

В.А.Резніченко

60 РОКІВ БАЗАМ ДАНИХ (заключна частина)

Наводиться огляд досліджень і розробок баз даних із моменту їх виникнення в 60-х роках минулого століття і по сьогодні. Виділяються наступні етапи: виникнення і становлення, бурхливий розвиток, епоха реляційних баз даних, розширені реляційні бази даних, постреляційні бази даних і великі дані. На етапі становлення описуються системи IDS, IMS, Total і Adabas. На етапі бурхливого розвитку висвітлені питання архітектури баз даних ANSI/X3/SPARC, пропозицій КОДАСИЛ, концепції і мов концептуального моделювання. На етапі епохи реляційних баз даних розкриваються результати наукової діяльності Е. Кодда, теорія залежностей і нормальних форм, мови запитів, експериментальні дослідження і розробки, оптимізація та стандартизація, управління транзакціями. Етап розширених реляційних баз даних присвячений опису темпоральних, просторових, дедуктивних, активних, об'єктних, розподілених та статистичних баз даних, баз даних масивів, машин баз даних і сховищ даних. На наступному етапі розкрита проблематика постреляційних баз даних, а саме: NoSQL, ключ-значення, стовпчикові, документні, графові, NewSQL, онтологічні. Шостий етап присвячений розкриттю причин виникнення, характерних властивостей, класифікації, принципів роботи, методів і технологій великих даних. Нарешті, в останньому із розділів подається короткий огляд досліджень і розробок баз даних у Радянському Союзі.

Ключові слова: Типи баз даних: ієрархічна, мережева, реляційна, навігаційна, темпоральна, просторова, просторово-темпоральна, просторово-мережева, об'єктів, що переміщуються, дедуктивна, активна, об'єктно-орієнтована, об'єктно-реляційна, розподілена, паралельна, масивів, статистична, багатовимірні, машина баз даних, сховища даних, NoSQL, ключ-значення, стовпчикова, документно-орієнтована, графова, мультимодельна, хмарна, наукова, багатозначна, XML, NewSQL, онтологічна, великі дані.

Бази даних зображень

База даних зображень (БДЗ) – це база даних, яка надає ефективні й розвинуті засоби й технології для підтримання процесів моделювання, збереження, індексування, пошуку і маніпулювання зображеннями та їхніми метаданими [1045].

Текстовий пошук зображень (text based image retrieval – TBIR). Діяльність щодо використання зображень в базах даних почалася в середині 70-х років минулого століття в зв'язку із появою розвинутих промислових СУБД. Основні роботи цього періоду були зосереджені на текстовому підході до пошуку зображень. Його суть полягала в анотуванні й пошуку зображень на основі текстової інформації. Зображення описувалися набором ключових слів, або текстовими дескрипторами і за допомогою закладених у СУБД засобів пошуку за текстом шукалися потрібні зображення. 1979

року відбулася міжнародна конференція з питань використання технологій баз даних у графічних додатках, де було підбито підсумки щодо БДЗ. У статтях [1047, 1048] подано ґрунтовні огляди праць із текстового опису й пошуку зображень у БД станом на 1984 і 1992 роки.

Згодом у зв'язку із складністю і різноманіттям описових елементів зображень дослідники усвідомили необхідність створення універсальних керованих словників, класифікаційних схем та інших підходів до термінологічного впорядкування опису зображень. Уперше це питання порушила Сара Шатфорд (Sara Shatford) у праці [1049], яка з часом опосередковано привела до появи різних тезаурусів, дотичних до зображень.

Контентний пошук зображень (content-based image retrieval - CBIR). На початку 90-х років з'явилися праці з проблем пошуку зображень за їхнім вмістом. Цей

Опишемо отримані результати в області СВІR на основі цієї схеми.

Обробка зображень – сегментація.

Сегментація зображень є важливим кроком у вирішенні задач розпізнавання, стиснення, візуалізації й пошуку зображень.

Початковим етапом аналізу (розпізнавання вмісту) зображення є його сегментація. Суть сегментації – розбиття зображення на численні області/об'єкти для спрощення й/або зміни представлення зображення, аби його було простіше і легше аналізувати [1061]. Використовується переважно для визначення об'єктів і меж (ліній, кривих тощо) у зображенні.

Дослідження у сфері сегментації зображень були ініційовані на початку 90-х років практично одночасно з появою напрямку СВІR.

Існує гостра потреба визначення, які ж фічери має зображення. Зазвичай це досягається розбиттям зображення на велику кількість однорідних (відносно певної властивості) прямокутних областей, кожна з яких називається сегментом, а сам процес – сегментацією. Існує багато методів сегментації, огляд і аналіз яких наведено в [1061 – 1064]. «Найсвіжішим» на момент написання статті є огляд [1065], де аналізуються 11 методів 4-х категорій та їх застосування в 10-и предметних областях.

Добування фічерів. Із зображень добуваються низькорівневі характеристики (фічери) і запам'ятовуються в БД у вигляді фічерних векторів. Добування фічерів складає основу СВІR. До базових фічерів зображень належать колір, текстура, фігура, місце розташування. Коротко обговоримо існуючі методи добування й представлення цих фічерів.

Колір. Колір є найбільш використовуваним фічером під час пошуку зображень. Кольори визначаються відповідно до обраного кольорного простору. Існує багато кольорних просторів і використовуються вони в різних додатках. Опис кольорних просторів можна знайти в [1066]. Стандарт MPEG-7 [1067] до складу дескрипторів кольору включив: основний колір, прямий колір, кольорний простір, квантування кольору, кольорову розкладку, масштабований колір, структуру кольору. Найбільш уживаним

способом представлення є колірна гістограма [1068]. Здебільшого це спільне розподілення інтенсивності трьох кольорних каналів (RGB). З огляду на те, що більшість гістограм є розрядженими й тому чутливими до шуму, була запропонована кумулятивна гістограма [1069], яка продемонструвала свої переваги порівняно зі звичайною кольорною гістограмою. В праці [1070] пропонується також метод квантування кольорного простору, що підвищує ефективність пошуку. Запропоновані також інші способи представлення кольору для цілей пошуку. Зокрема, коваріаційна матриця кольору (color covariance matrix), моменти кольорів (color moments) [1069], набори кольорів (color sets) [1071, 1072], вектори когерентності кольору [1073].

Текстура. Текстура – зображення, що відтворює візуальні властивості будь-яких поверхонь чи об'єктів. Вона містить важливу інформацію про структурне впорядкування поверхонь і їхній взаємозв'язок із навколишнім середовищем. Свого часу вона виявилась доволі корисною у вирішенні задач розпізнавання образів і комп'ютерного зору, а також плідно використовується в СВІR.

На початку 70-х років був запропонований метод представлення текстурних фічерів у вигляді матриці співпадіння (co-occurrence matrix) [1074]. У подальшому він був розвинутий у праці [1075]. У [1076] було запропоновано варіант представлення текстурних фічерів на основі психологічних досліджень візуального сприйняття зображень людиною. Цей метод виявився доволі привабливим у СВІR, адже сприяв створенню зручнішого для користувача інтерфейсу. Вдосконалений варіант був застосований в системах QVIC [1077] і MARS [1078].

На початку 90-х багато вчених почали використовувати вейвлет-перетворення у вивченні способів представлення текстури. У праці [1079] пропонується застосовувати добути з піддіапазонів вейвлета статистику як представлення текстури. За допомогою цього підходу було досягнуто 90% точності на 112 текстурах Бродаца. У праці [1080] з метою подальшого покращення класифікаційної точності була використана дерево-видна структура вейвлет-перетворення. Для

підвищення продуктивності були запропоновані підходи, які поєднували вейвлет-перетворення з іншими методами. Наприклад, розширення Карунена – Лоева (Karhunen – Loeve expansion) і карти Кохонена (Kohonen maps) [1081], матриця співпадінь [1082, 1083]. Були запропоновані й досліджені представлення текстури марковськими випадковими полями [1084], багатоканальною фільтрацією [1085], фільтрацією Габоора [1086], а також фрактальне [1087].

Протягом багатьох років публікувались оглядові й порівняльні статті [1088 – 1090]. Однією з останніх статей цього типу є [1091]. В ній аналізуються 22 методи/моделі представлення текстур, які належать до 4-х класів і 18-ти підкласів.

Фігура. Існують два способи представлення фігур – у вигляді контуру й у вигляді площини [1092]. Найбільш використовуваними методами цих двох представлень є дескриптори Фур'є і моментні інваріанти (moment invariants).

Основна ідея використання дескрипторів Фур'є полягає в представленні фічера фігури у вигляді контурного перетворення Фур'є. Перші праці в цьому напрямку належать до початку 70-х років [1093, 1094]. У праці [1092] запропонований модифікований дескриптор Фур'є для усунення перешкод під час оцифрування зображень. Основна ідея використання моментних інваріантів полягає у представленні фічера фігури за допомогою поверхневих моментів, інваріантних до перетворення. 1962 року в праці [1095] було виділено 7 таких моментів. У [1096] висунутий швидкий метод обчислення моментів у бінарних зображеннях. Для представлення фігур були також запропоновані: метод кінцевих елементів (finite element method – FEM) [1097], функція повороту (turning function) [1098], гістограма направлення граней (edge directions histogram) [1099], вейвлет-дескриптор [1100]. Були також опубліковані оглядові статті [1101, 1102]. Здійснювалися дослідження із представлення об'ємних фігур [1103 – 1106].

Колірне розподілення (color layout). У середині 90-х років у зв'язку зі збільшенням розмірів зображень дійшли висновку, що хорошим рішенням для представлення

й пошуку зображень є використання не точного попиксельного представлення зображень, а так званого колірнього розподілення (як з точки зору фічерів кольору, так і просторових взаємозв'язків). Ідея полягала в розділенні зображення на підблоки й добуванні колірних фічерів із кожного з них [1106, 1107]. Концепція колірнього розподілення була досліджена й удосконалена в наступних працях [1073, 1108-1112], а також застосована до текстури й інших фічерів зображень.

Просторове розміщення. Також є важливим фічером зображення й зазвичай застосовується до просторових об'єктів. Визначається переважно у вигляді понять «вгорі, вниз, ліворуч, праворуч тощо» згідно із розміщенням конкретного об'єкта [1113]. В [1086] для представлення просторової інформації використовується центр ваги об'єкта і мінімальний обмежувачий його прямокутник. В інших працях пропонується використовувати лише центр об'єкта [1114].

Взаємне розміщення об'єктів важливіше за їхні абсолютні координати. Для зазначення взаємного розміщення частіше використовуються рядки символів (ліворуч/праворуч > зверху/знизу) [1115]. У праці [1116] для підтримки семантичного пошуку зображень представлений алгоритм моделювання просторового контенту. Оригінальний метод просторового взаємного розміщення об'єктів прописується в [1117].

Багатовимірне індексування. Аби СВІР був дійсно масштабованим для колекції зображень великого розміру, необхідні ефективні методи багатовимірного індексування. З цього погляду значну роль відіграють наступні два фактори:

- Велика розмірність. Розмірність фічерних векторів зазвичай приблизно 10^2 .
- Неевклідова міра подібності. Оскільки евклідова міра не здатна ефективно імітувати людське сприйняття певного візуального контенту, варто використовувати інші міри подібності.

Тож для вирішення цих проблем-спершу необхідно зменшити розмірність векторів, а потім застосувати відповідні методи багатовимірного індексування, здатні підтримувати неевклідові міри подібності.

Що ж до зменшення розмірності, то було запропоновано два основні підходи: перетворення Карунена – Лоєва (Karhunen-Loeve transform – KLT) і кластеризація по стовпчиках (column-wise clustering). Зокрема, метод KLT досліджується й використовується в працях [1121-1124]. Метод кластеризації по стовпчиках був запропонований 1983 року [1125].

Існуючі популярні методи багатовимірної індексування включають блочний (bucketing) алгоритм, k-d – дерева, пріоритетні k-d – дерева, квадродрова, K-D-B-, hV-, R-, R+, R*-дерева. Історія методів багатовимірної індексування сягає середини 1970-х років, коли вперше були розроблені кліткові методи, квадродрова (quad-tree) і k-d – дерева (k-d – tree). Однак їхня продуктивність була далека від задовільної. Враховуючи потреби індексації просторових даних, тобто багатовимірної інформації, що виникають у системах ГІС і САПР, 1984 року в праці [1126] Антонін Гуттман (Antonin Guttman) вперше запропонував структуру індексації R-дерева. На основі цієї праці було висунуто багато інших варіантів R-дерева: R+ -дерево [1127], удосконалене R-дерево [1128], R*-дерево [1129]. Проте було показано, що вони перестають масштабуватися за умови розмірів векторів вище 20. У статтях [1122, 1130] представлені огляди порівняльного аналізу різних методів індексування станом на 1996 рік.

Як ми вже зазначали, другим аспектом у проблемі пошуку зображень є неевклідова міра подібності. Було висунуто два підходи в цьому напрямку: кластеризація і нейронні мережі. В праці [1131] був запропонований метод інкрементної кластеризації для динамічного пошуку інформації. Цей метод уможлилював обробку багатовимірних даних і використання неевклідових мір подібності. В подальшому він був розвинутий у праці [1132].

У [1133] запропоновано використання нейронних мереж карт самоорганізації (self-organization map – SOM) як інструменту побудови індексів деревовидної структури у пошуку зображень. У працях [1118-1120] подано широкий огляд методів індексування й доступу. Детальніша інформація з цього приводу наведена в розділі «Просто-

рові бази даних».

Зіставлення зображень. Для порівняння зображення запиту із зображеннями бази даних слід визначити метрики їх подібності. Це робиться шляхом введення поняття відстані між зображеннями. Чим менша відстань між ними, тим ближчі вони один до одного.

У монографії [1134], обсягом у понад 750 сторінок, наводиться вичерпний перелік метрик відстаней, які використовуються в різних науках. Зокрема, там описано 150 метрик для виміру відстаней у комп'ютерних науках, із них 35 метрик, що застосовуються у зображеннях.

Наведемо кілька з них, котрі найчастіше використовуються в наукових статтях і системах для визначення подібності зображень.

Евклідова відстань – найчастіше використовувана метрика для визначення ступеня близькості зображень, визначається як відстань між двома точками евклідового простору, який вираховується за теоремою Піфагора.

Відстань міських кварталів (city block distance) – метрика, введена Германом Мінковським. Згідно з цією метрикою відстань між двома точками дорівнює сумі модулів різниці їхніх координат. Тобто це відстань між двома точками із нанесеною прямокутною сіткою, коли переміщатися можна лише по сторонах сітки. Вона також називається метрикою Манхеттена, прямокутною метрикою, метрикою сітки.

Відстань шахової дошки (chessboard distance). Припускає, що існує сітка і можна переміщатися сторонами сітки й по діагоналі, тобто як король у шахах. Також називається відстанню ходу короля й відстанню Чебишева.

Відстань Махаланобіса (Mahalanobis distance) – міра відстані між векторами випадкових величин, що узагальнює поняття Евклідової відстані. Запропонована індійським статистиком Махаланобісом 1936 року [1135]. З допомогою відстані Махаланобіса можна визначити подібність невідомої і відомої вибірки. Вона відрізняється від відстані Евкліда тим, що враховує кореляції між змінними та інваріантна до масштабу.

Косинусна відстань (cosine distance),

відома також як косинусна подібність, є мірою величини різниці між двома фічерними точками, використовуючи косинусне значення кута між двома векторами у векторному просторі.

Відстань Мінковського (Minkowski distance) – параметрична метрика на евклідовому просторі, яку можна розглядати як узагальнення евклідової відстані та відстані міських кварталів. Дістала назву на честь німецького математика Германа Мінковського, який уперше систематично вивчив дане сімейство функцій відстані.

Кореляційна відстань (correlation distance) – популярний спосіб виміру відстані у вигляді кореляції двох фізичних точок із кінцевими дисперсіями.

Системи CBIR. У типовій системі CBIR (див. рисунок вище) висхідні зображення зберігаються в БД із використанням відповідних методів. Із цих зображень добуваються візуальні фічери. Вони представляються багатовимірними фічерними векторами, запам'ятовуються в БД фічерів та індексуються для швидкого пошуку з використанням відповідних методів доступу. Для знаходження відповідного зображення користувач вказує його у запиті. Система добуває із цього зображення всі фічери й представляє їх у вигляді фічерних векторів. Далі з допомогою розроблених процедур подібності відбувається співставлення фічерних векторів зображення запиту та зображень бази даних із використанням індексів і методів доступу. Зображення бази даних, які задовольняють співставлення, передаються користувачеві.

