

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-42-23>

УДК 024.5

Лук'янюк Світлана Василівна, науковий співробітник

<https://orcid.org/0000-0002-7469-8144>

Український науково-дослідний інститут спеціальної техніки та судових експертиз Служби безпеки України, м. Київ

ПРИНЦИПИ ФОРМУВАННЯ ЕЛЕКТРОННОГО АРХІВУ ДАНИХ ЯК ОСНОВИ ДЛЯ ЗДІЙСНЕННЯ ІНФОРМАЦІЙНО-ПОШУКОВОГО ЗАПИТУ

Лук'янюк С. В. Принципи формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту. Розкрито принципи формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту. Визначено поняття електронного архіву даних, електронний архів документів це багатогранна система, призначена для зберігання електронних документів у будь-якому форматі в залежності від їх призначення і вимог користувачів. Підкреслено, що система архівації електронних документів дозволяє класифікувати документи за типами та видами залежно від їх призначення і затребуваності, а також проводити релевантний пошук документів за ключовими словами, реквізитами та іншими параметрами. Наголошується, що формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту умовно можна розбити на кілька етапів: визначення масиву документації, яка підлягає розміщенню в електронному архіві; оцифровка документів; структурування електронних документів на основі різного роду критеріїв; розробка регламентів роботи; визначення рівнів доступу співробітників до документів для їх перегляду або редагування і забезпечення безпеки. Наведено структурну схему цифрового архіву та схему реалізації інформаційно-пошукового запиту на базі електронного архіву даних. Описано стадії наповнення електронного архіву документів та обґрунтовано способи наповнення. Запропоновано алгоритм реалізації інформаційно-пошукового запиту на базі електронного архіву та проаналізовано моделі інформаційно-пошукового запиту на базі електронного архіву даних з відокремленням їх переваг та недоліків, що представлені у вигляді порівняльної таблиці. Зазначено, що практичний аспект реалізації інформаційно-пошукового запиту на базі електронного архіву даних ґрунтується на наборі термінів, що представляють собою окремі слова або словосполучення, а математична складова відображення електронного документу являє собою покриття аналізу частин (фрагментів) документу, де кожному окремому фрагменту співвідноситься набір даних.

Ключові слова: електронний архів, структура, принципи, інформаційно-пошуковий запит, інформація.

Лук'янюк С. В. Принципы формирования электронного архива данных как основы для осуществления информационно-поискового запроса. Раскрыты принципы формирования электронного архива данных как основы для осуществления информационно-поискового запроса. Определено понятие электронного архива данных, электронный архив документов это многогранная система, предназначенная для хранения электронных документов в любом формате в зависимости от их назначения и требований пользователей. Подчеркнуто, что система архивации электронных документов позволяет классифицировать документы по типам и видам в зависимости от их назначения и востребованности, а также проводить релевантный поиск документов по ключевым словам, реквизитами и другим параметрам. Отмечается, что формирование электронного архива данных как основы для осуществления информационно-поискового запроса условно можно разбить на несколько этапов: определение массива документации, подлежащей размещению в электронном архиве; оцифровка документов; структурирования электронных документов на основе различного рода критериев; разработка регламентов работы; определение уровней доступа сотрудников к документам для их просмотра или редактирования и обеспечения безопасности. Приведена структурная схема цифрового архива и схему реализации информационно-поискового запроса на базе электронного архива данных. Описаны стадии наполнения электронного архива документов и обоснованно способы наполнения. Предложен алгоритм реализации информационно-поискового запроса на базе электронного архива и проанализированы модели информационно-поискового запроса на базе электронного архива данных с отделением их преимуществ и недостатков, которые представлены в виде сравнительной таблицы. Отмечено, что практический аспект реализации информационно-поискового запроса на базе электронного архива данных основывается на наборе терминов, представляющих собой отдельные слова или словосочетания, а математическая составляющая отображения электронного документа представляет собой покритие анализа частей (фрагментов) документа, где каждому отдельному фрагменту соотносится набор данных.

Ключевые слова: электронный архив, структура, принципы, информационно-поисковый запрос, информация.

