

ПОНЯТІЙНО-КАТЕГОРІАЛЬНИЙ АПАРАТ АРХІВНИХ ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМ

О.М. Збанацька

У статті розглянуто понятійно-категоріальний апарат архівних інформаційно-пошукових систем. Акцентовано увагу на видах інформаційно-пошукових мов та індексування.

Ключові слова: *понятійно-категоріальний апарат, архівні інформаційно-пошукові системи, інформаційно-пошукові мови.*

THE CONCEPTUAL AND CATEGORICAL APPARATUS OF ARCHIVAL INFORMATION AND SEARCH SYSTEMS

O.M. Zbanatska

The article is devoted to the conceptual and categorical apparatus archival information and search systems.

Keywords: *the conceptual and categorical apparatus, archival information and search systems, information and search languages.*

Найважливішим компонентом системи логічно упорядкованого знання є понятійно-категоріальний апарат, що становить сукупність понять, категорій та термінів, які дозволяють в узагальненій формі відобразити явища певної сфери знання, а також зв'язки між ними шляхом фіксації суттєвих ознак та закономірностей. Мірою уточнення теоретичних уявлень, поняття набувають семантичної визначеності і в подальшому закріплюються в термінах, упорядкованих у вигляді терміносистеми. Понятійно-категоріальний апарат архівних інформаційно-пошукових систем – це система погоджених та взаємозалежних понять і категорій (позначених словами і словосполученнями), необхідних для теоретичного пізнання архівної інформаційно-пошукової системи (далі – АрІПС).

Мета статті – дослідити понятійно-категоріальний апарат АрІПС, представлений у законодавчих та нормативних документах, що регулюють питання інформаційних відносин, інформаційної діяльності та архівної справи.

Проведення дослідження потребує визначення таких дефініцій: *інформаційна система, архівна інформаційна система, інформаційно-пошукова система, інформаційний пошук, довідковий апарат архівів, комп'ютеризований довідковий апарат, інформаційно-пошукова мова, класифікаційна інформаційно-пошукова мова, вербальна інформаційно-пошукова мова, предметизаційна інформаційно-пошукова мова, дескрипторна інформаційно-пошукова мова, інформаційно-пошуковий тезаурус, пошуковий образ документа, пошуковий образ запиту, індексування, систематизація, предметизація, координатне індексування тощо.*

Базовим терміном дослідження є *інформаційна система*. Під інформаційною системою розуміють сукупність організаційних і технічних засобів для збереження й оброблення інформації з метою забезпечення інформаційних потреб споживачів. Як окремий специфічний вид слід розглядати *архівну інформаційну систему*, розробка якої концентрується у двох взаємопов'язаних напрямках. Перший напрям торкається практичних питань представлення архівної інформації як системи знань в інформаційних системах та передавання її у формалізованому вигляді. Це питання гуманітарного характеру – визначення об'єктів каталогізації для введення до системи; розроблення уніфікованих методик описування архівних

документів; створення довідкового апарату (далі – ДА) системи, що передбачає розробку певних інформаційно-пошукових мов (далі – ПМ), єдиних класифікаційних систем документної інформації, уніфікацію архівної термінології тощо. Другий напрям стосується технологічних проблем, пов'язаних із реальним функціонуванням автоматизованих інформаційних систем. Концепційні засади такої системи почали розглядати з набуттям Україною незалежності [6].

Враховуючи багатофункціональність архівної інформаційної системи, її також можна назвати пошуковою системою. *Інформаційно-пошукова система* (далі – ППС) забезпечує «пошук документів або відомостей про них [11, 11]». ППС – програмна система для зберігання, пошуку і видачі користувачеві інформації.

Поєднання всіх функціональних особливостей системи (збирання, зберігання, опрацювання, пошук, видача) дозволяє запропонувати дефініцію *архівна інформаційно-пошукова система* і визначити її як ППС, що реалізується на базі використання обчислювальних систем, функціонує як одне ціле та призначена для автоматизації збирання, зберігання, опрацювання, пошуку та видачі ретроспективної архівної інформації.