Перша комерційна система CBIR була розроблена 1995 року в IBM і мала назву QBIC (Query By Image Content) [1077]. Відтоді була розроблена велика кількість комерційних і експериментальних систем CBIR. Такі, як MARS [1078], Photobook [1097], Virage [1137], Visual SEEK [1138], Netra [1139], SIMPLIcity [1140]. У звіті [1059] подано аналіз 58 систем і додатків CBIR, розроблених до 2002 року, із зазначенням їх першоджерел. Окрім цього, подана підсумкова таблиця із наведенням використаних у них фічерів. Стаття [1141] містить огляд понад 200 публікацій із CBIR станом на 2000 рік. Чудові огляди щодо

методів і принципів пошуку інформації в CBIR-системах опубліковані в статтях [1043, 1044, 1136].

Семантичний пошук зображень (Semantic-Based Image Retrieval – SBIR).

Недолік CBIR полягає у відсутності семантики. За допомогою низькорівневих фічерів неможливо описувати високорівневі поняття, що сприймаються людиною. Тобто існував «семантичний розрив» між низькорівневими фічерами зображення та використовуваними людиною високорівневими поняттями предметної області. У статті [1141] семантичний розрив визначається як «неспівпадіння інформації, яку можна виділити з візуальних даних, та інтерпретації цих же даних користувачем у конкретній ситуації». У зв'язку з цим на початку 2000-х років з'явилися дослідження, а згодом і розробки із семантичного пошуку зображень (Semantic-Based Image Retrieval – SBIR).

У звіті [1142] виділяються три рівні мов.

Рівень 1. Пошук за низькорівневими фічерами, такими, як колір, текстура, фігура, просторове розміщення. Типовим є запит «знайти зображення, схожі на задане».

Рівень 2. Пошук об'єктів заданого типу, що ідентифікуються вказаними фічерами, з можливим застосуванням логічних правил виведення. Приклад: «знайти зображення з автомобілями».

Рівень 3. Пошук за абстрактними характеристиками, які включають високорівневі міркування про цілі, способи, методи представлення зображених предметів або сцен. Може включати пошук названих подій або зображення з емоційним чи релігійним значенням тощо. Наприклад, «знайти зображення радіючого натовпу».

Рівні 2 і 3 належать до класу SBIR, а різниця між рівнями 1 і 2 характеризує «семантичний розрив».

Натепер пропонуються такі 5 методів зменшення семантичного розриву:

- використання онтологій для визначення високорівневих концептів/понять;
- методи машинного навчання для встановлення взаємозв'язків між низькорівневими фічерами й високорівневими концептами запиту;
- кластеризація даних;

- зворотній зв'язок по релевантності (relevance feedback – RF) у пошуковий цикл для безперервного вивчення намірів користувачів;

- семантичні шаблони (semantic template – ST) для підтримки високорівневого пошуку зображень.

Стисло опишемо ці методи, детальніше можна ознайомитися з ними в [1143 – 1146].

Онтологія об'єктів. Онтологічний підхід передбачає створення таксономічної онтологічної структури понять щодо фічерів зображень. У таких системах спочатку визначаються різні інтервали для низькорівневих фічерів. Ці інтервали визначають дескриптори зображень проміжного рівня, як-от, «світло-зелений, зелений, темно-зелений». Вони також можуть узагальнюватись із побудовою в кінцевому варіанті онтології понять і можуть використовуватись для визначення високорівневих понять запитів. Наприклад, «хмара» може бути визначена як «довільна опукла» (фігура), світло-блакитного» (кольору), «однотонна» (текстура), «вгорі» (просторове розміщення). На онтологіях можна визначити спеціальні правила виведення. Найпростішими з них є таксономічні правила. Прикладом такої онтологічної системи є [1114].

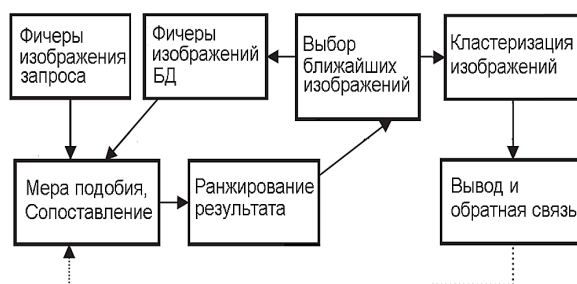
Розбиття фічерів на інтервали потребує використання єдиних правил іменування вершин створюваної онтології. Так, зокрема, в [1147] пропонується система іменування кольорів. У праці [1148] запропонована система з 12 основних кольорів, 5 рівнів яскравості й 3 рівнів насиченості. Всього 180 варіантів кольору. Для пошуку картин визначаються поняття: теплий колір, холодний колір, контрастність (світло-темний, тепло-холодний, доповнюючий). Приміром, знайти картини, написані в світло-темних тонах. За аналогією з кольором виникає необхідність створення системи іменування текстури, яка б стандартизувала опис і представлення текстур [1149]. Як виявилось, іменування текстур доволі складна задача, і досі немає єдиної системи іменування.

Машинне навчання. В багатьох випадках для отримання високорівневих семантичних фічерів варто використовувати

такі формальні засоби, як навчання з учителем і без нього.

У SBIR використовуються такі методи навчання з учителем, як метод опорних векторів (support vector machine – SVM), а також байєсівський класифікатор, штучна нейронна мережа, бутстрепінг, дерево рішень. Скажімо, в [1150] SVM використовується для анотування зображень, в [1151] з допомогою бінарного байєсівського класифікатора високорівневі поняття виводяться з низькорівневих фічерів зображень. У [1152] штучна нейронна мережа також застосовується для встановлення взаємозв'язку між фічерами зображень і попередньо вибраними високорівневими поняттями. В [1153] для анотування великої колекції зображень було застосовано метод бутстрепінгу. В [1154] спершу будується дерево рішень на основі набору зображень, релевантних запиту, й потому використовується як модель для класифікації зображень бази даних відповідно до двох класів: релевантних і нерелевантних.

Кластеризація зображень. Це типовий метод навчання без учителя для цілей пошуку зображень. Мета кластеризації – розбиття багатовимірної неструктурованої множини даних на ряд підмножин даних зі схожими характеристиками [1155, 1156]. Він передбачає групування наборів даних зображень таким чином, щоб максимізувати схожість зображень всередині кластерів і мінімізувати схожість поміж різними кластерами. Для кластеризації зображень найпопулярнішими є методи К-середніх (K-means) [1143] та його варіанти, Ncut [1157], нечіткої кластеризації с-середніх (fuzzy c-means) [1158]. Показовим прикладом використання кластеризації є автоматичне уточнення міри подібності для співставлення зображень за схемою, наведеною на рисунку нижче й запропонованої в [1159].



Вона функціонує так. На підставі зображення запиту і обраної міри подібності відшукуються схожі зображення, вони ранжируються й обираються найближчі з них. Потім на основі гіпотези, що зображення з однаковою семантикою мають тенденцію до групування, здійснюється кластеризація для розподілу результуючих зображень за різними семантичними класами. Далі система виводить кластери зображень і уточнює міру подібності згідно зворотнього зв'язку користувача.

Для пошуку зображень застосовуються також методи розпізнавання об'єктів.

Зворотній зв'язок за релевантністю (relevance feedback – RF). На відміну від попередніх підходів, метод RF передбачає онлайн-обробку, що забезпечує оперативну реакцію на наміри користувача. Традиційно RF використовувався в текстових інформаційно – пошукових системах, а від середини 90-х років почав застосовуватись у СВІР для включення користувача в пошуковий цикл з метою зменшення «семантичного розриву» між тим, що формулюється в запиті й наміром пошуку користувача. Дослідження показали, що застосування RF суттєво підвищує продуктивність СВІР-систем [1160]. Двома найбільш використовуваними стратегіями RF є зміна ваги фічерів (re-weighting) і переміщення точки запиту (query-point-movement) [1161].

Семантичні шаблони (ST). Семантичний шаблон – це відображення між високорівневими поняттями й низькорівневими візуальними фічерами. Зазвичай ST визначається як «репрезентативний» фічер поняття/концепта, що вираховується із колекції зразків зображень. Використанню ST у СВІР присвячені праці [1162 – 1165].

Мультимодельні бази даних

Концепція мультибаз даних, тобто систем, які уможливають роботу з багатьма БД, зародилася наприкінці 70-х років у зв'язку з широким упровадженням БД на виробництві й створенням комп'ютерних мереж. Ця концепція постійно розвивалася протягом наступних 40 років. Виділяються наступні етапи її розвитку:

- інтеграція неоднорідних баз даних;
- федеративні бази даних;

- багатоваріантне зберігання;
- мультимодальні бази даних;
- полі-бази даних.

Зупинимося докладніше на мультимодальних БД, стисло описавши всі інші.

Інтеграція неоднорідних баз даних.

Ідея інтеграції баз даних із різними моделями даних зародилася на початку 80-х років (тоді панували три моделі даних – ієрархічна, мережева й реляційна) з метою забезпечення одночасного й спільного використання прикладною програмою кількох баз даних, організованих у рамках різних СУБД. Інтеграція спрямовувалась на подолання програмної та інформаційної несумісності баз даних. Суттєві результати в цьому напрямку були отримані радянським ученим Калініченком Л.А.



Калініченко Л.А.

1983 року він опублікував монографію [1166] із представленням методів вирішення проблеми інтеграції баз даних, що базуються на створенні загальної моделі даних високого рівня і перетворенні довільних моделей даних на загальну модель.

Федеративні бази даних. *Федеративна база даних (ФБД) – це віртуальна база даних, що являє собою прозору інтеграцію багатьох автономних, можливо, неоднорідних і розподілених БД у логічно єдину БД для спільного використання й обміну даними.*

Уперше ідея федеративності була висловлена 1979 року у звіті [1168] й згодом детально пропрацьована Хеймбігнером і

Маклеодом (Heimbigner, McLeod) 1985 року в праці [1169]. ФДБ надає єдиний зовнішній інтерфейс, який дозволяє запам'ятовувати й вишукувати дані в автономних БД із використанням спільної мови запитів. Для цього ФДБ декомпозує запит на підзапити для їх відпрацювання складовими БД й відтак об'єднує отримані результати з використанням так званих «посередників» (wrappers).

Дуже важливим аспектом ФДБ є автономність їхніх БД-компонентів, тобто рівень їх самостійності. В працях [454, 1170] була запропонована і досліджена наступна класифікація автономності:

- *проектна* – здатність самостійно ухвалювати проектні рішення будь-якого плану;

- *комунікаційна* – здатність ухвалювати рішення щодо того, чи варто взаємодіяти з іншими БД-компонентами і як саме;

- *виконавча* – здатність виконувати власні локальні операції, ініційовані локальними користувачами або подіями, без будь-якої взаємодії із зовнішнім оточенням федерації;

- *асоціативна* – здатність ухвалювати рішення щодо того, чи варто «ділитися» своїми функціональними можливостями й ресурсами з іншими учасниками федерації і, якщо так, то якою мірою, аж до самостійного виходу з федерації або входу до неї.

У працях [1171 – 1174] були також запропоновані й досліджені інші види автономності.

Особливою характеристикою ФДБ є їхня *гетерогенність* (неоднорідність), яка належить до моделей даних, семантики даних, до обмежень цілостності та мов запитів.

Важливою особливістю ФДБ є їхня здатність підтримувати правила відображення/співставлення схем баз даних федерації. Загальноприйнятим рішенням є використання глобальної схеми, що містить релевантні складові схем – членів федерації і опис відображень у вигляді поглядів (views). Водночас пропонується два принципових рішення залежно від напрямку відображень [1175]:

- *Global as View (GaV)*: глобальна схема визначається в термінах локальних схем;

- *Local as View (LaV)*: локальні схеми визначаються в термінах глобальної схеми.

Наприклад, у відомій федеративній системі Multibase [1176] підтримується глобальна схема і єдиний інтерфейс для формування запитів. Запити, сформульовані відносно глобальної схеми, декомпозуються відповідно до існуючих підсхем і обробляються локальними базами даних.

Виділяють три категорії ФДБ [1177]: *слабопов'язані* (loosely coupled), *сильнопов'язані* (tightly coupled) й *гібридні* залежно від того, хто керує федерацією і яким чином інтегруються компоненти. В слабопов'язаній ФДБ саме адміністратор БД відповідальний за включення БД до федерації й глибину її інтеграції з іншими БД федерації. БД-компоненти такої федерації не підлягають управлінню адміністратора ФДБ. У деяких джерелах слабопов'язані ФДБ називаються *інтероперабельними* БД (interoperable database). У сильнопов'язаній ФДБ на її адміністратора покладається відповідальність за створення й управління федерацією та активним контролем за доступом до компонентів-БД. У гібридному підході робиться спроба об'єднати переваги попередніх двох підходів. Зокрема, можливість прямого доступу до багатьох інших сховищ даних та використання глобальної схеми для отримання інформації про локальні схеми.

Система є *однофедеративною* (single federation), якщо в ній можна представити не більше однієї федеративної схеми. В іншому випадку система є *багатофедеративною* (multiple federation).

Були здійснені дослідження з архітектури ФДБ, всебічний аналіз яких подано в огляді [454].

У 80-х роках розроблено ряд слабопов'язаних [1178–1180], сильнопов'язаних однофедеративних [1181, 1182] і сильнопов'язаних багатофедеративних ФДБ [1183, 1184]¹.

¹ Як було зазначено в [454], у 80-х роках термін мультибази даних широко використовувався дослідниками в різних контекстах. Зокрема, в роботах [1178, 1179] під ними малися на увазі слабопов'язані ФДБ, у [1185, 1186] – сильнопов'язані багатофедеративні БД, а в [1187 – 1189] – сильнопов'язані однофедеративні.

У другій половині 80-х років була опублікована ціла серія аналітичних оглядів на дану тематику [454, 1190 – 1194], що свідчить про її велику на той час популярність.

Багатоваріантне зберігання. Відродження ідеї мультимодельності баз даних припало на початок 2000-х років. 2006 року Ніл Форд (Neal Ford) висловив ідею багатомовного програмування (Polyglot Programming) [1195]. Її суть полягала в наступному. Кожна мова програмування найкраще придатна для вирішення задач певного класу. Через це, створюючи крупні системи, бажано використовувати не одну мову, а кілька, програмуючи кожну функціональну задачу тією мовою, яка реалізує її найефективніше.

На підставі цієї ідеї 2008 року Скотт Лебернайт (Scott Leberknight) ввів поняття *багатоваріантного зберігання* (polyglot persistence) [1196] для баз даних.

Мається на увазі можливість надавати прийнятні способи представлення, зберігання і маніпулювання даними для різних класів задач розроблюваної системи з використанням багатьох одномодельних різнотипних баз даних і програми-посередника для інтеграції цих баз даних.



Скотт Лебернайт

Було розроблено ряд дослідницьких прототипів, підтримуючих концепцію багатоваріантного зберігання [1197 – 1199]. Так у системі Spark SQL [1199]

надається API, де з допомогою мов DataFrames і SQL можна працювати з такими сховищами даних, як JSON, JDBC, Hive, ORC і Parquet.

