv Natsionalnoho arkhivnoho fondu do unikalnykh : Nakaz Ministerstva yustytzii Ukrainy vid 16.08.2016 № 2505/5. (2016). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/z1148-16#Text>.
 7. Polozhennia pro Tsentralnyi derzhavnyi arkhiv vyshchykh orhaniv vlady ta ukrainiky : Nakaz Ministerstva yustytzii Ukrainy vid 21.07.2022 № 3061/5. (2022). *Tsentralnyi derzhavnyi arkhiv hromadskykh obiednan ta ukrainiky*. Available at: <https://tsdahou.archives.gov.ua/polozhennya-pro-arhiv/>.
 8. Pro zatverdzhennia Poriadku korystuvannia dokumentamy Natsionalnoho arkhivnoho fondu, shcho належат derzhavi, terytorialnym hromadam : Nakaz Ministerstva yustytzii Ukrainy vid 19.11.2013 № 2438/5. (2013). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/z1983-13#Text>.
 9. Pro zatverdzhennia Pravyl roboty arkhivnykh ustanov : Nakaz Ministerstva yustytzii Ukrainy vid 10.04.2013 № 584/23116. (2013). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/z0584-13#Text>.
 10. Pro zatverdzhennia Poriadku derzhavnogo obliku dokumentiv Natsionalnoho arkhivnoho fondu : Nakaz Ministerstva yustytzii Ukrainy vid 27.09.2013 № 2045/5. (2013). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/z1695-13#n14>.
 11. Rozmir platy za nadannia platnykh posluh tsentralnymy derzhavnymy arkhivamy : Nakaz Ministerstva yustytzii Ukrainy vid 05.12.2024 № 3511/5. (2024). *Tsentralnyi derzhavnyi arkhiv hromadskykh obiednan ta ukrainiky*. Available at: https://tsdahou.archives.gov.ua/wp-content/uploads/2024/12/3.rozmir_platy.pdf.
 12. Papakin H. V. (2009). Suchasnyi stan zakonodavcho-normatyvnoho zabezpechennia arkhivuvannia dokumentiv: pohliad arkhivista. *Bibliotekoznavstvo. Dokumentoznavstvo. Informolohiia*, 4, pp. 7—16.
 13. Pyrih R. Ya. (2013). Tsentralnyi derzhavnyi arkhiv hromadskykh obiednan. *Entsyklopediia istorii Ukrainy*. T. 10. Kyiv. Available at: http://resource.history.org.ua/cgi-bin/eiu/history.exe?&I21DBN=EIU&P21DBN=EIU&S21STN=1&S21REF=10&S21FMT=eiu_all&C21COM=S&S21CNR=20&S21P01=0&S21P02=0&S21P03=TRN=&S21COLORTERMS=0&S21STR=Tsentralnyj_arhiv.
 14. Pro provedennia ekspertizi cinnosti dokumentiv : Postanova Kabinetu Ministriv Ukrainy vid 8.09.2007 № 1004. (2007). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/1004-2007-%D0%BF#Text>.
 15. Pro zatverdzhennia pereliku platnykh posluh, yaki mozhut nadavatysia arkhivnymy ustanovamy, shcho utrymuutsia za rakhunok biudzhetykh koshtiv : Postanova Kabinetu Ministriv Ukrainy vid 7.05.1998 № 639. (1998). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=639-98-%EF#n12>.
 16. Pro optymizatsiiu systemy tsentralnykh derzhavnykh arkhiviv : Postanova Kabinetu Ministriv Ukrainy vid 28.06.2022 № 732. (2022). *Verkhovna Rada Ukrainy*. Available at: <https://zakon.rada.gov.ua/laws/show/732-2022-%D0%BF#Text>.
 17. Strukturni pidrozdily Tsentralnoho derzhavnogo arkhivu hromadskykh obiednan ta ukrainiky. (2025). *Tsentralnyi derzhavnyi arkhiv hromadskykh obiednan ta ukrainiky*. Available at: <https://tsdahou.archives.gov.ua/struktura/>.

Надійшла до редакції 16.05.2025

[https://doi.org/10.36273/2076-9555.2025.10\(351\).38-45](https://doi.org/10.36273/2076-9555.2025.10(351).38-45)

УДК 021.61+930.255]:025.4]:303.4](477)(045)

Василь Курило,

аспірант НУ "Львівська політехніка",

спеціальність "Інформаційна, бібліотечна та архівна справа"

vasyl.h.kurylo@lpnu.ua

<https://orcid.org/0000-0003-1819-2269>

Систематизація підходів до класифікації документів в архівах

У статті проведено систематизацію традиційних і сучасних підходів до класифікації документів в архівних установах. Розглянуто теоретичні основи, ручні й автоматизовані методи класифікації. Констатовано, що до традиційних методів класифікації в архівах належать ієрархічні класифікаційні системи, тематичне та функціональне групування, архівний опис. Автоматизовані методи охоплюють системи на основі правил, статистичні методи класифікації, методи машинного навчання, глибоке навчання, гібридні підходи.

