НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
Національна бібліотека України ім. В.І. Вернадського

На правах рукопису

СОХАНЬ Олег Васильович

Методи побудови архівної малтигіпертекстової інформаційно-пошукової системи

Спеціальність 05.25.05 – Системи інформації та обладнання архівів, бібліотек і музеїв

Автореферат дисертації
на здобуття наукового ступеня
кандидата технічних наук

Київ 1996

581.5

ЛННБ України ім.В.Стефаника
00759910 (V)

Дисертацією є рукопис.

Роботу виконано в Національній бібліотеці України ім. В.І. Вернадського.

Наукові керівники: доктор технічних наук, професор
Гриша С. М.
доктор історичних наук
Дубровіна Л. А.

Офіційні опоненти: доктор технічних наук, професор
Корінний О. О.
кандидат технічних наук
Костенко Л. Й.

Провідна організація: Інститут проблем реєстрації інформації НАН України

Захист відбудеться "17" жовтня 1996 р. о 14 годині на засіданні спеціалізованої ради Д 01.31.01 по захисту дисертацій на здобуття наукового ступеня доктора наук (кандидата наук) в Національній бібліотеці України ім. В.І. Вернадського за адресою: 252039, Київ-39, проспект 40-річчя Жовтня, 3.

З дисертацією можна ознайомитися в Національній бібліотеці України ім. В.І. Вернадського (м. Київ).

Автореферат розіслано "10" вересня 1996 р.

Вчений секретар
спеціалізованої ради,
кандидат економічних наук А. О. Чекмарьов

Актуальність теми: Розвиток теорії інформації, поява нових гіпертекстових інформаційних систем і технологій, масове використання аудіовізуальних комп'ютерних систем значною мірою сприяло перегляду концепцій подання інформації.

Оригінальні точки зору на гіпертекст і гіпермедіа, сформульовані В. Бушем[1] в 1945 році, і відповідно – Т. Нельсоном[2] в 1964 році, були занадто широкими та всеосяжними. У відповідності до них в гіпертекстовій технології асоціативні зв'язки повинні охоплювати всю інформацію, до якої будь-хто хотів би мати доступ. При цьому користувачам інформації повинна надаватися достатня свобода вибору навігаційних маршрутів у представленому матеріалі. В противному разі – дані будуть зорганізовані в традиційному лінійному виді, що являє собою попередньо встановлену послідовність елементів інформації, а не набір фактів і ідей, у якій користувач зможе досліджувати варіанти, які становлять для нього певний інтерес.

Але з точки зору реальної практики, проблема дослідження гіперінформаційного простору і визначення в ньому навігаційних маршрутів була покладена на авторів (розробників) гіпертексту. Усталенню цього погляду вирішальним чином сприяла думка про те, що навігаційний маршрут можна розглядати як результат певного інформаційного пошуку або множини пошуків з наступним особливим аранжуванням цих результатів, в якому вибір форми представлення інформації залежить від технології її сприймання деякою групою користувачів. Деякі спеціалісти в галузі гіпертекстових технологій стверджують, що дослідники частіше

---

1 *Bush V.* As May Think? // Atlantic Monthly. – 1945. Vol. 176, № 1. P. 101-109.

2 *Нельсон Т.* Информационные системы будущего // Информационный поиск: Сб. материалов. – М.: Воениздат, 1970. С. 217-228.

бажають, щоб їх направляли, не даючи відхилятися занадто далеко від обраного навігаційного маршруту[3]. Але при цьому виникають такі проблеми:

1. Розробники гіпертекстових документів повинні приймати до уваги ступінь знання предметної галузі і власного контролю за використанням інформації.

2. Гіпертекст, що є зручним для однієї групи користувачів, може бути незручним для іншої або мати опосередковану структуру і бути за якимись ознаками однаково незручним для всіх груп користувачів.

3. Класичний гіпертекст дає певну множину інформаційних траєкторій, але суттєво підкреслити, що ця множина траєкторій є підмножиною повної множини траєкторій. До того ж остання має експоненціальну потужність.

Інші спеціалісти виступають проти жорсткого визначення навігаційних маршрутів. Наприклад, М. Бернстайн стверджує, що "ефективність побудови гіпертексту залеже від напруженого стану між систематизацією і величезною кількістю інформації, між передбачуваністю і невизначеністю"[4].

Без перебільшення можна стверджувати, що останнє десятиріччя розвитку світової інформатики проходить під знаком розвитку глобальних комп'ютерних мереж Internet і найбільш популярного ресурсу Word-Wide Web (WWW), який базується на технології гіпертекстів. Активна робота з гіпертекстовими

---

[3] *Grice R.A.* Linking to hypertext: A hypercritical analysis // Proceedings of the 37th International Technical Communication Conference (May 1990), P. RT53-RT54. Washington, DC: Society for Technical Communication.

[4] *Bernstein M.* Deeply intermingled hypertext: The navigation problem reconsidered // Technical Communication. – 1991. – Vol. 38, № 1. – P. 41-47, Washington, DC: Society for Technical Communication.

