

ОПЕРАЦІЇ РЕЛЯЦІЙНОЇ АЛГЕБРИ

ВИЗНАЧЕННЯ ТА ІСТОРІЯ РЕЛЯЦІЙНОЇ АЛГЕБРИ

Реляційна алгебра (relational algebra) – це система позначень, прийнята для опису запитів до змісту відношень.

На відміну від ER-моделі, яка взагалі не передбачає операцій над даними, реляційна модель містить конкретний набір „стандартних” операцій над відношеннями.

У теорії обчислюваності виконавець (множина обчислюючих елементів) називається *повним по Т'юрінгу*, якщо на ньому можна реалізувати будь-яку обчислювану функцію. Дивно, але набір операцій реляційної алгебри не є функціонально повним по Т'юрінгу (Turing complete), хоча мови програмування зазвичай повні по Т'юрінгу.

Тому засобами реляційної алгебри можуть бути описані далеко не всі можливі операції (і тому виникає необхідність розширень, як-от PL-SQL та Transact SQL). У цьому є перевага: спрощення множини операторів дозволяє суттєво оптимізувати запити до БД, які в цьому разі можуть бути написані на високорівневій мові, як-от SQL.

НАЙВАЖЛИВІШІ ОПЕРАЦІЇ РЕЛЯЦІЙНОЇ АЛГЕБРИ

Від початку реляційна алгебра була розроблена Е.Ф.Коддом (E.F.Codd) як сукупність операторів, які виконуються над *множинами кортежів* (тобто відношеннями) та які виражають типові запити щодо змісту відношень. Алгебра охоплювала 5 операцій над множинами:

- *Об'єднання (union);*
- *Різниця (difference);*
- *Декартів добуток (Cartesian product);*
- *Вибір (selection);*
- *Проекція (projection).*

Потім в алгебру увійшли деякі операції, які можна виразити в термінах початкового набору. Найважливіші з них – різні види *з'єднання (join)*.

В ході розвитку комерційних реляційних СУБД алгебру відношень довелося поповнити ще цілим рядом операцій, найважливіші з них пов'язані з виконанням функцій *агрегування (aggregation)*, тобто знаходження сумарних, середніх, максимальних, мінімальних та інших значень у певному стовпці таблиці-відношення.

Реляційна алгебра, як і всі інші алгебри, дозволяє створювати вирази довільної складності шляхом застосування операцій не тільки до початкових відношень, а й до відношень, отриманих в результаті виконання інших операцій.

Вирази реляційної алгебри можуть містити вирази, застосовані до часткових виразів, та дужки, які задають порядок виконання операцій. Припустимо надати вирази у вигляді дерев часткових виразів (рис.4.35).

Перелічимо найважливіші операції реляційної алгебри.

Таблиця 1 - Найважливіші операції реляційної алгебри

№ пп	Назва	Позначення	Визначення	Примітка
1	Об'єднання (union)	$R \cup S$	Множина елементів, присутня в R, S або в обох множинах одночасно	R та S повинні мати ідентичні схеми та атрибути, типи (домени) яких попарно співпадають. Порядок атрибутів має бути однаковий.
2	Перетин	$R \cap S$	Множина елементів, присутня в R та S одночасно	
3	Різниця	$R - S$	Множина елементів, присутня в R та відсутня в S	
4	Проекція (projection)	$\pi_{A_1, A_2, \dots, A_n}(R)$	Відношення, що містить явно перелічені атрибути A_1, A_2, \dots, A_n відношення R	У відношенні-результаті кортежі не дублюються
5	Вибір (selection)	$\sigma_c(R)$	Підмножина кортежів R, які задовольняють певній умові C	
6	Декартів добуток (Cartesian product)	$R \times S$	Множина пар кортежів, де першим елементом є будь-який кортеж R, а другим - будь-який кортеж S.	Схема результату є об'єднанням схем R і S. В разі співпадання імен треба перейменувати хоча б один елемент пари.
7	Природне з'єднання (natural join)	$R \bowtie S$	Включення в результат тих кортежів з R і S, у яких співпадають значення атрибутів, спільних для R і S.	Схема результату є об'єднанням без дублювань схем R і S

8	Тета-з'єднання	$R \bowtie_C S$	З Декартового добутку $R \times S$ відбираються ті кортежі, які задовольняють заданій умові C .	
---	----------------	-----------------	---------------------------------------------------------------------------------------------------	--