Запропоновано послідовність етапів класифікації в архівах: передопрацювання, виділення ознак, етап побудови або застосування моделі класифікації, оцінювання точності класифікації, інтеграція автоматизованої класифікації в системи управління документами.

Проаналізовано головні проблеми, що постають у процесі впровадження автоматизованих рішень, зокрема багатомовність, неструктурованість даних, низька якість оцифровування, етичні виклики.

Стаття базується на сучасних дослідженнях і практичних прикладах, що підтверджує актуальність теми в умовах цифрової трансформації архівної справи.

Ключові слова: класифікація документів; методи класифікації; автоматизовані системи; цифрове опрацювання документів; структурування архівних фондів; цифрові архіви

Постановка проблеми. В умовах цифрової трансформації архівної справи класифікація документів набуває особливого значення. Зростання обсягів цифрових архівів, різноманітність форматів і джерел інформації ускладнюють процес упорядкування, доступу та довготривалого зберігання документів. Традиційні методи класифікації не завжди відповідають вимогам швидкості, точності й масштабованості. Водночас сучасні автоматизовані підходи, зокрема із застосуванням машинного та глибокого навчання, відкривають нові можливості для вдосконалення пошуку, індексації й управління архівними даними.

Актуальність дослідження. Процес класифікації документів в архівах стикається з низкою викликів, які ускладнюють ефективне впорядкування, пошук і використання інформації. Однією з основних проблем є багатомовність документів, зумовлена історичним і регіональним розмаїттям архівних фондів. Другою перешкодою є неструктурованість даних, коли документам бракує єдиного формату, логіки побудови й схеми опису. Третій важливий виклик — низька якість сканування та розпізнавання тексту (OCR), що призводить до втрати чи спотворення змісту. Не менш суттєвими є етичні та правові аспекти автоматизованої класифікації, пов'язані з конфіденційністю, автентичністю джерел і збереженням історичної достовірності. Крім того, в умовах обмежених людських і технічних ресурсів архіви часто не мають змоги оперативно впроваджувати інноваційні підходи. Ці проблеми є актуальними не лише в теоретичному, а й у практичному сенсі, адже без належної класифікації цифрові архіви втрачають функціональність як наукові й культурні ресурси.

Актуальність теми зумовлена завданням адаптації інноваційних технологій до потреб архівних установ задля ефективного опрацювання та класифікації значних обсягів інформації.

Аналіз досліджень і публікацій. Останніми роками зростає увага фахівців до автоматизованої класифікації документів, зокрема в архівній сфері, що зумовлено потребою ефективного управління великими обсягами неструктурованих даних. У контексті архівної справи автоматизована класифікація набуває особливої актуальності через потребу управління великими обсягами історичних і культурних документів. У дослідженні, проведеному в архіві Національного літературного музею та дослідницького центру африканс (NALN), було застосовано методи машинного навчання для класифікації архівних документів. Автори наго-

лошують на важливості адаптації алгоритмів до специфіки архівних матеріалів, зокрема врахування багатомовності та різноманітності форматів документів. Результати дослідження свідчать, що використання машинного навчання дає змогу значно підвищити ефективність класифікації в архівних установах [1].

У межах вивчення методів машинного навчання, які застосовують для автоматизованого опрацювання історичних документів, розроблено спеціалізовану програму. Вона допомагає конвертувати паперові документи у структурований формат XML, що сприяє підвищенню рівня доступності та збереженню архівних матеріалів [2].

Упровадження методів глибокого навчання передбачає використання моделей U-Net та Siamese Neural Networks, призначених для розпізнавання та класифікації рукописних текстів, що є важливим для цифрових архівів з історичними документами [3].

У дослідженні, присвяченому захисту конфіденційності в архівах, розглянуто аспекти використання контрольованого машинного навчання для класифікації цифрових записів. Наголошено на важливості забезпечення конфіденційності й етичних аспектів під час автоматизованого опрацювання архівних документів [4].

Отже, результати сучасних наукових пошуків демонструють потенціал застосування цифрових технологій для оптимізації процесів класифікації в архівній справі. Водночас залишаються невирішеними питання адаптації моделей автоматизованої класифікації до специфіки архівних даних, дотримання конфіденційності й етичних аспектів. Наукові розвідки свідчать про активний розвиток нормативно-правової бази процесів оцифровування архівних документів в Україні та потребу подальших досліджень для забезпечення ефективного збереження й доступу до цифрової архівної спадщини.

