

ДОСВІД ЗАСТОСУВАННЯ ЕЛЕКТРОННИХ БАЗ ДАНИХ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ

У статті розглядаються інформаційні бази даних, історія виникнення, класифікація та їх застосування в процесі історичного дослідження в рамках проектів Інституту історії України НАН України.

Доступ великої кількості людей до електронних інформаційних масивів прискорив процеси структуризації суспільства, формування нових товариств, учасники яких об'єднуються на базі окремих інтересів. Реалізація цих інтересів пов'язана з використанням певних інформаційних баз даних – організованих масивів тематичної інформації, представлених в електронному виді. Таким чином, бази даних – це певним чином відібрані, організовані й упорядковані масиви інформації. При цьому мова йде про свідомо підготовлену і організовану інформацію¹.

Ще академік В. І. Вернадський зазначав, що інформація в ноосфері виробляється, зберігається, кодується, декодується, розмножується, відтворюється, втілюється в матеріальні цінності і предмети культури². Дане уявлення було першим чітким визначенням інформаційного ресурсу з інформаційною базою загальноцивілізаційного масштабу та з механізмами створення і управління інформацією.

Серед основних процесів, необхідних для підтримки і розвитку інформаційних баз дослідник М. Баград виділяє наступні: нагромадження і збереження інформації, створеної попередніми поколіннями; збір і оцінка інформації; аналітична обробка інформації й т. ін.³.

Відповідно до розподілу впровадження нових інформаційних технологій на три основні етапи, а саме, створення бібліографічних баз даних, запровадження процедур для їхньої автоматичної класифікації та розробка програм для кількісного аналізу є змістом першої стадії⁴. Глибока й усебічна переробка старих сутностей за допомогою нових технологій відкриває можливості переходу від управління документами до управління інформацією і знаннями.

Використання систем баз даних значно прискорює створення бібліографічної бази історичного дослідження, а комплексні системи, створені на їх основі, у кілька разів підвищують ефективність пошуку

тематичної інформації. Однак інтеграція знань неможлива без попереднього аналізу, що передбачає глибокий аналіз предметної галузі, її структуризацію і врахування інформаційних потреб дослідників. Найважливішим питанням, на наш погляд, є правильний вибір основних критеріїв структуризації інформації і шляхів її здійснення, для максимального задоволення вимог дослідників історичної науки.

Перші перспективи у використанні структурованої інформації в історичній науці намітилися вже у 60-і рр. XX ст. Хоча поняття бази даних тоді ще не ввійшло до наукового обігу історика, доцільність і важливість їх застосування дослідники розуміли повною мірою. Слід зазначити, що вже на ті часи використання баз даних в історичній науці цілком пов'язували з комп'ютерною технікою. Одним з перших науковців, хто своєчасно усвідомив важливість розробки даної проблематики і розпочав займатися цими питаннями був І. Д. Ковальченко. Він всіляко намагався привернути увагу істориків до того факту, що групування даних за певними критеріями, розташування матеріалу в будь-якій потрібній послідовності, усілякого роду вибірки, зіставлення і пошук даних з подальшим підбиттям підсумків, тобто всі так звані комбінаторні завдання, можуть успішно вирішуватись на обчислювальних машинах⁵.

Перевага використання електронних машин полягала ще й у тому, що відкривалися можливості для випробування різних принципів обробки даних і на основі зіставлення підсумків відбувався вибір найдоцільніших з них (звичайними методами це зробити практично неможливо). З таких структурованих форм зберігання інформації найбільшого поширення набули табличні форми, а своє завершене концептуальне вирішення вони знайшли в реляційній моделі даних, запропонованій наприкінці 60-х рр. XX ст. науковим співробітником компанії ІВМ І. Ф. Коддом. Реляційна модель відіграла свого часу значну роль в упорядкуванні організації інформації в базах даних і доступу до них. Вона була покладена в основу концепції доступу до інформаційних ресурсів низкою великих компаній. В результаті, на сьогодні, за оцінками експертів, сховища даних, які створені по реляційній моделі, охоплюють до 99% баз даних у світі.

Протягом декількох десятиліть сектор використання баз даних в історичній науці розвивався уповільненими темпами. Це відбувалося через низку причин, одна з яких полягала у відсутності широкого доступу до ЕОМ, інша – у недостатньому інформуванні істориків про широкі можливості і перспективи роботи з інформаційними системами.

Тільки у 80-і рр. XX ст., разом з появою досить великих масивів структурованої інформації, стало можливим створювати бази даних зі складною структурою та ефективно використовувати їх у роботі науковців різних галузей, зокрема, істориків.. Відкрилися можливості об'єднання декількох програм для відпрацювання даних із ними самими, а також викликати програми, вже виходячи з даних, а не навпаки, як це було раніше⁶. Зазначені процеси стали поштовхом до широкого впровадження в історичну науку систем управління базами даних (СУБД).

Наприкінці 1980-х рр. істориками були розроблені бази даних, що включали біографічні відомості про сотні тисяч персоналій. Створення численних баз даних потребувало їх стандартизації та концентрації⁷. До середини 1990-х рр. створення баз даних стало одним із пріоритетних напрямків розвитку історичної інформатики в СНД. Останнє привело до того, що значну частину наукової продукції членів кліометричного товариства почали становити сюжети, пов'язані саме із цим напрямком. А на наукових конференціях представлялися доповіді, теми яких П. Доорн визначив як «Я і моя база даних» або «Як я перетворив моє джерело на базу даних»⁸.