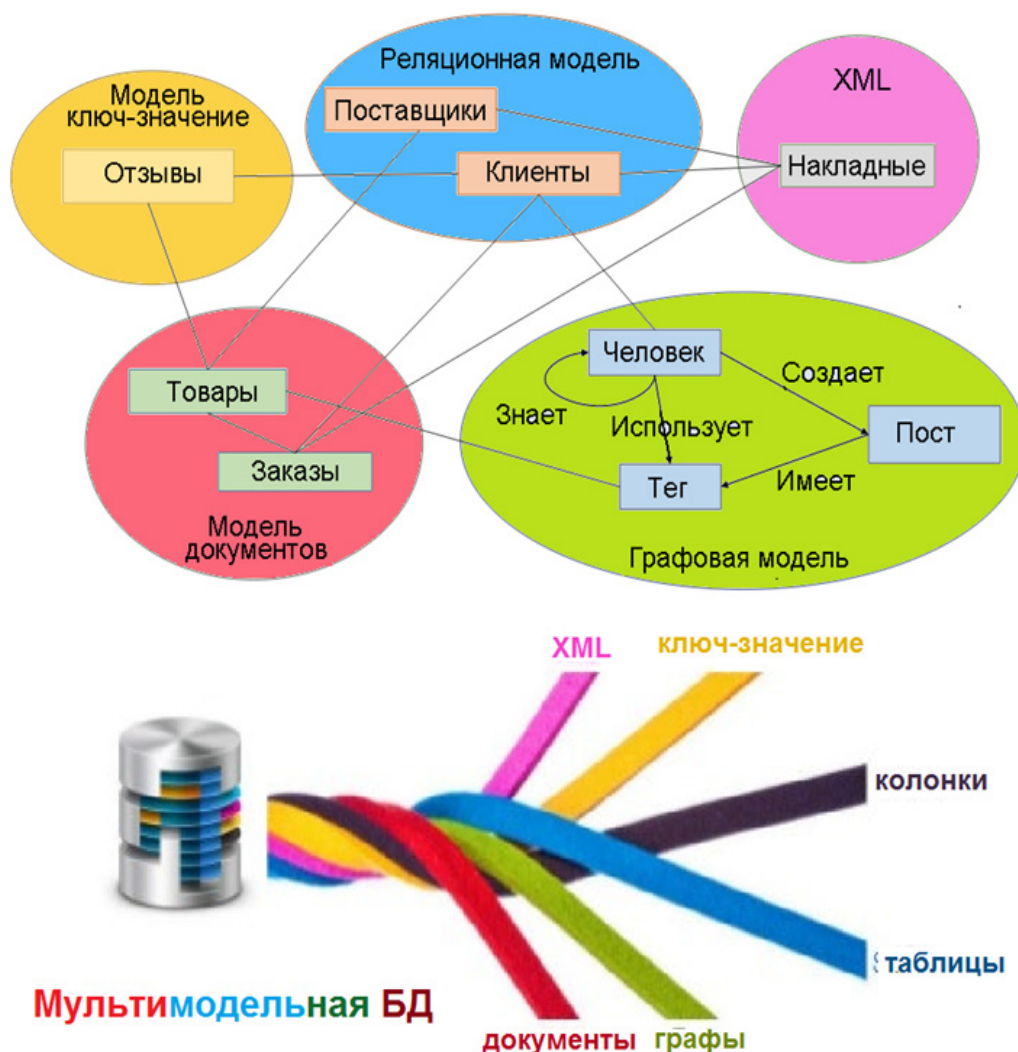
Мультимодельні БД. 2009 року Лука Гаруллі (Luca Garulli) вперше розробив і випустив на ринок мультимедійну БД Orient DB. Згодом, 2012 року, на конференції “Nosql Matters 2012” у Кельні (Німеччина) він уперше застосував термін *багатомодельність* відносно баз даних і передбачив розвиток систем класу NoSQL у нові системи з додатковою функціональністю та підтримкою різних моделей даних включно [1200]. Він запропонував створювати єдині інтегровані багатфункціональні NoSQL-продукти замість того, щоб збирати різні, окремі системи NoSQL разом для забезпечення аналогічного результату. Відтоді саме в розумінні Гаруллі почали застосовувати термін «мультимодельна база даних». Тобто єдина система баз даних, що підтримує велику кількість моделей даних. Зазначимо, що Гаруллі розробив також СУБД Arcade DB і Arcade Trader.



Лука Гаруллі

Мультимодельна БД (ММБД) – це БД, яка підтримує велику кількість моделей даних у межах однієї інтегрованої СУБД і забезпечує стандарти даних і стандарти мов запитів кожної з моделей.

На рис. нижче наведено графічну інтерпретацію мультимодельної БД.



Теорія категорій. Було здійснено ряд досліджень [1201 – 1203], де обґрунтовується застосування теорії категорій для формального опису відображень між моделями даних у ММБД.

ММБД мають усі характерні для БД властивості:

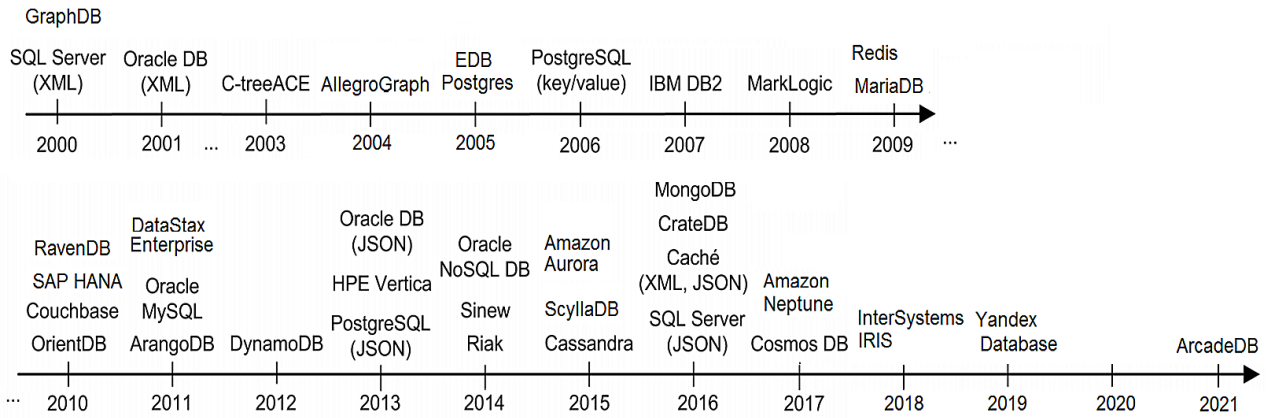
- зберігання даних, їх резервне копіювання й відновлення;
- розвинуті мови запитів і механізми індексування;
- підтримку ACID транзакцій;
- «безшовну» інтеграцію різних моделей;
- розвиток механізмів захисту й обмежень прав доступу.

Системи ММБД. На рис. нижче наведено часову вісь із зазначенням років появи мультимодальних систем завдяки, або розширенню вихідного формату новим, або початковій реалізації як мультимодальної СУБД. Еволюція систем природним чином відпо-

відала зростанню популярності відповідних моделей. Так, зокрема, перша хвиля появи ММБД припадає на початок 21-го століття у зв'язку з появою XML. Реляційні СУБД почали включати XML, застосовуючи стандарт SQL/XML, або ж його діалекти. Друга хвиля пов'язана із другим десятиліттям у зв'язку із настанням епохи NoSQL і великих даних. Уже 2015 року провідні аналітики Gartner заявили: «*Майбутнє СУБД, їхніх архітектур і способів використання — мультимодельність*» [1204]. Від 2017 року всі лідируючі виробники СУБД пропонують мультимодальні рішення, реляційні і NoSQL, на основі єдиної платформи.

Існують три стратегії підтримки мультимодельності [1205]:

- розробка принципово нової стратегії зберігання для підтримки багатьох моделей даних;
- розширення вихідної стратегії зберігання для підтримки додаткових моделей даних;



– створення нового зовнішнього інтерфейсу на основі вихідної стратегії зберігання.

В наступній таблиці наводяться ММБД відповідно до цієї класифікації.

Далі наводиться таблиця із зазначенням моделей/структур даних, що підтримуються відповідними ММБД. У списку додаткових моделей є стовпчик, який включає об’єктну модель, визначені користувачем типи й вкладені структури даних. Популярність СУБД взято із сайту DB-Engines

Ranking (<https://db-engines.com/en/renking>) станом на березень 2022 року.

Порівняльний аналіз ММБД. Із зростанням кількості платформ мультимодальних баз даних почалися дослідження з їхнього порівняльного аналізу. До прикладу, в працях [1206 – 1208] подається аналіз існуючих ММБД і порівняльні оцінки ММБД з іншими SQL і NoSQL БД.

Огляди з ММБД. Упродовж останніх років було написано кілька аналітичних оглядів і монографій із ММБД [454, 1177,

Стратегія підтримки мультимодельності	Тип исходной модели	СУБД
Новая стратегия хранения	реляционная	PostgreSQL
		SQL server
		IBM DB2
		Oracle DB
	колоночная	Cassandra
		CrateDB
		DynamoDB
ключ/значение	Riak key	
документная	Cosmos DB	
Расширение исходной стратегии хранения	реляционная	MySQL
	колоночная	HPE Vertica
	документная	ArangoDB
		MongoDB
	графовая	OrientDB
объектная	Cache	
Новый внешний интерфейс на основе исходной стратегии хранения	реляционная	Sinew
	ключ/значение	C-treeACE
		Oracle NoSQL Database
	документная	Couchbase
	MarkLogic	

1205, 1209 – 1211]. Даний розділ написаний переважно на основі матеріалу оглядів [454, 1210].

Полісховища (polystore). Полі-БД (poly-database) – це мультибаза даних, яка інтегрує багато гетерогенних баз даних і надає велику кількість інтерфейсів для формулювання запитів [1209]. Полі-БД поєднує властивості мультимодальних і багатоваріантних БД. Як і мультимодальна, вона підтримує багато гетерогенних моделей даних, водночас, як і багатоваріантна, надає велику кількість зовнішніх інтерфейсів під кожну з підтримуваних моделей даних. Вважеться, що першою системою полі сховищ була BigDAWG, представлена 2015 року [1198]. Крім того, до цього класу належать також CloudMdsQL [1212], Myria [1213], Apache Drill [1214], QoX [1215], Musketeer [1216], Rheem [1217], AWESOME [1218].

Етап 6. Великі дані (2010 – 2020+)

Світовий обсяг оцифрованої інформації зростає по експоненті. Від початку 1980-х років цифрова інформація подвоюється щокожні 40 місяців. Згідно з даними компанії IBS, до 2003 року світ накопичив 5 ексабайтів даних (1 ЕБ = 1 млрд. гігабайтів), а нині така кількість відтворюється кожні два дні. До 2008 року цей обсяг зріс до 0,18 зетабайта (1 ЗБ = 1024 ексабайта), до 2011 року – до 1,76 зетабайта, до 2013 – до 4,4 зетабайта. В травні 2015 року глобальна кількість даних перевищила 6,5 зетабайта. 2025 року, за прогнозами, людство сформує 163 зетабайта інформації.

Наведемо цитату з [1219], яка розкриває суть проблеми великих даних: «Даних стає все більше й більше, але водночас з поля зору випадає та обставина, що проблема ажніяк не зовнішня, вона викликана не стільки невимовно великою кількістю даних, як неспроможністю старими методами впоратися з новими обсягами. Спостерігається дисбаланс – здатність породжувати дані виявилася сильнішою за здатність їх переробляти... За ім'ям Big Data приховується якісний перехід у комп'ютерних технологіях, здатний викликати серйозні зміни, Не випадково цей перехід називають новою технічною революцією».

Широке використання терміну «великі дані» пов'язують із Кліффордом Лінчем (Clifford Lynch), редактором журналу Nature. 3 вересня 2008 року він підготував спеціальний випуск номера найстарішого британського наукового журналу, присвячений пошуку відповіді на питання: «Як можуть вплинути на майбутнє науки технології, які уможливають роботу з великими обсягами даних?» [1220].



Кліффорд Лінч

Варто також підкреслити, що вперше цей термін усе ж озвучив Джон Р. Маші (John R. Mashey) 1998 року [1221, 1222], який з приводу застосування терміну сказав: «Мені була необхідна найпростіша й коротка фраза, аби вказати, що межі обчислювальної техніки продовжують розширюватися».



Джон Р. Маші

Исходная модель	СУБД	Дополнительные модели/структуры							Популярность (2022)	Адрес	
		Реляционная/SQL	Колоночная	Ключ/значение	JSON	XML	Графовая	RDF			Вложенная/ОПТ/объектная
Реляционная	PostgreSQL	P		P	P	P			P	*****	https://wiki.postgresql.org/wiki/Main_Page
	SQL Server	P			P	P	P		P	*****	https://ru.wikipedia.org/wiki/Microsoft_SQL_Server
	IBM DB2	P				P	P	P	P	*****	https://en.wikipedia.org/wiki/IBM_Db2
	Oracle DB	P		P	P	P	P	P	P	*****	https://en.wikipedia.org/wiki/Oracle_Database
	Oracle MySQL	P		P					P	*****	https://en.wikipedia.org/wiki/MySQL
	Sinew	P		P						*	
	SAP HANA	P			P		P			*****	https://en.wikipedia.org/wiki/SAP_HANA
Колоночная	Cassandra		P				P		P	*****	https://ru.wikipedia.org/wiki/Apache_Cassandra
	CrateDB	P	P		P		P			***	https://en.wikipedia.org/wiki/CrateDB
	DynamoDB		P	P	P		P		P	*****	https://en.wikipedia.org/wiki/Amazon_DynamoDB
	HPE Vertica		P		P		P			***	https://en.wikipedia.org/wiki/Vertica
Ключ/ значение	Riak			P	P	P	P			*****	https://en.wikipedia.org/wiki/Riak
	c-treeACE	P		P			P			*	https://en.everybodywiki.com/C-treeACE
	Oracle NoSQL DB	P		P			P	P		****	https://en.wikipedia.org/wiki/Oracle_NoSQL_Database
	Datastax	P		P			P			*****	https://en.wikipedia.org/wiki/DataStax
	Redis			P	P		P			*****	https://uk.wikipedia.org/wiki/Redis
Документная	ArangoDB			P	P		P			****	https://en.wikipedia.org/wiki/ArangoDB
	Couchbase			P	P					*****	https://en.wikipedia.org/wiki/Couchbase_Server
	MongoDB			P	P				P	*****	https://en.wikipedia.org/wiki/MongoDB
	Cosmos DB	P	P	P	P		P		P	*****	https://en.wikipedia.org/wiki/Cosmos_DB
	MarkLogic	P			P	P	P	P	P	*****	https://en.wikipedia.org/wiki/MarkLogic
	AllegroGraph				P		P			***	https://en.wikipedia.org/wiki/AllegroGraph
	ArcadeDB	P		P	P		P			*	https://en.wikipedia.org/wiki/ArcadeDB
	EnterpriseDB (EDB Postgres)	P		P	P	P				****	https://en.wikipedia.org/wiki/EnterpriseDB
Графовая	OrientDB	P		P	P		P			****	https://en.wikipedia.org/wiki/OrientDB
	GraphDB						P	P		****	https://db-engines.com/en/system/GraphDB
	Amazon Neptune						P	P		****	https://en.wikipedia.org/wiki/Amazon_Neptune
Объектная	InterSystems Caché	P			P	P			P	****	https://en.wikipedia.org/wiki/InterSystems_Cach%C3%A9

Цей термін одразу ж призвичаївся в академічному середовищі, передусім щодо проблеми росту і різноманітності наукових даних, а згодом широко розповсюдився в діловому світі. 2010 року з'являються перші продукти й технології, які безпосередньо стосуються проблеми обробки великих даних. До 2011 року найбільші постачальники інформаційних технологій у своїх ділових стратегіях починають використовувати поняття «Великі дані». Це, зокрема, стосується IBM, Oracle, Microsoft, Hewlett-Packard, EMC. А основні аналітики ринку інформаційних технологій присвячують концепції спеціальні дослідження. В цьому ж 2011 році аналітична компанія Gartner відзначила великі дані як тренд номер два в інформаційно-технологічній інфраструктурі (після віртуалізації). Із 2013 року великі дані як академічний предмет починають вивчати у вузівських програмах щодо науки про дані, обчислювальних науках та інженерії. 2015 року Gartner відзначила, що технологія великих даних перейшла від етапу галасу до практичного застосування.

Існує багато визначень великих даних [1223, 1224]. Узагальнюючи ці та інші матеріали, дамо наступне визначення.

Великі дані (Big Data) – це величезні обсяги неоднорідної, неструктурованої або слабо структурованої, суттєво розподіленої та інтенсивно зростаючої, мінливої й використовуваної цифрової інформації, яку неможливо обробити традиційними засобами. Це також методи, технології й засоби їх збору, зберігання, обробки й аналізу з метою отримання сприйнятних людиною результатів.

Характеристичні властивості великих даних. 2001 року Дуглас Лейні (Douglas Laney), аналітик Gartner Inc., сформулював визначальні характеристики сучасних даних [1225], які дістали назву «Три V»: Volume, Velocity, Variety (обсяг, швидкість, різноманітність). І хоча він не говорив про великі дані, а просто про дані, однак у науковому середовищі ці три властивості почали розглядатися як визначальні характеристики саме великих даних.



Дуглас Лейні

Згодом Зікопулос (Zikopoulos) [1226] запропонував додати ще 2 ознаки – достовірність і цінність (Veracity, Value), отримавши таким чином «П'ять V». З часом були висунуті додаткові визначальні характеристики Big Data [1227 - 1230], що дістали назву «Сім V» і «Десять V». Серед науковців заведено вважати, що великі дані починаються з обсягів у петабайти і з інформаційними потоками у 100 ГБ на добу.

Класифікація великих даних. Редактор журналу Web 2.0 Journal Дайон Хінчкліф (Dion Hinchcliffe) дав класифікацію Big Data [1231, 1232], яка дозволяє співвідносити технологію з результатом, на який чекають від обробки Big Data. Хінчкліф поділяє підходи до Big Data на три групи: Fast Data (швидкі дані), їх обсяг вимірюється терабайтами – петабайтами; Big Analytics (велика аналітика – петабайтні – екзабайтні дані та Deep Insight (глибоке розуміння) – екзабайти – зетабайти. Групи відрізняються між собою не лише обсягами даних, а й якістю рішення задач із їхньої обробки.

