

УДК 930.25:004.775

Вовк Наталія Степанівна,
кандидат історичних наук,
старший викладач кафедри соціальних комунікацій та
інформаційної діяльності
Національного університету «Львівська політехніка»
Nataliia.S.Vovk@lpnu.ua
<https://orcid.org/0000-0002-2470-7188>

АРХІВНІ ІНФОРМАЦІЙНО-ПОШУКОВІ СИСТЕМИ: ШЛЯХИ ОПТИМІЗАЦІЇ ПОШУКУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Метою роботи є визначення основних шляхів ефективної роботи архівної інформаційно-пошукової системи. **Методологія дослідження** використовує інструментарій методу моделювання: постановка задачі, створення та дослідження моделі. Застосування принципів цілісності, ієрархічності та структурності дало змогу розглядати архівну інформаційно-пошукову систему як цілісну структуру. **Наукова новизна роботи** полягає у створенні основного алгоритму роботи з клієнтським запитом в архівній інформаційно-пошуковій системі. **Висновки.** Запропонований алгоритм обробки користувачького запиту в архівній ІПС дозволяє уникнути дублювання запиту та забезпечує якісну обробку інформацію адміністраторами пошукової системи.

Ключові слова документ, текстова інформація, інформаційно-пошукова система (ІПС), архів, інформація, користувачький запит.

Вовк Наталья Степановна,
кандидат исторических наук,
старший преподаватель кафедры социальных коммуникаций
и информационной деятельности
Национального университета «Львовская политехника»

АРХИВНЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ: ПУТИ ОПТИМИЗАЦИИ ПОИСКА ТЕКСТОВОЙ ИНФОРМАЦИИ

Целью работы является определение основных путей эффективной работы архивного информационно-поисковую системы. **Методология исследования** использует инструментальный ме-

тода моделювання: постановка задачі, створення та дослідження моделі. Застосування принципів цілостності, ієрархічності та структурності дозволило розглядати архівну інформаційно-пошукову систему як цілостну структуру. **Наукова новизна роботи** полягає в створенні основного алгоритму роботи з клієнтським запитом в архівній інформаційно-пошуковій системі. **Висновки** Представлений алгоритм обробки користуваческого запиту в архівній інформаційно-пошуковій системі дозволяє уникнути дублювання запиту та забезпечує якісну обробку інформації адміністраторами пошукової системи.

Ключові слова: документ, текстова інформація, інформаційно-пошукова система (ІПС), архів, інформація, користуваческий запит

Vovk Natalia,
Candidate of Historical Sciences
senior lecturer at the Department of Social Communications and
information activities
National University "Lviv Polytechnic"

ArchiVAL informAtion retrieVAL systems: WAys to optimiZe the seArch for the teXt informAtion

The purpose of the article is to determine the main ways of effective operation of the archival information and retrieval system. **The methodology** of the research uses the toolkit of the method of modeling: statement of the problem, creation, and study of the model. The application of the principles of integrity, hierarchy, and structuring allowed us to consider the archival information retrieval system as an integral structure. **Scientific novelty** of the work consists in the creation of the basic algorithm of work with the client request in the archival information retrieval system. **Conclusions.** The proposed algorithm for processing a custom query in the archive information retrieval systems avoids duplication of the question and provides quality processing information by the administrators of the search engine.

Key words: document, text information, information retrieval system (IRS), archive, information, custom query.

Актуальність теми дослідження. Розвиток інформаційних технологій призводить до стрімкого використання засобів інформатизації у багатьох галузях. Чільне місце у інформатизації суспільства займає оцифрування ретроспективної інформації та швидкий спосіб її доведення до суспільства шляхом використання інтернет-засобів та інших електронних засобів комунікації. Доступ до відкритої архівної інформації згідно із Законом України «Про Національний архівний фонд та архівні установи» є вільним. Проте, часто отримати таку багатопланову ретроспективну інформацію з архівних документів є досить важко. Першопричиною цього є відсутність однієї електронної бази архівних документів із зазначенням їх місця розташування. У зв'язку з цим, користувач неодноразово змушений звертатися із запитом декілька разів. Початковим етапом на шляху вирішення цієї пробле-

ми може стати аналіз архіву як інформаційно-пошукової системи.