Ще наприкінці 1980-х рр. головну увагу історики приділяли структуруванню первинних даних суто кількісного характеру, що не викликало проблем з переведенням інформації в електронний вигляд. Це найчастіше перекладалось у просту однотобличну базу даних, або електронну таблицю. На початку 90-х рр. XX ст. ситуація змінилась, адже найбільшої популярності досягли технології створення баз даних, які використовувались для обробки і аналізу структурованих джерел⁹.

До середині 1990-х рр. головною сферою роботи з базами даних стала розробка спеціалізованого програмного забезпечення, що враховує особливості історичних джерел насамперед систему KLEIO, розроблену М. Талером. Особливо активно велися дослідження, пов'язані з розбудовою та аналізом колективних біографій спеціалізованих груп людей. На той час у різних країнах функціонувало більше 20 великих банків даних або архівів електронних даних в галузі історико-соціальних досліджень. Вони мають певний проблемний профіль, організаційну структуру і займаються розширенням колекцій баз даних, їх розподілом і публікацією каталогів.

З удосконалюванням програмного забезпечення «стандартних» систем керування базами даних, типу dBASE та інших, і розробкою

спеціалізованого програмного забезпечення, що враховує особливості історичних джерел, розпочалася активізація діяльності цілої низки наукових центрів по створенню великих баз даних, що містять відомості з популярних історичних джерел¹⁰.

Багато дослідників можливостей ЕОМ стверджували, що саме бази даних виступають найважливішим проявом використання комп'ютера для управління інформацією, особливо коли вони пов'язані один з одним за допомогою передачі даних і доступні віддаленим користувачам¹¹. Саме в цей період історики, що працювали в галузі розробки і створення баз даних, почали розділяти інформацію у формі, близькій до оригіналу, а також у формі, пристосованій до потреб наступної обробки. З того часу питання підтримки перетворення даних розглядається як окремий напрямок, що заслуговує на спеціальну увагу.

Особливо варто звернутися до використання мережі Інтернет, як до основного фактору реалізації всіх можливостей описаної технології. З розвитком мережних технологій користувач одержав можливість працювати з базою даних у діалоговому режимі і з віддаленого терміналу формувати на екрані свій запит, переглядати результати пошуку (реферати, тексти документів) і вибирати необхідні, змінювати запити в процесі пошуку, одержувати статистичні й фактографічні дані, виводити на локальні принтери всі або вибіркові результати пошуку. Такий процес істотно скорочує час, необхідний для пошуку інформації, а також підвищує його ефективність. Саме мережну версію бази даних можна назвати сучасною і найбільш перспективною формою її існування, що може стати основою для розробки складніших аналітичних систем.

Першими мережними базами даних, з якими ознайомились історики, стали ресурси бібліотек. Якщо ще десять років тому електронні каталоги, що підключають до мережі Інтернет, розробляли лише деякі найбільші бібліотеки, то в наш час технологія створення онлайн-бібліографічних банків даних зробила великий крок уперед. Кількість комп'ютеризованих бібліотек, підключених до мережі Інтернет, продовжує зростати зі значною швидкістю.

Певний інтерес представляють бібліографічні системи. Деякі підходи до структуривання інформації в цій галузі стали основою для розробки баз даних, що використовуються на українському історичному порталі. Особливої уваги заслуговує Інтернет-портал Національної бібліотеки України ім. В. І. Вернадського (www.nbuv.gov.ua), розвиток якого здійснюється на глибокому науковому підґрунті, а Центр

комп'ютерних технологій, створений у 1992 р. на базі відділу автоматизації бібліотечно-інформаційних процесів є флагманом створення інформаційних ресурсів і підготовки методичних матеріалів з питань інформатизації¹².

З 1998 р. здійснюється створення Національної електронної бібліотеки, що містить значний обсяг наукових праць, нормативних актів, творів класичної літератури, періодичних видань. Проводиться комплекс заходів, спрямованих на розвиток онлайн-ресурсів як повноцінного наукового Internet-портала¹³.

Слід зазначити, саме структура сайту Національної бібліотеки України ім. В.І. Вернадського стала основою розробленої в 2003 р. концепції створення Інтернет-бібліотеки Інституту історії України НАН України. Так, була запозичена, перероблена і адаптована до запитів істориків база посилань на ресурси українських і світових бібліотек і архівів. За аналогією з порталом був організований сектор тематичних семінарів і конференцій. Подальший розвиток Інтернет-ресурсу в домені www.history.org.ua, відбувався відповідно до вдосконалювання технічної та програмної бази Інституту історії і ключових напрямків його досліджень

Оскільки електронні бібліотеки, представлені винятково в мережі Інтернет, досі не мають значних обсягів спеціалізованої історичної літератури, вони орієнтовані, головним чином, на популярні видання і не ставлять високих вимог до якості тексту і бібліографічних даних, основну увагу варто звернути на сервери традиційних бібліотек. Такі ресурси надають відомості про книги, журнали, дисертації та інші матеріали, що зберігаються в бібліотеці, інформацію про наявність документу у книгозховищі, а в окремих випадках дозволяють безпосередньо через мережу здійснити замовлення книги на абонементі.

Практичний досвід розробки баз даних можна простежити на прикладі порталу Інституту історії України НАН України, на якому розміщені чотири бази даних: три тематичних і одна інтеграційна. В огляді тематичних баз даних представлені Інтернет-бібліотека Інституту історії України НАН України, «Електронна енциклопедія історії України», «Історики України: онлайн-довідник». Найбільшу увагу приділено етапу проектування і можливостям роботи з інформацією на рівні користувача.

