

МІНІСТЕРСТВО НАУКИ І ОСВІТИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

М'якшило О.М.

«Організація баз даних та знань»
Тлумачний словник

Всі цитати, цифровий та фактичний матеріал, бібліографічні відомості перевірені. Написання одиниць відповідає стандартам

Підпис(и) автора(ів) _____
«____» _____ 200__ р.

Київ НУХТ 2011

УДК 681.3

Рецензент: канд.. техн.наук, доцент **О.А. Хлобистова**

О.М. М'якшило Організація баз даних та знань: Тлумачний словник. ISBN - К. НУХТ, 2010. - 22 с.

Наведено найбільш уживані терміни та поняття з теорії баз даних і знань, роз'яснено їхні значення та зв'язок з іншими поняттями. Володіння цим матеріалом забезпечує формування у студентів базисних уявлень про дисципліну, заохочує до подальшого поглибленого ознайомлення з предметом.

Для студентів денної і заочної форм навчання напряму «Комп'ютерні науки» . Може бути корисним для студентів усіх спеціальностей, які вивчають теорію баз даних і знань.

УДК 681.3

ISBN

© О.М. М'якшило, 2010
© НУХТ, 2010

Вступ

Головною умовою коректного й ефективного оперування знаннями про певну предметну область (ПрО) є досконале володіння її понятійним апаратом. Основою понятійного апарату є термінологічний простір, що утворюється термінами — позначеннями сутності предметної області, і визначеннями, які описують зміст цих сутностей. Одна з найважливіших її цілей навчальної діяльності студента — вироблення «правильного» особистого уявлення про термінологічний простір ПрО, що вивчається. Початковим «матеріалом» для побудови цього уявлення є навчальні повідомлення, які отримує студент у процесі навчання з різних джерел, різного обсягу і в різний час. Осмислюючи або інтерпретуючи ці повідомлення, студент вибирає з них інформацію, на основі якої поступово «викреслює свою карту понять», де відображається його індивідуальне уявлення про склад і взаємозв'язки сутностей дисципліни, що вивчається, і, отже, її понятійний апарат.

Суб'єктивність, властива процесу інтерпретації навчальних повідомлень, вносить, певні «погрішності» у власне уявлення термінологічного простору, що виробляється студентом. Природно, що за цих умов, чим раніше будуть внесені потрібні корективи, тим якіснішим стане кінцевий результат навчання. Таким чином, доходимо до висновку про важливість і доцільність надання людині на всіх стадіях навчання можливості порівнювати своє поточне уявлення про термінологічний простір ПрО, яку вивчають, з певним зразком, що в концентрованому вигляді становить суспільно-визнану інтерпретацію цього простору.

На роль такого зразка претендує цей тлумачний словник.

Дисципліна «Організація баз даних та знань» належить до базових дисциплін для наукового напрямку «Комп'ютерні науки». Справді, жодна інформаційна система будь-якого рівня не обходиться без бази даних або бази знань. Електронні бази даних впевнено увійшли в життя

пересічної людини: жовті сторінки, карти міст та автомобільних шляхів, звід законів і бухгалтерські нормативи, кулінарна книга чи телефонна книга, тощо. Ще більше бази даних потрібні фахівцям. Жоден облік не виконати без баз даних. Більше того, інформація, що знаходиться в облікових базах даних підприємства, містить в собі додаткові знання: кореляційні зв'язки між даними за певний період часу можуть відкрити нові не очевидні залежності, які допоможуть знайти шляхи для підвищення ефективності роботи підприємства. Бази даних і знань зберігають у собі безліч фактів та даних, інформаційні поклади, які можна порівняти з покладами корисних копалин. Це цілий світ створений людиною, і основні поняття баз даних та знань — стовпи, на яких він тримається.

«А»

Алгоритмічні знання — це алгоритми та обчислювальні функції, які виконують визначені перетворення, розв'язують певні задачі. Наприклад, тексти прикладних програм, операційні системи .

Альтернативні ключі — унікальні ключі, що не стали первинними, але за якими у разі потреби можна індексувати базу даних.

Артність (див. *МД реляційна*) — кількість складових в кортежі, що відповідають атрибутам (характеристикам) певного об'єкта і визначають кількість колонок таблиці в базі даних (БД).

Архітектура БД (Арх) — нині, як стандартну архітектуру БД використовують трирівневу систему організації БД, запропоновану американським комітетом із стандартизації ANSI. Вона складається з зовнішньої схеми даних (як їх бачить користувач), внутрішньої схеми даних (концептуальної моделі) та безпосередньо фізичної БД .

Арх. файл-серверна (Ф-С) централізованих БД з мережевим доступом передбачає функціонування одного з комп'ютерів мережі як виділеного сервера, на якому зберігаються файли централізованої БД.