Швидкі дані. Розуміючи, що традиційні методи зберігання, переміщення, обробки й вибірки даних недостатні, індустрія великих даних створила абсолютно новий набір методів і адаптувала деякі з існуючих, що дозволило обробляти всю сукупність інформації за прийнятний час. Обробка для Fast Data не передбачає отримання нових знань, її результати співвідносяться із апріорним знанням і дозволяють робити висновки про те, як відбуваються ті чи інші процеси. Вона дозволяє краще й детальніше

побачити все, що відбувається, підтвердити або ж заперечити певні гіпотези. Лише невелика частина з існуючих наразі технологій придатна для вирішення задач Fast Data. До цього списку потрапляють деякі технології роботи зі сховищами (продукти Hadoop, MapReduce, Greenplum, Netezza, Oracle Exadata, Teradata, СУБД типу Verica й kdb). Швидкість роботи цих технологій має зростати синхронно із зростанням обсягів даних.

Велика аналітика. Задачі, які вирішуються засобами Big Analytics, помітно відрізняються, до того ж не лише кількісно, а і якісно. А відповідні технології мають допомагати в отриманні нових знань. Вони слугують для перетворення зафіксованої в даних інформації в нове знання. Однак на цьому середньому рівні не передбачається наявність штучного інтелекту у виборі рішень або будь-яких автономних дій аналітичної системи – вона будується за принципом «навчання з учителем». Інакше кажучи, весь її аналітичний потенціал закладається в неї у процесі навчання. Класичними представниками такої аналітики є продукти MATLAB, SAS, Revolution R, Apache Hive, SciPy Apache і Mahout.

Глибоке розуміння. Могутніх, але не сфокусованих інструментів Big Analytics недостатньо, аби зажити слави великих даних. Deep Insight передбачає цілеспрямоване навчання без учителя (unsupervised learning) і використання сучасних методів аналітики, застосовуваних у конкретних областях, а також різні способи візуалізації. На цьому рівні можливе виявлення знань і закономірностей, апіорно невідомих. Методи глибокого проникнення дозволяють перетворити всю інформацію в оперативно діючий колективний інтелект.

Принципи роботи. Виходячи з визначення Big Data, можна сформулювати наступні основні принципи роботи з такими даними [1233].

Розподіленість. Зберігати інформацію в одному місці безглуздо і практично неможливо. Тому технологія роботи з Big Data має використовувати розподілене зберігання, управління, обробку й аналіз даних, що зберігаються в різних сховищах даних в усьому світі.

Горизонтальна масштабованість. Оскільки даних може бути як завгодно багато – будь-яка система, яка має на меті обробку великих даних, повинна бути розширюваною. Вдвічі зріс обсяг даних – удвічі збільшився кластер, і все продовжує працювати з такою ж продуктивністю.

Відмовостійкість. Принцип горизонтальної масштабованості припускає, що машин у кластері може бути багато. Як-от, Hadoop –кластер має понад 42000 машин. Це означає, що частина цих машин гарантовано буде виходити з ладу. Методи роботи з великими даними мають враховувати можливість таких збоїв і переживати їх без будь-яких значних наслідків.

Локальність даних. У великих розподілених системах дані розподілені серед великої кількості машин. У разі, якщо дані фізично знаходяться на одному сервері, а обробляються на іншому, то витрати на передачу даних можуть перевищувати витрати на саму обробку. Тому одним із найважливіших принципів проєктування Big Data-рішень є принцип локальності даних. За можливості обробляємо дані на тій же машині, де вони зберігаються.

Інтерпретація даних у процесі їх обробки (schema-on-road). Традиційні бази даних працюють за принципом schema-on-road – спочатку необхідно визначити схему і лише після цього можна вводити дані і працювати з ними. В Big Data робота з даними може здійснюватися за принципом schema-on-road – дані надходять до сховища такими, як є, без будь-якого попереднього опису, без зазначення їхньої структури й семантики. І лише в процесі вибірки для обробки відбувається їх «осмислення».

Усі сучасні засоби роботи з великими даними так чи інакше дотримуються цих п'яти принципів.

Методи і технології. Дотепер створено і адаптовано велику кількість методів і технологій для збору, агрегування, маніпулювання, аналізу й візуалізації великих даних. Ці методи й технології запозичені з різних сфер, із статистикою, інформатикою, прикладною математикою й економікою включно. Це означає, що для отримання вигоди з великих даних, необхідно використовувати гнучкий міждисциплінарний підхід.

Деякі методи й технології були розроблені для оперування значно меншими обсягами і розмаїттям даних, але були успішно адаптовані для Big Data. Інші були розроблені останнім часом, зокрема, для збору й аналізу великих даних. У звіті [1234] підрозділу McKinsey Global Institute (MGI) міжнародної аудиторсько – консалтингової компанії McKinsey & Company наводяться методи й технології аналізу і візуалізації, придатні для Big Data. В ньому наводяться:

- *методи аналізу* (інтелектуальний аналіз даних – Data Mining, краудсорсинг (crowdsourcing), машинне навчання, штучні нейронні мережі, розпізнавання образів, імітаційне моделювання, просторовий аналіз, генетичні алгоритми тощо);

- *технології* (бізнес-аналітика – Business intelligence, хмарні обчислення, сховища даних, розподілені системи тощо);

- *засоби* (Big Table, Cassandra, Dynamo, Google Fill System, Hadoop, MapReduce);

- *методи візуалізації* (хмара тегів (Tag cloud), кластерграма (Clustergram), історичний потік (History flow), просторовий інформаційний потік (Spatial information flow)).

Модель великих даних. Реляційна модель даних (РМД) не застосовується для великих даних. Її структура строго формалізована, в свою чергу великі дані можуть бути слабоструктурованими, або ж взагалі не мати структури. РМД передбачає існування схеми, а великі дані можуть бути безсхемними. Реляційна алгебра для великих даних абсолютно не придатна, Проблема незалежності даних взагалі не постає перед великими даними. Тому класичні архітектурні рішення РМД у вигляді архітектури ANSI/X3/SPARC не застосовні. Ідея концептуальної інформаційної моделі у великих даних відсутня. Гордість РМД – теорія залежностей і нормальних форм абсолютно не придатна через те, що нова декомпозиція стає згубною для великих даних. Тож для них більш природна концепція існування єдиного універсального відношення. Ще одна гордість РМД – принцип ACID для транзакцій є задорогим, неефективним і непотрібним задоволенням.

У зв'язку з цим у великих даних застосовують моделі даних NoSQL. Найпопулярнішою є модель ключ – значення. На цій моделі визначена модель обчислень MapReduce – модель розподіленої обробки даних, запропонована компанією Google для обробки великих обсягів «сирих» даних на комп'ютерних кластерах (великої кількості комп'ютерних вузлів). MapReduce була розроблена співробітниками Google Джеффри Діном (Jeffrey Dean) і Санджайем Гемаватом (Sanjay Ghemawat) [1235].



Джеффри Дін



Санджай Гемават

Стаття має величезну популярність. На момент підготовки даного матеріалу вона була опублікована в двох джерелах, і на неї було зроблено понад 33000 посилань.

Ця модель обчислень передбачає виконання трьох етапів.

Етап Map – попередня обробка і фільтрація вхідних даних у вигляді великого списку значень. При цьому головний вузол кластера отримує цей список, ділить його на частини й передає робочим вузлам. Кожен робочий вузол задіює функцію Map до локальних даних і в результаті видається велика кількість пар «ключ – значення». Що саме буде знаходитися в ключі і в значенні – вирішувати користувачеві.

Етап Shuffle. Відбувається непомітно для користувача. На цій стадії на кожному робочому вузлі на основі ключів, створених функцією Map, значення «розбираються по корзинах» (сортуються). Кожна корзина відповідає одному ключу виводу стадії Map. Ці корзини слугують входом для Reduce.

Етап Reduce. Кожна корзина зі значеннями, сформованими на етапі Shuffle, потрапляє на вхід функції Reduce. Ця функція задається користувачем і обчислює фінальний результат для окремої «корзини». Множина всіх значень, повернутих функцією Reduce, є фінальним результатом MapReduce задачі.

Попри простоту MapReduce, її перевага в тому, що це архітектура, яка забезпечує:

- автоматичне розпаралелювання даних із величезного масиву по великій кількості вузлів обробки, що виконують процедури MapReduce;

- ефективне балансування завантаження цих обчислювальних вузлів, яке не дозволяє їм простоювати або бути занадто перевантаженими;

- технологію відмовостійкої роботи, яка передбачає, що під час виконання загального завдання частина вузлів обробки може вийти з ладу, або з якоїсь іншої причини припинить обробку даних.

Дослідження й розробки баз даних у колишньому СРСР (1970 – 1991)

Опис досліджень і розробки в галузі баз даних буде неповним без урахування

результатів, отриманих в цій царині в колишньому СРСР. На початку цього сторіччя відомий учений – енциклопедист в галузі баз даних М.Р. Когаловський опублікував фантастичну монографію «Енциклопедія технологій баз даних»². В ній є розділ «Вітчизняні дослідження і розробки» у главі «Короткий нарис еволюції технологій баз даних». Тут міститься ґрунтовний матеріал з історії баз даних в СРСР. Власне, весь наступний текст є суттєво скороченим і дещо переробленим змістом вище згаданого розділу. Довелося опустити значну частину бібліографії і численні посилання на розділи енциклопедії. За можливості посилання замінені прізвищами авторів. Єдине, що додано нами, це останній розділ із персоналіями.

Щоб уявити обсяг здійснених на той час досліджень, наведемо фрагмент із монографії М.Р.Когаловського, який містить бібліографію в 380 назв: «... ми змушені переважно обмежитися тут мінімальною кількістю бібліографічних посилань, бо досить значна бібліографія зайняла б сотні сторінок. Навіть порівняно фрагментарна бібліографія, укладена за ініціативою члена-кореспондента Академії наук СРСР А.О.Стогнія, опублікована 1984 року, являє собою видання обсягом понад 200 сторінок».³

Організація й інфраструктура досліджень і розробок

Попри те, що потреби й передумови розвитку досліджень і прикладних розробок, пов'язаних із технологіями баз даних, існували в країні й раніше, активна діяльність у цій сфері розгорнулася лише на початку 70-х років. Саме в цей період почалося масове виробництво обчислювальних систем третього покоління ЕС ЕВМ, що мали дискове обладнання зовнішньої пам'яті прямого доступу, без якої неможливе створення систем баз даних.

Першим великим форумом зацікавлених у цій галузі спеціалістів стала Всесоюзна конференція з автоматизованих сис-

² Когаловский М. Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002г., 800с.

³ Банки данных и информационно-поисковые системы. Библиографический указатель. Киев: АН УССР. Институт кибернетики им. В.М.Глушкова, 1984. 231 с.

тем управління, яка відбулася восени 1973 року в Ташкенті і проводилася в Інституті кібернетики Узбецької академії наук. На конференції працювала спеціалізована секція «Банки даних». Конференція привернула увагу до проблематики баз даних в країні. До цього часу в кількох організаціях уже здійснювалися розробки інструментального програмного забезпечення.

Конференція виявила гостру потребу в створенні постійно діючого науково-суспільного організаційного ядра вітчизняної спільноти спеціалістів у галузі баз даних. Цю функцію було покладено на засновану Державним комітетом з науки і техніки 1974 року Робочу групу з програмного забезпечення банків даних (РГБД), пізніше (1984) реорганізовану в Науково-технічну комісію ДКНТ із баз даних. Ця комісія функціонувала до 1987 року.

Головою РГБД протягом усього періоду її функціонування був Г.К.Столяров (Інститут математики АН БРСР), заступниками голови – Л.А.Калініченко (ІНЕУМ), В.М.Савинков (ВНІПОУ) і А.О.Стогній (Інститут кібернетики АН УРСР), ученим секретарем – В.П. Дерибас (Інститут математики АН БРСР).

РГБД проводила всесоюзні конференції із банків даних і всесоюзні піврічні семінари, створювала тимчасові цільові й експертні підгрупи. Вона також видавала «Звіти РГБД» і методичні матеріали, керувала редакційною радою основного періодичного збірника із СУБД і БД «Прикладна інформатика», співпрацювала з міжнародними профільними групами USA CODASYL DBTG і British Computer Society DBAWG.

Пізніше, 1978 року, для розгортання й координації робіт у даній галузі в рамках Академії наук СРСР була створена Комісія з банків даних та інформаційно – пошукових систем при Президіумі Академії наук, що функціонувала до 1991 року.

Головою Комісії протягом усього цього періоду був член-кореспондент АН СРСР А.О.Стогній, заступником голови – В.І.Філіпов (ОЦ АН СРСР), а вченим секретарем – П.І.Андон (СКБ програмного забезпечення ІК АН України).

Спільними зусиллями РГБД і Комісії було організовано п'ять Всесоюзних кон-

ференцій із банків даних. Здійснювалося формування державних планів наукових досліджень із профілю цих організацій, розроблялися методичні матеріали, проводилася експертиза розробок великих систем програмного забезпечення.

Зрештою, тут необхідно відзначити важливу роль низки вітчизняних видавництв. Провідну роль у виданні монографічної літератури вітчизняних і зарубіжних авторів із проблематики систем баз даних відіграло видавництво «Фінанси і статистика». Чималим також є внесок видавництв «Наука» і «Світ».

Створення програмного інструментарію

Однією з необхідних умов практичного використання технологій баз даних є оснащення організацій – розробників і користувачів додатків необхідним програмним інструментарієм, передусім системами управління базами даних. Як уже зазначалося, на початковому етапі розвитку технологій баз даних в країні не існувало таких програмних засобів і можливостей їх придбання за кордоном. Тому необхідно було здійснити самостійні розробки СУБД попри відсутність досвіду створення таких складних програмних систем.

Перші кроки у вирішенні цієї задачі здійснювалися на початку 70-х років, коли почалося виробництво обчислювальних машин сімейства ЄС ЕВМ. Роботи спрямовувалися у двох напрямках. Передусім були здійснені спроби створення власних оригінальних вітчизняних СУБД. Водночас у пришвидшеному режимі розроблялися аналоги деяких широко розповсюджених за кордоном СУБД, здатних функціонувати на вітчизняних апаратно-програмних платформах. Подібний підхід було застосовано також під час створення СУБД для апаратних платформ, серійне виробництво яких почалося в країні після появи платформи ЄС ЕВМ, - для СМ ЕВМ, АСВТ, ІВМ-сумісних персональних комп'ютерів тощо.

Імовірно, першим проєктом в країні, спрямованим на створення оригінальної вітчизняної СУБД, сумірної із передовими досягненнями міжнародного рівня, була СУБД типу CODASYL НАБОБ для платформи ЄС ЕВМ, розробка якої почалась у

вказаний період у ВДПТІ ЦСУ СРСР.

Поряд із СУБД НАБОБ згодом були розроблені також і інші оригінальні вітчизняні системи. СУБД типу CODASYL ПАРМА для платформи ЄС ЕВМ із операційною системою ОС ЄС була створена НДІ УМС (м. Перм). В Обчислювальному центрі Академії наук було розроблено СУБД типу CODASYL КОМПАС для платформи БЕСМ-6 із операційною системою ДІС-ПАК.

В Інституті проблем управління було розпочато роботи з реалізації СУБД ієрархічного типу для платформи ЄС ЕВМ, згодом продовжені і ВНДІСІ й завершені створенням системи ІНЕС користувальницьких пристроїв. В Інституті кібернетики АН УРСР було створене сімейство сумісних реляційних СУБД ПАЛЬМА для платформ ЄС ЕВМ, СМ ЕВМ і ІВМ-сумісних персональних комп'ютерів. У Міністерстві легкої промисловості Латвії була розроблена розвинута реляційна СУБД ВІРА. Інститут системного програмування РАН створив мобільний SQL-сервер на платформі UNIX і як засіб вільно поширюваного програмного забезпечення передав його разом із вихідними текстами до консорціуму Free Software Foundation. Інститутом системного аналізу РАН розроблена мультимедійна СУБД НІКА для персональних комп'ютерів. Варто відзначити також створення у Воронежському СКТБ «Системпрогра» сімейства сучасних реляційних СУБД ІНТЕГРАЛ для різних програмно-апаратних платформ.