Технічна складова, пов'язана з реалізацією розроблених схем в Microsoft Access, конвертуванням баз даних, передачею запитів за допомогою sql у СУБД MySQL, формуванням і виведенням інформації

в браузер за допомогою мов html і php значною мірою пов'язана із програмуванням. Оскільки її програмну реалізацію можна здійснити на різних мовах програмування і за допомогою різних СУБД, безпосередньо в тексті статті ці питання розглядатися не будуть.

Першою найбільшою базою стала «Онлайнова бібліотека видань інституту історії України НАН України». Нагромадження великої кількості pdf-документів, які були структуровані винятково відповідно до номерів журналів і збірників, призвело до проблем, пов'язаних з пошуком необхідних матеріалів, що й стало основною причиною організації бібліографічної бази даних.

Найголовнішими критеріями відбору матеріалів мали стати максимальна точність збігу копії з оригіналом документу, що повною мірою забезпечується за рахунок використання формату pdf, а також якомога глибше розкриття інформаційного потенціалу збереженого матеріалу, доступне вже на етапі пошуку. Для цього необхідно було всебічно розглянути питання тематичного поділу статей і організації складного пошуку по базі даних, що зробило класифікацію і систематизацію статей першочерговим завданням.

Було доопрацьовано і перероблено досвід різних підходів до класифікації, заснованого не на простому розподілі на групи, а шляхом розкриття тематичних зв'язків та запровадження додаткових типів, родів, видів і категорій¹⁴. На розробку структури інформації істотно вплинув аналіз балок (зовнішніх підключень). За його результатами було визначене ядро користувальницьких інтересів, відповідно до якого відбувалося подальше структурування бази даних.

Можливість роботи з базою даних одночасно може бути доступною значній кількості користувачів. З огляду на структурування інформації, головна цінність бази даних полягає в тому, що за допомогою комп'ютерних систем вона надає неоціненну допомогу користувачеві в пошуку необхідної і всебічної інформації в галузі його інтересів, оскільки головні складнощі виникають не стільки із замовленням потрібних книг, журналів, статей, скільки з їхнім наступним переглядом, випискою необхідних фактів і відомостей.

При організації баз даних виникає проблема відбору інформації для зберігання. Якщо для документальних систем, у принципі зрозуміло, які об'єкти підлягають зберіганню, то для фактографічної та енциклопедичної інформації питання надмірності стоїть надзвичайно гостро. Широко розповсюджений у точних науках так званий метод «завдань»¹⁵, що полягає у відборі і зберіганні тільки тих параметрів,

які використовуються в завданнях, можна реалізувати при створенні бібліографічної бази, в якій кожен з запис містить основні вихідні дані (автор, назва, видавництво, рік випуску, випуск або том, кількість сторінок, адреса електронної копії). Вся додаткова інформація може бути отримана безпосередньо з електронної версії, що істотно заощаджує ресурси сервера. Інформаційний метод полягає в розробці інформаційної моделі досліджуваного об'єкта і виділенні збережених параметрів на основі цієї інформаційної моделі. Тут виникають проблеми оцінки цінності інформації, старіння інформації. Найбільш успішна стратегія використання двох методів: методу «завдань» – при базовому пошуку, і інформаційного – при розширеному.

На рівні базового пошуку до системи може бути сформований запит, в якому зазначене ім'я автора і назва статті. В результаті користувач довідається про кількість матеріалів у базі, пов'язаних із запитом, одержує список статей по темі, або усі роботи даного автора. Також можливий і реферативний опис окремих праць. Рівень розширеного пошуку шляхом залучення додаткової інформації про видавництво, рік випуску і, що особливо важливо підключенні додаткової інтегруючої бази даних, дозволяє більш глибоко зробити вибірку та одержати матеріали, найбільш близькі за змістом до теми пошуку.

Від січня 2002 р. в Інституті історії України НАН України діє робоча група з наукової інформації суспільних наук. Перше завдання, яке було покладено на цю групу – створення електронної версії «Українського історичного журналу», що здійснювалося за фінансової підтримки Міжнародного фонду «Відродження». Завдяки сучасним електронним технологіям журнал став доступним для користувачів Інтернет в усьому світі. Протягом першого року існування електронного ресурсу з ним ознайомилися більше 9 тис. читачів із 64 країн світу. У вільному доступі тут розташований архів журналів починаючи з 1998 р. Формат pdf дозволяє зберегти всі особливості друкованого видання при використанні програми AcrobatReader; до того ж ця програма є вільно поширюваним програмним забезпеченням¹⁶.

На початковому етапі створення електронної версії «Українського історичного журналу» весь матеріал був елементарно структурований за часом виходу номерів. Для більш глибокої орієнтації пропонувалося звертатися до бібліографічних списків статей, які розміщалися в останньому номері за рік. Такі списки являли собою pdf-документ, що передбачало витратити додатковий час на його завантаження і запуск, а також необхідною умовою була обов'язкова наявність Adobe Reader

для його відтворення. Будь-які можливості пошуку були відсутні ще й у тому випадку, якщо останнього номеру не було, тому для отримання результату потрібно було послідовно переглядати всі статті, що вийшли в журналі того року.