Відповідно до запитів користувачів файли з файл-сервера передаються на робочі станції користувачів, де здійснюється основне оброблення даних.

Арх. клієнт-серверна (К-С) централізованих БД з мережевим доступом на відміну від Ф-С передбачає, що центральний сервер виконує основну частину оброблення даних. Користувачі звертаються до центрального сервера за допомогою спеціальної мови запитів — SQL. Звернення користувачів ініціюють на сервері процеси оброблення даних. У результаті користувач отримує вже оброблений набір даних. На відміну від технології Ф-С, між клієнтом і сервером передається не весь набір даних, а лише дані, потрібні користувачеві.

Атрибут (Атр.) — визначена характеристика об'єкта дані про який зберігаються у БД.

Атр. ключовий — унікальна характеристика об'єкта, дані про який зберігаються в БД, значення ключових атрибутів не повторюються у БД. Якщо таку властивість виділити важко, то використовують сукупність атрибутів, щоб досягти унікальності. Ключовий атрибут можна створити штучно приписуючи кожному об'єкту унікальний номер (ідентифікатор — ID). Після видалення об'єкта з БД його ідентифікатор більше не використовується.

Атр. не ключовий — властивість об'єкта, що не є унікальною і може бути властива іншим об'єктам в БД.

«Б»

Бази даних (БД) — це структурована певним чином сукупність даних, що належать до конкретної предметної області, і використовуються для вирішення чітко окресленого кола завдань організації. База даних може бути локальною, централізованою (на сервері мережі з багатокористувацьким доступом), розподіленою (на багатьох комп'ютерах з багатокористувацьким доступом).

БД локальні — на одному комп'ютері для одного користувача, або в локальній мережі для декількох користувачів. Відрізняються від БД з розподіленим доступом тим, що не мають чітко виділених механізмів виконання транзакцій (змін у БД), не мають збережуваних процедур, що готують набори даних для широкого кола користувачів, скриптів, для модифікування структури БД.

БД настільні — те саме, що локальні БД, крім серверних БД, що працюють у локальному режимі. Найчастіше входять до складу офісних пакетів. До сучасних настільних СКБД відносяться: MS Access, Paradox, Visual dBASE, Visual FoxPro. Потужність настільних СКБД значно менша за потужність промислових, хоча вони пристосовані для вирішення широкого кола завдань керування.

БД промислові — сучасні промислові БД реалізовані на платформі «клієнт-сервер», що дає можливість зберігати на сервері саме базу даних, а засоби подання даних та алгоритми їх оброблення (бізнес-логіку) перенести до клієнтської частини системи. Існують системи, в яких частину бізнес-логіки реалізовано вже на сервері БД, або на проміжному сервері, що спрощує структуру клієнтського додатку і залишає серверу БД традиційні функції збереження та підтримки цілісності даних. До клієнт-серверних СКБД належать широко відомі СКБД Oracle, MS-SQL-Server, Interbase, Informix та ін.

БД розподілені — розташовані на кількох комп'ютерах, але для користувачів усе виглядає так, наче всі дані розміщуються на одному сервері. Для роботи з розподіленими БД використовують системи керування розподіленими БД (СКРБД).

База знань (БЗ) — пакет знань, занурений у систему, або система, що ґрунтується на знаннях.

Банк Даних (БнД) — це інформаційна система, що охоплює інформаційні, математичні (алгоритмічні), програмні, мовні, організаційні та технічні засоби (апаратна платформа й операційна система), що

забезпечують в сукупності централізовану підтримку і колективне використання даних користувачем БнД.

«В»

Вбудований SQL — оператори мови запитів SQL, вбудовані у вихідний текст програми на базовій мові програмування (див. *мова запитів — SQL*)

Відношення — базове поняття реляційної алгебри. Відношенням називають деяку підмножину декартового добутку одного чи кількох доменів. Декартовий добуток доменів $D_1, D_2, \dots, D_k$ позначається як $D_1 \times D_2 \times \dots \times D_k$. Елементи відношення називають *кортежами*. Тоді відношенням називатимемо множину всіх кортежів $(v_{1i}, v_{2i}, \dots, v_{ki})$ довжини k , таких, що

$$v_{1i} \in D_1, v_{2i} \in D_2, \dots, v_{ki} \in D_k.$$

Зручно подавати відношення як *таблицю*, де кожний рядок є кортеж і кожна комірка відповідає одному компоненту кортежу. При цьому стовпчики називають *атрибутами*. Один кортеж описує один об'єкт предметної області, а кожна його складова — певну властивість об'єкта (атрибут).

Власник набору — тип логічного запису в мережній моделі БД. Перебуває у відношенні головний — підлеглий до логічних записів типу «член набору». Набір найвищого рівня в базі даних називається «сингулярним набором», а його власником є система.

«Д»

Дані — це інформація зафіксована в певній структурованій формі, пристосованій для подальшого зберігання, обробки та передавання.