технологіями в глобальних мережах Internet приводить до їх удосконалення як засобу подання інформації, основні напрями розвитку яких сформульовані М.М. Субботіним (Російський державний науково-технічний центр гіперінформаційних технологій)[5]. Це, у свою чергу, висуває перед спеціалістами в галузі гіпертекстових технологій грунтовні завдання, від вдалого розв'язання яких залежить якість швидкого і ефективного пошуку та доступу до інформації в глобальних мережах, що зробить миттєвим анонсування нових унікальних інформаційних ресурсів.

Об'єкт дослідження: Об'єктом дослідження є процеси задоволення інформаційних потреб користувачів-дослідників, які опрацьовують великі масиви архівної та бібліотечної інформації в ході наукових розробок.

Гіперінформаційні технології в цьому відношенні виступають ефективним засобом, які дозволяють розробку таких технологій обробки інформації, в яких наукові абстракції, навіть найвищого рівня складності, можливо трансформувати в реально відчутні та зримо сприйнятливі образи, що дають можливість користувачу комп'ютера при певних умовах визначити нові низки іманентного значення.

Предмет дослідження: Предметом дослідження є методи оптимального та адаптивного візуального подання даних у архівно-бібліотечних інформаційно-пошукових системах (ІПС) та реалізація швидкого і ефективного пошуку потрібної інформації як безпосередньо даних, так і конкретних фактів, що стосуються даних.

---

5 *Субботин М.М.* Эвристический эффект гипертекста // НТИ. Сер. 1. – 1994. – № 4. С. 9-14.

Завдання дослідження полягає у створенні комп'ютерних технологій інтелектуалізованого інформаційного забезпечення користувачів-дослідників, а саме введення елементів поточного управління структурою гіпертексту до складу засобів кінцевого користувача.

Конкретні завдання дисертації полягали в тому, щоб:

1. Обгрунтувати ефективність застосування гіпер-інформаційних технологій в архівних і бібліотечних інформаційно-пошукових системах.

2. Виконати аналіз класичних гіпертекстових технологій та визначити наявність цілого ряду інформаційних проблем, пов'язаних з цими технологіями, які не позбавлені недоліків.

3. Довести актуальність розробки нової технології оптимального та адаптивного гіпертекстового подання інформації.

4. Запропоновувати альтернативну (відмінну від класичної) концепцію гіпертекстового подання інформації з можливістю управління структурою гіпертексту з боку користувача та виконати дослідження її властивостей.

5. Застосувати певну технологію структуризації даних для динамічної побудови гіпертекстових структур великих обсягів інформації.

6. Побудувати моделі динамічного синтезу гіпер-інформаційного простору та розробити алгоритми адаптивної організації індексної інфраструктури, яка не потребує прямої відповідності "індекс – запит".

7. Розробити архітектуру архівної інформаційно-пошукової системи, в якій передбачено компоненти самоорганізації даних.

Метою дисертаційної роботи є підсилення інтелектуального потенціалу дослідників, які працюють із значними масивами архівно-бібліотечних інформаційних ресурсів, що входять до Національного архівного фонду України, а також перехід від звичного пошуку необхідної інформації до стратегічного використання інформаційних ресурсів.

Методи дослідження: Теорія систем, об'єктно-орієнтована методологія, дискретна оптимізація, теорія баз даних та теорія штучного інтелекту.

Наукова новизна дослідження:

1. Обгрунтовано ефективність застосування сучасних гіпертекстових технологій в архівній і бібліотечній справі та виявлені фактори їх обмеженості.

2. Запропоновано концепцію малтигіпертекстового подання інформації в інформаційно-пошукових системах та вперше виконано дослідження її властивостей.

3. Застосовано асоціативну технологію структуризації даних для динамічної побудови гіпертекстових структур великих обсягів інформації[6].

4. Побудовано моделі динамічного синтезу гіперпростору.

5. Розроблено оригінальні алгоритми адаптивної організації індексів, які не потребують прямої відповідності "індекс – запит", для підтримки процесів динамічної побудови гіпертекстів.

6. Розроблено архітектуру архівної інформаційно-пошукової системи на основі малтигіпертекстової технології, в якій вперше передбачено компоненти самоорганізації даних.

---

[6] Таким чином, досягається ефективність наявності цілої множини гіпертекстів у розпорядженні користувача, тобто є підстави говорити про малтигіпертекст (множинний гіпертекст).

Практична цінність результатів, одержаних у роботі, полягає в тому, що зазначені ідеї були використані при розробці алгоритмічного та програмного забезпечення інтелектуалізованого інтерфейсу користувача в рамках проекту "Розробка програмного забезпечення підтримки Національної архівної інформаційної системи (НАІС) – клієнтська частина", що впроваджено в дослідну експлуатацію в Інституті української археографії та джерелознавства ім. М.С. Грушевського НАН України. Передбачається використання здобутих наукових результатів при впровадженні в промислову експлуатацію проекту "Національна архівна інформаційна система (НАІС)" в установах Головного архівного управління при Кабінеті Міністрів України.

Реалізація результатів роботи виконувалась у рамках програми 8.4.2-92 "Архівна та рукописна україніка" Державного комітету України з питань науки та технологій за темою "Архівна та рукописна україніка: Національна зведена система документальної інформації" (напрям Національної архівної інформаційної системи), а також у рамках проекту створення комп'ютерного довідника "Археобібліобаза" як складової "Архівної та рукописної україніки" та ін.