Зі збільшенням кількості номерів, доступних в електронному вигляді і включенням у портал ще декількох періодичних видань («Соціум», «Рутеніка», «Україна в Центральній-східній Європі») успішний тематичний пошук став ще складнішим. Про підключення можливостей пошуку свідчили і спеціальні журнали (балки), в яких висвітлювалася історія підключень користувачів. Більшість користувачів послідовно переходили від одного номера до іншого, переглядаючи списки статей у пошуках необхідного матеріалу, не маючи чіткого уявлення про те, де він може знаходитись.

Пошук матеріалів журналу по його електронних версіях став нагальною потребою. Ще одним фактором, що істотно вплинув на структуру розроблюваної бази даних, став початок створення електронного каталогу бібліотеки Інституту історії України НАН України.

Основна таблиця бази даних містила наступні пункти: порядковий номер – унікальне ключове поле, автор, назва статті, назва журналу або іншого видання, номер або том, рік видання, видавництво, html-адреса на порталі, тематичний індекс, сторінки розміщення статті, коментарі. Така структура основної таблиці дозволила сформулювати рp-запит до SQL, що виводить всю необхідну інформацію для складання посилання на тверду копію. З цією таблицею за принципом рівноправності пов'язані дві додаткових таблиці: одна містить розгорнуту інформацію про тематичні індекси¹⁷, інша – загальну кількість сторінок у виданні. Ще одна таблиця в SQL-базі являє собою згруповані дані і є залежною: у ній зберігається інформація про загальну кількість номерів і статей видання, отримана на основі підрахунку значень деяких полів з основної таблиці.

Запровадження двох додаткових логічних полів «наявність/відсутність електронної версії» і «наявність/відсутність у бібліотеці Інституту історії НАН України» дозволило підключити до порталу бібліографічну базу книг бібліотеки, створену в Інформаційному центрі в системі «Ір-біс», що дозволило співробітникам Інституту одержати повний список всіх видань, що надходили до бібліотеки, починаючи з 2003 р.

На основі електронної версії «Довідника УІЖу» і переробки довідника 1995 р., бібліографічна база журналу була істотно розширена і вмщувала інформацію про всі статті, що вийшли з 1985 р.

нана зі складним механізмом оплати, особливо для більшості вітчизняних користувачів. Ще одним недоліком є непостійна адреса того чи іншого ресурсу.

Однією з головних переваг Інтернет-версій енциклопедій є можливість їх постійного поновлення відповідно до побажань користувачів. Іншою перевагою таких енциклопедій стала відсутність фактичних обмежень за кількістю наданої інформації, тоді як програми, розміщені на дисках, мають обмежений обсяг. Швидкість роботи з енциклопедіями на дисках значно вища, тому що не має потреби в підключенні до мережі. Однак виправлення неточностей, зміна і доповнення таких робіт призводить до появи нових версій, зміна яких призводить до додаткових витрат. Різні браузері (програми перегляду ресурсів мережі Інтернет) відображають інформацію по-різному, що часом стає причиною зміни або втрати частини тексту, або некоректного відображення графічних файлів.

Нерідко виникають складнощі у зв'язку з відсутністю необхідних шрифтів для прочитання польської, грецької, німецької та будь-якої іншої мови, що містить унікальні символи. При інсталяції програми з компакт-диску всі необхідні шрифти автоматично встановлюються в систему. Таким чином, на даний момент не можна віддати перевагу жодному способу. Самі творці енциклопедій намагаються розвивати обидва напрями, і більшість вдалих проєктів, випущених на дисках, мають посилання на сайт, що містить коротку Інтернет-версію продукту.

Не можна не відзначити й успішний проєкт створення «Історії Києва» на двох дисках, у розробці якого брали участь і співробітників Інституту історії України НАН України. Цікаво, що в Інтернеті абсолютно незалежно був створений і продовжує розвиватися сайт «Енциклопедія Києва» (<http://wek.kiev.ua>).

Приблизно повний обсяг «Енциклопедії історії України» складе більше 13 тис. статей. В Інституті історії України НАН України статті зберігаються у форматі doc, а після передачі у видавництво перетворюються у формат програм верстки. Статті з версток не використовуються як звичайні тексти, оскільки не можуть бути відкриті за допомогою стандартних текстових процесорів, а деякі потребують ще й обов'язкового підключення додаткових шрифтів.

Розробка і створення електронної версії «Енциклопедії історії України» передбачає переклад статей з формату верстки у формат текстового документа. Але пряме перетворення текстів верстки у звичайний текстовий формат є неможливим, оскільки відбувається втрата

цилому ряду нестандартних символів, що не входять до англійського і російського алфавітів. Такі шрифти, як грецький, німецький, польський і т.п. відображаються не правильно.

Щоб уникнути залежності від використовуваних шрифтів, тексти переводяться в кодування Юнікод. UNICODE (UTF-8), що дозволяє підтримувати багато мов і не вимагає установки в системі додаткових шрифтів. Особлива перевага полягає в тому, що такі тексти можуть бути використані при розробці продуктів на більшості мов програмування. Кодування Юнікод підтримується як Інтернет-броузерами (починаючи з 4-ої версії Explorer), так і практично всіма сучасними операційними системами.

Зміна кодування для більшості статей (близько 90 %) не представляє складності. Достатньо лише файли з верстки експортувати в текстовий формат. Складнощі виникають при використанні в тексті мов з унікальними символами (грецька, німецька, польська та ін.). У цьому випадку такі частини тексту можна перетворити такими способами: а) скопіювати через буфер обміну; б) скористатися спеціальними конверторами; в) набрати текст заново.