Мета статті — провести системний аналіз методологічних підходів до класифікації архівних документів та обґрунтувати ефективність упровадження автоматизованих методів у контексті сучасної цифрової трансформації архівної справи.

Для досягнення поставленої мети визначено **завдання:**

— розкрити теоретичні засади класифікації документів в архівній, бібліотечній та інформаційній практиці, зокрема з'ясувати відмінності між поняттями класифікації, категоризації й індексації;

— розглянути структуру, переваги та недоліки традиційних ручних методів класифікації документів в архівах;

— проаналізувати сучасні автоматизовані методи класифікації, зокрема статистичні моделі, машинне навчання (контрольоване й неконтрольоване), глибоке навчання та гібридні підходи;

— систематизувати етапи класифікації архівних документів: від попереднього опрацювання до інтеграції в системи електронного документообігу (DMS);

— визначити переваги, ризики та перспективи впровадження автоматизованих систем у практику архівних установ.

Виклад основного матеріалу дослідження.

Класифікація документів — це процес систематичного впорядкування інформації за певними ознаками з метою її зберігання, пошуку та подальшого використання. У контексті архівної, бібліотечної та інформаційної справи це поняття має спільні основи, але різну практичну реалізацію залежно від типу інформаційного ресурсу й функціонального призначення системи. В архівній справі класифікація документів полягає у визначенні їх місця у фонді за тематикою, походженням, структурою та функцією. Основними одиницями класифікації є фонд, опис, справа та документ. Упорядкування здійснюється не лише за змістом, а й за походженням документів — принципом дотримання фондової належності (принцип *provenance*), що є основою архівного опису. У бібліотечній практиці класифікація передбачає розміщення джерел інформації (переважно друкованих видань) за галузями знань, для чого використовують ієрархічні системи, наприклад Універсальну десяткову класифікацію (УДК), класифікацію Бібліотеки Конгресу США. Головною метою бібліотечних класифікацій є забезпечення тематичного доступу користувача до ресурсу. В інформаційній аналітиці класифікація часто виконує функцію категоризації цифрових документів, даних чи об'єктів у системах управління знаннями, пошукових системах чи базах даних. У цій сфері активно застосовують автоматизовані алгоритми класифікації, які базуються на метаданих, контенті чи структурі інформації.

Отже, мета класифікації в архівній, бібліотечній справі, інформаційній аналітиці споріднена — організація доступу та впорядкування інформації, проте методи, критерії й інструменти її реалізації суттєво відрізняються.

В архівній справі терміни "класифікація", "категоризація" та "індексація" часто використовують як синоніми, проте кожен із них має

спеціальне визначення, яке конкретизує його функції та сферу застосування. Розмежування є критично важливим для побудови ефективних систем упорядкування документів. Згідно з термінологією, *класифікація* — це процес систематичного групування документів за певними ознаками (тематика, функція, походження) з метою створення впорядкованої структури архівного фонду. Передбачає побудову ієрархічної системи, де кожен об'єкт належить до певного класу або підкласу, що дає змогу відобразити логічні зв'язки між документами. У класифікації важливою є фіксована структура та чітко визначені правила введення об'єктів. *Категоризація* є менш формалізованим процесом, що полягає у віднесенні документів до однієї чи кількох тематичних груп. Категорії не завжди мають ієрархічну будову й можуть бути створені відповідно до оперативних чи функціональних потреб архівної установи. Категоризація зручна для навігації й фільтрації, особливо в цифрових середовищах. *Індексація* — це присвоєння ключових слів, термінів чи кодів, які спрощують пошук документа в інформаційній системі. Індексація зазвичай не змінює структури зберігання документа, але значно підвищує його доступність через пошукові засоби.

Класифікація документів в архівах не лише виконує функцію впорядкування, а й слугує основою для забезпечення ефективної взаємодії користувача з інформаційним ресурсом. Основною метою класифікації є доступність, керованість і пошук архівних документів. *Доступність* — це можливість швидко та зручно знайти потрібний документ незалежно від його фізичного чи цифрового формату. Якісно спроектована класифікація забезпечує прозору навігацію у фонді й надає можливість дослідникам, архівістам чи автоматизованим системам точно ідентифікувати місце зберігання інформації. Особливо важливою є доступність у цифрових архівах, де брак чіткої структури може зробити навіть оцифрований фонд непридатним до використання. *Керованість* стосується здатності архівної установи ефективно організовувати, контролювати та підтримувати документи протягом усього життєвого циклу — від надходження до довготривалого зберігання або знищення. Класифікація забезпечує логічну й послідовну структуру, що допомагає архівістам координувати описи, оновлювати метадані та відстежувати зміни у фонді. *Пошук* передбачає використання класифікаційних ознак (класів, індексів, категорій) для швидкого виявлення документів за змістом, датою, автором, темою. Ефективний пошук знижує на-

вантаження на персонал, підвищує якість обслуговування користувачів і створює основу для аналітичного опрацювання документних масивів.