Домен — множина значень будь-якої характеристики (атрибута) об'єкта. Домен задається типом даних, кількістю символів та логічними або природними обмеженнями на значення атрибута.

«Е»

Евристики (продукції) — «якщо, то» правила, за якими побудована *продукційна* модель знань.

Екземпляр сутності — об'єкт предметної області, що належить до одного класу і має спільну сутність з об'єктами свого класу. Спільні ознаки (характеристики) об'єктів одного класу називають *атрибутами*.

Експертні системи (ЕС) — системи, які забезпечують створення і використання баз знань експертів .

Екстент (Е.) — безперервна область дискового простору. Екстенти складаються з чотирьох типів сторінок: сторінки даних, сторінки індексів, бітові сторінки та сторінки blob-об'єктів (див. *сторінка*). Таблиця даних моделюється сукупністю екстентів. Для моделювання кожної таблиці використовують два типи екстентів: перший і наступний.

Е. перший — задається СКБД при створенні об'єкта типу таблиця. Його розмір EXTENTSIZE визначають під час створення — мінімальний розмір екстента в кожній системі свій, але здебільшого становить 4 сторінки, максимальний — 2 Гбайта.

Е. наступний — новий екстент, створюваний після заповнення попереднього, зв'язаний з ним спеціальним посиланням, яке розташоване на останій сторінці екстента. NEXT SIZE — розмір кожного наступного екстента. Розмір екстента подвоюється, якщо кількість виділених під базу екстентів збільшується в пропорції кратній 16, тобто після заповнення кожних 16 екстентів розмір наступного екстента стає вдвічі більшим за розмір попереднього.

«Ж»

Журнал транзакцій — журнал змін у базі даних. Забезпечує можливість відтворення узгодженого стану бази даних після будь-якого збою. Основою відтворення БД є резервне збереження даних та записи у журналі транзакцій, що показують послідовність змін у БД.

«3»

Залежні сутності (3.с.) — дочірні сутності, екземпляри яких логічно пов'язані з екземплярами батьківських сутностей.

3.с. асоціативна — сутність, яка пов'язана з кількома батьківськими сутностями. Така сутність містить інформацію про зв'язки сутностей.

3.с. категоріальна — дочірня сутність в ієрархії наслідування. Ієрархія наслідування (чи ієрархія категорій) — тип об'єднання сутностей, які мають загальні характеристики. Загальні характеристики виокремлені у батьківську сутність.

3.с. поіменована — та, що має тільки посилання на батьківські сутності й не має власних атрибутів (окремий випадок асоціативної).

3.с. характеристична — залежна дочірня сутність, яка пов'язана тільки з однією батьківською сутністю і за змістом зберігає інформацію про характеристики батьківської сутності.

Запис (3.) — фрагмент бази даних, що описує один об'єкт. В реляційній базі даних запис відповідає рядку таблиці.

3. віртуальний — елемент ієрархічної моделі даних, містить адресу іншого запису, в якому власне і зберігаються дані. Призначений для запобігання перетинам між гілками ієрархічного дерева.

3. змінної довжини — характерні для мережної та ієрархічної моделей даних.

3. логічний — елемент мережної моделі даних. Бувають двох типів: T1 — член набору, T2 — власник набору. Перебувають у

відношенні «багато до одного». Цей зв'язок можна зобразити у форматі $T2(T1)^*$ (знак * означає «повторити 0 або більше разів»), тобто після кожного запису типу T2 наводяться всі пов'язані з ним записи типу T1. Така організація називається *мультисписком* і передбачає у довільному записі один показник для кожного зв'язку, в якому він бере участь.

3. постійної довжини — характерні для реляційної моделі даних.

Збережувані процедури (Stored Procedure) — підпрограми мовою SQL, що виконуються на сервері. Використовують у клієнт-серверних БД для зменшення обсягів інформації, що передається по мережі між клієнтом і сервером. Забезпечують економний мережевий трафік.

Зв'язок (Зв.) — логічна залежність між об'єктами.

Зв. багато до багатьох — вказує на те, що багатьом об'єктам однієї сутності відповідає багато об'єктів іншої сутності. Допускається тільки на логічному рівні для відтворення логіки взаємодії об'єктів. На фізичному рівні реалізується за допомогою введення додаткової сутності, яка містить ключі двох попередніх сутностей.

Зв. ідентифікований — одному об'єкту батьківської сутності відповідає один або багато об'єктів дочірньої сутності. Атрибути первинного ключа батьківської сутності мігрують до області первинного ключа дочірньої сутності.

Зв. категоріальний — Зв'язок реалізує ієрархію спадковості, де одному екземпляру батьківської сутності відповідає один екземпляр будь-якої дочірньої сутності. Потужність зв'язку — один до одного.