Електронна версія «Енциклопедії історії України» наповнюється як матеріалами вже опублікованих томів, так і текстами повного циклу редактування. При розміщенні матеріалів у мережі Інтернет до будь-якої статті можна зробити коментар, який автоматично вводиться в базу даних та індексується за обсягом і часу внесення. Окрім універсальної бази, яка містить всі статті, організовані за алфавітним принципом, можуть бути запроваджені й інші класифікації, такі як біографічний і термінологічний словники (як додатковий варіант хронологічний і (або) архітектурний розділи). Якщо енциклопедія буде розміщена в мережі Інтернет, пошук за ключовими словами бажано здійснювати переважно в назвах статей, тому що робота із власне самими текстами істотно збільшить час виконання запиту і кількість знайдених матеріалів може вираховуватися десятками або навіть сотнями.

Розроблено і введено в дію сайт, на якому розміщується інформація про вихід нових томів енциклопедії із прес-релізами, дані про роботу з електронним проектом, поновлення для придбаних компакт-дисків з «ЕІУ». Для забезпечення достатнього рівня захисту, кожний користувач має одержати унікальний код, що забезпечує доступ до поновлень для сайту даного проекту.

Карти, включені до енциклопедії, можуть бути перетворені на геоінформаційні системи (системи керування базами даних, призначені

для роботи з територіально-орієнтованою інформацією – далі ГІС), що дозволить пов'язувати картографічні об'єкти (тобто, об'єкти, що мають форму й місце розташування) з описовою, атрибутивною інформацією (текстами і зображеннями), що відноситься до цих об'єктів і описує їхньої властивості. За допомогою ГІС, самі карти можуть бути доповнені звітними документами, тривимірними зображеннями, графіками, таблицями, діаграмами, фотографіями та іншими мультимедійними засобами.

При розробці даної бази, вперше була випробувана технологія об'єктно-орієнтованих банків даних, що забезпечує використання прямих адресних посилань та відкриває широкі можливостях багаторазового перевизначення структурних відносин між об'єктами, і забезпечення підтримки взаємозв'язків на кшталт «значні-до-значних»¹⁸. На наш погляд, саме такий підхід є оптимальним для енциклопедичного формату.

Онлайн-версія «Енциклопедії історії України» розроблялася на основі Accesses-бази. Для організації Інтернет-версії, текстові матеріали переводилися в html-формат, здійснювалася вставка перехресних посилань, підключення унікальних стилів для кожного з підрозділів (назва, основний текст, бібліографія, автор) енциклопедичної статті, системна реорганізація назв і форматів ілюстрацій. Особливістю бази є її інтерактивність: будь-який користувач, може доповнити існуючі матеріали, розширити бібліографію, вказати на ресурси пов'язані з даною тематикою.

Одним з фундаментальних стало питання про можливість одночасної модифікації даних і знань у випадку їхнього взаємного перетинання. Знання, застосовані до даних, постійно змінюються, а отже, зв'язок між ними в історичних інформаційних базах є динамічним¹⁹. Саме енциклопедична база даних забезпечує гнучкість, адже у випадку значних перетворень її статті, шляхом динамічного створення інтеграційною базою відповідних гіперпосилань, пов'язуються з будь-якою інформацією, що дозволяє одержати найбільш повне уявлення про ту чи іншу тему.

На окрему увагу заслуговує база зображень (близько 5 тис. одиниць), яка в перспективі може бути розширена і виділена в самостійну базу. В 90-і роки ХХ ст. творці колекцій зображень були вимушені робити вибір між якісними і кількісними характеристиками матеріалів²⁰. Тоді в більшості випадків при створенні баз дані, які нараховували десятки, або й сотні тисяч одиниць, використовувалася відносно низь-

ка розв'язна здатність. Це стало причиною повернення до повторного процесу оцифровки, але вже на підставі інших вимог до якості зображень. Однак уже в той час професійні історики вказували на те, що якщо матеріали відносно низької якості ще можуть використовуватися при організації презентацій, або для освітніх функцій у музеях та навчальних закладах, то для систематичних досліджень, переставлення зображення з більш низькою якістю, ніж на фотографіях, є недопустимим²¹.

Третя база даних розробляється у співробітництві із проектом «Українські історики», а її теперішня структура визначена з опорою на видання «Інститут історії України». Вона містить інформацію біографічного характеру про українських дослідників історії, а також бібліографічні дані про найбільш визначні роботи.

В історичній науці розвиток баз даних, пов'язаних із просопографічними дослідженнями, метою яких є створення і аналіз колективних біографій специфічних груп людей, відбувався особливо інтенсивно. Наприкінці 80-х рр. ХХ ст. істориками були розроблені бази даних, до яких входили біографічні відомості про сотні тисяч персоналій. Створення численних інформаційних баз потребувало їх стандартизації і концентрації. До середини 90-х рр. минулого століття у різних країнах функціонувало більше 20 великих банків даних або архівів електронних даних в галузі історико-соціальних досліджень, що мають певний проблемний профіль, організаційну структуру і займаються розширенням колекцій баз даних, їхнім розподілом, та публікацією каталогів²².

Таким чином, методологія проектування баз даних колективних біографій досить глибоко розроблена, основні вимоги до таких баз сформовані. Завданням проекту «Історики України: онлайн-довідник» було створення довідкового ресурсу, присвяченого українським історикам, що містить інформацію про біографії дослідника, його наукові інтереси і основні роботи. Був розроблений модуль що дозволяє додавати нові записи в базу через мережу Інтернет, а спеціальна розмітка тексту біографії дозволяє проводити вибірки за додатковими ознаками (місце навчання, наукова школа, колеги й співавтори).