Будь-яка ППС призначена для здійснення *інформаційного пошуку* у великій кількості саме тих документів, котрі відповідають інформаційному запиту. Інформаційний пошук заснований на використанні пошукових образів документів. Сукупність понять про головні предмети (теми) документа, представлені штучною пошуковою мовою, називаються *пошуковим образом документа* (далі – ПОД). Тема, за якою вибирають документи з масиву, виражається тією самою мовою, є *пошуковим образом запиту* (далі – ПОЗ). А.І. Чорний визначає, що ПОД – це складений за певними правилами текст, в якому виражена центральна тема документа і лише частково – супутні їй теми [10, 23]. Чим більш стисло формулюються пошукові образи, тим вище швидкість пошуку, але одночасно нижчі його точність і повнота.

Залежно від виду ППС, інформаційний пошук може бути документальним або фактографічним. Об'єктами документального інформаційного пошуку є архівні описи, або самі архівні документи, або їхні копії. Документальний інформаційний пошук здійснюється за допомогою документальної ППС. «Документальна система – це система, одним з головних елементів якої є документ (чи система документів)... [11, 11]». ППС функціонують як у традиційному, так і в автоматизованому режимах. У нашому дослідженні це будуть каталоги, картотеки, бази даних, архівні довідники.

Інформатика, що заявила про себе в середині ХХ ст., не тільки принесла нову термінологію (*дескриптор, тезаурус, індексування*) і задовольнила звичайну потребу в протиставленні нового традиційному, але й спричинила більш широкий підхід до явищ і принципів. Приміром, поняття ПМ дозволило розглядати архівне описування, предметизацію, систематизацію, координатне індексування як процеси, що використовують штучні мови, семантичну силу яких можна виміряти в порівнянні з можливостями природної мови. Поняття ППС об'єднало предмети, які раніше розглядалися ізольовано.

Дескрипторні ППС відкрили принципово нову можливість пошуку необхідних документів й інформації, що в них міститься, за будь-яким поєднанням раніше не передбачених ознак [1, 161].

Основною складовою АрППС є *довідковий апарат архівів*. Даний термін було введено до наукового і практичного обігу новою редакцією Закону України «Про Національний архівний фонд та архівні установи» (2001 р.) [8]. Відповідне поняття містять ДСТУ 2732:2004 «Діловодство й архівна справа. Терміни та визначення», чинний від 1 липня 2005 р. [3] та «Основні правила роботи державних архівів» (2004) [7]. До цього часу в практиці роботи архівних установ традиційно застосовувалися терміни *науково-довідковий апарат* та *система наукового довідкового апарату*. Система наукового довідкового апарату визначалася як «сукупність взаємопов'язаних облікових документів, архівних довідників, механізованих і автоматизованих ППС, створених на єдиних методологічних і науково-методичних основах для забезпечення збереженості і пошуку архівних документів і ретроспективної документної інформації [2, 8]».

Проте, на сьогодні визначення *довідкового апарату архівів* має таку дефініцію: це «сукупність архівних описів у вигляді взаємопов'язаних архівних довідників, у т. ч. електронних, і баз даних, призначених для пошуку документів і архівної інформації [7, 182]». І основними типами архівних довідників відповідно до їхнього цільового призначення є архівний опис, архівний каталог, путівник, огляд, покажчик. Кожний тип архівного довідника має свої види і різновиди. Паралельно з традиційним ДА архіву функціонує *комп'ютеризований ДА* – це «структурована сукупність взаємозв'язаних і взаємодоповнюваних інформаційних баз даних різних рівнів та довідкових баз даних, які забезпечують тематичний і адресний пошук архівної інформації, виведення її на екран дисплею, паперовий носій та у файл [7, 226]».