Lukianiuk Svitlana. Principles of creation of the electronic data archive as a background for retrieval requests. The principles of forming an electronic data archive as a basis for the implementation of information retrieval are revealed. The concept of electronic data archive is defined, electronic document archive is a multifaceted system designed to store electronic documents in any format depending on their purpose and user requirements. It is emphasized that the system of archiving electronic documents allows to classify documents by types and types depending on their purpose and demand, as well as to conduct a relevant search of documents by keywords, details and other parameters. It is emphasized that the formation of the electronic data archive as a basis for the implementation of information retrieval can be divided into several stages: determining the array of documentation to be placed in the electronic archive; digitization of documents; structuring electronic documents based on various criteria; development of work regulations; determining the levels of access of employees to documents for viewing or editing and ensuring security. The structural scheme of the digital archive and the scheme of realization of the information search query on the basis of the electronic data archive are given. The stages of filling the electronic archive of documents are described and the methods of filling are substantiated. An algorithm for implementing an information retrieval query based on an electronic archive is proposed and models of an information retrieval query based on an electronic data archive are analyzed, separating their advantages and disadvantages, which are presented in the form of a comparative table. It is noted that the practical aspect of the implementation of information retrieval on the basis of

electronic data archive is based on a set of terms representing individual words or phrases, and the mathematical component of electronic document display is a failure to cover the analysis of parts (fragments) data.

Keywords: electronic archive, structure, principles, information retrieval request, information.

Вступ та постановка проблеми дослідження. В умовах стрімкого розвитку цифрового простору, сучасний стан автоматизованих інформаційних систем та технологій, комп'ютерних мереж та програмного забезпечення, спрямованого на автоматизацію роботи з документами будь-якого типу і формату, обумовлює необхідність впровадження цифрових документів та ведення електронних архівів.

На сьогодні, найбільш поширеним інструментом роботи з цифровими документами є електронний архів як «система структурованого зберігання електронних документів, що забезпечує надійність зберігання, конфіденційність і розмежування прав доступу, відстеження історії використання документа, швидкий і зручний пошук» [1].

Точніше, електронний архів документів це багатогранна система, призначена для зберігання електронних документів в будь-якому форматі в залежності від їх призначення і вимог користувачів.

Крім того, це програмне забезпечення, інтегроване у відповідну технічну оболонку, можна розглядати як універсальну автоматизовану систему управління документообігом. Система архівації електронних документів дозволяє класифікувати документи за типами та видами залежно від їх призначення і затребуваності, а також проводити релевантний пошук документів за ключовими словами, реквізитами та іншими параметрами [2]. Вона забезпечує цілісність електронних документів і можливість спільної роботи багатьох користувачів з документами відповідно до функцій кожного окремого користувача та рівня доступу. Разом з тим електронний архів даних можна використовувати для здійснення інформаційно-пошукового запиту, а також перегляду, редагування та друку шуканого документа.

Аналіз останніх досліджень і публікацій. В останні кілька років, чимало, як зарубіжних так і вітчизняних вчених здійснило відкриття у сфері електронного архівування даних.

Питання застосування систем управління електронним документообігом на підприємстві розглянути Н.С. Козлова та Т.М. Білоусько [2]. Авторами було проаналізовано головні принципи роботи системи електронного документообігу та виконано аналіз прогресивного програмного забезпечення, які забезпечують автоматизацію всієї діяльності підприємства.

В.Я. Плаксієнко та К.В. Черненко [3] розглянуто переваги електронного безпаперового документообігу, визначено основні цілі автоматизації обліку, розкрито структурні елементи безпаперової бухгалтерії в умовах розвитку ринку. Науковцями підкреслено, що без надійної системи документообігу жоден суб'єкт господарювання не може якісно та ефективно функціонувати, оскільки оперативність, надійність та віддача всього господарського механізму залежить від об'єктивності та своєчасності інформаційного забезпечення.

Представили основні напрями діяльності служб науково-технічної інформації державних архівних установ України, визначили форми і методи їх інформаційної, науково-методичної роботи та довідково-інформаційного обслуговування такі автори як С. С. Артамонова, А. А. Майстренко, Л. П. Одинока, Р. В. Романовський [4].