При роботі над базою «Українські історики» іноді досить складно визначити, які джерела вказувати, оскільки найчастіше дані, що заносять у запис бази, інтегровані з фрагментів різних джерел. Обсяг створюваних у ході написання біографії «запитально-перехресних» таблиць може перевищувати обсяг кінцевого продукту в десятки, а

іноді й у сотні разів. У такому випадку необхідна одночасна публікація остаточної, фактологічної бази даних і допоміжної, взаємопов'язаної з першою, бази даних, створеної на основі всієї залученої до дослідження літератури і джерел із посиланням на їх вихідні дані, або дані архівного зберігання.

Неабиякий інтерес для дослідників представляють енциклопедії із графічною складовою, як джерела портретів, карт, різноманітних фото- і відео- матеріалів. Що ж стосується текстів, то тут, виходячи з власного досвіду, перевагу варто віддавати сайтам бібліотек і періодичних видань. Взагалі, саме текстові матеріали є найпопулярнішими Інтернет-ресурсами.

Враховуючи той факт, що кількість інформації в Мережі подвоюється кожні 9 місяців, досить важливим фактором є формування загальних принципів організації у розміщеній інформації. Тексти можуть бути представлені як у формі, що читається машиною (HTML, текстовий формат), так і у формі малюнків (у першу чергу мова йде про рукописні документи). Поряд з вищевказаними формами стрімко збільшується кількість документів у спеціальних форматах (одним з найбільш відомих і розповсюджених форматів на сьогоднішній день є PDF).

Процеси інтеграції й синтезу, є однією з визначальних характеристик сучасного етапу розвитку науки. Успішно вирішувати завдання інтеграції можливо лише при чіткому виявленні узагальнюючих закономірностей і використанні єдиного методологічного підходу до об'єктів дослідження²³.

Експериментальною перевіркою ефективності використання інтеграційних рішень, стало створення бази даних «Наукові напрямки». Наукова тема завжди є набагато ширшою за її представлення в бібліографічному каталозі. Головна проблема полягає в тому, що формування ключових слів та інших змістовних критеріїв, за якими буде можливий пошук у базі даних, розглядається як проста формально-технічна процедура, що не вимагає високої кваліфікації. Такий підхід призводить до втрати «прикордонних» тем для одного або декількох напрямків. У результаті, навіть наявність у базі даних великого обсягу інформації не гарантує можливість її одержання. При такому підході може відбуватися повторне «відкриття» тем, що може стати особливо актуальним на рубежах кількох галузей знань²⁴.

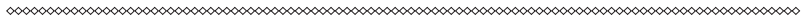
Навіть використання всіляких класифікацій, і створення на їхній основі систематичних (тематичних) каталогів, не виключають втрати матеріалу для багатьох зацікавлених у його одержанні дослідників.

Для більш глибокої структурної розбудови необхідна участь на етапі класифікації вузькопрофільних фахівців, що чітко представляють увесь спектр тем, пов'язаних з певною проблематикою. Так, визначення тематичної області матеріалів повинно належати безпосередньо авторові. Крім того, текст має пройти ряд етапів автоматизованого аналізу з використанням спеціальних алгоритмів.

Об'єднання трьох існуючих баз даних потребувало спеціального дослідження теоретичних питань, пов'язаних з логікою та методологією створення інтегрованих баз даних. Прийнято виділяти п'ять основних напрямів інтеграції знань у системах баз даних: здійснення теоретичного синтезу знань на методологічному рівні, розробка математичної теорії СБД, створення програмного забезпечення СБД, інтеграція знань за формами подання інформації та через техніку (інтеграція локальних баз даних в інформаційно-обчислювальних мережах)²⁵. Оскільки експериментальні бази являють собою, головним чином, масиви текстової інформації, головна увага приділяється питанням методології, формам подання матеріалів і механізмам реалізації специфічних вимог історичної науки.

Інтеграційна база даних «Наукові напрямки» є центром забезпечення комплексного пошуку, а її основними елементами є систематизаційні й класифікаційні матриці, які забезпечують складний пошук відповідно до запитів, генерованих на основі даних користувача. До систематизаційних матриць відносяться «Хронологія», «Персоналії», «Події», «Географія», «Спеціальні історичні дисципліни». Класифікаційними матрицями є «Періодизація», «Проблеми», «Міжнародні відносини».

Основною цінністю даної бази є можливість її модифікації користувачами, що забезпечує більш гнучку реакцію при підборі матеріалів відповідно до вузькоспеціалізованих запитів. При її розробці особлива увага приділялася фіксації основних результатів розвитку кожного конкретного наукового напрямку в інституті історії України НАН України і чітке визначення областей перетинання різних класифікацій відповідно до обраної тематики. На даний час розроблена тільки базова модель, оскільки комплексна її реалізація можлива тільки при залученні вузькопрофільних фахівців, тому доступна на порталі база не має прикладного значення. Однією з невирішених проблем стала цілісність бази даних – відсутність суперечливих даних і відносин: в інтеграційній базі наукових напрямків, наявність суперечливих даних і відносин мало передбачатися.