Традиційні (ручні) методи класифікації документів були основним інструментом організації архівних фондів упродовж століть. Серед традиційних методів вагоме місце належить ієрархічній класифікації, тематичному та функціональному групуванню, побудові архівного опису (фонд — опис — справа — документ). Вони ґрунтуються на професійному аналізі змісту документа, його функціонального призначення, походження, форми та контексту створення. Таку класифікацію фахівці-архівісти проводять вручну, вона передбачає глибоке знання фонду, історії установи — джерела комплектування, а також нормативних актів і класифікаційних схем [5]. Ручні методи вирізняються високим ступенем точності, контекстуальної чутливості та збереженням логіки походження документа. Вони добре працюють з унікальними, історичними чи змістовно складними документами, які важко автоматизувати. Водночас ці методи є трудомісткими, потребують значних часових і кадрових ресурсів, складно масштабуються на великі цифрові масиви.

Ієрархічні класифікаційні системи є одними з найдавніших і найпоширеніших методів упорядкування документів в архівах. Їх суть полягає у побудові багаторівневої структури, кожен рівень якої деталізує чи уточнює попередній. Такий підхід дає можливість логічно організовувати великі масиви документів за тематикою, функцією або походженням. В архівній практиці одним із найважливіших принципів ієрархічної класифікації є дотримання фондової належності: документи групують відповідно до походження (фонд), далі — за структурою (опис), функцією (справа), відтак — за хронологією чи змістом (аркуш). Ця система забезпечує збереження історичного контексту документів і є основою архівного опису. Крім специфічно архівних систем, в установах також використовують універсальні ієрархічні класифікації, наприклад Універсальну десятикову класифікацію, Державну класифікацію управлінської документації (ДКУ), Класифікацію Конгресу США (Library of Congress Classification). Їх часто застосовують у бібліотечно-архівних установах, де важливо забезпечити тематичний доступ до змісту документів.

Переваги ієрархічних систем полягають у структурованості, прозорості та легкості масштабування. Водночас недоліком є обмежена гнучкість: документи, які не вписуються в жорстку

схему, важко класифікувати без порушення логіки системи. Крім того, ручне підтримання таких структур потребує значних людських ресурсів.

Тематичне та функціональне групування є важливими складниками традиційних методів класифікації в архівах. Вони допомагають упорядковувати документи за змістовими ознаками, адміністративним чи організаційним призначенням, що значно спрощує навігацію, пошук і подальшу роботу з архівними матеріалами. Тематичне групування полягає у формуванні документних масивів відповідно до основних тем, що розкривають зміст документів. Функціональне групування базується на врахуванні управлінських чи адміністративних функцій, які виконувала юридична особа — фондоутворювач. Обидва підходи застосовують окремо та в комбінації, особливо під час створення описів чи структур внутрішнього обліку. Вони дають змогу підтримувати логічну цілісність архівного фонду, не порушуючи принципу походження. У цифрових системах ці схеми часто переносять як категорії чи мітки в метаданих.

Архівний опис є базовою формою традиційної класифікації документів в архівній справі. Він передбачає ієрархічну структуру, що складається з чотирьох основних рівнів: фонд, опис, справа та аркуш. Ця модель класифікації дає змогу зберігати походження, контекст створення документа та забезпечувати контроль над документними масивами. Фонд — це сукупність документів, що виникли в результаті діяльності однієї юридичної чи фізичної особи. Він є головною одиницею обліку в архіві й відображає історичне походження документів. Опис — логічно структурований перелік справ у межах одного фонду. Його створюють для зручності навігації й доступу до документів, об'єднаних за певною ознакою (тематика, функція, хронологія). Справа — базова одиниця зберігання, що об'єднує документи, пов'язані між собою за змістом чи походженням. Кожна справа має заголовок, дати та обсяг. Аркуш — найнижчий рівень опису, який застосовують у виняткових випадках, наприклад, під час роботи з особливо цінними документами чи рукописами. Ця багаторівнева система класифікації допомагає архівам зберігати не лише зміст, а й історичний, функціональний і контекстуальний зв'язок між документами. Вона є основою для створення описових довідників та електронних каталогів, а також для ефективного управління архівним фондом. У цифрову епоху ці рівні переносять у метадані та структури електронних архівів, зберігаючи логіку традиційної класифікації.