Зв. не ідентифікований — одному об'єкту батьківської сутності відповідає нуль, один або багато об'єктів дочірньої сутності. Атрибути первинного ключа батьківської сутності мігрують до складу неключових атрибутів дочірньої сутності.

Зв. один до багатьох — вказує на те, що одному об'єкту однієї сутності відповідає багато (більше ніж один) об'єктів іншої сутності.

Зв. один до одного — вказує на те, що певному об'єкту однієї сутності відповідає тільки один об'єкт іншої сутності. Є окремим випадком зв'язку один до багатьох.

«І»

Ім'я ролі — додаткове найменування ключових атрибутів при перенесенні їх у дочірню сутність у наслідок утворення зв'язків між об'єктами. Ім'я ролі присвоюють атрибуту в таких випадках: за наявності кількох однотипних зв'язків між двома сутностями; за наявності рекурсивного зв'язку між екземплярами однієї сутності.

Інверсні входи — не унікальні атрибути за якими створюють індекси в базі даних.

Інвертовані списки — вид індексації за не ключовими атрибутами.

Індексація — упорядкування даних у БД за зростанням або спаданням значення обраного поля або полів. Відрізняється від сортування створенням спеціального індексного файлу, який містить індекс та номер запису, що відповідає цьому індексу. Використовують для прискореного пошуку даних. В сучасних БД замість індексних файлів створюють сторінки індексів.

Індекс (Інд.) — атрибут або сукупність атрибутів за якими відбувається індексація даних у БД.

Інд. щільний — індексний файл містить 2 поля: значення первинного ключа та номера запису в БД. На зовнішньому магнітному (лазерному) носії пошук відбувається до запису.

Інд. нещільний — індексний файл, що містить значення первинного ключа та номер блока на зовнішньому носії. Відбувається пошук блока за номером індексу, пересилання блока в оперативну пам'ять і послідовний пошук запису в блоці.

Інд. хешований — індексний файл, що містить значення первинного ключа та номер блока на зовнішньому носії, який

вираховано як *h*-функцію первинного ключа. Як *h*-функцію (див. *хеш-функція*) вибирають таку, щоб давала приблизно однакову щільність розподілу даних між блоками.

Інструментальні засоби побудови БЗ — програмні продукти, що дають можливість наповнити БЗ знаннями щодо певної предметної області, використовуючи задану модель знань. До них можна віднести засоби побудови онтологій, оболонки експертних систем, інтегровані гібридні середовища, інструментальні засоби побудови автоматизованих навчальних систем.

«К»

Категорії (Кат.) — (див. *зв.категоріальний*)

Кат. повна — у повній категорії одному запису батьківської сутності відповідає запис у будь-якій дочірній сутності.

Кат. неповна — якщо категорія побудована не повністю, то в батьківській сутності можуть існувати значення атрибута *Тип*, які не мають відповідних екземплярів у дочірніх сутностях.

Ключі (Кл.) — спеціальні атрибути БД, за якими відбувається зв'язок записів у різних таблицях, або індексація.

Кл. альтернативні — потенційні ключі, що не стали первинними, можуть бути задані як альтернативні, тобто, ключі за якими, в разі потреби, можна індексувати БД.

Кл. зовнішній — первинний ключ батьківської сутності, що мігрував до дочірньої сутності для забезпечення зв'язку.

Кл. первинний — атрибут чи група атрибутів, які ідентифікують об'єкт. Вимоги до первинного ключа: унікальність, компактність, не може мати нульового значення та змінюватися впродовж усього часу існування БД.

Кл. потенційні — унікальні атрибути, що претендують на роль первинного ключа (Candidate key).

Кл. сурогатний — додатковий штучний атрибут, який має тип даних — лічильник і використовується для ідентифікації запису. При вилученні запису значення його сурогатного ключа також знищується і ніколи не присвоюється іншому запису.

Колонка — частина таблиці в реляційній базі даних, що відповідає одному атрибуту й відображає одну спільну характеристику групи об'єктів однієї сутності.

Комірка — частина таблиці в реляційній БД, що знаходиться на перетині стовпчика і рядка. В комірці має зберігатися єдине значення однієї характеристики одного об'єкта. Тоді БД вважають нормалізованою (див. *Нормалізація даних*).

Конструктивні знання — це знання про набори можливих структур об'єктів і взаємозв'язки між їх компонентами. Цей тип знань одержують у технічних і більшій частині прикладних наук.

Концептуальна схема БД — показує класи об'єктів предметної області (сутності) та зв'язки між ними. Це найбільш узагальнена логічна структура БД.

Кортеж — див. *МД. реляційна*.

«М»

Метазнання — знання про порядок і правила застосування знань.

Механізм одержання рішень (МР) — забезпечує використання знань з БЗ для розв'язання поставлених задач. Принцип роботи МР тісно пов'язаний зі способом подання знань у базі знань. Якщо знання подано рівняннями, то МР є процедурою розв'язання цих рівнянь, якщо евристичними, то набором евристик. У будь-якому разі МР містить алгоритм одержання знань.