За своєю суттю база даних «Наукові напрямки» є семантичною оболонкою, що забезпечує успішний пошук при вирішенні конкретних завдань. Під семантичною оболонкою ми розуміємо передачу знань у вигляді набору категорій, об'єднаних в ієрархічні структури, пов'язані з текстами за рахунок організації спеціального форматування. Деякі елементи тексту (слова, словосполучення, імена і прізвища, дати, географічні назви) набувають статусу визначальної значимості для теми запиту. При цьому кожна категорія може мати більше однієї «батьківської», тобто ті самі матеріали можуть бути подані при формуванні різних запитів.

Використання такого підходу забезпечує «доступ до атомарної інформації», дозволяючи тим самим здійснювати пошук як по вертикалі – шляхом переміщення від більше загальних категорій до більше часток, так і по горизонталі – від однієї родинної категорії до іншої²⁶. У випадку запиту до одного з базових напрямів, система трансформує його в набір текстових шаблонів, що входять у визначення даного напрямку, для організації подальшого пошуку. Семантичні системи, побудовані на основі таких принципів, представляють цінність ще й за рахунок того, що можуть бути використанні з іншими текстами і базами даних без суттєвої модифікації²⁷.

Таким чином, можна зробити висновки, що, хоча сховища даних, засновані на реляційній моделі є загальноприйнятим стандартом, існує низка принципових недоліків, властивих табличній формі зберігання даних. Серед цих недоліків відокремлюємо наступні:

- неможливість забезпечення повної підтримки об'єктів;
- недостатня гнучкість і продуктивність роботи;
- низький відсоток аналітичного потенціалу реляційної моделі даних²⁸.

Останній пункт пов'язаний з тим, що таблична форма призначена для реєстрації інформації, тож існує як реєстраційна форма подання даних. Вона не пристосована до того, щоб враховувати складні структури об'єктів і численні, динамічно мінливі зв'язки між ними. Окрім того, вона не дозволяє реалізувати складні механізми інформаційно-логічної обробки даних на принципах структурної ієрархії, властивій високоорганізованим інтелектуальним системам. Ці обставини ускладнюють можливості створення системи збору, обробки й аналізу неструктурованої вербалізованої інформації.

Поряд із названими недоліками технічного плану, в середині 90-х рр. XX ст. у середовищі істориків звернули увагу на ще одну про-

блему: загальну захопленість створенням банків даних. Ряд дослідників указували на переключення інтелектуальних сил на досягнення суто прагматичних цілей. Це, у свою чергу, призвело до обмеження дослідницького арсеналу стандартними пакетами програм²⁹. У результаті такого процесу значна кількість членів комп'ютерного співтовариства істориків (у тому числі й через об'єктивні причини) була задіяна у створенні баз і банків даних, удосконалюванням СУБД і програмного забезпечення, і відійшла від завдань вивчення історії³⁰. Не історики, що підключилися до вищевказаного процесу, не безпідставно почали сприймати комп'ютеризацію як спробу перетворення історії з гуманітарної науки на якийсь математично-комп'ютерний додаток.

Ще один недолік баз даних, що обмежує їхнє широке використання, полягає в тому, що по-справжньому оцінити можливості баз даних можна лише в тому випадку, якщо кількість записів у них досягає декількох тисяч, а їх структура досить глибока й широко продумана. Відзначимо також і той факт, що поза комп'ютерною технікою бази даних існувати не можуть, а комп'ютери й мережі без баз даних втрачають свою наукову значимість, залишаючись лише інструментом обчислення і засобом зберігання інформації. Бази даних є невід'ємним елементом функціонування комп'ютерної системи, але при вибірці вони формують результати, глибоко заховані в масивах інформації і ті, які іноді неможливо отримати шляхом прямого перебору. Певним чином закладене в базу даних знання є частиною всього об'єктивно існуючого масиву знань і безпосередньо з нею зв'язане.

Базам даних властиві, в першу чергу, наступні риси:

- швидкість пошуку й одержання інформації;
- економія часу користувачів;
- збільшення кількості інформації з досліджуваної проблеми;
- зростання кількості користувачів через підвищення зручності одержання і обробки інформації;
- робота з інформацією на відстані;
- можливість працювати з однією й тією ж інформацією одночасно великій кількості користувачів.

Використання баз даних в Інтернеті не тільки істотно розширюють пошукові можливості для користувачів мережі, але й забезпечують інтерактивність: можна одержати дані про найбільш популярні теми, організувати коментарі, розширити їхнє інформаційне наповнення. У випадку пошуку літератури, робота у такому режимі дозволяє організувати поряд із алфавітним, тематичний, хронологічний та інші

принци пошуку матеріалів. Технології баз даних покладені в основу пошукових систем, довідкових ресурсів, а на деяких Інтернет ресурсах, вони лежать в основі всієї структури сайту.

У такий спосіб бази даних, безумовно, істотно розширюють можливості дослідників, але вирішення складних завдань можуть забезпечити тільки у випадку обширної інтерактивної взаємодії користувача з сервером, шляхом організації додаткових інтеграційних сервісів. Однак головна цінність баз даних полягає у тому, що саме вони є основною складовою онлайн-аналітичних комплексів, геоінформаційних і експертних систем.

Труднощі, що виникають при розробці і створенні баз даних, не можуть бути успішно вирішені в рамках однієї наукової спеціальності й навіть у рамках всього спектру класичних наук. Необхідна кооперація між фахівцями з інформатики, історії, соціології, географії. Широкий діапазон типів джерел (тексти, малюнки, карти) даних, що входять у бази, вимагає залучення експертів з кожної дисципліни, що передбачає різноманітні підходи до обробки даної інформації, дозволяючи знаходити оптимальні шляхи інтеграції та аналізу візуальних, текстових і картографічних матеріалів.