Автоматизовані методи класифікації документів відкривають нові можливості для архівної справи, особливо в умовах зростання обсягів цифрових документів та оцифрованих фондів. На відміну від традиційних ручних методів, автоматизовані забезпечують високу швидкість опрацювання, масштабованість і здатність працювати з неструктурованими даними.

Ці методи поділяються на кілька основних груп: системи на основі правил (Rule-Based), статистичні алгоритми, машинне навчання (контрольоване та неконтрольоване), глибоке навчання (нейронні мережі) та гібридні моделі, що комбінують різні підходи [6]. Останні спрямовано на автоматичне віднесення документів до відповідних класів чи тем із використанням текстового наповнення, метаданих чи контексту. В архівах їх застосовують для попереднього опрацювання надходжень, оптимізації пошуку чи формування електронних описів. Автоматизація не лише підвищує ефективність роботи, а й надає можливості доступу до документів для широкого кола користувачів.

Системи на основі правил (rule-based системи) — це один із найпростіших типів автоматизованої класифікації, який працює за принципом "якщо... то...", використовуючи фіксовані набори ключових слів, термінів чи фраз для віднесення документа до певної категорії. В архівній справі такі системи застосовують для автоматичного попереднього сортування документів за темами, типами чи функціональними ознаками. Переваги: прозорість логіки, контрольованість і простота реалізації; недоліки: потребують ручного налаштування правил і слабо адаптуються до нових типів даних чи контекстів. Rule-based системи оптимально придатні для стабільних, формалізованих документних масивів.

Статистичні методи класифікації базуються на аналізі частотності слів та ймовірнісному підході до визначення тематики документа. Одним із найпоширеніших алгоритмів є наївна байєсівська класифікація, що обчислює ймовірність належності документа до певного класу, виходячи з частоти появи ключових слів. Цей метод оптимально працює в умовах великого масиву даних і демонструє високу ефективність під час класифікації текстів із типовою структурою. В архівах статистичні підходи застосовують для автоматизованого тематичного розподілу оцифрованих справ чи попереднього сортування нових надходжень.

Методи машинного навчання надають можливість системам самостійно навчатися класифікувати документи на основі заздалегідь під-

готовлених даних. У контрольованому навчанні (Supervised Learning) моделі тренуються на розмічених прикладах і використовують алгоритми, як-от SVM, дерева рішень. У неконтрольованому навчанні (Unsupervised Learning) система самостійно виявляє структуру в даних, наприклад через кластеризацію. В архівній справі машинне навчання використовують для виявлення прихованих тематичних груп, автоматичного маркування документів чи оптимізації доступу до великомасштабних цифрових фондів. Недоліком є чутливість до якості оцифрованих даних.

Глибоке навчання застосовує багаторівневі нейронні мережі для аналізу складних текстових структур. Моделі типу BERT (двонаправлене розуміння контексту) та GPT (мовне моделювання) демонструють високу точність під час класифікації документів, зокрема в умовах багатомовності та неструктурованих текстів. Для архівної справи це надає можливості автоматичного розпізнавання змісту історичних або малоструктурованих документів. Глибокі моделі дають змогу враховувати контекст, синонімію й стилістичні особливості текстів, проте водночас потребують великих обчислювальних ресурсів і ретельного налаштування, що ускладнює їх упровадження в архівних установах.

Гібридні підходи поєднують різні методи класифікації — rule-based системи, статистичні моделі, машинне та глибоке навчання — для досягнення вищої точності й гнучкості. Наприклад, ключові слова можуть використовуватися для попереднього фільтрування, після чого машинне навчання здійснює остаточну класифікацію. В архівах гібридні рішення надають можливість зберегти переваги традиційних підходів (контрольованість, прозорість) і поєднати їх із можливостями сучасних моделей (швидкість, адаптивність). Ці системи краще впораються з багатомовністю, варіативністю документів і складною ієрархією, однак потребують складніших механізмів розроблення й тестування.

Класифікація архівних документів, особливо в умовах цифрової трансформації, є поетапним процесом, що охоплює підготовку, опрацювання й інтеграцію даних в інформаційні системи. Основні етапи: попереднє опрацювання документів (очищення, нормалізація, розпізнавання тексту); виділення ознак (ключові слова, метадані, структура); вибір і застосування моделі класифікації; оцінювання точності результатів; інтеграцію в систему управління документами (DMS). Кожен із них є критично важливим для досягнення надійної, масштабованої та достовірної класифікації у цифрових архівах [7].