Мова запитів SQL — SQL (Structured Query Language) — структурована мова запитів — стандартна мова запитів до реляційних БД. Мова SQL з'явилася після реляційної алгебри. Її прототип було розроблено

наприкінці 70-х років 20 ст. у компанії IBM Research. Перший міжнародний стандарт мови SQL було прийнято у 1989 р. Наприкінці 1992 р. було прийнято новий міжнародний стандарт мови SQL — SQL/92 або SQL2. У 1999 р. з'явився стандарт SQL3. У ньому введено нові типи даних, при цьому передбачено можливість задання складних структурованих типів даних, які більш точно відповідають об'єктній орієнтації.

Мовні засоби — мови, за допомогою яких описується структура даних (DDL) та мови маніпулювання даними (DML).

Модель даних (МД) — опис об'єктів і зв'язків між ними. Модель даних визначається трьома компонентами: допустимою організацією даних, обмеженнями, що підтримують цілісність даних та множиною операцій, допустимих над об'єктами БД .

МД ієрархічна — є лісом (сукупністю дерев), де всі зв'язки розташовані у напрямку від сина до батька. В ієрархії використовують мережну термінологію: «*тип логічного запису*» замість «*запис*». Крім того, введено додаткове поняття — «*тип віртуального запису*». Такий запис є показником на логічний запис. Типи віртуальних записів потрібні в тих випадках, коли інтуїтивно ми хотіли б розташувати який-небудь запис у двох або більше деревах, чи у кількох місцях на одному дереві.

МД логічна (див. *Концептуальна схема БД*) . Існує три рівня логічної моделі даних:

- Діаграма об'єкт-зв'язок (ERD);
- Модель даних, що базується на ключах (Key Based model (KB);
- Повна атрибутивна модель (Fully Attributed model (FA).

Діаграма об'єкт-зв'язок являє собою модель даних верхнього рівня — Definitions model. Вона призначена для обговорення структури даних з експертами з боку замовника. KB показує за рахунок яких атрибутів об'єкти пов'язані між собою. FA — вміщує повний набір сутностей Про, їх атрибути та зв'язки.

МД мережна — модель ПРО у вигляді у графа. Вузлами графа є об'єкти (сутності) ПРО. Поняття «зв'язок» для мережної моделі є синонімом бінарного зв'язку «багато до одного». Цей зв'язок можна зобразити записами змінної довжини з форматом T2(T1)* (див. 3. логічний).

МД реляційна. Основою реляційної моделі є на теоретико-множинне поняття відношення, що відображається у вигляді таблиці, елементами якої є дані про характеристики об'єктів.

МД трансформаційна — один із видів фізичної моделі даних, що використовують у середовищі CASE засобу ERWin для налаштування на обрану СКБД.

МД фізична — фізична модель даних у середовищі ERWin будується автоматично на основі логічної моделі при виборі певної СКБД.

Існує два рівня фізичної моделі даних:

- трансформаційна модель;
- модель системного каталогу БД у вибраній СКБД.

Мультисписки — див. *МД мережна*.

«Н»

Нормалізація даних — процес перевірки і реорганізації сутностей та атрибутів з метою задоволення вимог до реляційної моделі даних.

Нормальна форма (НФ) — стан бази даних, що відповідає вимогам до реляційної моделі даних.

Перша НФ. Відношення перебуває у першій нормальній формі тоді і тільки тоді, коли всі атрибути містять атомарні значення, тобто, на перетині кожного стовпчика і рядка є тільки одне значення.

Друга НФ. Відношення перебуває у другій нормальній формі, якщо воно є у першій нормальній формі і кожний неключовий атрибут

повністю залежить від первинного ключа (не може бути залежності від частини ключа). Друга нормальна форма існує тільки для відношень, які мають складний первинний ключ.

Третя НФ. Відношення перебуває у третій нормальній формі тоді і тільки тоді, коли воно є у другій нормальній формі і не містить транзитивних залежностей між атрибутами. Транзитивна залежність існує тоді, коли атрибут не залежить безпосередньо від первинного ключа, а пов'язаний з ним через інший атрибут.

Четверта НФ. Відношення перебуває у четвертій нормальній формі, коли воно є у третій нормальній формі і не має багатозначних залежностей між атрибутами.

«О»

Онтології — бази знань, що підтримують фреймову модель знань, тобто мають ієрархічну структуру.

Операції реляційної алгебри (ОРА) — операції над відношеннями.

Е. Кодд визначив усього вісім операцій. Усю сукупність операцій можна розбити на два типи: теоретико-множинні та спеціальні операції.

ОРА Теоретико-множинні — до теоретико-множинних операцій належать такі: *об'єднання, перетин, різниця, зчеплення та розширений декартів добуток*.