В.І. Тимофеев, Ю.А. Тимошин, В.П. Ярченко та С.О. Шуст [5] розробили схеми архітектури обробки та архівації електронних даних на прикладі бібліотеки із застосуванням гетерогенного середовища.

Із зарубіжних авторів варто відмітити роботи А. Sophian, G.Y. Tian, D. Taylor, J. Rudlin [6], S. Pittner, S.V. Kamarthi [7], С. Huang, X.Wu, Z.Xu, Y.Kang [8], X. Chen, Y. Lei [9], С. Huang, X. Wu [10], Sophian A., Tian G.Y., Taylor D., Rudlin J. [11] та інші.

Проте, враховуючи описані наукові набутки, за темою, питання дослідження принципів формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту залишається відкритим та потребує детального опрацювання.

Мета статті. Розкрити принципи формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту.

Викладення основного матеріалу дослідження. Формування електронного архіву даних як основи для здійснення інформаційно-пошукового запиту умовно можна розбити на кілька етапів:

- визначення масиву документації, яка підлягає розміщенню в електронному архіві;
- оцифровка документів;
- структурування електронних документів на основі різного роду критеріїв;
- розробка регламентів роботи;

– визначення рівнів доступу співробітників до документів для їх перегляду або редагування і забезпечення безпеки.

В умовах сьогодення, на практиці прийнято створювати електронні образи документів при вхідній реєстрації або з моменту їх підписання, і таким чином формувати, по суті, версію аналогового документа. Тому актуальним залишається питання щодо формування самостійного електронного документу, гострота якого визначається відсутністю чіткої нормативно-правової бази, яка регламентує коло документів, дозволених до функціонування виключно в електронному вигляді, що часто пов'язано з масштабністю різновидів документів [1]. Крім того, будь-який носій, на якому розміщується той чи інший електронний документ, має обмежений термін служби, що тягне за собою необхідність його дублювання, а також збереження його описових, структурних та ідентифікаційних метаданих. Структурна схема цифрового архіву наведена на рисунку 1.

Обробка і структурування масиву електронних матеріалів зв'язується з визначенням життєвого циклу документів для своєчасного виведення їх із системи у зв'язку із закінченням терміну зберігання. Введення довідково-пошукового апарату дозволяє визначити системи пошуку, адаптовані під практичні потреби тієї чи іншої організації. При коректно сформованій пошуковій системі нівелюється можливість уявної втрати документа, оскільки вона дає можливість здійснювати пошук як по постійній частині реквізитів документа, яка відображає сутність документа та його індивідуальні особливості, довідковим і сервісним реквізитами, так і у відповідності до вихідних відомостей електронного документа [5].

У зв'язку з тим, що основним завданням електронного архіву є повне забезпечення схоронності даних, важливим в роботі електронного архіву є питання безпеки інформації, що зберігається як від несанкціонованого прочитання третіми особами, так і від будь-якого виду модифікації. Слід розуміти, що забезпечення такого типу безпеки має на увазі збереження не тільки інформації, а й фактичного інформаційного середовища, в якому забезпечується функціонування електронного архіву.