Передопрацювання є першим і надзвичайно важливим етапом у процесі автоматизованої класифікації архівних документів. Його мета — підготувати документи до аналізу, забезпечивши чистоту, узгодженість і машинну зчитуваність даних. Для оцифрованих архівів цей етап часто починають з оптичного розпізнавання тексту, особливо якщо йдеться про скани рукописів чи друканих джерел. Після цього проводять очищення тексту: видалення зайвих символів, пунктуації, артефактів сканування. Далі виконують етапи токенизації (розбиття тексту на окремі слова чи фрази), нормалізації (приведення слів до стандартної форми — лематизація або стемінг), перекодування й уніфікації форматів. Особливу увагу архіви приділяють багатомовності, історичному правопису, нестандартним скороченням і варіативності форматів. Без якісного передопрацювання навіть найточніші алгоритми можуть надавати хибні результати. Цей етап є основою надійної класифікації.

Виділення ознак (Feature Extraction) — це основний етап, на якому з тексту документів здобувають інформацію, котру використає модель класифікації. В архівній справі ознаками можуть слугувати ключові слова, назви установ, дати, географічні назви, тематика, тип документа (наказ, звіт, лист), а також структурні характеристики: кількість сторінок, розмір, формат. Для цього використовують прості методи (наприклад, TF-IDF) й складніші підходи, як-от векторизація слів (Word Embeddings), контекстні подання (BERT, FastText). У структурованих архівах важливу роль відіграють метадані: шифр справи, опис, авторство, категорія зберігання. Коректне виділення ознак суттєво впливає на точність класифікації, адже саме вони формують уявлення моделі про зміст документа. Невдалий вибір чи обмежений набір ознак може спотворити результати навіть найкращого алгоритму, тому на цьому етапі доцільно поєднувати експертний аналіз архівістів з автоматизованими інструментами.

Після завершення етапів передопрацювання та виділення ознак розпочинається *етап побудови або застосування моделі класифікації*. Якщо модель створюють без використання готових моделей, її навчають на розміченому архівному корпусі, де кожен документ уже віднесено до певної категорії чи тематики. У разі використання готових моделей, наприклад на основі BERT чи GPT, відбувається їх адаптація до специфіки архівних матеріалів (Fine-Tuning). У процесі побудови важливо обрати відповідний алгоритм, наприклад наївний байєсівський класифікатор, SVM, дерева рішень або глибокі нейрон-

ні мережі. Рішення залежить від обсягу даних, складності ознак та мети класифікації. Цей етап критично важливий для забезпечення якості й достовірності результату.

Оцінювання точності класифікації є вагомим етапом для перевірки ефективності обраної моделі в архівній практиці. Він дає змогу визначити, наскільки правильно система віднесла документи до відповідних категорій чи тематичних груп. Основні метрики охоплюють: точність (Precision) — частку правильно класифікованих документів серед усіх, що були віднесені до певного класу; повноту (Recall) — частку правильно розпізнаних документів серед усіх документів цього класу; F1-міру — гармонійне середнє між точністю й повнотою. Для оцінювання використовують тестові набори даних, які не брали участі в навчанні моделі. В архівній справі важливо тестувати моделі на реальних фондах, щоб перевірити їх здатність враховувати багатомовність, варіативність форматів та історичні особливості текстів. Оцінювання точності допомагає своєчасно виявити помилки, перенавчання або нерелевантні результати і є основою для подальшого вдосконалення моделей класифікації.

Інтеграція автоматизованої класифікації в системи управління документами (DMS) надає можливість забезпечити цілісне керування архівними фондами в цифровому середовищі. Завдяки цьому документи автоматично отримують відповідні індекси, тематики чи категорії ще на етапі завантаження в систему, що суттєво скорочує час опрацювання, спрощує пошук і підвищує достовірність опису. Для архівів важливо, щоб DMS підтримували метадані, багаторівневу структуру (фонд — опис — справа), а також можливість вбудовування алгоритмів класифікації. Успішна інтеграція забезпечує не лише зручність для користувача, а й сталість архівних стандартів у цифровій екосистемі.