ДА архівів, залежно від цільового призначення, поділяється на типи: опис, каталог, путівник по фондах, покажчик, огляд. Залежно від об'єкта, аспекту описування, ступеня деталізації, кожен тип довідника має свої види та різновиди: опис – здавальний, інвентарний, архівний; каталог – систематичний, предметний (різновиди – іменний, географічний), предметно-тематичний, хронологічний; путівник – по фондах, короткий путівник, тематичний путівник; огляд – тематичний, фондový; покажчик – до описів фонду, покажчик справ (документів) (різновиди – систематичний, предметний, хронологічний) [9].

Для представлення *архівної інформації в Інтернет*, Держкомархівом України¹, іншими архівними установами та архівами почали створюватися і підтримуватися відповідні веб-сайти або сторінки на сайтах (порталах) інших юридичних осіб. Обов'язковими елементами структури веб-сайту архіву є історична довідка, дані про склад і зміст фондів, перелік та характеристика основних фондів, інформація про унікальні та особливо цінні документи, характеристика довідкового апарату, відомості про режим роботи архіву і читальної зали та порядок доступу до документів [7].

Не всі явища, пов'язані з розвитком архівної справи в минулому, вкладаються в рамки сучасної архівознавчої термінології. Приміром, ППМ у сучасному тлумаченні з'являються лише на певному історичному етапі. Поняття *інформаційно-пошукова мова* можна використовувати в широкому значенні по відношенню до всієї сукупності елементів ДА архівів. В основі будь-якого з елементів ДА архівів лежить ППМ – «формалізована штучна мова, призначена для індексування документів, інформаційних запитів і описів фактів з метою наступного зберігання і пошуку [4, 2]».

Інформаційно-пошукова мова – штучна мова, створена людиною, призначена для позначення інформації (документів, фактів) і для наступного економічного і швидкого пошуку та розподілення цієї інформації у відповідь на запит.

У рамках усталеної термінології ППМ поділяються на класифікаційні та вербальні. *Класифікаційні ППМ* – це ППМ, призначені для індексування документів (частин документів) та інформаційних запитів за допомогою понять і кодів будь-якої класифікаційної системи [4, 24; 5, 4]. *Вербальні ППМ* – це ППМ, які використовують для представлення своїх лексичних одиниць слова і вирази природної мови в їхній орфографічній формі [4, 3]. Вербальні ППМ поділяються на предметизаційні і дескрипторні. *Предметизаційні ППМ* (мова предметних рубрик) – це ППМ, призначені для індексування документів (частин документів) та інформаційних запитів за допомогою предметних рубрик. *Дескрипторні ППМ* – це ППМ, призначені для координатного індексування документів та інформаційних запитів засобами дескрипторів та/або ключових слів [4]. *Мова ключових слів* – ППМ, призначена для індексування документів та інформаційних запитів засобами ключових слів [4, 3].

Кожний з ППМ відповідає певний вид індексування. *Індексування* – вираження центральної теми або предмета будь-якого тексту або описування будь-якого об'єкта на ППМ. *Предмет* – об'єкт (матеріальна річ, поняття, властивість або відношення), який розглядається або згадується в документі/інформаційному запиті. *Тема* документа/інформаційного запиту – розділ науки або техніки, галузь практичної діяльності або проблема, якій присвячено документ/інформаційний запит.

* Нині – Державна архівна служба України

Класифікаційна ПІМ використовується при *систематизації* – присвоєнні даним або документам класифікаційних індексів, відповідно до правил будь-якої класифікаційної ПІМ; предметизаційна ПІМ використовується при *предметизації* – предметному індексуванню засобами мови предметних рубрик. Дескрипторна ПІМ використовується при *координатному індексуванні*, – індексуванні, яке передбачає багатоаспектне вираження основного смислового змісту документа або смислового змісту інформаційного запиту множиною ключових слів і/або дескрипторів [4, 19].