Апробація результатів роботи: Основні положення та результати дослідження були викладені на Міжнародному семінарі "Проблеми повернення культурних цінностей в Україну" (Чернігів, 1994); Всеукраїнській нараді "Актуальні проблеми розвитку архівної справи" (Київ, 1995); Міжнародній нараді "Гіпертекстові технології і міжнародні комп'ютерні мережі Internet" (Бібліотека конгресу США, Вашингтон, США, 1995); Міжнародному науковому семінарі "Археобібліобаза (АББ): проблеми загально-

доступності українських архівних матеріалів" (Український науковий інститут Гарвардського університету, Кембридж, США, 1995).

На захист виносяться такі положення:

1. На основі аналізу класичних гіпертекстових технологій пропонується концепція малтигіпертекстового подання інформації.
2. Запропонований метод асоціативної організації сховища даних для малтигіпертекстової технології.
3. Побудовані моделі оцінки реактивності малтигіпертекстової інформаційно-пошукової системи.
4. Запропоновані алгоритми самоорганізації індексної підтримки малтигіпертекстової технології.
5. Розроблена архітектура асоціативної технології.

Структура і обсяг роботи: Дисертаційна робота складається з вступу, чотирьох розділів, висновків, переліку використаної літератури і додатку.

Перший розділ: Сформульовані сучасні вимоги до використання архівних інформаційних ресурсів; зроблено огляд проблем інформаційного пошуку в класичних архівно-бібліотечних інформаційно-пошукових системах; обгрунтовано ефективність застосування сучасних гіпертекстових технологій в архівній та бібліотечній справі та виявлені фактори їх обмеженості. Визначені та обгрунтовані завдання дослідження, показано їх місце у загальному колі питань, що вирішують проблеми усунення обмежень на пошукову навігацію в інформаційному просторі та вирішення проблеми релевантності інформаційного пошуку в ІПС.

Другий розділ: Зроблений опис основних властивостей класу "малтигіпертекст" та показані основні елементи малтигіпертекстової архітектури; визначено технологію формування ієрархічних сегментів (агрегатів) малтигіпертексту; описано

архітектуру асоціативної бази даних як основи малтигіпертекстової технології; сформульовані моделі оцінки ефективності використання ресурсу пам'яті і моделі оцінки ефективності швидкості пошуку для асоціативної бази даних (АсБД).

Третій розділ: Охарактеризовані проблеми оптимального асоціювання для реалізації динамічної побудови малтигіпертекстів; розроблено модель організації асоціатора АсБД; описано алгоритмічну схему вибору плану побудови асоціатора; зроблено опис сімейств планів побудови асоціатора та доведено обгрунтування повноти моделі опису сімейств; розроблено механізм вибору сімейства планів асоціатора для розгалуження, і виконані оцінки якості сімейств.

Четвертий розділ: Запропоновано архітектуру асоціативної технології для реалізації пошуку в інтелектуальній архівній інформаційно-пошуковій системі з малтигіпертекстовим інтерфейсом користувача; наведено і описано сукупність компонентів асоціативної технології, що дозволяють реалізувати ефективний пошук та генерацію гіпертекстових структур у відповідності до творчої поведінки користувача.

ЗМІСТ РОБОТИ

У першому розділі зазначено, що робота пов'язана з проблемами розробки інформаційно-пошукових систем на основі гіпертекстових технологій, в яких закладена певна множина напрямів використання інформаційних ресурсів шляхом моделювання процесів творчого сприйняття користувача. Є очевидним, що гіпертекст, який зорієнтований на деяку групу користувачів інформації та має асоціативні зв'язки з іншими фрагментами інформації, позбавляє їх необхідності користування

складними і не завжди якісними можливостями класичних інформаційно-пошукових систем, і це є головною метою створення гіпертексту. Але класичні гіпертекстові технології – суб'єктивні, один і той самий гіпертекст за своєю структурою не може однаково добре задовольняти різні категорії дослідників-користувачів інформації. Щоб зробити гіпертекстовий матеріал корисним для дослідника, автор (розробник) гіпертексту повинен ретельно продумати, як встановити зв'язки, що поєднують різні частини даних інформаційного простору, і як користувачі будуть переходити від однієї частини інформації до іншої. Яким чином розробники можуть знайти правильне рішення? Відповідь на це питання залежить від ряду факторів: природи поданого матеріалу, аудиторії, для якої цей матеріал подається, і мети його використання. Очевидно, що актуальність інформації можна збільшити, якщо дозволити надати користувачеві-дослідникові певну свободу вибору, коли він сам може контролювати спосіб подання йому інформації. Іншими словами, інтерактивні зв'язки форми гіпермедіа мають успіх тому, що користувачі, вільно виконуючи навігацію у гіперінформаційному просторі, можуть самостійно вирішувати, що вони бажають і в якому порядку.