ОРА Спеціальні. Першою спеціальною операцією реляційної алгебри є *операція фільтрації, або операція обмеження відношень*. Друга — *операція проектування*. Проекцією відношення R на набір атрибутів B , що позначається як $R[B]$, називають відношення зі схемою $S_r[b] = B$, яке містить кортежі, отримані з кортежів відношення R видаленням з них атрибутів підмножини B_1 , які нас не цікавлять. За визначенням відношень усі дубльовані кортежі вилучаються з результуючого відношення.

Третьою спеціальною операцією реляційної алгебри є *операція умовного об'єднання*. На відміну від розглянутих вище операцій фільтрації і проектування, які є унарними, тобто відбуваються над одним відношенням, операція умовного об'єднання є бінарною, в якій приймає участь пара відношень. У результаті виконання операції об'єднання утворюється одне відношення.

Четвертою операцією, що входить до набору спеціальних операцій реляційної алгебри, є *операція ділення*.

«П»

Поле — частина логічного запису, що описує одну з властивостей об'єкту і в реляційній моделі відповідає атрибуту.

Понятійні знання — це набір понять, якими користуються при формулюванні та розв'язанні певної задачі. Понятійні знання відтворюють модель ПРО й складаються з уявних об'єктів, які називають *поняттями*. Поняття зазвичай мають ім'я, визначення, структуру, вони пов'язані між собою і входять до якоїсь системи понять. Уявні об'єкти, на яких ґрунтуються понятійні знання, певним чином співвідносяться з реальними об'єктами — в усіх галузях, крім математики. В математиці понятійні знання пов'язані лише між собою.

Потужність зв'язку (Cardinality) — показує скільки екземплярів дочірньої сутності відповідає одному екземпляру батьківської сутності для зв'язку «один до багатьох».

Правила валідації — це логічні обмеження на значення атрибутів. Наприклад: дата накладної, по якій вже одержано товар не може бути більшою за поточну; вік людини не перебільшує 150 років; паспорт людини отримує не раніше ніж у 16 років; продукти мають певний термін придатності і т.д. Правила валідації допомагають уникнути помилок при введенні даних у БД. Вони можуть задаватись у вигляді допустимого інтервалу, перелічення значень атрибута, виразу.

Правила підтримки цілісності даних (ППЦД) — правила доповнення, оновлення та видалення даних за рахунок RI-тригерів — логічних конструкцій SQL коду, що відображують умови використання даних.

ППЦД обмеження (RESTRICT) — коли слід заборонити дії над записом у батьківській сутності, якщо є хоча б один пов'язаний з ним запис у дочірній сутності. Для обмеження операції видалення діє правило Parent Delete RESTRICT. Обмеження на видалення з боку дочірньої сутності встановлюється правилом Child Delete RESTRICT.

Для операцій Insert та Update діють правила Parent Insert RESTRICT, Parent Update RESTRICT, Child Insert RESTRICT, Child Update RESTRICT, відповідно.

ППЦД каскад (CASCADE) — правило, коли усі дії над записом у батьківській сутності переносяться на усі, пов'язані з ним, записи у дочірній сутності. Існують правила: Parent Delete CASCADE, Parent Insert CASCADE, Parent Update CASCADE. Аналогічно, для дочірньої сутності: Child Delete CASCADE, Child Insert CASCADE, Child Update CASCADE.

ППЦД SET NULL — за наявності неідентифікованого зв'язку, коли запис в дочірній сутності може набувати значення NULL використовують правило SET NULL.

ППЦД SET DEFAULT — використовують, тоді, коли атрибуту зовнішнього ключа надають значення за замовчуванням. Діє тільки у випадку не ідентифікованого зв'язку.

ППЦД NONE — для обох випадків ідентифікованого і неідентифікованого зв'язку можна задати правило, коли зміни в кожній сутності відбуваються незалежно від інших.

Предметна область (ПРО) — частина реального світу, дані про який мають бути збережені в БД.

Продукційна модель знань — це модель у якій пошук та додавання знань відбуваються згідно з правилами «якщо,то», що названі продукціями:

Якщо <умова>, то <умова>інакше <умова>.

Процедурні знання — методи, алгоритми і програми, які можна використовувати, передавати і об'єднувати у бібліотеки.

«Р»

Реляційна алгебра — алгебра над відношеннями. *Алгебра* — це множина об'єктів, яку називають *основною множиною*, із заданою на ній сукупністю операцій, замкнених відносно цієї множини (тобто, результати операцій також є об'єктами розглядуваної множини). В реляційній алгебрі основною множиною є множина відношень, кожне з яких описує об'єкти одного класу. У фізичній моделі БД відношенню відповідає таблиця.