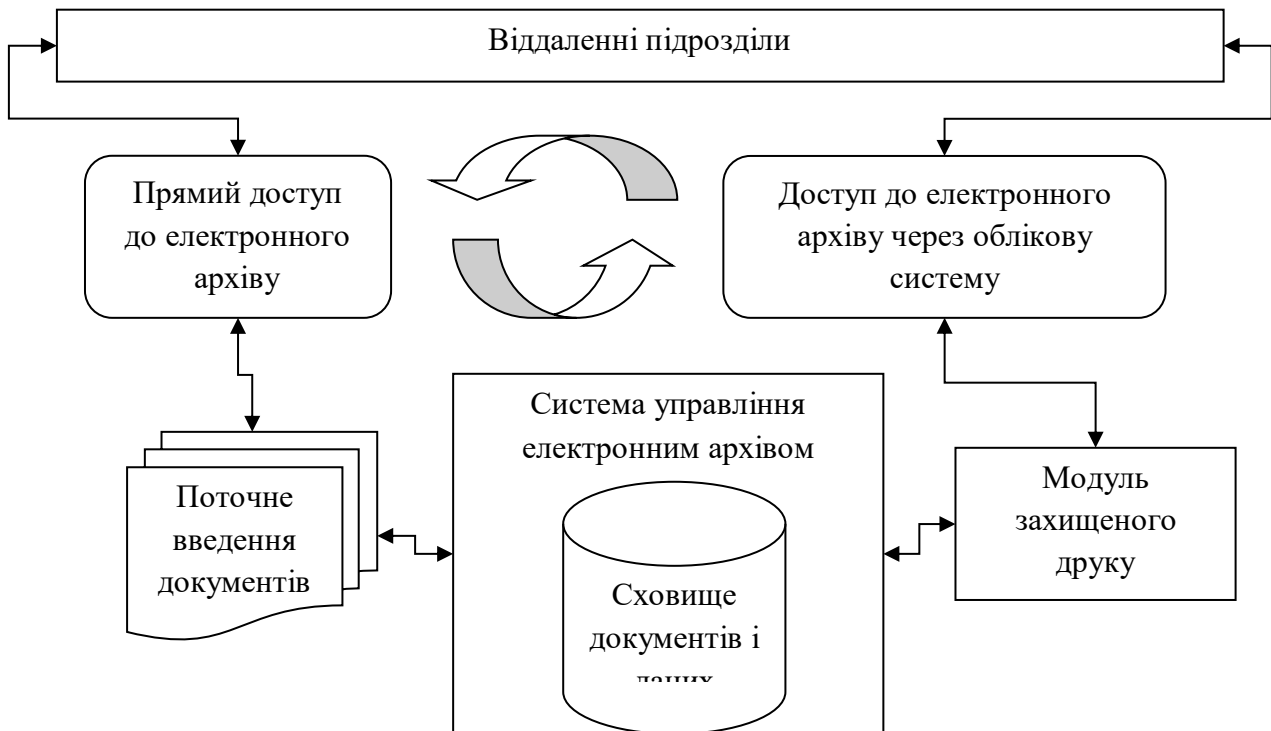


Рис. 1. Структурна схема цифрового архіву

Способи наповнення електронного архіву документів наведено на рисунку 2. Вони умовно поділяються на два види:

- сканування документів та введення їх в електронну базу;
- електронний обіг документів за умови отримання останніх з електронних джерел.

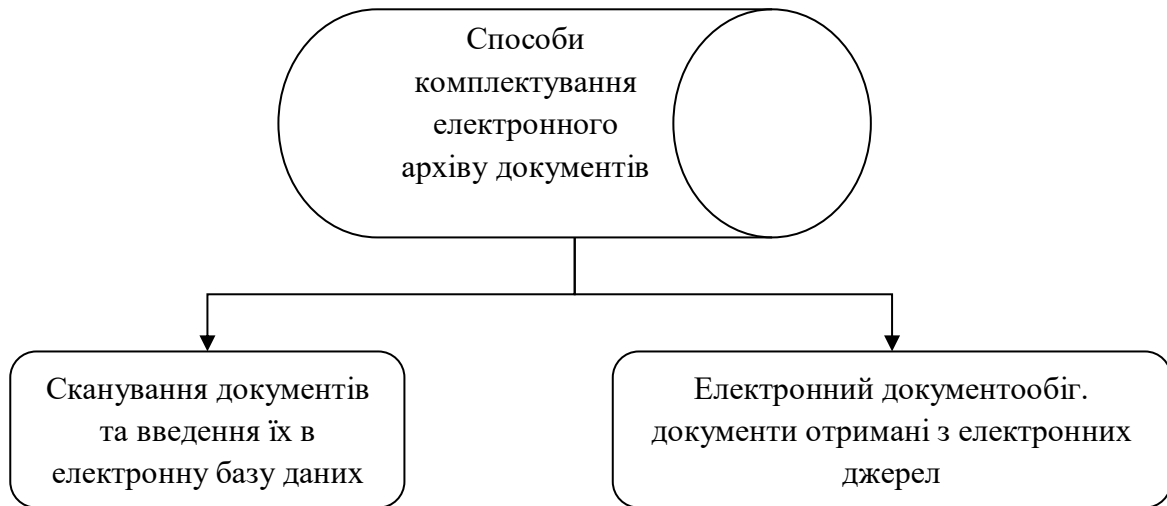


Рис. 2. Способи наповнення електронного архіву документів

Головна стадія формування електронного архіву – внесення впорядкованої документації на цифрові носії. В даному питанні також є ряд вимог – вони пред'являються до відповідності носіїв стандартам ДСТУ, до програмного забезпечення (повинно бути ліцензованим), до шифрування і формату запису файлів, і т.д.