Мета дослідження – визначення основних шляхів ефективної роботи архівної інформаційно-пошукової системи.

Дослідженнями інформаційно-пошукових систем займалися: Васильєв О., Чьочь В., Струнгар В., Сазонець О., Збанацька О. та інші.

Виклад основного матеріалу. Для огляду зазначеної проблематики необхідно визначити ряд термінологічних понять.

Інформаційно-пошукова система (ІПС) – це сукупність довідково-інформаційного фонду і технічних засобів інформаційного пошуку в ньому. У свою чергу, довідково-інформаційний фонд (ДІФ) – це сукупність інформаційних масивів (упорядкованих сукупностей документів, фактів або відомостей про них) і пов'язаного з ними довідково-пошукового апарату (тобто даних про адреси

зберігання документів з певними пошуковими образами документа). Пошуковий образ документа – це текст, що складається з лексичних одиниць інформаційно-пошукової мови (тобто спеціального формалізованої штучної мови), що виражає основний смисловий зміст документа і призначений для реалізації інформаційного пошуку. Процес вираження змісту документа на інформаційно-пошуковій мові називається індексуванням [7].

Інформаційно-пошукова система (ІПС) – це різновид автоматизованих інформаційних систем, в яких завершальна обробка даних не передбачається [6]. Призначенням таких систем в архівах буде пошук документів, їх частин та інших фактографічних записів в сховищах (за наявності – у базах даних) за формальними характеристиками.

На рис. 1 зображено класифікаційні ознаки інформаційно-пошукових систем.

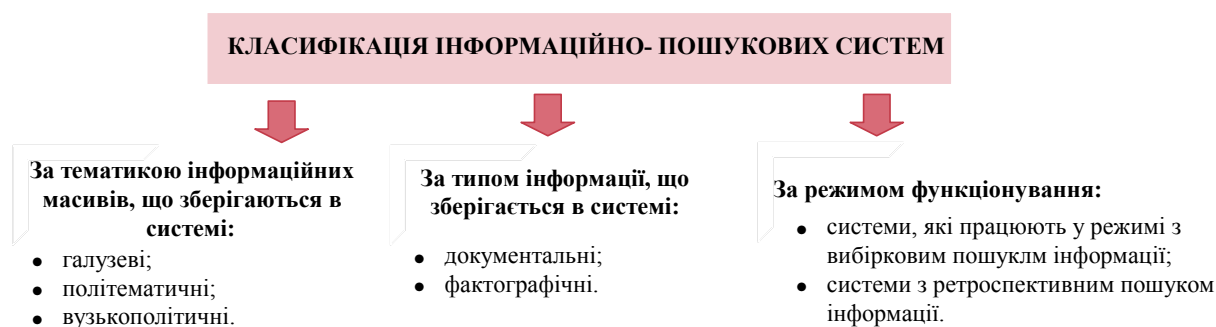


Рис. 1. Класифікація інформаційно-пошукових систем

Аналізуючи рисунок, можна стверджувати, що архівна ІПС за тематикою інформаційних масивів належить до політематичних ІПС, оскільки вона акумулює інформацію різних галузей в одному сховищі даних. За типом інформації, яка зберігається в досліджуваній ІПС, архів – це документальна ІПС: об'єктом зберігання та пошуку є документи. І за третьою ознакою класифікації, архівна ІПС – це система з ретроспективним пошуком інформації.

Для пошук документів, їх частин та інших фактографічних записів в архівній ІПС варто виділити два основних етапи: збір і зберігання інформації; пошук та видача інформації користувачам. Перший етап передбачає збір інформації у вигляді різноманітних джерел (записів, згадок, довідок, листів тощо), аналіз зібраної інформації та її систематизація в архівних фондах. При другому етапі роботи архівної ІПС головним буде оброблення користувацьких запитів та видача самої інформації користувачам.

