

использование экспертного знания // История и компьютер/ Тр. Ужгород. междунар. семин. "Нов. инф. технол. в ист. исслед. и образ." (июнь 1992).— М.: Геттинген, 1993.— С.181—184.

8. Гусар-Струк Д. Усне повідомлення на засіданні Науково-методичної ради Інституту біографічних досліджень ЦНБ 24 червня 1994 р.

9. Сторчак В.В. Знакомтесь — MULTIMEDIA // Компьютеры + программы.— 1993.— № 4(5).— С.22—23.

10. Phelan Colleen A. The Dickens House Museum Library: A hypertext prototype//ASLIB Proc.—1992.— Vol. 44, № 9.— P.309—318.

11. Berg Donna Lee. The research potential of the electronic OED database at the University of Waterloo: A case study //Libr. Hi. Tech.— 1991.— Vol. 9, № 4.— P.37—52.

12. Лаврищева Н., Штонда В. CASE — Method. Основные понятия ER-моделирования //SoftReview: Компьютерное обозрение.— 1994.— № 3(5).— С.34—40.

13. Кукуруза П.В. Проблемы проектирования реляционных баз данных //Компьютеры + программы.— 1993.— № 4(5).— С.6—13.

14. Кукуруза П.В. Полнота реляционных систем //Компьютеры + программы.— 1993.— № 2(3).— С.4—9.

15. Кукуруза П.В. Реляционный язык SQL //Компьютеры + программы.—1993.— № 3(4).— С.4—12.

16. Мамешин И.Ю. Знакомство с Oracle //Компьютеры + программы.— 1993.— № 4(5).—С.56—61.

17. Четыре новые версии Paradox //SoftReview: Компьютерное обозрение.— 1994.— № 3(5).— С.30—31.

Брайчевський С.М.

ВИКОРИСТАННЯ ІЕРАРХІЧНИХ ФАЙЛОВИХ СТРУКТУР ДЛЯ ОРГАНІЗАЦІЇ БАЗ ДАНИХ У БІОГРАФІЧНИХ ДОСЛІДЖЕННЯХ

Поява на ринку потужних персональних комп'ютерів з гігабайтовими вінчестерами та високою швидкодійністю уможливила використання в широкій практиці наукових установ сучасних ідей стосовно зберігання та обробки великих масивів інформації, що раніше були доступні лише нечисленним власникам стаціонарних ЕОМ. Йдеться, зокрема, про створення різноманітних баз даних (БД) різного рівня: автономних, призначених для роботи в мережах тощо [1, 2].

Проте, досить швидко з'ясувалося, що дана ситуація породжує цілий ряд нових проблем як технічного, так і принципового характеру [3—4]. Вони пов'язані головним чином з необхідністю забезпечення можливості ефективного обміну інформацією між різними користувачами, в тому числі і електронною поштою. Оскільки такий обмін здійснюється на файловому рівні, проблема організації структури БД набула нетривіального характеру.

В запропонованій статті ми розглянемо деякі аспекти даної проблеми стосовно побудови розвинених БД в галузі біографічних досліджень.

Біографічні дослідження з точки зору розвитку сучасних інформаційних технологій становлять унікальну сферу діяльності. Перш за все, це зумовлюється тим, що вони передбачають комплексну обробку великих масивів даних різного рівня формалізації: від повністю формалізованих записів анкетного типу до повністю неформалізованих текстів (власне біографій). Отже, проблеми, що виникають у зв'язку із створенням БД для біографічних досліджень, належать до широкого спектру, починаючи від суто технічних задач організації зберігання інформації на зовнішніх носіях і закінчуючи концептуальними проблемами гіпертексту.

Крім того, до специфічних особливостей даної галузі слід віднести також те, що атрибути по кожній персоналії розподіляються принаймні на три групи за рівнем загальності:

- 1) загальні атрибути, які мають всі персоналії;
- 2) окремі атрибути, які мають тільки представники певної групи (страги);
- 3) унікальні атрибути, наявні лише у одній конкретній особі.

Дійсно, всі без винятку персоналії мають прізвище, ім'я, дату народження тощо. В якомусь конкретному випадку вони можуть бути невідомими, але об'єктивно вони існують, отже, у структурі записів мають бути зарезервовані відповідні поля. Оскільки такі поля повністю уніфіковані, робота з ними не викликає жодних проблем.