Серед СУБД та інших засобів програмного забезпечення систем баз даних, створених в країні, що мають зарубіжні аналоги, найширшого розповсюдження набули СУБД СІНБАД (МНІПІ АСУ ГХ); СУБД ОКА і телемонітор КАМА Інституту кібернетики АН УРСР; СУБД ДІСОД, розроблена НДІ «Восход»; система БАНК Пермського НДІУМС; створена радянсько-болгарським інститутом ІНТЕРПРОГРАМА в Софії система СЕДАН; система РЕБУС Всесоюзного науково-дослідного інституту непромислової сфери та ряд інших розробок. Зауважимо, що деякі зі згаданих систем, зокрема, ОКА і ДІСОД мали надзвичайно розвинуте оточення, функціонально значно багатше, ніж у систем прототипів.

Розробка додатків

Головною сферою застосування технологій баз даних у 70-і – 80-і роки в СРСР були автоматизовані системи управління різного рівня в економіці. Розроблялися такі унікальні системи макrorівня, як Автоматизована система планових розрахунків (АСПР) Держплану країни і планових органів республік та Автоматизована система Державної статистики (АСДС). Дещо згодом СУБД стали невід'ємним компонентом програмного забезпечення числених галузевих систем управління.

Одначе наймасовішою сферою застосування були автоматизовані системи управління підприємствами. Типову архітектуру таких систем і комплекс типових додатків розробив інститут «Центропрограмсистем» (м. Калінін). Цей інструментарій використовувався на практиці численними промисловими підприємствами країни.

Активну підтримку діяльності у вказаному напрямку надавав міжнародний Радянсько – Болгарський інститут «Інтерпрограма» (Софія), який створив різноманітне типове програмне забезпечення, широко розповсюджене в обох країнах.

У 80-і роки на основі технологій баз даних було створено ряд інформаційних систем центральних організацій різних відомств – патентної служби, Держстандарту, Вищої атестаційної комісії, Всесоюзного науково-технічного інформаційного центру тощо. СУБД почали використовуватися для створення інформаційних систем на транспорті та будівництві, у найбільших державних бібліотеках, в системах управління складними технічними системами, в багатьох інших галузях. Однак усі ці розробки були доступні лише великим організаціям, спроможним утримувати в своїй структурі центри обробки даних.

Радикальна зміна ситуації відбулася у другій половині 80-х років, коли в країні почали з'являтися персональні комп'ютери. Навіть доволі скромні за своїми функціональними можливостями й надзвичайно прості в користуванні СУБД, створені для цієї швидко прогресуючої апаратної платформи, уможливили застосування найпростіших технологій баз даних у системах обробки даних для задоволення інформа-

ційних потреб практично в будь-якій сфері життєдіяльності.

Деякі додаткові відомості про розвиток додатків технологій баз даних в СРСР 70-х -80-х років можна знайти в оглядах⁴.

Наукові дослідження

в галузі систем баз даних

Дослідження, пов'язані з розробкою нових СУБД. Фактор абсолютної новизни проблеми для вітчизняних спеціалістів вимагав здійснення досліджень на багатьох етапах реалізації першої вітчизняної СУБД типу CODASYL НАБОБ. При створенні системи ІНЕС, що стала прототипом СУБД ІНЕС, була розроблена деревовидна структура індексу з блоками, що розщеплюються. В ІНЕС уперше серед ієрархічних СУБД застосована ідея самоописуваності баз даних, пропонується раніше для реляційних систем. На основі досвіду реалізації системи КОМПАС її авторами була запропонована інтегрована реляційно – мережева модель даних.

У процесі створення SQL-серверу Інституту системного програмування було узагальнено й ефективно використано досвід реалізації передових дослідницьких прототипів реляційних СУБД – System R і Ingres, втілені концепції відкритих систем і мобільності програмного забезпечення, використані деякі принципи об'єктного підходу. Творці системи ПАЛЬМА використали в своєму проєкті принципи багаторівневої архітектури СУБД і техніку відображення моделей даних.

Розвиток теорії реляційних баз даних. Проблеми математичної теорії реляційних баз даних викликали в країні такий же значний інтерес, як і за кордоном. Їм була присвячена велика кількість досліджень, здійснених переважно в 70-і – 80-і роки.

Більшість праць цього напрямку пов'язана з дослідженнями у галузі теорії залежностей і теорії нормалізації; з оцінкою можливостей реляційних мов, із питаннями

еквівалентності реляційних баз даних, з алгебраїчними аспектами реляційної моделі даних. Досліджувалися також аксіоматичні підходи в області реляційної моделі, формальні методи синтезу схем і логічного проєктування реляційних баз даних. Досліджувалися також взаємозв'язок логіки і реляційної моделі, питання обчислюваності реляційних запитів. Велику увагу привертала проблеми неповноти інформації в реляційних базах даних.

Моделювання даних. Вітчизняні роботи в цій царині почалися ще в 70-і роки. До цього напрямку належать, зокрема, дослідження, пов'язані зі створенням канонічної моделі даних для систем інтеграції неоднорідних баз даних [Л.А.Калініченко та інші] і моделей даних концептуального рівня архітектури мультимодельної багаторівневої СУБД [М.Р. Когаловський та інші]. Деякими авторами вводяться різні розширення реляційної моделі.

У середині 70-х років у мовах програмування сформувалася концепція абстрактного типу даних, яка вплинула на подальший розвиток підходів у сфері моделювання даних [О.В. Замулін та інші].

Потужніші моделі потребувалися в системах інтеграції неоднорідних інформаційних ресурсів. Одна з таких моделей визначається мовою Синтез [Л.А. Калініченко].

Дослідженню логіко-математичних основ моделювання даних присвячені праці В.І.Філіпова, В.А.Крахта, М.Ш.Цаленка.

Відображення моделей даних. У зв'язку з розробками розподілених систем баз даних, систем інтеграції неоднорідних баз даних і СУБД із багаторівневою архітектурою, серед них і мультимодальних систем, виникли проблеми відображення моделей даних. Їх вирішенню були присвячені дослідження радянських авторів, спрямовані на створення методів перетворення моделей даних і конструювання комутативних відображень [Л.А.Калініченко], на розробку архітектурних аспектів відображення

⁴ Перевозчикова О.Л., Ющенко Е.Л. Тенденции развития систем обработки данных// Программирование. 1977. №5. С.70-90.

Dale A.G. Database Management Systems Development in the USSR. Computing Surveys, Vol. 11, No. 3, 1979, pp.213-226.

моделей даних [М.Р.Когаловський] та специфікацію визначення відображень для конкретних моделей даних [Р.П.Крамаренко, А.Л.Віллемс].

СУБД із мультимодельним зовнішнім рівнем. У радянських дослідницьких проєктах, пов'язаних із розробками мультимодельних СУБД, існували два підходи. В першому з них [М.Р.Когаловський, М.М.Виноградов та інші] роль концептуальної моделі даних відіграє функціонально достатньо розвинута модель, яка забезпечує можливість відображення широко розповсюджених моделей.

Другий підхід орієнтувався на нові дослідження в мовах програмування. Водночас концептуальна модель, насправду, не фіксується в системі. В системі програмування баз даних АТЛАНТ [О.В.Замулін] передбачається можливість її специфікації як певної системи типів даних, котрі визначаються користувачем засобами інструментальної системи. Аналогічний підхід фактично застосовується в [Х.-М.Х Хаав], де інструментальною системою є система програмування ПРИЗ, на основі якої реалізована СУБД DABU.

Управління конкурентним доступом. Серед ранніх публікацій в галузі управління конкурентним доступом у системах баз даних передусім можна назвати працю [М.В.Оленін та інші], в якій досліджена і запропонована модель паралельних транзакцій для розподіленого об'єктного середовища. Варто також згадати здійснену в рамках проєкту вільно розповсюдженого мобільного SQL-серверу реалізацію методу серіалізації транзакцій, заснованого на двофазному протоколі предикатних блокувань [С.В.Кузнєцов та інші].

Г.Г.Домбровська досліджувала техніку підтримки вкладених транзакцій і транзакцій інших типів на рівні механізмів управління буферизацією в середовищі зберігання бази даних. Вона ж показала, що внесення деякої додаткової інформації в

дерево активних транзакцій дозволяє суттєво розширити сферу застосування техніки управління транзакціями.

Оптимізація запитів у системах баз даних. Варто відзначити фундаментальний аналітичний огляд⁵ і окремі статті С.Д.Кузнєцова, а також огляд В.І.Задорожного⁶ з оптимізації рекурсивних запитів у системах дедуктивних баз даних.

Системи програмування баз даних і знань. Ідеї створення мов програмування, котрі забезпечували б єдине ефективне середовище як для розробки додатків, так і для управління даними, вперше були висловлені О.В.Замуліним. Система програмування із вхідною мовою Бояз була реалізована на платформі БЕСМ-6 і використовувалася в деяких організаціях. Групою О.В.Замуліна було виконано великий комплекс досліджень в царині мов програмування баз даних, розроблено й реалізовано сучаснішу (порівняно з Бояз) мову Атлант.

Аналогічний доволі цікавий підхід у сфері створення мови програмування і бази даних був запропонований пізніше, наприкінці 70-х років в Інституті кібернетики АН Естонії. Засобами системи програмування високого рівня ПРИЗ, яку її ідеолог Е.Х.Тиугу кваліфікує як інструмент концептуального програмування, можна не лише програмувати додатки, а й описувати та підтримувати на стадії виконання потрібну модель даних для цього додатку. Саме так авторами була реалізована, зокрема, СУБД DABU.

Машини баз даних. Перші вітчизняні дослідження в цій сфері з'явилися наприкінці 70-х – початку 80-х років. Новий сплеск розробок на цю тему був пов'язаний із заснуванням в країні в середині 80-х років програми НІР зі створення засобів обчислювальної техніки нового покоління.

У монографії Л.А.Калініченка⁷ представлені результати здійсненого в Інституті проблем інформатики РАН комплексного дослідження представлення даних і знань в

⁵ Кузнєцов С.Д. Методы оптимизации выполнения запросов в реляционных СУБД// Сб. Итоги науки и техники. Вычислительные науки. Т.1. – М.: ВИНТИ, 1989. – С.76–145.

⁶ Задорожный В.И. Методы вычисления и оптимизации рекурсивных запросов в дедуктивных базах данных. Препринт докл.//V Всесоюз. конф. «Системы баз данных и знаний». Львов, 1991. 47 с.

⁷ Калиниченко Л.А., Рывкин В.М. Машины баз данных и знаний. М.: Наука, 1990. 296с.

машинах баз даних, їхньої архітектури і методів ефективної апаратної реалізації.

Об'єктні бази даних. У ранній праці С.Д.Кузнецова аналізуються найважливіші принципи об'єктивного підходу і концепції об'єктних СУБД. Водночас приділяється особлива увага аспектам моделювання даних, мовам запитів у таких системах та оптимізації об'єктних запитів. Принципи відображення розвинутих об'єктних моделей досліджувалися в рамках проєкту СИНТЕЗ Інституту проблем інформатики РАН.

Дедуктивні бази даних. До ранніх досліджень у цьому напрямку, можна віднести розробки Інституту прикладної математики Академії наук, в результаті якої була створена діюча система «Запит – відповідь» [Е.З.Любимський та інші]. Великий цикл теоретичних досліджень у напрямку дедуктивних баз даних здійснений спільно М.І.Дехтярем та А.Я.Діковським. Ґрунтовний аналіз і класифікація відомих методів обчислення й оптимізації рекурсивних запитів у системах дедуктивних баз даних подано в працях В.І.Задорожного. Йому належать також інші результати у сфері мов запитів і оптимізації в дедуктивних базах даних.

Розподілені бази даних. У 70-і роки в країні активізувалися роботи зі створення обчислювальних мереж. Водночас у стилі, цілком адекватному централізованому характеру управління економікою й іншими сферами життєдіяльності радянської держави, було поставлено масштабне завдання створення Державної мережі обчислювальних центрів (ДМОЦ). У проєкті такої мережі передбачалося й створення функціонуючих в її середовищі розподілених баз даних. Основні дослідження в цьому напрямку були розгорнуті в наукових закладах Москви (ІНЕУМ, ІПМ, ВДПТІ ЦСУ СРСР), Києва (ІК АН УРСР), Риги (ІЕВТ АН Латвійської РСР).

Публічні обговорення науково-технічних проблем розподілених баз даних почалися 1975 року, коли в Інституті кібернетики АН УРСР відбувся семінар «Принципи побудови РАБД державної мережі ВЦ». За рік ширший семінар із цих проблем організувала в Паневежисі РГБД спільно з Інститутом фізики й математики АН Литовської РСР.

Один із напрямків досліджень був присвячений розробці математичних моделей, які дозволяли б оптимізувати організацію і функціонування систем розподілених баз даних [Е.М.Беніамінов та інші].

Важлива проблема – організація неоднорідних розподілених баз даних із можливостями інтеграції даних – розглядалася відповідно до концепцій дослідницького проєкту СІЗІФ, який виконувався на той час в ІНЕУМ.

Уже на ранньому етапі досліджень розроблялися конкретні інструментальні засоби для створення розподілених баз даних. Прикладом може бути гібридна СУБД Базис (ІНЕУМ), яка підтримувала інтегровані бази даних із фактографічними й текстовими даними.

Інтеграція інформаційних ресурсів. Дослідження у сфері інтеграції інформаційних ресурсів почалися в СРСР у середині 70-х років у рамках робіт із створення розподілених баз даних.

Найяскравішим прикладом вітчизняних розробок цього періоду безумовно є передовий проєкт СІЗІФ Інституту електронних управляючих машин. У проєкті була розроблена архітектурна концепція системи інтеграції неоднорідних баз даних, заснована на інтегруючій канонічній моделі, яка забезпечує єдине представлення даних для всіх включених у систему даних. Це представлення – схема віртуальної бази даних – описується за допомогою спеціальної мови. Була запропонована також заснована на логіці предикатів мова маніпулювання даними, представленими в термінах цієї схеми.

Окрім цього авторами проєкту був розроблений метод побудови комутативних відображень моделей даних, що забезпечує підтримку відповідності між даними інтегрованих баз даних і даними віртуальної бази даних. Застосування цього методу було продемонстровано на прикладі відображення мережевої моделі даних CODASYL у реляційну модель.

У низці публікацій за матеріалами проєкту були продемонстровані можливості використання мови Синтез для однорідного опису інформаційних ресурсів, представлених засобами різних моделей структурованих і слабоструктурованих даних.

Проектування баз даних і розробка додатків. Чи не найпопулярніша сфера досліджень і розробок в галузі технологій баз даних пов'язана з проблемами проектування систем баз даних, вирішення яких має доволі важливе значення для забезпечення ефективного практичного використання цих технологій.

Пік активності вітчизняних досліджень у цій сфері припав на 80-і роки. Не випадково на 2-ій Всесоюзній конференції «Банки даних» для обслуговування проблем проектування баз даних було організовано спеціальну секцію. В цей період у різних наукових центрах країни над вказаною проблематикою успішно працювало кілька груп дослідників. Звісно, досить привабливим був напрямок, пов'язаний із формальними методами синтезу схем реляційних баз даних. Однак розроблялися й інші підходи, головною метою яких було створення засобів інфологічного моделювання предметної області системи бази даних і відображення його результатів у середовище конкретних СУБД.

Один із напрямків цих робіт був пов'язаний із створенням «інженерної» методики проектування концептуальної схеми бази даних у термінах, близьких до ER-моделі, і перетворення її в схему бази даних, вибраної проектувальником СУБД [В.В.Бойко та інші].

Більш формалізований підхід із використанням спеціально розроблених розвинутих засобів інформаційного моделювання був запропонований групою дослідників із ВНДПІ АСУ Газпрому і ВНДППОУ ДКНТ. За задумом авторів це дослідження мало б стати теоретичним базисом автоматизованої системи проектування баз даних. Було розроблено прототип такої системи – Омега-1.

Інший підхід, також націлений на автоматизовану технологію проектування був запропонований в ІК УРСР. Розроблена модель для опису предметної сфери підтримує ієрархію різного роду абстракцій. На її основі створено мову опису концептуальних схем. Реалізований прототип системи ПРОБАД, що базується на запропонованому підході.

У працях Г.І.Фурсіна та інших основні цілі полягали в створенні концептуальної моделі даних високого рівня, заснованої на обчисленні предикатів, технології її використання, а також інструментарію для підтримки процесу моделювання предметної області системи бази даних її засобами.

Цікавим є підхід В.М.Ветошкіна та інших, у якому джерелом інформації для формалізованого процесу синтезу схеми реляційної бази даних є вербальний опис предметної області. Розроблено також метод синтезу схеми бази даних, оптимальної щодо введеного автором критерію складності.

Поряд із вказаними підходами розвивався напрямок, пов'язаний із моделюванням семантики предметної області засобами, використовуваними в системах представлення знань [М.Ш.Цаленко, Е.Х.Тиугу, М.І.Кахро та інші].