У другому розділі, як наслідок проведеного аналізу існуючих гіпертекстових технологій, було запропоновано концепцію малтигіпертексту (множинного гіпертексту), що є технологією візуального подання інформації, яка може зберігатись у формах, відмінних від класичних гіпертекстів. Такий гіпертекст має назву "***малтигіпертекст***" у тому розумінні, що він фактично є сімейством різних гіпертекстів, заснованих на єдиному джерелі інформації. Малтигіпертекстовий підхід як технологія – заснований на обробці глибоко об'єктизованого інформаційного

простору, а не на обробці лінійних блоків текстових баз даних з подальшим синтезом з них гіпермережі, що суттєво різнить його від інших підходів. Фактично мова йде про малтигіпермережу або віртуальну гіпермережу, яка регенерується кожний раз відповідно до поведінки мислення користувача у відповідному k-му місці об'єктизованого інформаційного простору предметної галузі, що складається з різноманітних інформаційних об'єктів.

Отже, малтигіпертекст – це сукупність об'єктизованого інформаційного простору і асоціативного навігатора:

МНТ = {I, AsN},

де: $I = \{i_{00}, \ldots, i_{nk}\}$ – об'єктизований інформаційний простір; i_{nk} – n-інформаційний об'єкт k-го класу (наприклад, об'єкти, що визначають: Назви архівних інституцій класу "*Архівосховище*"; Назви і/або Номери архівних фондів класу "*Архівні фонди*"; Прізвище або Назва установи класу "*Фондоутворювачі*"; Архівний опис класу "*Обліково-пошукова документація*"; Путівники і/або довідники класу "*Бібліографія*" та ін.).

AsN – асоціативний навігатор, що є об'єктно-орієнтованим механізмом, який дозволяє з об'єктизованої структури інформаційного масиву синтезувати n-рівневі гіпертекстові структури, що побудовані на відношеннях входження і наслідування.

Основними типами навігаційних компонентів у малтигіпертексті слід вважати:

1. *Горизонтальні асоціації* (асоціації з іншим фактом на подібні теми). Це так звані спектральні зв'язки, які визначають рух між компонентами одного рівня ієрархії гіпертекстової мережі.

2. *Асоціації по ієрархії входження*. Це асоціації до нижчестоячого і вищестоячого об'єктів.

3. *Асоціації по спорідненості структур і спорідненості значень.* У даному випадку спорідненість вживається в контексті успадкування.

4. *Немарковські асоціації.* Це асоціації, котрі інтерпретуються в контексті маршруту навігації в гіпертекстовій мережі.

5. *Асоціації по підструктурам гіпертекстової мережі.* Це так звані асоціації до інших рубрик поточного факту.

Виходячи з того, що для динамічного генерування малтигіпертекстової мережі потрібно мати деяку низку програмних засобів, доцільною є асоціативна технологія структуризації даних, яка б моделювала процеси предметного середовища і забезпечувала синтез малтигіпертекстів.

Асоціативна база даних первісно була задумана як платформа для CASE-технологій в частині збереження даних в процесі їх нормалізації. Проблема полягає в тому, що для конструювання нормалізованих даних необхідно спочатку зібрати інформацію про загальні властивості предметного середовища. Це можливо зробити лише в процесі комп'ютерної обробки даних. Таким чином, для класичної технології виникає парадокс, а саме — щоб спроектувати базу, треба мати базу. Класична теорія нормалізації відносин Е. Кодда декомпозує опис об'єкта на змістовному рівні і дозволяє виділити нові відношення між даними, що дають можливість уникнути аномалій, мінімізувати дублювання інформації і спростити процедури обробки та оновлення. Тобто необхідно у кожному випадку вирішити питання про розміщення атрибутів (рубрик) або фрагментів змісту у кількох відношеннях (реляційних таблицях). Асоціативна база даних вирішує ці протиріччя шляхом параметризації структурних рішень. Іншими словами, індивідуальні запити дослідника щодо бажаної структури гіпертекстового

подання можливо трактувати як зміни параметрів предметного середовища. Тобто асоціативний підхід досягає тієї ж мети шляхом декомпозиції за структурною ознакою, де структурні рішення (логічна та фізична структура) подаються як частина даних. Структурні ознаки більш стабільні.

У цьому випадку структура інформаційного опису включає такі об'єкти: значення; поняття; рубрики; факти; їх асоціації.

Для створення асоціацій всі об'єкти ефективно кодуються за принципом:

– поняття складаються з рубрик;

– рубрики складаються із значень;

– факти - це відношення кількох значень рубрик.

У загальному випадку структура АсБД складається з довідника рубрик, що виконує функцію класифікатора рубрик; входження рубрик (ВР), що виконує функцію відображення структури складних рубрик; довідника значень (ДЗ), що виконує функцію глобального словника системи; асоціатора (Ас), що виконує функцію збереження асоціативних зв'язків семантичних фрагментів з різними позиціями різних фактів; довідника місць (ДМ), що є допоміжною структурою, потрібною для зменшення розмірів Ас, яка кожній парі *"код зв'язку – код рубрики"*, що відображає певний шлях на дереві рубрик, ставить у відповідність "код місця".

Алгоритм пошуку у АсБД заснований на маніпулюванні множинами асоціацій. Результатом пошуку по елементу замовлення є множина кодів фактів, які асоціюються з даним значенням. Перетнувши згадані множини для всіх елементів запиту, отримуємо множину релевантних інформаційних об'єктів.