Реляційні БД — створені на основі реляційної моделі даних, яка містить таблиці, зв'язки, правила підтримки цілісності даних, правила валідації. Нині реляційні БД є промисловим стандартом баз даних.

Рядок — елемент таблиці, що містить дані, які описують один об'єкт, шляхом переліку його характеристик. Одній характеристиці відповідає одна комірка рядка.

«С»

Семантична мережа — це модель знань, що відображає сукупність об'єктів предметної області і відносини між ними. При цьому об'єктам відповідають вершини (або вузли) мережі, а відносинам — з'єднувальні дуги.

Система баз знань (СБЗ) — система, що забезпечує створення і використання баз знань, або система, яка ґрунтується на знаннях.

Система керування базами даних (СКБД) — комплекс інструментальних засобів (програмних і мовних), що реалізують

централізоване керування БД і забезпечують доступ до даних та маніпулювання ними (внесення нових даних, коригування даних, вилучення даних, резервне копіювання і т. д.).

Слот (сторінкова модель фізичної БД) — показує розміщення рядків даних на сторінці, це 4 байтове слово, в якому 2 байта відповідають зсуву рядка на сторінці, а інші 2 байта - довжині рядка.

Слот (фреймова модель БЗ) — див. *фрейм*.

Спеціальні операції реляційної алгебри — операції реляційної алгебри, що дали підставу для розроблення та обґрунтування міжплатформової мови запитів SQL, а саме оператора SELECT з усіма реченнями, які забезпечують умови вибору, роботу зі з'єднаними таблицями, групування даних і. т.д. До спеціальних операцій реляційної алгебри належать такі операції: фільтрації, проектування, умовного об'єднання та ділення. Оперують з відношеннями, які мають різні схеми, тобто різний набір атрибутів.

Стовпчик (стовпець) — див. *колонка*

Сторінка — головна одиниця обміну даними між процесором та пристроями зовнішньої пам'яті. На одній сторінці зберігаються однорідні дані: або вміст таблиці, або індекси, Blob-об'єкти чи бітова інформація.

Сутність — підмножина об'єктів одного класу.

Схема БД — графічне зображення складових БД. Для реляційної БД — це набір таблиць і зв'язки між ними.

Схема відношення — перелік назв атрибутів відношення (у фізичній БД відповідає переліку назв стовпчиків таблиці).

«Т»

Теоретико-множинні операції (ТМО) реляційної алгебри — операції реляційної алгебри, які оперують з відношеннями, що мають однакову схему, тобто однаковий набір атрибутів. До теоретико-множинних

операцій належать такі: об'єднання; перетин; різниця; зчеплення; розширений декартів добуток.

ТМО об'єднання — *об'єднанням* двох відношень називають відношення, що містить множину кортежів, які належать або першому, або другому вихідним відношенням, або обом відношенням одночасно.

ТМО перехрещення — *перехрещенням* відношень називають відношення, яке містить множину кортежів, що належать одночасно і першому і другому відношенням R_1 і R_2 .

ТМО різниця — *різницею* відношень R_1 і R_2 називають відношення, що містить множину кортежів, які належать R_1 та не належать R_2 і навпаки.

Трансакції в БД — зміни в БД (редагування даних, доповнення новими записами, вилучення записів), що переводять БД з одного цілісного стану в інший, без порушення логічних зв'язків між даними та втрат даних.

Тригери — логічні конструкції мови SQL, що відображають бізнес-правила використання даних і є правилами доповнення, оновлення та вилучення даних. Зберігаються на сервері БД і запускаються автоматично під час виконання трансакції.

«Ф»

Файли індексні — використовують під час організації доступу за *первинним ключем*. У деяких комерційних системах індексними називають також файли, організовані у вигляді інвертованих списків, які використовують для доступу за *вторинним ключем (довільним атрибутом)*.

Файли послідовного доступу — Файли із змінною довжиною запису завжди є файлами послідовного доступу і розташовуються на пристроях

послідовного доступу (мережна або ієрархічна модель даних). Вони можуть бути організовані двома способами:

1. Кінець запису завжди відмічається спеціальним маркером:

Запис1	х	Запис2	х	Запис3	х
--------	---	--------	---	--------	---

2. На початку кожного запису записується його довжина:

LZ1	Запис1	LZ2	Запис2	LZ3	Запис3
-----	--------	-----	--------	-----	--------

Тут LZ_n - довжина *n*-го запису.

Файли прямого доступу. Файли з постійною довжиною запису, що розташовуються на пристроях прямого доступу, називають *файлами прямого доступу*. В цих файлах фізичну адресу розташування потрібного запису можна вирахувати за номером запису. Для файлів з постійною довжиною запису адресу розташування запису з номером *K* можна вирахувати за формулою:

$$BA + (K - 1) * LZ + 1,$$

Де BA — базова адреса; LZ — довжина запису.

Фактографічні знання — кількісні та якісні характеристики об'єктів і явищ.