До другої групи належать атрибути, так чи інакше пов'язані із практичною діяльністю особи. На певному рівні структурності вони начебто присутні у всіх персоналії, але в процесі деталізації, як правило, виявляється, що поза межами деякої групи досягти належної уніфікації практично неможливо. Наприклад, вчені видають друковані праці, але не здобувають військових перемог; з іншого боку, композитори, подібно до вчених, друкують свої твори, але характер, а отже і відповідна термінологія цих творів сут-

тево різняться.

Атрибути третьої групи взагалі не вписуються в жодну систематизовану схему.

Отже, в даному випадку сама предметна область виявляється структурованою таким чином, що вибір множини атрибутів для її адекватного відображення є неоднозначною процедурою. Слід також мати на увазі, що в біографічних дослідженнях, як ніде більше, уявлення про відносну значимість тих чи інших інформаційних елементів є суб'єктивними. А якщо ми до того ж врахуємо обсяг БД (кількість персоналій обраховується десятками тисяч, і по кожній з них є дані в середньому на 5–6 ст. друкованого тексту), то стане ясно, що оптимальне розміщення наявної інформації становить складну і відповідальну задачу.

Згадані проблеми (як і ряд суто технічних проблем, які ми тут не розглядатимемо) можуть бути ефективно вирішені за допомогою ієрархічної файлової системи [4], побудованої за наступними принципами:

- головний (керуючий) файл містить у собі, крім універсального ключа, лише повністю формалізовані атрибути першої категорії за рівнем загальності;

- наступний рівень (рівні) ієрархії включає файли, що містять у собі розширені відомості, частково формалізовані і різного ступеня загальності (скажімо, у вигляді мемо-полів);

- далі йдуть спеціалізовані файли, кожен з яких відповідає певній спеціальній групі (страті), а тому має свою власну структуру і містить лише ті персоналії, які належать до даної групи;

- нарешті, останній рівень файлової ієрархії містить текстовий файл (файли) із власне біографічними матеріалами;

- СКБД на програмному рівні встановлює на початку сеансу систему зв'язків між файлами різних рівнів (відповідно до поставленої конкретної задачі), але таким чином, що первинні пошук та обробка інформації здійснюються за допомогою головного файлу, а дані з інших файлів зчитуються лише в міру необхідності.

Неважко побачити, що така файлова система є досить гнучкою і дозволяє будувати розвинені БД будь-якого ступеня складності, забезпечуючи при цьому можливість:

- динамічної модифікації записів як усієї БД в цілому, так і окремих її частин;

- диференційованого захисту різних інформаційних блоків;

- практично вільного обміну цілими файлами з колегами із дружніх установ.

Викладені ідеї використовуються нами при створенні БД Українського біографічного словника (УБС) на ПК 486 DX2-66 в оточенні системи FoxPro/Lan 2. Відповідна файлова структура схематично зображена на Рис.1.