Завершальна стадія – перевірка носіїв, на яких вже знаходяться документи, формування та відправка файлів в систему на зберігання.

Програмне забезпечення на сервері, що забезпечує роботу архіву, має механізм авторизації за логіном і паролем, а також шифрування каналу (HTTPS протокол). Даний механізм забезпечує фіксацію часу внесення змін і логіну, застосованого для цього, а також дозволяє відстежувати деталі внесених змін шляхом порівняння з попередніми версіями файлу.

Назва файлів технічної документації в електронному архіві має строгий формат. Цей формат включає в себе: ідентифікатор проекту; характеристику децимальних номерів; порядковий архівний номер; тип документа; номер версії файлу; розширення файлу.

Інформаційно-пошуковий запит на базі електронного архіву даних здійснюється у відповідності до схеми наведеної на рисунку 3.

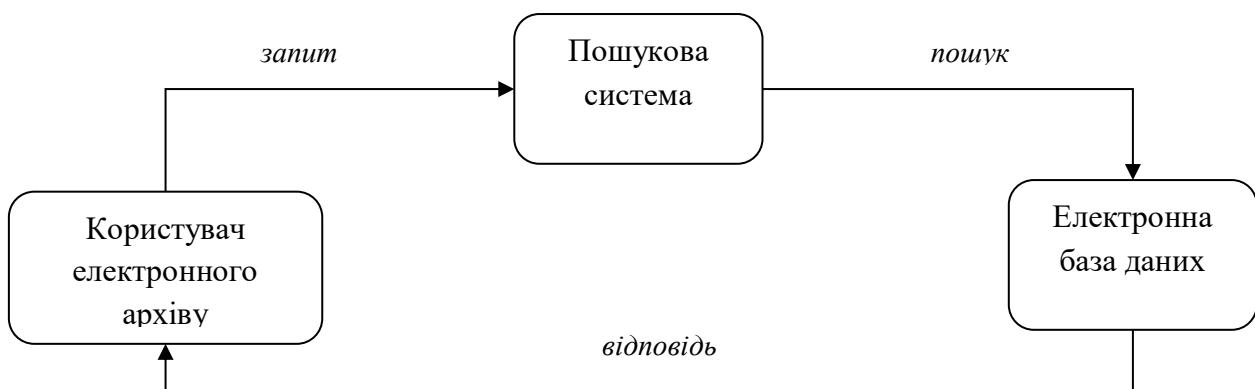


Рис. 3. Схема реалізації інформаційно-пошукового запиту на базі електронного архіву даних

В даний час роботу більшості пошукових систем можна спрощено представити у вигляді наступних кроків:

1. Користувач формує запит.
2. Запит передається пошуковій системі.
3. Система здійснює пошук ресурсу в базі даних на основі цього запиту.
4. Відображення списку результатів, що найбільш підходять до даного пошукового запиту.

На сьогодні, основними підходами до організації інформаційно-пошукового запиту на базі електронного архіву даних, є три моделі кожна з яких має свої переваги і недоліки (табл.1).

Таблиця 1

Моделі організації інформаційно-пошукового запиту на базі електронного архіву даних

Модель	Переваги	Недоліки
Булева модель	проста в реалізації	відсутність контекстних операторів
	дозволяє обробляти великі обсяги дани	невисока ефективність пошуку
Векторно-просторова модель	компактність подання	у класичному вигляді малоприсадибна для обробки великих обсягів даних у зв'язку із застосуванням масивів високої розмірності
Імовірнісна модель	можливість прямого введення	низька обчислювальна масштабованість
	можливість виконання ряду аналітичних операцій	необхідність навчання системи

Крім наведених вище, існують і інші підходи, але їх ефективність поки не досягла того рівня, який би сприяв їх поширенню.