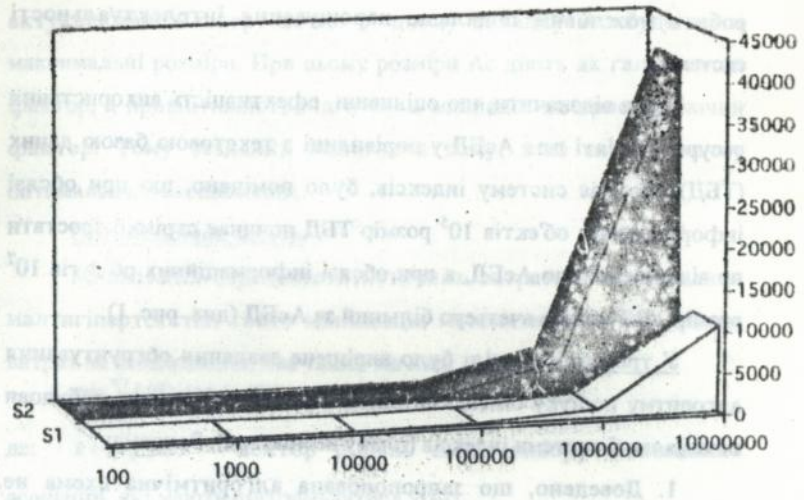

Рис. 1. Залежність розміру бази даних від потужності множини об'єктів для асоціативної БД (S1) і для текстової БД з системою індексів (S2) (у кілобайтах).

Помітним стає ряд побічних переваг такої структури, а саме:

– швидка побудова меню вибору значень як при заповненні описів, так і при формуванні запитів;

– виконання пошуку на фоні запиту з метою оперативної перевірки коректності останнього;

– розпаралелювання процесу пошуку шляхом виділення серверів для різних компонентів архітектури (ДР, Ас, ДЗ або навіть частини цих елементів);

– пристосування структури Ас до статистичних параметрів потоку запитів;

– ефективне ущільнення інформації шляхом кодування слів та часто вживаних фраз;

– близькість запропонованої технології пошуку до механізмів логічних висновків, вживаних у системах штучного інтелекту, що

робить можливим подальше нарощування інтелектуальності системи.

Треба відзначити, що оцінивши ефективність використання ресурсу пам'яті для АсБД у порівнянні з текстовою базою даних (ТБД), що має систему індексів, було помічено, що при обсязі інформаційних об'єктів 10^5 розмір ТБД починає стрімко зростати по відношенню до АсБД, а при обсязі інформаційних об'єктів 10^7 розмір ТБД майже вчетверо більший за АсБД (див. рис. 1).

У третьому розділі було вирішене завдання обгрунтування алгоритму пошуку сімейства планів асоціатора з метою побудови оптимальної системи індексів (плану асоціатора). Зокрема:

1. Доведено, що запропонована алгоритмічна схема не повертається до розгляду варіантів, що були проаналізовані на попередніх ітераціях.

2. Доведено, що не існує таких варіантів плану асоціатору (плану системи індексів), який в разі оптимальності не був вибраним. Тобто алгоритмічна схема не втрачає варіантів.

3. Числові експерименти довели, що для отримання досить ефективних планів складність пошуку не перевищує складності аналізу 30 відсотків інформації.

Навігація у гіперінформаційному просторі в малти-гіпертекстовій технології супроводжується виконанням асоціацій, де ефективність обслуговування запитів залежить від того, наскільки план Ас відповідає структурам множини гіпертекстів. Можливо визначити два таких крайніх випадки:

1. У плані Ас передбачені тільки одиночні асоціації (тільки по одній рубриці). В цьому випадку всі множини асоціації обробляються як перетин одиночних асоціацій.

2. У плані Ас підтримуються всі множини асоціацій для всіх актуалізованих гіпертекстів. У цьому випадку Ас буде мати максимальні розміри. При цьому розміри Ас діють як гальмуючий фактор, а привінтивність підготовки асоціацій як прискорюючий фактор. Тому завдання полягає в тому, щоб знайти точку оптимального асоціювання.

Оптимізаційна задача:

Мінімізація середньостатистичних витрат на асоціювання малтигіпертекстів, тобто мінімізація математичного очікування витрат на асоціювання, має такий вигляд:

$$\min_{\bar{x}} \sum_{j \in J} l_j(\bar{x}) \cdot \log_n(m(\bar{x})) \cdot \omega_j ,$$

де: $\bar{x} = \{x_j\}, j \in J$ – вектор рішень щодо вибору множинних асоціацій, які мають підтримуватися у Ас;

$L_j(\bar{x})$ – функція витрат на одне елементарне асоціювання j-го типу (J - множина типів);

$m(\bar{x})$ – функція розміру Ас залежно від плану $\bar{x}$;

$\log_n(m(\bar{x}))$ – витрати на пошук у індексі типу бінарного дерева;

ω_j – статистична оцінка математичного очікування елементарних асоціацій j-го типу протягом одного циклу експлуатації системи (рік або квартал);

Функцію $l_j(\bar{x})$ можливо визначити за такою формулою:

$$l_j(\bar{x}) = \sum_{i \in I} z_{ij} ,$$

де: $z_{ij} \in \{0,1\}$ – рішення про використання i-го масива для асоціацій j-го типу.