Вибір і оцінка СУБД. Проблеми вибору СУБД для конкретних додатків або для додатків у певній специфічній предметній області, а також для оцінки характеристик їхнього функціонування актуальні на всіх стадіях розвитку технологій. Особливо, коли йдеться про розробки великих систем і систем із критичними вимогами до продуктивності, ресурсів пам'яті, надійності.

У вітчизняних розробках систем баз даних були спроби визначення сукупності факторів, здатних стати основою вибору й оцінки СУБД для конкретного додатку. Здійснювався порівняльний аналіз характеристик різних СУБД, пропонувалися методики оцінки й вибору СУБД для конкретних додатків.

Робилися також спроби оцінки характеристик функціонування СУБД з допомогою методів імітаційного моделювання [Г.К.Столяров, О.М.Вейнеров та інші]. Однак застосування техніки імітаційного моделювання не мало подальшого розвитку. Ймовірно, одна з причин криється в тому, що отримувані з допомогою дороговартісних імітаційних моделей, оцінки виявляються доволі грубими. Ефективнішими виявилися підходи, засновані на використанні засобів збору статистики функціонування, якими оснащені сучасні СУБД. Для оцінки продуктивності СУБД в середовищі деяких

типових додатків консорціумом ТРМ розроблені еталонні тести. Ще до заснування цього консорціуму близький підхід використовувався в дослідженнях Центрпрограмсистем, пов'язаних із отриманням порівняльних оцінок продуктивності промислово-супроводжувальних СУБД.

Персоналії

В цьому прикінцевому розділі подається узагальнений опис внеску окремих науковців, які брали участь у розробці проблематики баз даних в колишньому СРСР. Автор статті або був знайомий з ними по спільній роботі в РГБД, або зачитувався їхніми монографіями і статтями.

Наперед перепошую, що багатьох достойників, які здійснили суттєвий внесок у розвиток баз даних, не згадано в цьому розділі. Персональні відомості наведені в алфавітному порядку.

Андон Пилип Іларіонович



Академік НАН України, доктор фізико-математичних наук, заслужений діяч науки і техніки України. Лауреат державних премій в галузі науки і техніки УРСР і України, премій Ради Міністерств СРСР, премій НАН України ім. В.М.Глушкова й імені С.О.Лебедєва.

Член РГБД, член програмних комітетів Першої і Другої Всесоюзних конференцій «Банки даних». Учений секретар Комісії з банків даних та інформаційно-пошукових систем Координаційного комітету Академії наук СРСР із обчислювальної техніки.

Підготував 11 кандидатів і 5 докторів наук. Опублікував понад 200 наукових праць, зокрема, 5 монографій.

Під його керівництвом розроблено цілий ряд систем загальнодержавного і галузевого рівня. Він був головним конструктором систем ІНФОР і ЮПІТЕР, а також науковим керівником систем ПАЛЬМА, ОКА, КАМА.

Дрибас Віктор Прокопович

Співробітник Інституту математики АН БРСР. Секретар РГБД ДКНТ протягом усього часу її існування.

У 70-х роках опублікував досить популярний на той час препринт з реляційної моделі даних⁸, а його монографія з реляційної моделі баз даних⁹ упродовж багатьох років користувалася заслуженим авторитетом.

В.П.Дрибас також працював над моделюванням даних із багатозначною класифікацією об'єктів, а також над рекомендаціями щодо вибору баз даних.

Замулін Олександр Васильович (1943 – 2006)



⁸ Дрибас В. П., Курскова Г. Л., Столяров Г. К. и др Введение в реляционные модели базы данных. Минск: Препринт/ Ин-т математики АН БССР, №4(20), 1977, 54 с.

⁹ Дрибас В.П. Реляционные модели баз данных. Минск: БГУ, 1982, 192 с.

Учень А.П.Єршова, доктор фізико-математичних наук, професор, головний науковий співробітник Інституту систем інформатики ім. А.П.Єршова СВ РАН, зав. кафедрою Новосибірського державного університету.

А.В.Замулін активно працював у галузі інформативно-пошукових систем і систем управління базами даних. Він очолював створення інформаційно-пошукової системи загального призначення Вега для ЕОМ БЕОМ, що була на той час однією з кращих ІПС у СРСР.

Вважається одним із засновників нового наукового напрямку – створення систем програмування баз даних. Під його керівництвом була розроблена перша в світі мова програмування баз даних БОЯЗ (1976) і заснована на ній система програмування баз даних БОЯЗ-6 (1979); мова програмування баз даних Атлант (1986) і заснована на ній система програмування баз даних (1989); мова специфікації баз даних Руслан (1994), яка була визнана за кордоном.

О.В.Замулін опублікував понад 100 праць, зокрема, 2 монографії, присвячені типам даних у мовах програмування і базах даних¹⁰ та системам програмування баз даних¹¹.

Був співголовою РГБД і членом Комісії по банках даних Координаційного комітету АН СРСР із обчислювальної техніки.

О.В.Замулін був також членом редколегії журналу «Програмування» й міжнародних журналів “Information Systems”, “Universal Computer Science”, “The Computer Journal”, членом редколегії періодичної збірки статей «Системна інформатика».

Калініченко Леонід Андрійович (1937 – 2018)

Доктор фізико-математичних наук, зав. лаб. Інституту проблем інформатики РАН, професор ВМК МДУ, лауреат Державної премії СРСР в галузі науки і техніки, заступник голови РГБД.



Сфера наукових інтересів – методи інтеграції неоднорідних баз даних і управління розподіленими базами даних.

Доволі змістовний огляд¹² і монографії з питань інтеграції неоднорідних баз даних¹³ та машин баз даних¹⁴ не втратили корисності й цитуються дотепер. Наукові дослідження Л.А.Калініченка були впроваджені в системах Базис і Сізіф, мові СИНТЕЗ.

Підготував 10 кандидатів наук. Член редколегії журналу “Distributed and parallel databases”. Засновник і голова московської секції ACM SIGMOD.

Когаловський Михайл Рувимович



¹⁰ Замулін А.В. Типы данных в языках программирования и базах данных. Новосибирск: Наука, 1987. 152 с.

¹¹ Замулін А.В. Системы программирования баз данных и знаний. Новосибирск: Наука, 1990. 352 с.

¹² Калиниченко Л.А. и др. Архитектура и алгоритмы систем управления распределенными базами данных / Л.А.Калиниченко, О.Е.Костромина, О.Н.Хитрова. М.: ИНЭУМ, 1982. 140с.

¹³ Калиниченко Л.А. Методы и средства интеграции неоднородных баз данных. М.: Наука, 1983, 140с.

¹⁴ Калиниченко Л.А., Рывкин В.М. Машины баз данных и знаний. М.: Наука, 1990, 296с.

Учений в галузі баз даних та інформаційних систем, кандидат технічних наук, старший науковий співробітник, доцент, член редколегій журналів «Програмування», «Інформаційне товариство», «Електронні бібліотеки», професійний член ACM, учений секретар Московської секції ACM SIGMOD, провідний науковий співробітник Інституту проблем ринку РАН. Науковий редактор і перекладач російських видань монографій із баз даних Джефрі Ульмана, Кристофера Дейта, Алана Саймона, специфікацій мови визначення даних CODASYL, а також знаменитого звіту ANSI/X3/SPARC.

М.Р.Когаловський – член робочої групи з питань програмного забезпечення банків даних (РГБД) при Держкомітеті з науки і техніки впродовж усього часу її існування (1974 – 1987). Член і співголова Програмних комітетів низки великих міжнародних і вітчизняних наукових конференцій, має понад 200 друкованих праць, зокрема, 6 монографій. Його монографія «Енциклопедія технологій баз даних»¹⁵ оцінюється спеціалістами як «фантастично тяжка праця, яка реально закриває діру в літературі, присвяченій базам даних», а видання книги вважається винятково корисним як для вітчизняних спеціалістів, так і для світової громадськості¹⁶.

Пасічник

Володимир Володимирович

Доктор технічних наук, професор Національного університету «Львівська політехніка».

Вихованець наукової школи Інституту кібернетики імені В.М.Глушкова. Учасник і керівник багатьох міжнародних наукових проєктів і перспективних науково-дослідницьких розробок.

Автор 14 монографій і навчальних посібників, серед яких особливо виділяється монографія¹⁷, в якій досліджуються питання реляційної моделі баз даних, теорії залежностей і нормальних форм. Лауреат

Державної премії України в галузі науки і техніки, відмінник освіти України.



Працював провідним експертом із технологій баз даних і знань ДКНТ СРСР та країн – членів Ради економічної взаємодопомоги.

В.В.Пасічник підготував понад два десятки кандидатів і докторів наук у галузі баз даних і знань, інформаційного аналізу й сучасних інформаційних технологій.

Савінков

Володимир Макарович

Заступник директора з наукової роботи ВДПТІ ЦСУ СРСР. Заступник голови РГБД. Член організаційних і програмних комітетів Всесоюзних конференцій «Банки даних».

Відповідальний редактор збірки «Алгоритми й організація вирішення економічних задач» та «Прикладна інформатика». На той час вони були найавторитетнішими періодичними виданнями в країні, які публікували праці на тему систем баз даних та інформаційних систем.

В.М.Савінков – співавтор підручника по Алголу¹⁸, тлумачного словника з інформатики¹⁹ й монографії, присвяченої проєк-

¹⁵ Когаловський М.Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800с.

¹⁶ Кузнецов С.Д. «Энциклопедия технологий баз данных» Михаила Рувимовича Когаловского. - http://www.citforum.perm.ru/book/encdbd/encdbd_vv.shtml

¹⁷ Пасичник В.В., Стогний АА. Реляционные модели баз данных. Киев : ИК АН УССР, 1983. 286 с.

¹⁸ В.М. Савинков, В.Д. Цальп. Программирование на АЛГОЛе (Учеб. пособие для втузов). М. : Высшая школа, 1975. - 215 с.

¹⁹ Першиков В.И., Савинков В.М. Толковый словарь по информатике: Более 10000 терминов. Москва: Финансы и статистика, 1991. – 536 с.

туванню баз даних²⁰. Редактор перекладу з англійської мови добре відомої монографії Чарльза Мідоу²¹.

Стогній
Анатолій Олександрович
(1932 – 2007)

Доктор фізико-математичних наук, професор, член-кореспондент АН СРСР і член-кореспондент НАН України, заступник директора Інституту кібернетики НАН України, директор Інституту прикладної інформатики (м.Київ).



Лауреат Державної премії СРСР в галузі науки і техніки 1968 року в складі колективу розробників ЕОМ МИР-1. Лауреат премії ім. М.Островського, премії ім. В.М.Глушкова.

Сфера наукових інтересів: системи обробки даних, інформаційні системи і бази даних.

Заступник голови РГБД, голова Комісії з банків даних інформаційно-пошукових систем Координаційного комітету Академії наук СРСР із обчислювальної техніки. Член організаційного й програмного комітетів Всесоюзних конференцій «Банки даних».

Столяров
Геннадій Костянтинович



Ініціатор і беззмінний голова створеної в країні національної Міжвідомчої «Робочої групи з програмного забезпечення банків даних (РГБД)» (1973 – 87 рр.), яка об'єднала провідних розробників банків даних СРСР. Наглядач від Академії наук СРСР у робочих групах із баз даних США й Великобританії.

Голова Комісії Президії АН БРСР із автоматизації (інформатизації) наукових досліджень. Керівник Робочої групи Комісії Академії наук соцкраїн із обчислювальної техніки.

Був заступником головного конструктора ЕОМ «Мінськ-1», «Мінськ-2», «Мінськ-23». Керував розробкою програмного забезпечення для ЕОМ «Мінськ». Лауреат Державної премії СРСР в галузі науки і техніки (1970) та Державної премії БРСР в галузі науки і техніки (1982).

Г.К.Столяров – ініціатор, науковий керівник і учасник розробки та впровадження сімейства сумісних документально-фактографічних інформаційних систем для великих, міні- й персональних комп'ютерів, баз даних і конвертерів.

Нагороджений медаллю «Піонер комп'ютерної техніки» (Computer Pioneer Award) – найпрестижнішою нагородою

²⁰ Бойко В.В., Савинков В.М. Проектирование баз данных информационных систем. Москва: Финансы и статистика, 1989 – 350 с..

²¹ Мидоу, Чарльз Т. Анализ информационных систем: сокр. пер. с англ. / Ч. Мидоу ; под общ. ред. и с послесл. канд. техн. наук В. М. Савинкова ; [пер. Б. В. Ананьев и др.]. - М. : Прогресс, 1977. - 400 с.

Комп'ютерної спільноти IEEE. Вручена за роботу над програмним забезпеченням комп'ютерів «Мінськ», програмним забезпеченням інформаційних систем, за поширення й просування концепцій систем управління базами даних.

Тиугу Енн Харальдович
(1935 – 2020)



Доктор технічних наук, професор, академік Естонської Академії наук, почесний професор Талліннського технічного університету, професор Королівського технологічного інституту в Стокгольмі.

Лауреат Державної премії СРСР в галузі науки і техніки. В Інституті кібернетики АН ЕРСР під керівництвом Е.Х.Тиугу в кінці 70-х років був розроблений підхід щодо створення єдиного середовища мови програмування і бази даних. У монографії²² запропоновані методологія концептуального моделювання предметної області й підтримуючий її інструментарій, які виявилися придатними для створення СУБД.

Засобами системи програмування високого рівня ПРИЗ²³, яку її ідеолог Е.Х.Тиугу кваліфікує як інструмент концептуального програмування, можна підтримувати потрібну модель даних. Таким чином авторами була реалізована СУБД DABU.

Філіпов Віктор Іванович

Співробітник ОЦ АН СРСР. Колектив, очолюваний В.І.Філіповим, розробив методи реалізації систем управління базами даних. Ним було розроблено інтерпретуючу систему ДІАЛОГ для ЕОМ ВЕОМ-6, з якою користувачі спілкувалися в інтерактивному режимі спочатку через телетайпи, а згодом і через дисплеї. Була розроблена також СУБД типу CODASYL КОМПАС для платформи ВЕОМ-6 із операційною системою ДІСПАК. Під керівництвом В.І.Філіпова була реалізована перша в країні СУБД СУРНА та інтерактивна реляційна СУБД ДІСУР. Він запропонував модель, інтегруючу функціональні можливості реляційної моделі й мережевої моделі CODASYL, а також теоретико-множинний підхід до моделей даних.

В.І.Філіпов був заступником голови Комісії із банків даних та інформаційно-пошукових систем Координаційного комітету АН СРСР у галузі обчислювальної техніки.

Цаленко

Михайло Шамшонович



Кандидат фізико-математичних наук, доктор технічних наук, професор. Завідувач кафедри математики РДГУ. Завідував науково-дослідницькими лабораторіями, викладав у МДУ і в Педагогічному інституті.

Автор монографій на тему сучасної алгебри і теорії баз даних, десятків статей з алгебри, інформатики і лінгвістики.

На початку 70-х років він перший в країні випустив препринт з реляційної мо-

²² Тыугу Э.Х. Концептуальное программирование. М.: Наука, Гл. ред. физ.-мат. литерат., 1984. 256 с.

²³ Кахро М.И. и др. Инструментальная система программирования ЕС ЭВМ (ПРИЗ) / М.И. Кахро, А.П. Калья, Э.Х. Тыугу. М.: Финансы и статистика, 1981, 158 с.

делі даних, написаний за працями Кодда, що став настільною книгою практично для всіх дослідників з баз даних. Згодом він у переробленому вигляді був надрукований у двох номерах збірки «Алгоритми і організація вирішення економічних задач» за редакцією В.М.Савінкова²⁴. Варто також особливо відзначити дві його монографії, де досліджуються математичні моделі баз даних²⁵ і методи моделювання семантики баз даних²⁶.

Післямова

На жаль, чимало питань, пов'язаних з історією баз даних, залишилися поза цим оглядом. До них належать матеріали щодо баз даних в інтернеті, бази даних XML, структури зберігання, методи доступу й питання оптимізації, мови й системи програмування баз даних, словники/довідники, робота конференцій і симпозіумів, видавнича діяльність. Сподіваюся, що ці питання в майбутньому все ж таки будуть висвітлені.

References

1044. Long F., Zhang H.J., Feng D.D. Fundamentals of content-based image retrieval. In: Feng, D.D., Siu, WC., Zhang, HJ. (Eds.), *Multimedia Information Retrieval and Management*, Springer, Berlin, 2003.