Виділяють такі основні процеси, які необхідно виконати при здійсненні згаданих двох етапах:

- аналіз документів і їх добір;
- створення пошукового образу документів (ПОД);

- запис документів і їх пошукових образів на прийнятні носії;
- зберігання документів і ПОД;
- аналіз запитів;
- видача документів користувачам [2].

Пошук інформації в архівних ІПС здійснюється трьома способами: простий пошук, розширений пошук та пошук з використання мови запитів.

Простий пошук характеризується пошуком за визначеними словами, які розділені пробілами. Негативною ознакою такого пошуку є неврахування розташування слів у тексті та їх місце знаходження. Результатом простого пошуку є великі групи документів, які вимагають додаткового доопрацювання і уточнення пошукового запиту. Для користувача необхідно чітко сформулювати фразу формального майбутнього запиту використовуючи лише слова природньої мови.

Розширений пошук дозволяє використовувати додаткові уточнюючі умови, які значно полегшують процес пошуку інформації. Для зручності умови вводяться в спеціальні діалогові форми, де зазначено готові списки, або набори полів для вибору [2].

Найпопулярнішим на сьогодні є пошук з використання мови запитів. При такому пошу-

ку використовують специфічні оператори мови – об'єднання слів природньої мови та спеціальних службових символів. Недоліком є наявність окремої мови запитів для кожної ПС.

Для ефективної роботи архівної ПС доцільно використовувати пошук з використанням мови запитів. Недоліком такого підходу буде незнання семантики та синтаксису користувачів. Цю проблему повинні вирішити безпосередньо архівні підрозділи. Є такі шляхи подолання цієї проблеми:

– формулювання єдиного можливого пошукового запиту для всіх архівних ПС;

– чітка структуризація мови запиту в окремих архівних установах.

Другий етап роботи архівної ПС – пошук та видача інформації користувачам, доцільно представити у вигляді блок-схеми.

Алгоритм відповіді на користувацький запит (рис. 1) демонструє процес формування та задоволення інформаційної потреби користувача шляхом вибору необхідної архівної установи та отримання відповіді на запит.

На перших двох етапах виникає поняття пертинентності інформаційної потреби та релевантності запиту.