Фактуальні знання — ґрунтуються на фактах і правилах. Факти виконують роль елементарних одиниць знань, а правила призначені для виявлення зв'язків, залежностей між фактами та їхніми комбінаціями.

Фасет — визначає граничні значення заповнювача слота, задає діапазон або перелік його можливих значень. Використовують у базах знань на основі фреймів (*див. фрейм*).

Фізичні моделі БД — визначають способи розміщення даних в середовищі зберігання і способи доступу до даних, які підтримуються на фізичному рівні.

Формат логічного запису — показує зв'язок між записами в мережній моделі БД, що позначається у вигляді формату $T2(T1)^*$ (знак * означає «повторити 0 або більше разів»), тобто після кожного запису типу T2 наводяться всі пов'язані з ним записи типу T1. Таким чином утворюється набір записів змінної довжини. Тип запису T2 називають *власником набору*, а тип запису T1 — членом набору.

Фрейм — це структура для опису стереотипної ситуації, що успадковує характеристики іншої ситуації та їх значення. Характеристики називаються слотами, а значення — заповнювачами слотів. Фрейми утворюють ієрархічну модель знань. Фрейми мають здатність успадковувати значення характеристик своїх батьків, що перебувають на вищому рівні ієрархії. Значення характеристик фреймів можуть передаватися за замовчуванням фреймам, що розміщуються нижче за них в ієрархії. Проте якщо останні мають власні значення даних характеристик, то за істинні беруть саме їх. Слот може містити не тільки конкретне значення, а й ім'я процедури, що дає змогу обчислювати його за заданим алгоритмом, а також одну чи кілька продукцій (еврістик), за допомогою яких ці значення можна знайти. У слот може входити кілька значень. Іноді слот містить компонент, так званий фасет, який задає діапазон або перелік його можливих значень. Фасет вказує також граничні значення заповнювача слота.

«Х»

Хешовання — спосіб збереження даних, в якому адреса блока вираховується як функція від значення ключа.

Хешований файл. Головна ідея організації файлів з хешованим доступом полягає в розподілі записів файла між ділянками, кожна з яких складається з кількох блоків пам'яті. Якщо v — значення ключа, то $h(v)$ (див. *хеш-функція*) вказує номер ділянки, де зберігається запис з цим значенням ключа.

Хеш-функція (h-функція) — функція, яка використовує в як аргумент значення ключа файла і продукує деяке число від 0 до максимального значення. Формування $h(v)$ функції є складним завданням, яке не має однозначного вирішення. Головна вимога до $h(v)$ функції полягає в тому, що всі її значення повинні мати приблизно однакову ймовірність, для забезпечення рівномірного розміщення записів БД по всіх ділянках.

«Ц»

Цілісність даних — відсутність протиріч між даними в БД. У реляційних БД цілісність даних підтримується за рахунок RI (Referational Integrity) тригерів — фрагментів SQL коду, які розміщують на сервері БД для автоматичного запуску під час виконання транзакцій (змін) у БД.

«Ч»

Чанк (chunk) — частина диска, фізичний простір на диску, який асоціюється з одним процесом (on line) оброблення даних.

Член набору (див. *Формат логічного запису*) — один із двох видів записів у мережній моделі бази даних у зв'язку «головний — підлеглий». Є підлеглим відносно власника свого набору.

Л і т е р а т у р а

1. ДСТУ ISO/IEC 2382-17:2005 — Інформаційні технології. Словник термінів. Частина 17. Бази даних (ISO/IEC 2382-17:1999, IDT) — Чинний від 01.08.2007 — К.: Держспоживстандарт, 2007. — 42 с.
2. Карпова Т.. Базы данных. Модели, разработка, реализация. — СПб.: Питер, 2001. —304 с.
3. Маклаков С.В. CASE — средства разработки информационных систем. — М. : Диалог — МИФИ, 1999. — 406 с.
4. Маклаков С.В. Создание информационных систем с ALLFusion Modeling Suite. — М.: ДИАЛОГ МИФИ, 2005. — 427 с.
5. Мартин Дж. Организация баз данных в вычислительных системах. — М. Мир, 1978. — 615 с.
6. Ульман Дж. “Основы систем баз данных. М : Финансы и статистика, 1983. — 333 с.
7. Цикритзис Д., Лоховски Ф. Модели данных. М : Финансы и статистика, 1985. — 342 с.

Наукове видання

М'якшило Олена Михайлівна

Організація баз даних та знань
Тлумачний словник

Редактор *Л.М. Орішич*
Художнє оформлення *Є. В. Чурія*
Комп'ютерна верстка

Підписано до друку _____. Формат 60x84/16.

Папір офс. № 1 Гарнітура _____.

Ум. друк. арк. _____. Обл.- вид. арх. Наклад 30 прим. Вид. № 26/10.

Зам. №