Процес упровадження автоматизованих систем класифікації документів у практику архівних установ має як очевидні переваги, так і ризики, що потребують усвідомленого керування. Водночас такі системи надають нові перспективи розвитку архівної справи в умовах цифрової трансформації. Переваги автоматизації передбачають значне підвищення швидкості опрацювання документів, зменшення навантаження на працівників архіву та розширення можливостей доступу користувачів до інформації. Машинне навчання та глибокі моделі уможливають високоточну класифікацію навіть неструктурованих чи фрагментарних текстів. Окрім того, автоматизовані системи дають змогу гнучко оновлювати

класифікаційні схеми, інтегрувати метадані й забезпечувати аналітику на основі вмісту архівів. Водночас є ризики, що полягають у зниженні точності класифікації в умовах низької якості сканів або розпізнавання тексту, втраті контексту під час автоматичного аналізу, а також у порушенні етичних та юридичних норм у разі неправильного опрацювання конфіденційних матеріалів. Не менш важливим є питання розроблення стандартизованих підходів до класифікації в цифрових середовищах архівів. У найближчому майбутньому можна очікувати широкого впровадження гібридних моделей, де архівіст виконує роль експерта, а система — технічного асистента.

Перспективним напрямом у класифікуванні архівних документів є розвиток національних платформ на базі відкритого програмного забезпечення, що дають змогу інтегрувати класифікацію в архівні DMS з урахуванням мовної, історичної та культурної специфіки фондів.

Висновки. Класифікація документів в архівах — це структуроутворювальний процес, що визначає якість збереження, пошуку й осмислення документальної спадщини. Традиційні методи, зокрема ієрархічна система "фонд — опис — справа — аркуш" і тематичне групування, залишаються основою архівної логіки, оскільки віддзеркалюють контекст походження й змістову

єдність документів. Зі зростанням обсягів цифрових фондів ці методи втрачають ефективність у самостійному застосуванні. Автоматизовані системи класифікації, що базуються на ключових словах, статистичних моделях, машинному та глибокому навчанні, дають змогу масштабувати процес класифікації, враховувати приховані зв'язки між документами та працювати з неструктурованими текстами. Їх використання відкриває шлях до створення інтелектуальних архівних систем, здатних до адаптації, самонавчання й підтримки складних запитів. Перспективним підходом стає поєднання людського експертного аналізу з гнучкістю автоматизованих інструментів. Класифікація у цифрову епоху має стати не фіксованою схемою, а динамічною моделлю, що змінюється разом із потребами користувачів, формами зберігання й типами інформації. Інтеграція традиційного змістового підходу та сучасних технологій дасть змогу забезпечити сталий розвиток архівів і зберегти їх суспільне значення в майбутньому.

Перспективи подальших досліджень полягають у вивченні проблематики розроблення засад національних і міжархівних стандартів автоматизованої класифікації, що враховуватимуть мовну, історичну й контекстуальну специфіку документів.

Список бібліографічних посилань

1. Brokensha S., Kotzé E., Senekal B. Machine learning for document classification in an archive of the National Afrikaans Literary Museum and Research Centre. *Journal of the South African Society of Archivists*. 2023. Vol. 56. P. 134—147. DOI: <https://doi.org/10.4314/jsasa.v56i.9> (accessed: 18.05.2025).
2. Esposito F. [et al.]. Machine learning methods for automatically processing historical documents: from paper acquisition to XML transformation. *First International Workshop on Document Image Analysis for Libraries*. 2004. Palo Alto, CA, USA. [S. l.]. DOI: <https://doi.org/10.1109/dial.2004.1263262> (accessed: 18.05.2025).
3. Lombardi F., Marinai S. Deep Learning for Historical Document Analysis and Recognition—A Survey. *Journal of Imaging*. 2020. Vol. 6, no. 10. P. 110. DOI: <https://doi.org/10.3390/jimaging6100110> (accessed: 18.05.2025).
4. Hutchinson T. Protecting Privacy in the Archives: Supervised Machine Learning and Born-Digital Records. 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10—13 December 2018. [S. l.], 2018. DOI: <https://doi.org/10.1109/bigdata.2018.8621929> (accessed: 18.05.2025).
5. Trace C. B. Maintaining Records in Context: A Historical Exploration of the Theory and Practice of Archival Classification and Arrangement. *The American Archivist*. 2020. Vol. 83, no. 1. P. 91—127. DOI: <https://doi.org/10.17723/0360-9081-83.1.91> (accessed: 18.05.2025).
6. Jiang Sh. [et al.]. Deep Learning for Technical Document Classification. *IEEE Transactions on Engineering Management*. 2022. P. 1—17. DOI: <https://doi.org/10.1109/tem.2022.3152216> (accessed: 18.05.2025).
7. Barros T. H. B., Moraes J. B. E. Archival Classification and Knowledge Organization: Theoretical Possibilities for the Archival Field. *Categories, Contexts and Relations in Knowledge Organization*. [S. l.]. 2012. P. 272—276. DOI: <https://doi.org/10.5771/9783956504402-272> (accessed: 18.05.2025).