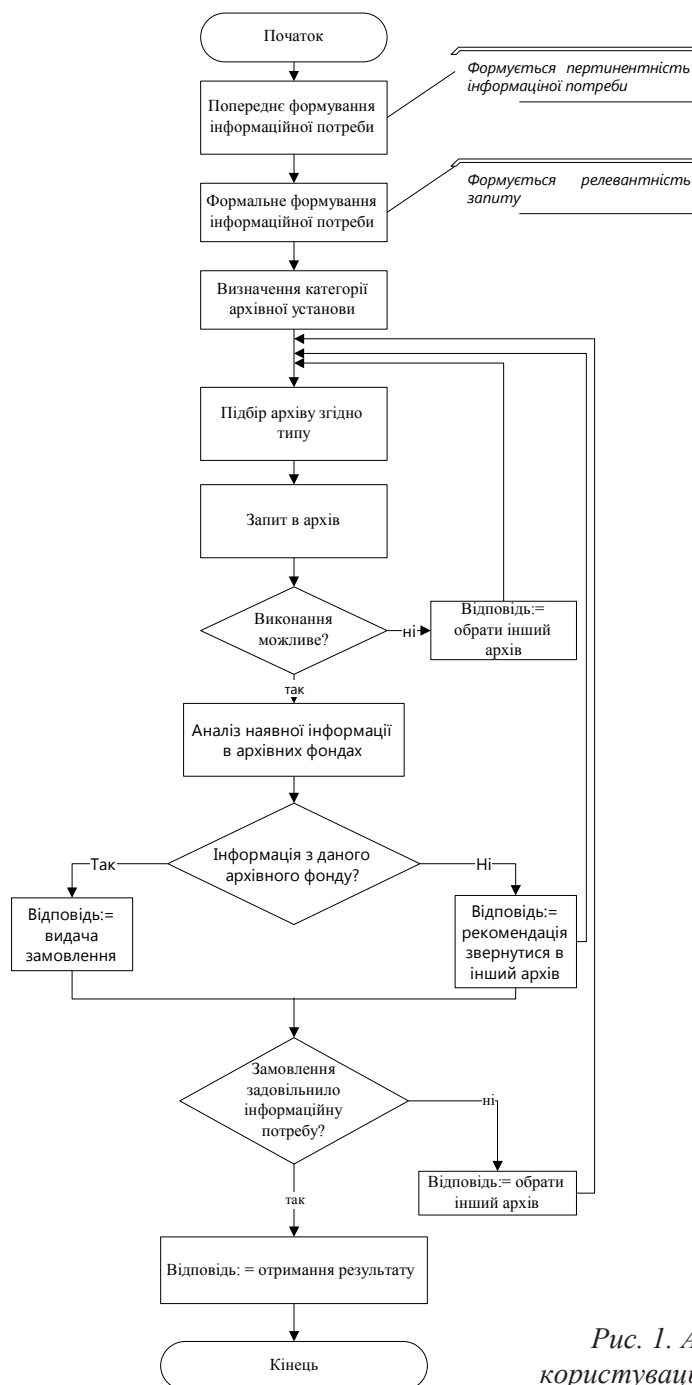


Рис. 1. Алгоритм відповіді на користувацький запит в архівній ПС

Пертинентність інформаційної потреби, у даному випадку, варто розглядати як відповідність сформованого тексту інформаційного запиту фактичній інформаційній потребі споживача [6]. Тобто, при формуванні самої інформаційної потреби користувач повинен чітко визначити хронологічні рамки інформаційного запиту, його суть та можливі джерела.

Релевантність запиту передбачає відповідність тексту запиту тексту документів, наявних в архівних фондах. Якщо на пертинентність інформаційної потреби може вплинути сам користувач, то ступінь релевантності запиту залежить лише від документного масиву архівних фондів.

Алгоритм вибору архіву передбачає виконання окремого підалгоритму – відповіді на запит відвідувача. Важливим етапом у цьому алгоритмі є розбір синтаксису і семантики запиту: від правильності формулювання тексту запиту залежатиме відповідь архівної установи: рекомендація уточнити запит або аналіз наявної інформації в архівних фондах для подальшого видачі замовлення або рекомендацію звернутися в інший архів.

Найважливішою характеристикою даного алгоритму є наявність інформації, яку важливо чітко прокласифікувати за різними ознаками.

Ознака класифікації	Вид інформації
Напрямок руху	Вхідна
	Вихідна
Ступінь опрацювання	Первинна
	Вторинна
Власник інформації	Користувацька
	Архівна
Рівень доступу до інформації	Відкрита
	Таємна
Цілі використання:	Особиста
	Навчальна
	Наукова
	Комерційна
	Довідкова

Як правило, при формуванні користувачького запиту та отриманні відповіді на нього, завжди присутня як вхідна, так і вихідна інформація. Користувацькою інформацією виступає безпосередньо запит відвідувача, а архівною – наявна інформація в архівних фондах. Важливо при розробці алгоритму врахувати і рівень доступу до інформації. При умові запиту за таємною інформацією, архів може відмовити користувачу або затребувати додаткові документи для виконання запиту.

Ступінь опрацювання інформації в архівній ПІС визначається на рівні аналізу наявної інформації в архівному фонді. Якщо в результаті виконання зазначеної умови в алгоритмі (Інформація з даного архівного фонду?) адміністратор здійснює видачу замовлення, то воно буде наповнене вторинною інформацією.

Висновки. Отже, архівна інформаційно-пошукова система – це сукупність архівної документної (текстової) інформації та засобів її пошуку та опрацювання. Важливим компонентом у архівних ПІС є користувачький запит, який повинен бути сформульований із дотриманням правил синтаксису і семантики. Одну з ключових ролей в алгоритмі опрацювання користувачького запиту належить адміністратору інформаційно-пошукової системи, який не лише перевіряє власне сам запит, а й формулює чітку відповідь на нього. Виконання запропонованого алгоритму під час пошуку текстової інформації у архівних ПІС скоротить час опрацювання користувачького запиту та покращить якість контролю управління.