¹ Горюхов В.М. Особливості розвитку інформаційних баз сучасного українського суспільства. – К., 2005. – С. 18-19.

² Вернадский В. И. Биосфера. – М.: Мысль, 1967. – С. 356.

³ Баград М. Научное управление обществом в аспекте классификации его структурных элементов. – Рига, 1973. – С. 11-12.

⁴ Livari J. Hierarchical spiral model for information system and software development. Part 1: Theoretical background. // Information and Software technology. – 1990, Vol. 3, No. 6. – PP. 386-399.

⁵ Ковальченко И.Д., Устинов В.А. Применение электронных вычислительных машин в исторической науке // Вопросы истории. – 1964. – № 5. – С. 67.

⁶ Поспелов Г.С. Искусственный интеллект – основа новой информационной технологии. – М., 1988. – С. 30.

⁷ Бородин Л.И. Историческая информатика: этапы развития // Новая и новейшая история. – 1997. – № 1. – С. 4.

⁸ Доорн П. Европейские традиции в истории и компьютеризации // Источниковедческие и историографические проблемы истории Украины. – Вып.20. Теория и методы. – Днепропетровск, 1993.

⁹ Липницкая О.Л., Попова Е.Э. Учебно-методический комплекс по исторической информатике. Базовый курс: Обработка и анализ статистических и

структурированных исторических источников средствами электронных таблиц и систем управления базами данных. – Мн.: БГУ, 2002. – С. 13.

¹⁰ Бородкин Л.И. Историческая информатика: этапы развития // Новая и новейшая история. – 1997. – № 1. – С. 4.

¹¹ Героименко В.А., Лазаревич А.А., Титаренко Л.Г. Знание. Компьютер. Общество / Под ред. М.А. Слемнева. – Мн.: Навука і тэхніка, 1992. – С. 51.

¹² Соловяненко Д.В. Проблемы впровадження та розвитку інтернет-проектів у бібліотеках України // Наук. праці Національної бібліотеки України імені В. І. Вернадського. Вип. 14. – К.: НБУВ, 2005.

¹³ Баркова О.В. Перспективи створення національного наукового порталу в мережі Internet на базі Web-сайта НБУВ // Нові обличчя бібліотек та організацій: від надій до партнерства і профусіоналізму: Фокусний семінар. – Б.м. – 2001. – С. 24-28.

¹⁴ Липницкая О.Л., Попова Е.Э. Учебно-методический комплекс по исторической информатике. Базовый курс: Обработка и анализ статистических и структурированных исторических источников средствами электронных таблиц и систем управления базами данных. – Мн.: БГУ, 2002. – С. 11.

¹⁵ Стогний А.А., Глазунов Н.М. Современные проблемы создания интегрированных баз данных / Кибернетика. Становление информатики. – М.: Наука, 1986. – С. 136.

¹⁶ Плахонін А.Г. Електронні проекти Інституту історії України НАН України. // Хроніка та інформація. – № 1. – К., 2005.

¹⁷ За основу було взято тематичний розподіл статей «УІЖ» в бібліографічному довіднику.

¹⁸ Сменцарев Г.В. О представлении и использовании гуманитарных знаний в сети Интернет // Круг идей: Историческая информатика в информационном обществе. Труды VII конференции Ассоциации «История и компьютер». – М.: Мосгорархив, 2001, – С. 328.

¹⁹ Леверман В. Данные и знания в исторических базах данных. История и компьютер: новые информационные технологии в исторических исследованиях и образовании. – М.: МГУ, 1993. – С. 26

²⁰ Greenhalgh M. An interactive text-and-image database for the general public implemented on a low-cost micro computer for the Australian National Gallery. // Computer and the History of Art. – 1991, Vol. 1. – No. 2. – Pp. 26-29.

²¹ Яриц Герхард. К новому образу образов. // Круг идей: Историческая информатика в информационном обществе. Труды VII конференции Ассоциации «История и компьютер». – М.: Мосгорархив, 2001, – С. 212.

²² Бородкин Л.И. Историческая информатика: этапы развития. // Новая и новейшая история. – 1997. – №1. – С. 4.



²³ Стогний А.А., Глазунов Н.М. Назв. праця. – С. 131.

²⁴ Гутнов Д.А. Опасности глобальной информатизации гуманитарной науки. // Историк, источник и Интернет. – Новая и новейшая история. – 2001. – № 2. – С. 68.

²⁵ Стогний А.А., Глазунов Н.М. Назв. праця. – С. 132.

²⁶ Белова Е., Лазарев В. Система управления историческими текстами. // Круг идей: Историческая информатика в информационном обществе. Труды VII конференции Ассоциации «История и компьютер». – М.: Мосгорархив, 2001, – С. 194.

²⁷ Белова Е., Лазарев В. Назв. праця. – С. 194.

²⁸ Сменцарев Г.В. О представлении и использовании гуманитарных знаний в сети Интернет. // Круг идей: Историческая информатика в информационном обществе. Труды VII конференции Ассоциации «История и компьютер». – М.: Мосгорархив, 2001, – С. 327.

²⁹ Интервью с профессором Л.А.Миловым // Информационный бюллетень. - 1992. – № 6. – С. 91-92.

³⁰ Підгаєцький В.В. Основи теорії та методології джерелознавства з історії України ХХ століття. Навч. посіб. – Д.: Вид-во Дніпропетр. ун-ту, 2000. – С. 171.