Необхідність в ПІМ виникає тоді, коли пошуковий документальний масив достатньо великий. Для прискорення пошуку його розбивають на класи, які поєднують документи за однією темою, і це слугує пошуковим образом для відповідного масиву документів.

Процес створення ПОД – індексування – складається з двох етапів: 1) аналітико-синтетичного перероблення інформації, в результаті якої індексатор визначає основний зміст документа; 2) переведення основного змісту на ПІМ, кодування його за допомогою лексичних і граматичних засобів ПІМ, результатом чого і є створення ПОД.

Найдавнішою з ПІМ є класифікаційна, яка функціонує в певній документальній ПІС і засобами функціонування якої є процес систематизації та систематичні каталоги. Результатом систематизації є *класифікаційний індекс* – пошуковий образ, побудований засобами класифікаційної ПІМ. Практичними посібниками для систематизаторів, матеріальним представленням класифікаційної системи є *класифікаційні таблиці, схеми класифікації*.

Предметизаційна ПІМ функціонує в певній документальній ПІС і засобами функціонування її є процес предметизації, предметні каталоги. Проблеми предметизації і предметного каталогу виникали і частково вирішувалися протягом більше ніж 130-річної історії. Деякі з проблем, які виникли ще за радянських часів, залишаються актуальними і сьогодні.

Результатом предметизації є *предметна рубрика* – елемент ПІМ, який представляє собою коротке формулювання теми природною мовою [4, 12]. Предметні рубрики бувають простими (складаються з однієї лексичної одиниці) і складними (складаються з кількох лексичних одиниць). Предметна рубрика складається із заголовка предметної рубрики – першого елемента багаточленної предметної рубрики, відокремленого від наступних розділовим знаком, і підзаголовку предметної рубрики – другого і наступного елемента багаточленної предметної рубрики, відокремленого від інших розділовим знаком [4, 13]. Підзаголовки предметної рубрики, залежно від змісту, можуть бути тематичними, географічними, хронологічними і формальними. Для уніфікованого розташування лексичних одиниць у предметній рубриці нами виведено таку предметизаційну формулу: ЗПР – ТП – ГП – ХП – ФП, де ЗПР – заголовок предметної рубрики, ТП – тематичний підзаголовок, ГП – географічний підзаголовок, ХП – хронологічний підзаголовок, ФП – формальний підзаголовок.

Практичними посібниками для предметизаторів є *словники (списки) предметних рубрик*. Словник (список) предметних рубрик – це сукупність предметних рубрик і пов'язаного з ними посилально-довідкового апарату предметного каталогу або покажчика.

Результатом координатного індексування є *ключові слова* – інформативні слова, приведені до стандартної лексикографічної форми та *дескриптори* – лексичні одиниці, виражені інформативним словом (вербально) або кодом і є ім'ям класу синонімічних або близьких за значенням ключових слів [4, 16]. Ключові слова без урахування всіх видозмін їхньої форми і відношень між ними вилучаються безпосередньо з тексту документа. Такий вид координатного індексування називається вільним. Контрольоване координатне індексування – коли в ПОД включаються тільки слова, зафіксовані в *інформаційно-пошуковому тезаурусі*, де зазначені їхні синонімічні, морфологічні й асоціативні відношення. Існує багато визначень поняття *інформаційно-пошуковий тезаурус*. Наведемо визначення, вироблене в результаті дискусії на Міжнародній конференції з принципів побудови тезаурусів (Варшава, 1970 р.): «Тезаурус є лексичним інструментом інформаційно-пошукових систем. Він складається з контрольованого, змінюваного словника термінів, між якими зазначені смислові відношення. Такий словник повною мірою охоплює певну специфічну галузь знань, містить перелік