Vasyl Kurylo,

Postgraduate Student of the Department
of Social Communications and Information Activity,
Lviv Polytechnic National University

Systematization of approaches to document classification in archives

The article presents a systematization of traditional and modern approaches to document classification in archival institutions. It examines the theoretical foundations of classification, manual and automated methods. The theoretical

foundations of classification, manual and automated methods are considered. It is stated that traditional methods of classification in archives include: hierarchical classification systems; thematic and functional grouping; archival description. Automated methods include rule-based systems; statistical classification methods; machine learning methods; deep learning; hybrid approaches.

A step-by-step sequence of classification stages in archives is proposed: preprocessing, feature extraction, stage of construction or application of a classification model, assessment of classification accuracy, integration of automated classification into document management systems.

Key challenges in implementing automated solutions are analyzed, such as multilingualism, data unstructuredness, low digitization quality, and ethical concerns.

The material is based on contemporary research and practical examples, highlighting the relevance of the topic in the context of digital transformation of archival practices.

Keywords: document classification; classification methods; automated systems; digital document processing; archival fonds structuring; digital archives

Надійшла до редакції 29.05.2025

ДОКУМЕНТОЗНАВСТВО



[https://doi.org/10.36273/2076-9555.2025.10\(351\).45-47](https://doi.org/10.36273/2076-9555.2025.10(351).45-47)

УДК 930.25НАФ+025.177:[004.738.5]:303.4(477)(045)

Леся Ковальська,

докторка історичних наук, професорка кафедри
документознавства та інформаційної діяльності
Державного університету
інформаційно-комунікаційних технологій
dreamlife.lesya@gmail.com
<https://orcid.org/0000-0002-1579-7708>

Григорій Ковальський,

кандидат філософських наук, доцент,
в. о. директора Українського науково-дослідного
інституту архівної справи та документознавства
g.kovalsky.undiasd@arch.gov.ua
<https://orcid.org/0000-0002-3352-4754>

Організація архівного зберігання аудіовізуальних документних ресурсів: чинник соціокомунікаційних технологій

Мета дослідження полягає у визначенні ролі соціокомунікаційних технологій в організації архівного зберігання аудіовізуальних документних ресурсів ініціативного походження. **Методологічну основу дослідження** становлять загальнонаукові та спеціальні методи, які застосовують в архівознавстві, документознавстві й соціальних комунікаціях. У процесі дослідження використано методи наукового аналізу й синтезу для виявлення сутності ініціативного документування як джерела аудіовізуальних ресурсів. Застосовано оглядово-аналітичний метод для систематизації джерельної бази та сучасного досвіду організації аудіовізуальних документів, а також метод класифікації для визначення типів ресурсів ініціативного походження й підходів до їх обліку, організації та використання в межах соціокомунікаційних платформ. **Наукову новизну дослідження** визначають тенденції активного впровадження соціокомунікаційних технологій у практику фіксації та документування інформації, архівування аудіовізуальних документних ресурсів. Наукова новизна полягає у визначенні ролі соціокомунікаційних технологій як інструменту організації аудіовізуальних документів ініціативного походження, а також функціональних моделей взаємодії між користувачами, фахівцями й інформаційними установами в умовах цифровізації суспільства. Запропоновано концептуальне поєднання ініціативного документування із соціокомунікаційною практикою, що розглядається як новий вектор розвитку архівної справи у цифрову епоху. Наголошено, що важливим завданням архівістів є аналіз сутності й особливостей ініціативного документування як джерела формування аудіовізуальних документів, налагодженої методології опису архівних документів. З огляду на це слід розкрити особливості поняття аудіовізуальних документних ресурсів ініціативного походження та їх роль у системі соціальної пам'яті. Термінологічне погодження дасть змогу дослідити потенціал соціокомунікаційних технологій у процесах збирання, організації, зберігання й популяризації аудіовізуальних документів. У процесі архівування аудіовізуальних документних ресурсів визначено основні форми й інструменти взаємодії користувачів і фахівців через соціокомунікаційні платформи (сайти, соціальні мережі, онлайн-архіви тощо). Окремо обґрунтовано доцільність інтеграції соціокомунікаційних технологій у практику архівних та інформаційних установ для роботи з документами ініціативного походження. **Висновки.** Подальші наукові рефлексії слід спрямувати на вдосконалення методології архівного зберігання аудіовізуальних документів, що враховує технічні цифрові реалії. Архіви мають не лише зберігати, а й забезпечувати якісний доступ до цих документів, сприяючи їх інтеграції у глобальний інформаційний і культурний простір.