1045. Döller M., Kosch H. Image Database. In *Encyclopedia of Database Systems*, Ling Liu, M. Tamer Özsu Editors, pp. 1761 - 1766

1046. Blaser A. *Data Base Techniques for Pictorial Applications*, Florence, Italy, June 20-22, 1979, *Proceedings. Lecture Notes in Computer Science 81*, Springer 1980,

1047. Tamura H, Yokoya N. Image database systems: a survey. *Pattern Recognition*, 1984., Vol. 17, No 1, pp 29-43

1048. Chang S.-K., Hsu A. Image information systems: Where do we go from here? *IEEE Trans. on Knowledge and Data Engineering* Vol. 4 No. 5, 1992 pp. 431–442

1049. Shatford S. Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly*. 1986, Vol. 6, No 3, pp. 39–62.

1050. Content-based image retrieval. - https://en.wikipedia.org/wiki/Content-based_image_retrieval

1051. Kato T. (April 1992). "Database architecture for content-based image retrieval". *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, International Society for Optics and Photonics, 1992, pp. 112–123.

1052. Lew M.S., Sebe N., Djeraba Ch., Jain R. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2006, Vol. 2, No. 1, pp. 1–19

1053. Jain R., Guest E., Special Issue on Visual Information Management, *Communications of ACM*, 40(12), 30-32, Dec. 1997.

1054. Gudivada V.N., Raghavan J.V. Special issue on content-based image retrieval systems, *IEEE Computer Magazine*, Vol. 28, No. 9, September 1995.

1055. Narasimhalu A.D. Special section on content-based retrieval. *Multimedia Systems*, 1995, 3 (1): 1-2.

1056. Pentland A., Picard R., Special issue on digital libraries, *EEE Transactions on Pattern Analysis and Machine Intelligence*, 1996. No 8,

1057. Schatz B., Chen H., Building large-scale digital libraries, *Computer*, 1996, Vol. 26, No. 5, pp. 22-26

1058. Linda G. Shapiro, George C. Stockman. *Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001, 608 p.

1059. Veltkamp R.C, Tanase M. *Content-Based Image Retrieval Systems: A Survey*, Dept. Computing Science, Utrecht University, Utrecht, The Netherlands, Tech. Rep. UU-CS-2000-34, 2002

²⁴ Цаленко М.Ш. Реляционные модели баз данных (обзор) // Алгоритмы и организация решения экономических задач / Под ред. В.М. Савинкова. Вып. 9. М.: Статистика, 1977. С. 18-36

Цаленко М.Ш. Реляционные модели баз данных (обзор) // Алгоритмы и организация решения экономических задач / Под ред. В.М. Савинкова. Вып. 10. М.: Статистика, 1977. С. 16-29

²⁵ Цаленко М.Ш. Семантические и математические модели баз данных // Ито-ги науки и техники. Сер.: Информатика. Т.9, 1985. 208 с.

²⁶ Цаленко М.Ш. Моделирование семантики в базах данных. М.: Наука. Гл. ред. физ.-мат. лит., 1989. 288 с.

1060. Frank Y. Shih. Image Segmentation. In Encyclopedia of Database Systems, Ling Liu, M. Tamer Özsu Editors, pp. 1795 - 1803
1061. Linda G. Shapiro and George C. Stockman (2001): "Computer Vision", New Jersey, Prentice-Hall, 609 p.
1062. Image segmentation. - https://en.wikipedia.org/wiki/Image_segmentation
1063. Pal N.R., Pal S.K. A review on image segmentation techniques. Pattern Recognition. 1993, vol. 26, No. 9, pp. 1277–1294.
1064. Manpreet Kaur, Lal Chand. Review of image segmentation and its techniques. Journal of Emerging Technologies and Innovative Research (JETIR), 2018, Vo. 5, No. 7, pp. 974-981
1065. Salwa Abdulateef, Mohanad Salman. A Comprehensive Review of Image Segmentation Techniques. Iraqi Journal for Electrical And Electronic Engineering, 2021, vol. 17, No. 2, pp. 166-175
1066. Plataniotis K.N., Venetsanopoulos A.N. Color Image Processing and Applications. Springer, Berlin, 2000.
1067. Manjunath B.S., et al. Introduction to MPEG-7. Wiley, New York, 2002.
1068. Zhao Q., Yang J., Liu H. Stone Images Retrieval Based on Color Histogram. In IEEE International Conference on Image Analysis and Signal Processing, 2009, pp. 157-161.
1069. Stricker M., Orengo M. Similarity of color images. In Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995.
1070. Wan X., Kuo K. Color distribution analysis and quantization for image retrieval. In SPIE Storage and Retrieval for Image and Video Databases IV, vol.SPIE 2670, 1996, pp. 9- 16
1071. Smith J.R., Chang S.-F. Single color extraction and image query, in Proc. IEEE Int. Conf. on Image Proc., 1995.
1072. Smith J.R., Chang S.-F. Tools and techniques for color image retrieval, in IS & T/SPIE Proceedings, Vol. 2670, Storage & Retrieval for Image and Video Databases IV, 1995.
1073. Pass G., Zabih R., Miller J. Comparing images using color coherence vectors. In MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia, 1997, pp. 65–73
1074. Haralick R.M., Shanmugam K., Dinstein I. Texture features for image classification, IEEE Transactions on Systems, Man, and Cybernetics, 1973, Vol. SMC-3, No. 6, pp.: 610-621
1075. Gotlieb C.C., Kreyszig H.E. Texture descriptors based on co-occurrence matrices, Computer Vision, Graphics, and Image Processing, Vol. 51, No. 1, 1990, pp. 70–86
1076. Tamura H., Mori S., Yamawaki T. Texture features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics, 1978, Vol.8, No. 6, pp. 460-473
1077. Flickner M., Sawhney H., Niblack W., Ashley J., Qian Huang, Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele D., Yanker P. "Query by image and video content: the QBIC system". Computer. 1995, Vol. 28, No.9, pp. 23–32.
1078. Huang T.S., Mehrotra S., Ramchandran K. Multimedia Analysis and Retrieval System (MARS) project. In P.B. Heidorn. B. Sandore (eds) Proceedings of the 33rd Annual Clinic on Library Application of Data Processing: Digital Image Access and Retrieval, Urbana, IL, March 1996, pp. 100-117. University of Illinois, 1997.
1079. Smith J.R., Chang S.-F. Transform features for texture classification and discrimination in large image databases. In Proceedings of 1st International Conference on Image Processing, 1994, pp. 407-411.
1080. Chang, Kuo C.-C.J. Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions on Image Processing, 1993, Vol. 2, No.4, pp. 429–441
1081. Gross M.H., Koch R., Lippert L., Dreger A. Multiscale image texture analysis in wavelet spaces, In Proceedings of 1st International Conference on Image Processing, 1994. pp. 412–416
1082. Kundu A. Chen J.-L. Texture classification using QMF bank-based subband decomposition. CVGIP: Graphical Models and Image Processing 54(5), 1992, pp. 369–384.
1083. Thyagarajan K.S., Nguyen T., Persons C.

- A maximum likelihood approach to texture classification using wavelet transform. In Proceedings of 1st International Conference on Image Processing, 1994, vol. 2, pp. 640–644
1084. Cross, Jain A.K. Markov random field texture models. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-5, No.1, 1983, pp.25–39
1085. Plataniotis K.N., Androutsos D., Venetsanopoulos A.N. Multichannel filters for image processing. Signal Processing: Image Communication, Vol 9, No. 2, 1997, pp.143-158
1086. Ma W.Y., Manjunath B. Netra: a toolbox for navigating large image databases. Proceedings of the IEEE International Conference on Image Processing, 1997, pp. 568–571.
1087. Pentland A.P. Fractal-based description of natural scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-6, No. 6, 1984, pp. 661–674.
1088. Weszka J., Dyer C., Rosenfeld A., A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man, and Cybernetics. Vol. SMC-6, No. 4, 1976, pp. 269 - 285
1089. Ohanian P.P., Dubes R.C. Performance evaluation for four classes of texture features. Pattern Recognition, vol. 25, no. 8, pp. 819–833, 1992.
1090. Ma W.Y. Manjunath B.S. A comparison of wavelet transform features for texture image annotation. In Proceedings Second International Conference on Image Processing (ICIP'95), 1995, vol. 2, pp. 256-259,
1091. Armi L., Fekri-Ershad S. Texture image analysis and texture classification methods - a review. International Online Journal of Image Processing and Pattern Recognition, Vol. 2, No.1, pp. 1-29, 2019
1092. Rui Y., She A.C., Huang T.S. Modified fourier descriptors for shape representation—a practical approach, in Proc. of First International Workshop on Image Databases and Multi Media Search, 1996, pp. 22-23
1093. Zahn C.T., Roskies R.Z. Fourier descriptors for plane closed curves, IEEE Transactions on Computers, 1972, Vol. C-21, No. 3, pp. 269 - 281
1094. Persoon E., Fu K.S. Shape Discrimination Using Fourier Descriptors. IEEE Transactions on Systems, Man, and Cybernetics, 1977, Vol. 7, No. 3, pp. 170 - 179
1095. Hu M. Visual Pattern Recognition by Moment Invariants. IRE Transactions on Information Theory, IT-08, 1962, Vol. 8, No.2, pp. 179-187.
1096. Yang M., Algrejtsen F. Fast computation of invariant geometric moments: A new method giving correct results, in Proceedings of 12th International Conference on Pattern Recognition, 1994, pp. 201-204.
1097. Pentland A., Picard, Sclaroff S. Photobook: Content-based manipulation of image databases, International Journal of Computer Vision, 1996, 18 (3), pp. 233–254.
1098. Arkin E.M., Chew L., Huttenlocher D., Kedem K., Mitchell J. An efficiently computable metric for comparing polygonal shapes, IEEE Trans. Patt. Recog. Mach. Intell. 13(3), 1991.
1099. Lin H.-C., Chiu C.-Y., Yang S.-N. Finding textures by textual descriptions, visual examples, and relevance feedbacks. Pattern Recognition Letters, 2003, vol. 24, No. 12, pp. 2255-2267
1100. Chuang G.C.-H., Kuo C.-C.J. Wavelet descriptor of planar curves: Theory and applications, IEEE Trans. Image Proc. 5(1), 56–70, 1996.
1101. Li B., Ma S.D. On the relation between region and contour representation. Proceedings of 12th International Conference on Pattern Recognition, 1995.
1102. Mehtre B.M., Kankanhalli M., Lee W.F. Shape measures for content based image retrieval: A comparison, Information Processing & Management 33(3), 1997.
1103. Taubin G. Recognition and positioning of rigid objects using algebraic moment invariants, in SPIE Vol. 1570, Geometric Methods in Computer Vision, 1991.
1104. Wallace I., Mitchell O. Three-dimensional shape analysis using local shape descriptors, IEEE Trans. Patt. Recog. and Mach. Intell., PAMI-3(3), May 1981.
1105. Wallace I., Wintz P. An efficient three-

- dimensional aircraft recognition algorithm using normalized Fourier descriptors, *Computer Graphics and Image Processing* 13, 1980.
1106. Faloutsos C., Flickner M., Niblack W., Petkovic D., Equitz W., Barber R. Efficient and Effective Querying by Image Content, *Journal of Intelligent Information Systems*, Vol. 3, No. 3-4, 1994, pp. 231–262.
1107. Chua T.S., Tan K.-L., Ooi B.C. Fast signature-based color-spatial image retrieval. In *ICMCS '97: Proceedings of the 1997 International Conference on Multimedia Computing and Systems*, 1997.
1108. Lu H., Ooi B., Tan K., Efficient image retrieval by color contents. In *Proc. of the 1994 International Conference on Applications of Databases*, 1994, pp 95–108
1109. Smith J.R. Chang S.-F. Single color extraction and image query. In *Proc. IEEE International Conference on Image Processing*, 1995.
1110. Rickman R.M., Stonham T.J. Content-based image retrieval using colour tuple histograms, In *Proc. SPIE Storage and Retrieval for Image and Video Databases IV*, 1996.
1111. Stricker M., Dimai A. Color indexing with weak spatial constraints. In *Proc. SPIE Storage and Retrieval for Image and Video Databases IV*, 1996.
1112. Huang J., Kumar S., Mitra M., Zhu W.-J., Zabih R. Image indexing using color correlogram. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '97)*, 1997.
1113. Mojsilovic A., Rogowitz B. ISee: perceptual features for image library navigation, *Proceedings of the SPIE, Human Vision and Electronic Imaging*, vol. 4662, 2002, pp. 266–277.
1114. Mezaris V., Kompatsiaris I., Strintzis M.G. An ontology approach to object-based image retrieval, *Proceedings of the ICIP*, vol. II, 2003, pp. 511–514.
1115. Chang S.K., Shi Q.Y., Yan C.W. Iconic indexing by 2D string. *IEEE Trans. Pattern Anal. Mach. Intell.* 9 (3) (1987) 413–428.
1116. Ren W., Singh M., Singh C., Image retrieval using spatial context, *Ninth International Workshop on Systems, Signals and Image Processing (IWSSIP'02)*, Manchester, November, 2002.
1117. Smith J.R., Li C.-S. Decoding image semantics using composite region templates, *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL-98)*, June 1998, pp. 9–13.
1118. Gaede V., Günther O. Multidimensional access methods. *ACM Comput Surv.* 1998;30(2):170–231.
1119. Ahn H.K., Mamoulis N., Wong H.M. A survey on multidimensional access methods. Technical report, *Utrecht University* (2001)
1120. Venkateswaran J. A Survey of Recent Multidimensional Access Methods. Technical Report, *University of Missouri-Rolla*. -2004. 162 p.
1121. Faloutsos C., Lin K.-I. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. *ACM SIGMOD Record*, Vol. 24, No 2, 1995, pp. 163–174
1122. Ng R.T., Sedighian A. Evaluating multidimensional indexing structures for images transformed by principal component analysis, in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.
1123. White D., Jain R. Similarity indexing: Algorithms and performance, in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.
1124. Chandrasekaran V., Manjunath B.S., Wang Y.F., Winkeler J., Zhang H. An eigenspace update algorithm for image analysis. *CVGIP: Graphical Models and Image Processing Journal*, Vol. 59, No. 5, 1997, pp. 321-332
1125. Salton G., McGill M.J. *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.
1126. Guttman A. R-tree: A dynamic index structure for spatial searching. *ACM SIGMOD Record*, Vol. 14, No. 2, 1984, pp. 47–57
1127. Sellis T., Roussopoulos N., Faloutsos C. The R+-tree: A dynamic index for multidimensional objects. *Proceedings of the 13th VLDB Conference*, Brighton 1987, pp. 507-518.