Треба відзначити, що:

$$z_{ij} \le a_{ij} ,$$

де: $a_{ij} \in \{0,1\}$ – характеристика i-го масиву відносно j-го запиту.

Тобто 0 – коли $val(i) \not\subset val(j)$, де $val()$ –множина рубрик, що входять до даного Ас.

Дане обмеження заперечує використання Ac, якщо він не є підмножиною асоціацій.

$$z_{ij} \le x_i, \forall j \in J \qquad (*)$$

Асоціація може бути використаною, якщо вона передбачена у плані Ac.

$$m(\bar{x}) = \sum_{i \in I} x_i m_i$$

Введемо таке обмеження:

$$b_{jr} = \begin{cases} 0 \\ 1 \end{cases},$$

де: 1 – якщо r-та рубрика включена в j-ий топік[7];

0 – в протилежному випадку.

$$a_{ir} = \begin{cases} 0 \\ 1 \end{cases},$$

де: 1 – якщо r-та рубрика включена в i-ту асоціацію;

0 – в протилежному випадку.

Відповідно можливо записати таке обмеження:

$$\sum_{i \in I} z_{ij} a_{ij} \ge b_{ir}, \forall j, r$$

Останнє обмеження потрібно, щоб для кожного запиту був обраний саме повний набір асоціацій. З іншого боку, через обмеження (*) це гарантує повноту плану Ac $\bar{x}$. Повнота полягає в тому, що для кожного запиту множини J можливий набір повної множини асоціацій I.

Як і будь-яка непроцедурна система, малтигіпертекст потребує автоматичного вирішення завдання підтримки структури. Ця задача має аналогію з задачею, яка розв'язується в SQL-сервері

---

[7] Реально малтигіпертекст складається із топіків (тем), які будуються динамічно на відміну від класичного гіпертексту. Топік – послідовний лінійний відрізок тексту, який зберігається як одна одиниця або його адреса, що віднесена у індекс. Найбільш важливою ознакою топіку є те, що він є лінійною структурою. Для того, щоб побудувати принципово новий гіпертекст з фактологічного матеріалу інформаційного простору, у гіршому випадку необхідно з підмножин різних топіків побудувати нові топіки.

стосовно оптимізації планів запитів. Тобто в системах подібного типу потрібен інтелектуальний агент, який би розв'язував ці задачі і знімав їх з користувача.

Враховуючи надзвичайну комбінаторну складність задачі, що визначена на множині підмножин, застосування універсальної стратегії висновків (метод резолюцій) для даного інтелектуального агента є неприйнятливим. Тому в даному випадку застосовується спеціалізована, цілеспрямована стратегія на основі методу "гілок та границь" (branch and bound).

Сімейство планів описується як система обмежень на елементи, що можуть входити до складу Ac. Опис сімейства планів можливо подати в вигляді такої моделі:

< I, J, K >,

де: **I** – обмеження включення, яке несе в собі елементи асоціатора, що обов'язково присутні в плані асоціатора;

J – обмеження виключення, що описує елементи асоціатора, які не будуть включатися в даному сімействі планів;

K – обмеження осуспільнення, що передбачає опис множин кінцевих асоціацій, які мають обслуговуватися спільним набором асоціацій.

Якщо до ІПС асоціативного типу надходить деякий запит на обслуговування, то спочатку в плані Ac визначаються множинні асоціації a(1), ... , a(n), що покривають даний запит.

Але при цьому виникає аномалія, що певні елементарні асоціації відповідної множинної асоціації, які покривають сферу запиту, присутні в іншій множинній асоціації. Для усунення цього виникає потреба в осуспільненні виявлених множинних асоціацій і синтезу спільного набору асоціацій b(1), ... b(n) з виключенням повтору елементарних асоціацій (див. рис. 2).

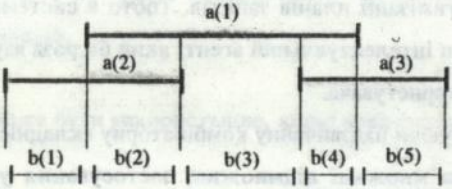

Рис. 2. Утворення спільного набору асоціацій.

Тобто, спільні асоціації утворюються як базис системи множин (підмножини, що неперехрещуються і з яких складаються інші множини). З точки зору швидкості доступу, якщо певна множинна асоціація є базовою по відношенню до всіх кінцевих запитів (тобто мається на увазі, що дана асоціація або цілковито входить до кінцевого запиту, або не має з ним ніякого перетину), то її подальша декомпозиція не доцільна.

Визначення 1: Сімейство називають закритим, якщо не існує способу його розподілу на сімейства, що мають кращі оцінки.

Визначення 2: Сімейство називають виродженим, якщо, в разі врахування всіх обмежень, воно подає єдиний план асоціатору.

Якщо сімейство невироджене і незакрите, то його можливо розбити на два підсімейства. Вибір підсімейства для розподілу має здійснюватися з урахуванням таких ознак:

1) верхньої оцінки якості сімейства;

2) нижньої оцінки якості сімейства;

3) міри різноманітності, котра визначається кількістю варіантів, що містяться в сімействі.