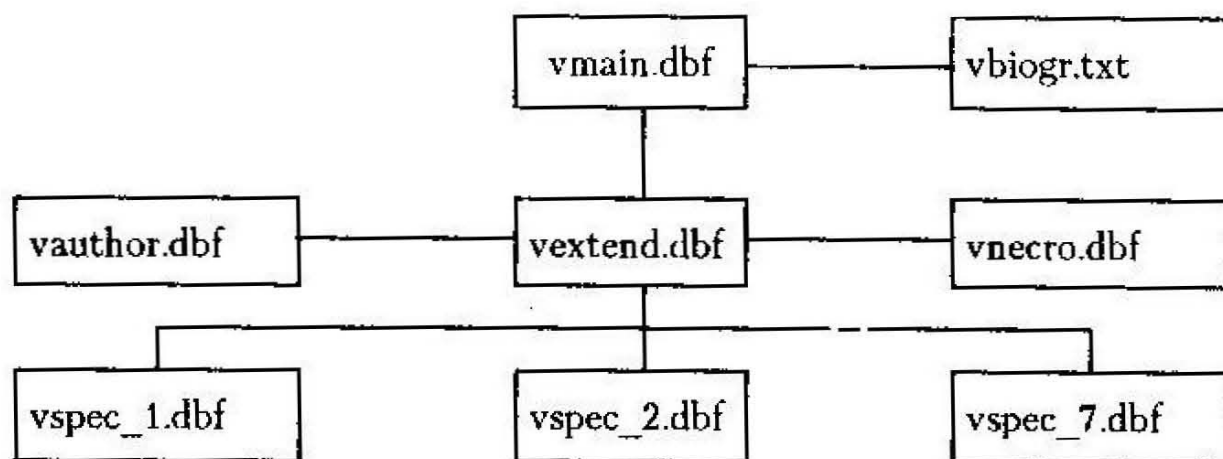


Рис.1

Головний файл (vmain.dbf) містить у собі основні анкетні дані (прізвище, ім'я, псевдонім, роки життя тощо) персоналії; універсальним ключем є прізвище, ім'я та по батькові.

На початку сеансу встановлюється постійний зв'язок між головним файлом та файлом vextend.dbf, в якому знаходяться додаткові відомості загального характеру (місця народження та поховання, учбові заклади, нагороди та відзнаки тощо). Всі подальші операції здійснюються саме через цей файл. На цьому ж рівні знаходяться файли vauthor.dbf та vnecro.dbf, що містять відповідно дані щодо авторів біографічних статей Словника і некрологи.

До третього рівня відноситься набір із 7 файлів vspec_1.dbf, vspec_2.dbf, ..., vspec_7.dbf, що відповідають наявним структурним підрозділам ІБД ЦНБ ім.В.І.Вернадського. Вони займають проміжне місце між даними анкетного типу та власне біографіями, оскільки містять спеціальну інформацію стосовно практичної діяльності людини в соціальному та фаховому аспектах. Кожен із цих файлів має власну структуру записів, що відбиває специфіку конкретного розділу УБС, а також включає в себе лише ті персоналії, які відносяться до нього (звичайно, як виняток, існують персоналії, що входять до кількох таких файлів).

Нарешті, біографії накопичуються в файлі vbiogr.txt, що є звичайним текстовим файлом. Він на даному етапі може оброблятися звичайними текст-процесорами. На Рис.1 зображено зв'я-

зок цього файла з файлом vmain.dbf. Такий зв'язок дійсно проектується, але його реалізація вимагає застосування спеціальних програмних методів, отже, обговорення цієї проблеми виходить за межі пропонованої статті.

Література:

1. Хаббард Дж. Автоматизированное проектирование баз данных.— М., 1984.
2. Мейер Д. Теория реляционных баз данных. — М., 1978.
3. Кукуруза П.В. Полнота реляционных систем // Компьютеры + Программы. — 1993. — № 2(3). — С.4–8.
4. Панченко Б.Е., Гайдабрус В.Н., Церковицкий С.Л. Создание сетевых информационных комплексов // Компьютеры + Программы. — 1993. — № 5(6). — С.46–50.

Кислов В.В., Яценко О.М.

НОВІ ПАРАДИГМИ ІНФОРМАТИКИ ТА ОСНОВИ КОМП'ЮТЕРНОГО ЗАБЕЗПЕЧЕННЯ БІОГРАФІЧНИХ ДОСЛІДЖЕНЬ

Розробка "Концепції програмно-комп'ютерного забезпечення підготовки та видання Українського біографічного словника" проходила в атмосфері постійної дискусії між її виконавцями, керівниками творчих груп Інституту біографічних досліджень, фахівцями-істориками та фахівцями з питань кібернетики та інформатики, програмістами-практиками. Компромісу було досягнуто, але це компроміс "мети і засобів". Дискусія з питань методології практичного втілення у життя основних положень затвердженої концепції триває. І це природньо, оскільки є відображенням суперечностей, які існують і постійно з'являються як наслідок швидкої еволюції суспільних зв'язків, різноманітних за природою та походженням. У нових концепціях менеджменту, комунікаційних та інформаційних технологій бачиться, як завжди, панацея вирішення багатьох суспільних проблем.

Затверджена концепція передбачає формування Баз Знань та перехід до експертних систем на ґрунті центральної і локальних регіональних та галузевих підсистем, які діють у єдиній мережі. Концепція є втіленням ефективного комп'ютерно зорієнтованого досвіду відомої парадигми інформатики, яка розглядає проблеми інформаційних систем (ІС) як тільки технічні, технологічні та внутрішні, вирішення яких можливе лише на базі апелювання до ринку існуючих та нових інформаційних технологій (ІТ). Розроб-