Практичний аспект реалізації інформаційно-пошукового запиту на базі електронного архіву даних ґрунтується на наборі термінів, що представляють собою окремі слова або словосполучення.

Електронний документ можна узагальнити у вигляді покриття:

$$F = \bigcup_i F_i$$

де F_i - це частини (фрагменти) документу, що підлягає пошуку.

Кожному окремому фрагменту F_i співвідноситься набір даних:

$$K_i = \{(K_l, z_l) | l = 1 \dots n_i\}$$

де K_l - ключове слово;

z_l - вага ключового слова.

Далі будемо припускати, що виконуються наступні співвідношення: якщо певна частина F_i міститься у фрагменті F_j , то набір ключових слів міститься в K_j . Більш того, будемо вважати, що фрагменту $F_i \cap F_j$ відповідає набір ключових слів $K_i \cap K_j$. Таким чином, визначається правило успадкування ключових слів.

Рішення про видачу результату інформаційно-пошукового запиту на базі електронного архіву даних приймається в результаті порівняння наборів термінів, які використовувалися відповідно до документів і запитів. До уваги користувача пропонуються ті документи, набори термінів яких збігаються з наборами термінів запитів.

Висновки та перспективи подальших досліджень. З практичної точки зору, застосування електронних архівів в роботі з документами в організації будь-якого типу на сучасному етапі розвитку не викликає питань ні у дослідників, ні у користувачів будь-якого рівня професійної підготовки. Однак питання введення і експлуатації електронних архівів як інструменту роботи тільки архівних підрозділів організацій є предметом наукових дискусій. На практиці оцифровані документи часто розглядаються як страхувальні, що є неприпустимим, а функції електронного архіву зводяться до фактичної зв'язки електронної копії документа з оригіналом на паперовому носії.

Разом з тим, незважаючи на ряд позначених організаційних проблем, можна стверджувати, що на сучасному етапі розвитку електронний архів є багатофункціональним продуктом, що дозволяє здійснювати різнопланову роботу з практично будь-якими масивами документів.

Список бібліографічного опису

1. Тихонов В. И., Юшин И. Ф. (2001). Современные концепции электронных архивов. *Отечественные архивы*, 2, 17-26.
2. Козлова Н.С., Білоусько Т. М (2021). Система управління електронним документообігом на підприємстві. URL : <http://dspace.uccu.org.ua/> (дата звернення: 11.03.2021).
3. Плаксієнко В. Я., Черненко К. В. (2019). Електронний (безпаперовий) документообіг. Електронний цифровий підпис. *Економіка. Управління. Інновації. Серія : Економічні науки*. Вип. 1. URL : http://nbuv.gov.ua/UJRN/eui_2019_1_9.

4. Служби науково-технічної інформації державних архівних установ України : інформаційна, організаційна та науково-методична діяльність : метод. рекомендації. Держ. архів. служба України (2012). УНДІАСД ; уклад.: С. С. Артамонова, А. А. Майстренко, Л. П. Одинока, Р. В. Романовський. Київ, 34 с.
5. Тимофєєв В. І., Тимошин Ю. А., Ярченко В. П., Шуст С. О. (2009). Архітектура систем обробки та архівації бібліотечного контенту в гетерогенному середовищі ВНЗ. *Інформ. технології в освіті*, 4, 41-48.

References.

1. Stiles T., Lawrence J. (2014). Neil Gow A Guide to Archiving of Electronic Records Paperback. Import, 24 February 2014. 26 p. URL : <https://the-hsraa.org/wp-content/uploads/2017/12/AGuidetoArchivingElectronicRecordsv1.pdf>
2. Chowdhury G. G. (2010). Information Retrieval, 3rd edition. London: Facet Publishing, 488 p.
3. Kirchgassner G. (2007). Introduction to Modern Time Series Analysis / G. Kirchgassner, J. Wolters. Berlin : Springer. 274 p.
4. Hegger R., Kantz H., Schreiber T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS* 9. P. 413–435.
5. Kuhlthau C. C. (2004). Seeking Meaning: A Process Approach to Library and Information Services. Westport, CT: Libraries Unlimited. 342 p.
6. Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001). Searching the web. *ACM Transactions on Internet Technology (TOIT)*, 1(1), 2-43.