Оцінка верхньої границі сімейства – це величина, яку не може перевищувати максимальна якість плану даного сімейства. Оцінка нижньої границі сімейства – це величина, про яку відомо, що найкращий план сімейства буде не гіршим. Тобто верхня границя – це ідеальна (оптимістична) оцінка, а нижня – песимістична. В свою

чергу, верхню границю можливо оцінити, перетворивши задачу у спеціальну задачу, що має нескладний комбінаторний розв'язок, а нижню границю можливо отримати за допомогою евристичного алгоритму відкидання різноманіття сімейств.

Визначення 3 (Правило відкидання сімейств): Якщо верхня границя сімейства А менша нижньої границі сімейства В, то сімейство В треба виключити з розгляду.

Очевидно, оптимістичну оцінку сімейства можливо оцінити за таким алгоритмом:

I. Взяти розмір Ас найменшим при даних обмеженнях сімейств.

II. Кількість елементарних пошуків, що визначаються структурою Ас, також взяти найменшим.

Останнє має місце в тому випадку, коли підтримуються множинні асоціації, що відповідають кінцевим запитам у всіх випадках, що не суперечать обмеженням сімейств.

Але при цьому залишається відкритою задача визначення мінімального розміру Ас. На думку автора, мінімальний розмір асоціатора можливо оцінити так:

1. Виконати остаточну базисну декомпозицію.

2. Якщо вона містить множинні асоціації, то кожну таку асоціацію треба розбити на дві асоціації, якщо сумарний розмір отриманих буде меншим за розмір первинної асоціації.

Це завдання хоч і має квадратичну трудомісткість, але враховуючи те, що вона застосовується до підмножин конкретних кінцевих запитів, комбінаторно вона є цілком сприйнятливою.

Отже, ІПС набуває характеру інтелектуального агента в тому розумінні, що вона, спостерігаючи за умовами роботи та оцінюючи ефективність згідно із закладеними критеріями, приймає рішення про перебудову своєї структури.

У четвертому розділі розроблена асоціативна технологія формування запиту в архівній малтигіпертекстовій ІПС, що наведена на рис. 3.

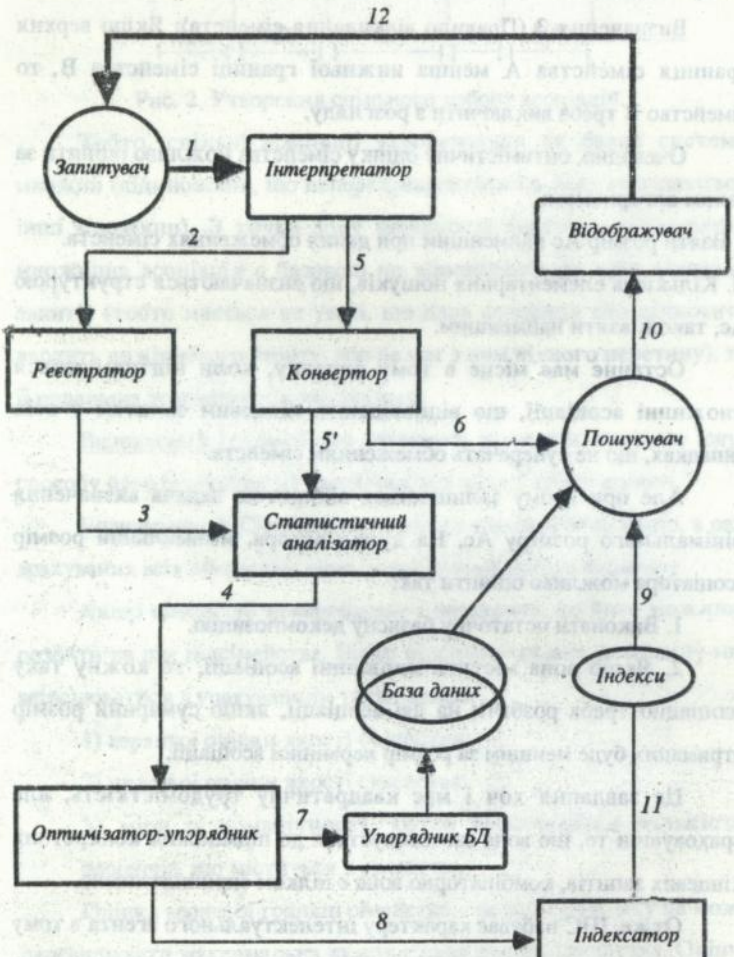

Рис. 3. Асоціативна технологія побудови архівної ІПС.

Направлені зв'язки між блоками означають:

1 - запит на інформацію у зовнішній мовній формі;

2 - індентифікуючі реквізити запиту, потрібні для його відображення в реєстрі;

3 - інформація для статистичного аналізу потоку запитів за встановлений проміжок часу;

4 - інформація для аналізу відповідності поточної структури БД та структури індексів потоку запитів, що склався на даний момент;

5 - запит, насичений контекстною інформацією із позаоб'єктної бази знань;

5' - запит, конвертований у внутрішні коди понять (індекси);

6 - запит, розкладений у вигляді потоку елементарних (канонічних) запитів для безпосереднього виконання пошукувачем;

7 - замовлення на зміну структури зберігання даних для її більшої відповідності характеристикам потоку запитів, що склався на даний момент;

8 - замовлення на зміну структури індексів для її більшої відповідності характеристикам потоку запитів, що склався на даний момент;

9 - інформація, подана у внутрішньому кодовому форматі для пошуку;

10 - інформація, подана у зовнішніх мовних формах для аналізу запитувачем;

11 - індексна інформація для ефективного звуження зони пошуку;

12 - звіт про пошук у вигляді гіпертексту, поданий за формою і змістом, замовленими запитувачем.

ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

1. Доведено існування цілого ряду інформаційних проблем, пов'язаних з класичними гіпертекстовими технологіями, які в повній мірі не забезпечують дослідників релевантною інформацією через їх фіксовану спрямованість на певну групу користувачів, які генерують опосередковану структуру запитів.

2. Запропоновано ідею малтигіпертексту як технологію візуального подання інформації, що може зберігатися у формах, відмінних від класичних гіпертекстів. Такий гіпертекст має назву "*малтигіпертекст*" у тому розумінні, що він фактично є сімейством різних гіпертекстів, заснованих на єдиному джерелі інформації, де користувач може змінювати структуру гіпертексту згідно з своїми поточними проблемами.

3. Запропоновано метод внутрішнього подання даних, що заснований на асоціативній технології структуризації даних, для динамічної побудови гіпертекстових структур.

4. Розроблені математичні моделі, що дозволяють оцінити ефективність динамічного синтезу гіпертекстів на платформі SQL-серверу, що працює з базою даних, побудованою за асоціативним принципом.

5. Доведено можливість ефективного керування процесом конструювання гіпертекстів шляхом організації індексної інфраструктури.

6. Запропоновано алгоритми автоматизованого налагодження індексної інфраструктури на структуру запитів користувачів, що можуть бути покладені в основу інтелектуальних агентів – упорядників "date warehouse".

ОСНОВНІ ПОЛОЖЕННЯ ДИСЕРТАЦІЇ ВИКЛАДЕНІ В ТАКИХ ПУБЛІКАЦІЯХ:

1. Гіпертекстові системи та технології (Перспективність використання їх в архівних інформаційних системах) // Національна архівна інформаційна система "Архівна та рукописна україніка" і комп'ютеризація архівної справи в Україні: Зб. наук. праць. – К., 1996. Вип. 1: Інформатизація архівної справи в Україні: Сучасний стан та перспективи. – С. 88-97.

2. Програмно-технологічні аспекти розбудови Національної архівної інформаційної системи (Асоціативний підхід) // Там само. (У співавторстві з Гришою С.М.). – С. 72-87.

3. Національна архівна інформаційна система: технічне завдання // Там само. (У співавторстві з Гришою С.М., Галяпою А.А., Пийтером М.В., Кіржаєвим С.М., Новохатським К.Є., Захарченко Т.М.). – С. 196-224.

4. Гіпертекстові технології і глобальні мережі Internet. Чи це потрібно сучасному архіву? // Актуальні проблеми розвитку архівної справи в Україні: Матеріали наради. Березень 1995 р. – К., 1996. – С. 135-137.

5. Національна архівна інформаційна система: структура даних (Матеріали для обговорення) // Упор. Сохань О.В. – К., 1994.– (У співавторстві з Дубровіною Л.А.). – 46 с.

6. Національна архівна інформаційна система та її місце у реституції культурної спадщини України : Матеріали Міжнародного семінару "Проблеми повернення культурних цінностей в Україну". Чернігів. Вересень 1994 р. – К., 1996. – С. 270-271.

Сохань О. В. Методы построения архивной малтигипертекстовой информационно-поисковой системы.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.25.05. – "Системы информации и оборудования архивов, библиотек и музеев". Национальная библиотека Украины им. В.И. Вернадского. Киев, 1996.

Предложена концепция малтигипертекстового представления информации и выполнено исследование её свойств. Применена ассоциативная технология структуризации данных для динамического построения гипертекстовых структур. Исследованы модели динамического построения гиперпространства. Разработаны алгоритмы управления системой индексной поддержки процессов динамического построения гипертекстов. Разработана архитектура информационно-поисковой системы, которая базируется на малтигипертекстовой технологии.

Sokhan O.V. Methods of forming archival multi-hypertext information searching systems.

Dissertation submitted for the granting of the academic degree Candidate in Technical Sciences in the specialization 05.25.05. "Information Systems and Equipment of Archives, Libraries and Museums". V.I. Vernadsky National Library of Ukraine, Kiev, 1996.

The thesis proposes a concept of multi-hypertext presentation of information and presents research on its characteristics. Associative technology for the structuralization of data is applied for the dynamic construction of hypertext structures. Studied are models of the dynamic construction of hyperfields. Algorithms for the control of indexing support of dynamic construction of hypertexts are developed. Also developed is an information searching systems architecture, which is based on multi-hypertext technology.

Ключові слова: гіпертекстові системи і технології, системи обробки інформації, інформаційно-пошукові системи, бази даних, архіви та бібліотеки.

Підписано до друку 07.09.96 Формат 60х84 1/16
Ум.друк.арк. 0,9 Обл.вид.арк. 1,0
Тираж 100. Зам.364 1996р. Безплатний
Поліграф.д-ця Ін-ту історії України НАН України
Київ-1, Грушевського, 4

U3+529

AB 35.539