1128. Greene D. An implementation and performance analysis of spatial data access methods. Proceedings of the Fifth International Conference on Data Engineering, 1989, pp. 606–615
1129. Beckmann N., Kriegel H.-P., Schneider R., Seeger B. The R*-tree: An efficient and robust access method for points and rectangles. SIGMOD '90: Proceedings of the 1990 ACM SIGMOD international conference on Management of data, 1990, pp. 322–331
1130. White D. A., Jain R. Similarity indexing: Algorithms and performance. Proc. SPIE 2670, Storage and Retrieval for Still Image and Video Databases IV, 1996
1131. Charikar M., Chekur C., Feder T., Motwani R. Incremental clustering and dynamic information retrieval. Proc. of the 29th Annual ACM Symposium on Theory of Computing, 1997, pp. 626–635.
1132. Rui Y., Chakrabarti K., Mehrotra S., Zhao Y., Huang T.S. Dynamic clustering for optimal retrieval in high dimensional multimedia databases. University of Illinois, Department of Computer Science Technical Report MARS-97-10. Urbana, IL: Department of Computer Science, 1997.
1133. Zhang H.J., D. Zhong. A scheme for visual feature based image retrieval. Proc. SPIE 2420, Storage and Retrieval for Image and Video Databases III, 1995
1134. Deza M.M., Deza E. Encyclopedia of Distances. Springer; Softcover reprint of the original 3rd ed. 2014. 753 p.
1135. Mahalanobis P.C. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India, 1936, vol. 2, No. 1, pp. 49—55.
1136. Patil R.S., Agrawal A.J. Content-based Image Retrieval Systems: A Survey. Advances in Computational Sciences and Technology, 2017, Vol. 10, No. 9, pp. 2773-2788
1137. Gupta A., Jain R. Visual information retrieval, Commun. ACM 40 (5) (1997) 70–79.
1138. Smith J.R., Chang S.F. VisualSeek: a fully automatic contentbased query system, Proceedings of the Fourth ACM International Conference on Multimedia, 1996, pp. 87–98.
1139. Ma W.Y., Manjunath B. Netra: a toolbox for navigating large image databases, Proceedings of the IEEE International Conference on Image Processing, 1997, pp. 568–571.
1140. Wang J.Z., Li J., Wiederhold G. SIMPLiCity: semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (9) (2001) 947–963.
1141. Smeulders A.W.M, Worring M., Santini S., Gupta A., Jain R. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, Vol. 22, No 12, pp.1349–1380.
1142. Eakins J., Graham M. Content-based image retrieval, Technical Report, University of Northumbria at Newcastle, 1999, 59 p.
1143. Ying Liua, Dengsheng Zhanga, Guojun Lua, Wei-Ying Mab. A survey of content-based image retrieval with high-level semantics. Pattern Recognition, Vol. 40, No. 1, 2007, pp. 262–282
1144. Mussarat Yasmin, Sajjad Mohsin, Muhammad Sharif. Intelligent Image Retrieval Techniques: A Survey. Journal of Applied Research and Technology, 2014, Vol. 12, No. 1, pp. 87-103
1145. Kherfi M.L., Ziou D., Bernardi A. Image Retrieval from the World Wide Web: Issues, Techniques, and Systems, ACM Computing Surveys, 2004, Vol. 36, No. 1, pp. 35–67
1146. Popescu A., Grefenstette G. A Conceptual Approach to Web Image Retrieval. Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 2008, pp. 297-304
1147. Berk T., Brownston L., Kaufman A. A new color-naming system for graphics language. IEEE Comput. Graphics Appl. 2 (3), 1982, pp. 37–44.
1148. Stanchev P.L., Green Jr. D., Dimitrov B. High level color similarity retrieval, Int. J. Inf. Theories Appl. 10 (3) (2003) 363–369.
1149. Rao A.R., Lohse M. Towards a texture naming system: identifying relevant dimensions of texture, IEEE Proceedings of the Fourth Conference on Visualization, 1993, pp. 220–227.
1150. Shi R., Feng H., Chua T.-S., Lee C.-H. An

- adaptive image content representation and segmentation approach to automatic image annotation. International Conference on Image and Video Retrieval (CIVR), 2004, pp. 545–554.
1151. Vailaya A., Figueiredo M.A.T., Jain H.J., Zhang A.K. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 2001, Vol. 10, No.1, pp. 117–130.
1152. Town C.P., Sinclair D. Content-based image retrieval using semantic visual categories. Society for Manufacturing Engineers, Technical Report MV01-211, 2001.
1153. Feng H., Chua T.-S. A bootstrapping approach to annotating large image collection. Workshop on Multimedia Information Retrieval in ACM Multimedia, November 2003, pp. 55–62.
1154. MacArthur S.D., Brodley C.E., Shyu C.-R. Relevance feedback decision trees in content-based image retrieval. *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'00)*, June 2000, pp. 68–72.
1155. Jain A., Dubes R. *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
1156. Baraldi A., Alpaydin M. Constructive feedforward ART clustering networks—Part I and II. *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp.645–677, May 2002.
1157. Shi J., Malik J. Normalized Cuts and Image Segmentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, Vol. 22, pp. 888-905
1158. Dunn J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 1973, Vol. 3, No. 3, pp. 32–57
1159. Chen Y., Wang J.Z., Krovetz R. An unsupervised learning approach to content-based image retrieval. *IEEE Proceedings of the International Symposium on Signal Processing and its Applications*, July 2003, pp. 197–200.
1160. Rui Y., Huang T.S., Ortega M., Mehrotra S. Relevance feedback: a power tool for interactive content-based image retrieval, *IEEE Trans. Circuits Video Technol.* 8 (5) (1998) 644–655.
1161. Jing F., Li M., Zhang L., Zhang H.-J., Zhang B. Learning in regionbased image retrieval, *Proceedings of the International Conference on Image and Video Retrieval (CIVR2003)*, 2003, pp. 206–215.
1162. Smith J.R., Li C.-S. Decoding image semantics using composite region templates, *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL-98)*, June 1998, pp. 9–13.
1163. Zhuang Y., Liu X., Pan Y. Apply semantic template to support content-based image retrieval, *Proceedings of the SPIE, Storage and Retrieval for Media Databases*, vol. 3972, December 1999, pp. 442–449.
1164. Chang S.-F., Chen W., Sundaram H. Semantic visual templates: linking visual features to semantics, *International Conference on Image Processing (ICIP), Workshop on Content Based Video Search and Retrieval*, vol. 3, October 1998, pp. 531–534.
1165. Zhuang Y, Liu X., Pan Y. Apply semantic template to support content-based image retrieval, *Proceedings of the SPIE, Storage and Retrieval for Media Databases*, vol. 3972, December 1999, pp. 442–449.
1166. Kalinichenko L.A. *Methods and Tools for Integration of Heterogeneous Databases (Rus)*. Moscow, Nauka, 1983, 424 p.
1167. Smith J.M., Bernstein P.A., Dayal U., Goodman N., Landers T., Lin K.W.T., Wong E. Multibase: Integrating heterogeneous distributed database systems. In *Proceedings of the May 4-7, 1981, National Computer Conference (AFIPS '81)*. ACM, New York, NY, USA, pp. 487-499.
1168. Hammer M., McLeod, D. 1979. On database management system architecture. Tech. Rep. MIT/LCS/TM-141, Massachusetts Institute of Technology, Cambridge, Mass.
1169. Heimbigner D., McLeod, D. (1985). A Federated architecture for information management. *ACM Transactions on Information Systems*, Volume 3, Issue 3. pp. 253–278
1170. Veijalainen J., Popescu-Zeletin R. 1988. Multidatabase systems in ISO/OSI environment. In *Standards in Information Technology and Industrial Control*,

- Malagardis, N., and-Williams, T., Eds. North-Holland. The Netherlands. DD. 83-97.
1171. Alonso R., Barbara D. 1989. Negotiating data access in federated database systems. In Proceedings of the 5th International Conference on Data Engineering (Feb.), pp. 56-65.
1172. Heimbigner D., McLeod D. 1985. A federated architecture for information management. *ACM Transactions on Information Systems*. 3, 3 (July), 253-278.
1173. Du W., Elmagarmid A., Kim W. 1990. Effects of local autonomy on heterogeneous distributed database systems. MCC Tech. Rep. ACT-OODS-EI-059-90, Microelectronics and Computer Technology Corp., Austin Tex.
1174. Garcia-Molina H., Kogan B. 1988. Node autonomy in distributed systems. In Proceedings of the International Symposium on Databases in Parallel and Distributed Systems (Austin, Tex., Dec.), pp. 158-166.
1175. Lenzerini M. Data integration: a theoretical perspective. Proceedings of the 21-st ACM SIGCAT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS 2002). New York: ACM Press, 2002 ,pp. 233–246. .
1176. Jer-Wen Huang. 1994. MultiBase: a heterogeneous multidatabase management system. In Proceedings Eighteenth Annual International Computer Software and Applications Conference (COMPSAC 94), pp. 332–339.
1177. Bondiombouy C., Valduriez P. Query processing in multistore systems: an overview. *International Journal of Cloud Computing* 5.4 (2016): 309-346
1178. Litwin W. 1985. An overview of the multidatabase system MRDSM. In Proceedings of the ACM National Conference (Denver, Oct.), pp. 495-504.
1179. Rusinkiewicz M., Elmasri R., Czejdo B., Georakopoulous D., Karabatis G., Jamoussi A., Loa L., LIY. 1989.-OMNIBASE: Design and implementation of a multidatabase system. In Proceedings of the 1st Annual Symposium in Parallel and Distributed Processing (Dallas, Tex., May), pp. 162-169.
1180. Jacobson G., Piatetsky-Shapiro G., Lafond C., Rajinikanth M., Hernandez J. 1988. CALIDA: A knowledge-based system for integrating multiple heterogeneous databases. In Proceedings of the 3rd International Conference on Data and Knowledge Bases (Jerusalem, Israel, June), pp. 3-18.
1181. Litwin W., Boudenant J., Esculier C., Ferrier A., Glorieux A., La Chimia J., Kabbaj K., Moulinoux C., Rolin P., Stangret C. 1982. SIRIUS Systems for Distributed Data Management. In *Distributed Data Bases*, H.-J. Schneider, Ed. North-Holland, The Netherlands, pp. 311-366.
1182. Dwyer P., Larson J. 1987. Some experiences with a distributed database testbed system. In *Proc. IEEE* 75, 5 (May), 633-647.
1183. Templeton M., Brill D., Chen A., Dao S., Lund E., McGregor R., Ward P. 1987. Mermaid: A front-end to distributed heterogeneous databases. In *Proc. IEEE* 75,5 (May), 695-708.
1184. Landers T., Rosenberg R. 1982. An overview of Multibase. In *Distributed Databases*, H.-J. Schneider, Ed., North-Holland, The Netherlands, pp. 153-184.
1185. Ellinghaus D., Hallmann M., Holtkamp B., Kreplin K. 1988. A multidatabase system for transaction autonomy. In Proceedings of the International Conference on Extending Database Technology (Venice, Italy, Mar.). In *Computer Science*, Vol. 303, Springer-Verlag, New York, pp. 600-605.
1186. Veijalainen J., Popescu-Zeletin R. 1988. Multidatabase systems in ISO/OSI environment. In *Standards in Information Technology and Industrial Control*, Malagardis, N., and-Williams, T., Eds. North-Holland. The Netherlands. DD.-83-97.
1187. Dayal U., Hwang H. 1984. View definition and generalization for database integration in a multidatabase system. *IEEE Trans. Soft. Eng.* SE-IO, 6 (Nov.), 628-644.
1188. Belcastro, V., et al. 1988. An overview of the distributed query system D&S. In Proceedings of the International Conference on Extending Data Base Technology (Venice, Italv. Mar.). In *Computer Science.-Vol. 303,* Spriger-Verlag, New York, pp. 170-189.

1189. Breitbart Y., Silberschatz A. 1988. Multidatabase update issues. In Proceedings of the ACM SZGMOD Conference (June), 135-142.
1190. Barker K., Ozsu T. 1988. A survey of issues in distributed heterogeneous database systems. Tech. Rep. TR 88-9, Univ. of Alberta Edmonton, Canada.
1191. Litwin W., Zeroual A. 1988. Advances in multidatabase systems. In Research into Networks and Distributed Applications (Proceedings of the EUTECO'88). Sneth. R.. Ed. Elsevier Science Publishers' B.V., North-Holland, pp. 1137-1151.
1192. Ram S., Chastain C. 1989. Architecture of distributed data base systems. Journal of Systems and Software, Vol. 10, No. 2, pp. 77-95.
1193. Siegel M. 1987. A survey on heterogeneous database systems. Tech. Note 87-174.1, GTE Laboratories, Waltham, Mass.
1194. Batini C., Lenzerini M., Navathe S. 1986. A comparative analysis of methodologies for database schema integration. ACM Computing Surveys, Vol. 18, No. 4, pp. 323-364.
1195. Ford N. Polyglot Programming. - <http://memeagora.blogspot.com/200/12/polyglot-programming.html>, December 05, 2006
1196. Leberknight S. Polyglot Persistence. - http://www.sleberknight.com/blog/sleberkn/entry/polyglot_persistence, October 15, 2008
1197. Harold Lim, Yuzhang Han, and Shivnath Babu. 2013. How to Fit when No One Size Fits. In CIDR. www.cidrdb.org.
1198. J. Duggan, A. J. Elmore, M. Stonebraker, M. Balzinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. G. Mattson, and S. B. Zdoni. 2015. The BigDAWG Polystore System. SIGMOD Record 44, 2 (2015), 11–16.
1199. Michael Armbrust, Reynold S. Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K. Bradley, Xiangrui Meng, Tomer Kaftan, Michael J. Franklin, Ali Ghodsi, and Matei Zaharia. 2015. Spark SQL: Relational Data Processing in Spark. In SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, May 2015, pp. 1383–1394.
1200. Garulli L. NoSQL adoption: what's the next step. - https://www.slideshare.net/lvca/no-sql-matters2012keynote/47-MultiModel_storage_12_one_product
1201. Zhen Hua Liu, Jiaheng Lu, Dieter Gawlick, Heli Helskyaho, Gregory Pogossiant, Zhe Wu. Multi-Model Database Management Systems - a Look Forward. In: Heterogeneous Data Management, Polystores, and Analytics for Healthcare, 2018, pp.16-29
1202. Valter Uotila, Jiaheng Lu. A Formal Category Theoretical Framework for Multi-Model Data Transformations. Rezig E.K. et al. (eds) Heterogeneous Data Management, Polystores, and Analytics for Healthcare. DMAH 2021, Poly 2021. Lecture Notes in Computer Science, vol 12921. pp. 14-28
1203. Henrik Forssell, Håkon Robbestad Gylterud, David I. Spivak. Type theoretical databases. Journal of Logic and Computation, Vol. 30, No 1, January 2020, pp. 217–238
1204. Zaidi E., Heudecker N., Adrian M. Market Guide for NoSQL DBMSs. - <https://www.gartner.com/en/documents/3105622>
1205. Jiaheng Lu, Irena Holubová, and Bogdan Cautis. Multi-model Databases and Tightly Integrated Polystores: Current Practices, Comparisons, and Open Challenges. In CIKM '18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management October 2018. pp. 2301–2302.
1206. Ewa Pluciennik and Kamil Zgorzalek. "The Multi-model Databases - A Review". BDAS 2017: 141–152. Ewa Pluciennik-Psota, Kamil Zgorzalek
1207. Fábio Roberto Oliveira, Luis del Val Cura. "Performance Evaluation of NoSQL Multi-Model Data Stores in Polyglot Persistence Applications". IDEAS '16: Proceedings of the 20th International Database Engineering & Applications Symposium, July 2016, pp. 230–235
1208. Zhang C., Lu J., Xu P., Chen Y. (2019) UniBench: A Benchmark for Multi-model Database Management Systems. In: Nambiar R., Poess M. (eds) Performance Evaluation and Benchmarking for the Era of Artificial Intelligence. TPCTC 2018. Lecture Notes in Computer Science, vol 11135. Springer, pp 7-23

1209. Ran Tan, Rada Chirkova, Vijay Gadepally, and Timothy G. Mattson. 2017. Enabling query processing across heterogeneous data models: A survey. 2017 IEEE International Conference on Big Data (Big Data). 3211–3220
1210. Jiaheng Lu, Irena Holubová. Multi-model Databases: A New Journey to Handle the Variety of Data. ACM Computing Surveys, Vol. 52, No 3, 2020, Article No.: 55, pp 1–38
1211. Aven P., Burley D. Building on Multi-Model Databases. O'Reilly Media, Inc., 2017, 96 p.
1212. B. Kolev, C. Bondiombouy, P. Valduriez, R. Jimenez-Peris, R. Pau, and J. Pereira, “The CloudMdsQL multistore system,” in Proc. ACM International Conference on Management of Data (SIGMOD’16), 2016, pp. 2113–2116.
1213. J. Wang, T. Baker, M. Balazinska, D. Halperin, B. Haynes, B. Howe, D. Hutchison, S. Jain, R. Maas, P. Mehta, D. Moritz, B. Myers, J. Ortiz, D. Suci, A. Whitaker, S. Xu. The Myria big data management and analytics system and cloud service. The 8th Biennial Conference on Innovative Data Systems Research (CIDR ‘17), 2017
1214. M. Hausenblas and J. Nadeau, “Apache Drill: Interactive adhoc analysis at scale,” Big Data, vol. 1, no. 2, pp. 100–104, 2013.
1215. A. Simitsis, K. Wilkinson, M. Castellanos, and U. Dayal . Optimizing analytic data flows for multiple execution engines, In Proc. ACM International Conference on Management of Data (SIGMOD’12), 2012, pp. 829–840.
1216. I. Gog, M. Schwarzkopf, N. Crooks, M. P. Grosvenor, A. Clement, and S. Hand, “Musketeer: all for one, one for all in data processing systems,” in EuroSys ‘15: Proceedings of the Tenth European Conference on Computer Systems, April 2015 Article No.: 2, pp. 1–16
1217. D. Agrawal, L. Ba, L. Berti-Equille, S. Chawla, A. Elmagarmid, H. Hammady, Y. Idris, Z. Kaoudi, Z. Khayyat, S. Kruse et al., “Rheem: Enabling multi-platform task execution,” in Proc. ACM International Conference on Management of Data (SIGMOD’16), 2016, pp. 2069–2072.
1218. S. Dasgupta, K. Coakley, and A. Gupta, “Analytics-

Про автора:

Резніченко Валерій Анатолієвич, кандидат фізико-математичних наук, заступник завідувача відділом.

Кількість публікацій в українських виданнях – 61.

Кількість зарубіжних публікацій – 4. Індекс Хірша – 12.

<http://orcid.org/0000-0002-4451-8931>.

Місце роботи автора:

Інститут програмних систем НАН України, 03187, м. Київ-187,

проспект Академіка Глушкова, 40.

Тел.: (044) 526 3559.

E-mail: reznich@isofts.kiev.ua