Список використаних джерел

1. Васильєв О. Системи пошуку наукової інформації / О.Васильєв, В.Чьочь // Бібліотечна планета. – 2009. – № 2. – С. 11-16.
2. Вовк Н. Алгоритм користувацького запиту в архівних інформаційно-пошукових системах / Н.Вовк // Інформація, комунікація, суспільство 2018: Матеріали 7-ї Міжнародної наукової конференції ICS-2018. – Львів: Видавництво Львівської політехніки, 2018. – С. 127-128.
3. Вовк Н. Використання архівних ресурсів у суспільстві: інформаційно-аналітичне дослідження / Н. Вовк, Х. Демків, О. Даниляк // Інформація, комунікація, суспільство 2016: Матеріали 5-ї Міжнародної наукової конференції ICS – 2016. – Львів: Видавництво Львівської політехніки, 2016. – С. 328-329.
4. Збанацька О. Архівні інформаційно-пошукові системи: розширення пошукових можливостей (1991-2008 рр.): автореф. дис. на здобуття наук. ступеня канд. істор. наук : спец. 27.00.02 «Документознавство, архівознавство» / О.Збанацька. – К.: 2010. – 20 с.
5. Про Національний архівний фонд та архівні установи: Закон України [Електронний ресурс]. – Режим доступу: <http://zakon2.rada.gov.ua/laws/show/3814-12>
6. Сазонець О. Інформаційні системи і технології в управлінні зовнішньоекономічною діяльністю / О.Сазонець. – К. : «Центр учбової літератури», 2014. – 256 с.
7. Струнгар В. Інформаційно-пошукова система бібліотеки як інструмент прискорення опрацювання даних / В.Струнгар // Вісник книжкової палати. – 2013. – № 10. – С. 1-3.
8. Тесленко Г. Інформаційні системи в аграрному менеджменті – К. : КНЕУ, 1999. – 232 с.
9. Vovk N. Archival Resources of Ukraine and Poland in the Global Information Network // Public Policy and Economic Development: Scientific and Practical Journal. – Poznan-Mykolayiv: Adam Mickiewicz University, Petro Mohyla Black Sea National University. – 2016. Issue 7 (11). – P. 192-205.

References

1. Vasyljjev, O. (2009). Retrieval system of scientific information. *Bibliotechna planeta*, 2, 11-16 [in Ukrainian].
2. Vovk, N.(2018). Algorithm of a custom query in archival information retrieval systems. *Proceedings from VII International Scientific Conference «Information, Communication, Society 2018»*. (pp. 127-128). Lviv: Publishing House of Lviv Polytechnic [in Ukrainian].
3. Vovk, N. & Demkiv, Kh. & Danyljak, O. (2016). Using archival resources in the community: information-analytical study. *Proceedings from V International Scientific Conference «Information, Communication, Society 2016»*. (pp. 328-239). Lviv: Publishing House of Lviv Polytechnic [in Ukrainian].
4. Zbanacjka, O. (2010). Archived information retrieval systems: expanding search capabilities (1991-2008). Extended abstract of candidate's thesis. Kyiv [in Ukrainian].
5. Law of Ukraine on National Archival Fond and Archival Institutions (1994). Retrieved from <http://zakon2.rada.gov.ua/laws/show/3814-12> [in Ukrainian].
6. Sazonecj, O. (2014). Information systems and technologies in the management of foreign economic activity. Kyiv, Center for Educational Literature [in Ukrainian].
7. Strunghar, V. Information retrieval system of the library as a tool for accelerating the processing of data. *Visnyk knyzhkovoji palaty*, 10, 1-3 [in Ukrainian].
8. Teslenko, Gh. (1999). Information systems in agrarian managementю . Kyiv, Kyiv National Economic University [in Ukrainian].
9. Vovk, N. (2016) Archival Resources of Ukraine and Poland in the Global Information Network. *Public Policy and Economic Development: Scientific and Practical Journal*. – Poznan-Mykolayiv: Adam Mickiewicz University, Petro Mohyla Black Sea National University. – 2016, 7 (11), 192-205.