дескрипторів і аскрипторів (допоміжних термінів), упорядкований за систематичним та абетковим принципами і має вказівки на смислові відношення між ними як ієрархічного (родовидового), так і неієрархічного типів [12, 160]». У державному стандарті ДСТУ 2398-94. «Інформація та документація. Інформаційні мови. Терміни та визначення» на форму представлення інформаційно-пошукових тезаурусів сформульована більш коротка дефініція: «Тезаурус – це контрольований словник термінів із зафіксованими семантичними відношеннями, який охоплює одну чи більше спеціальних галузей знання [5]». Проте, найбільш вдалим є визначення інформаційно-пошукового тезауруса як нормативного словника дескрипторної ІПМ із зафіксованими в ньому парадигматичними відношеннями лексичних одиниць [4, 16]. Лексичними одиницями ІПМ є дескриптори. Інформаційно-пошуковий тезаурус враховує семантичні зв'язки між словами: антоніми, синоніми, омоніми, асоціації.

Отже, ми розглянули категоріально-понятійний апарат, який є фундаментом дослідження проблем розвитку архівних інформаційно-пошукових систем. Характеристика основних дефініцій створює системне підґрунтя для розгорання подальшого наукового пошуку.

Використана література

1. Гиляревский Р.С. Основы информатики : курс лекций / Р.С. Гиляревский. – М. : Экзамен, 2003. – 320 с.
2. Делопроизводство и архивное дело. Термины и определения. ГОСТ 16487-83. – Взамен ГОСТ 16487-70, Введен с 01.01.85. – М. : Гос. ком. СССР по стандартам. – 12 с. – (Межгос. стандарт Союза ССР).
3. Діловодство й архівна справа. Терміни та визначення понять : ДСТУ 2732:2004. – [Чинний від 2005-07-01]. – К. : Держспоживстандарт України, 2005. – 35 с.
4. Информационно-поисковые языки. Термины и определения : ГОСТ 7.74-96. – [Введен 01.07.97]. – Минск, 1997. – 33 с. – (Межгос. стандарт). – (Система стандартов по информ., библ. и издат. делу). – Идентичен стандарту ISO 5127-6-83.
5. Інформаційні мови. Терміни та визначення : ДСТУ 2398-94. – [Чинний від 01.01.95]. – К., 1994. – 73 с. – (Держ. стандарт України). – (Інформація та документація). – Ідентичний стандарту ISO 5127-6:1983, NEQ
6. Національна архівна інформаційна система : технічне завдання // Національна архівна інформаційна система «Архівна та рукописна Україніка» і комп'ютеризація архівної справи в Україні : зб. наук. пр. – К., 1996. – Вип. 1. – С. 196–224.
7. Основні правила роботи державних архівів України / Держ. ком. архівів України, Укр. НДІ архів. справи та документознавства ; [Г.В. Боряк, О.П. Володіна, Л.О. Драгомірова та ін. ; за заг. ред. Л.О. Драгомірової, К.С. Новохатського]. – К., 2004. – 227, дод.[30].
8. Про Національний архівний фонд і архівні установи : Закон України від 13 грудня 2001 р. [зі змінами] // Відомості Верховної Ради. – 2002. – № 11. – Ст. 81.
9. Христова Н.М. Науково-довідковий апарат архівів та рукописних підрозділів бібліотек в Україні у 1920-1990 рр. : автореф. дис. ... канд. іст. наук : спец. 07.00.08 «Книгознавство, бібліотекознавство та бібліографознавство» / Н.М. Христова ; Нац. б-ка України імені В.І. Вернадського НАН України. – К., 1999. – 19 с.
10. Черный А.И. Введение в теорию информационного поиска / А.И. Черный ; ВИНТИ СССР. – М. : Наука, 1975. – 238 с.
11. Швецова-Водка Г. Документальні чи документні? / Г. Швецова-Водка // Бібліотечний вісник. – 1997. – № 3. – С. 10–11.
12. Proceedings of the international conference on general principles of thesauri building. – Warsaw, 1970. – 186 p.