Приклад 1 - Декартового добутку

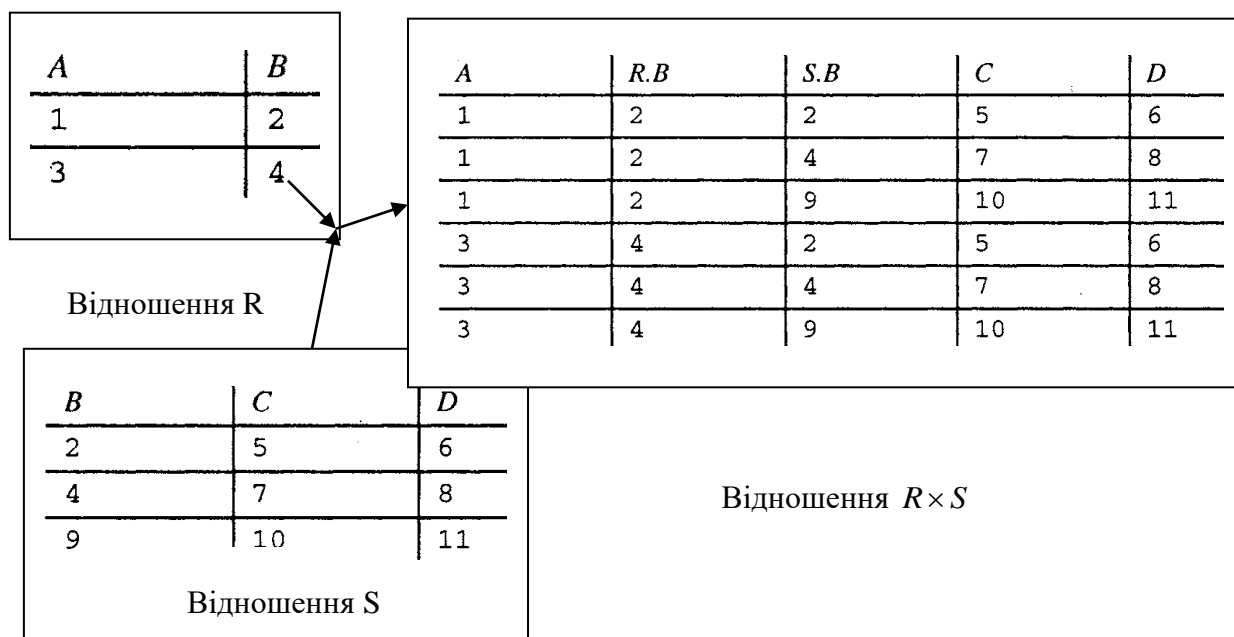


Рисунок 1 - Декартів добуток двох відношень

НАТУРАЛЬНЕ З'ЄДНАННЯ

Позначення: $R \bowtie S$

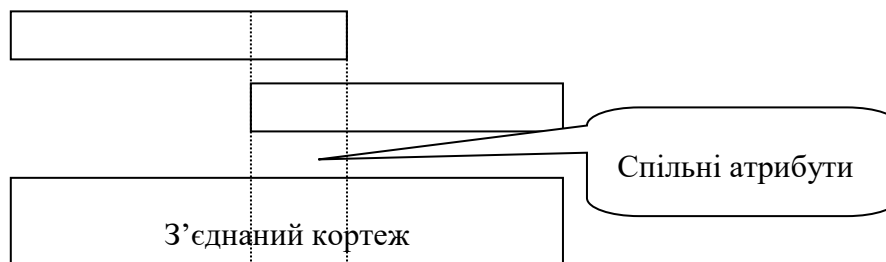


Рисунок 2 - Принцип натурального з'єднання

Приклад 2. Натуральне з'єднання

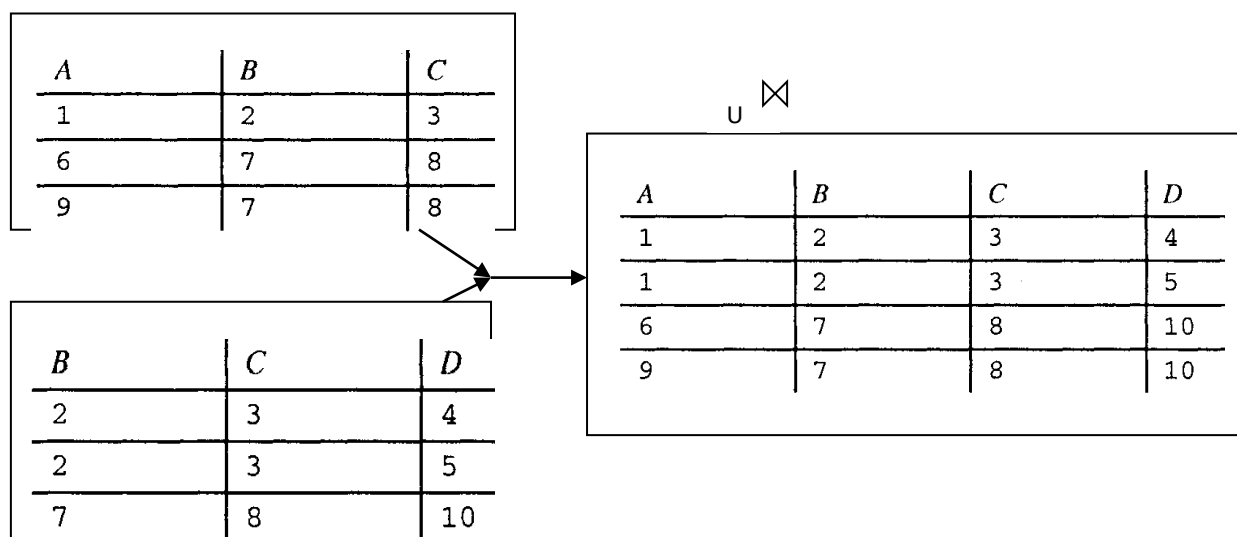


Рисунок 3 - Приклад натурального з'єднання відношень

ПРИКЛАД ТЕТА-З'ЄДНАННЯ

Приклад 3. Тета-з'єднання

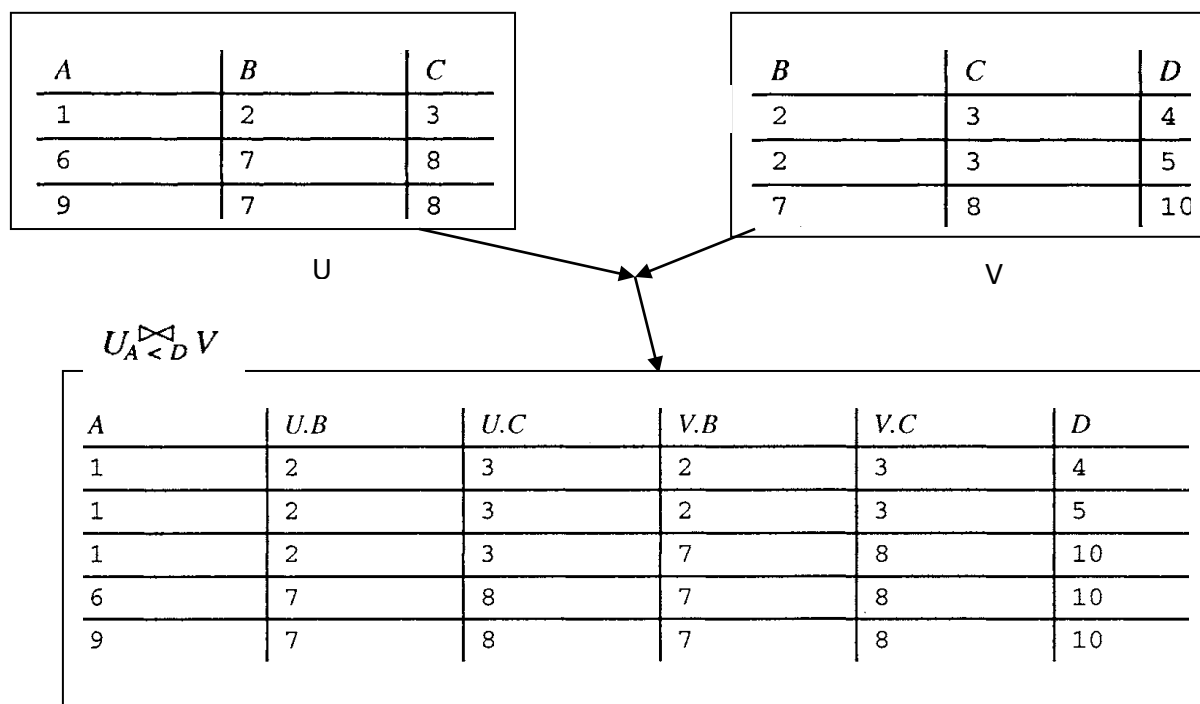


Рисунок 4 - Приклад тета-з'єднання відношень

ДЕРЕВО ЧАСТКОВИХ ВИРАЗІВ

Приклад 4. Розглянемо відношення Movie.

<i>title</i>	<i>year</i>	<i>length</i>	<i>filmType</i>	<i>studioName</i>
Star Wars	1977	124	color	Fox
Mighty Ducks	1991	104	color	Disney
Wayne's World	1992	95	color	Paramount
Addams Family	1991	102	color	Paramount

Рисунок 5 - Відношення Movie для дерева часткових виразів

Нехай треба в'яснити назви і роки випуску фільмів студії Fox, тривалість яких не менша за 100 хвилин. Послідовність операцій може бути наступна:

1. Вибрати з відношення Movie кортежі з $length \geq 100$;
2. Вибрати з відношення Movie кортежі з $studioName = 'Fox'$;
3. Обчислити перетин множин-відношень, отриманих у пп.1, 2.
4. Виконати проекцію відношення п.3 на атрибути title і year.

Цей алгоритм (вираз) можна представити у вигляді дерева виразів:

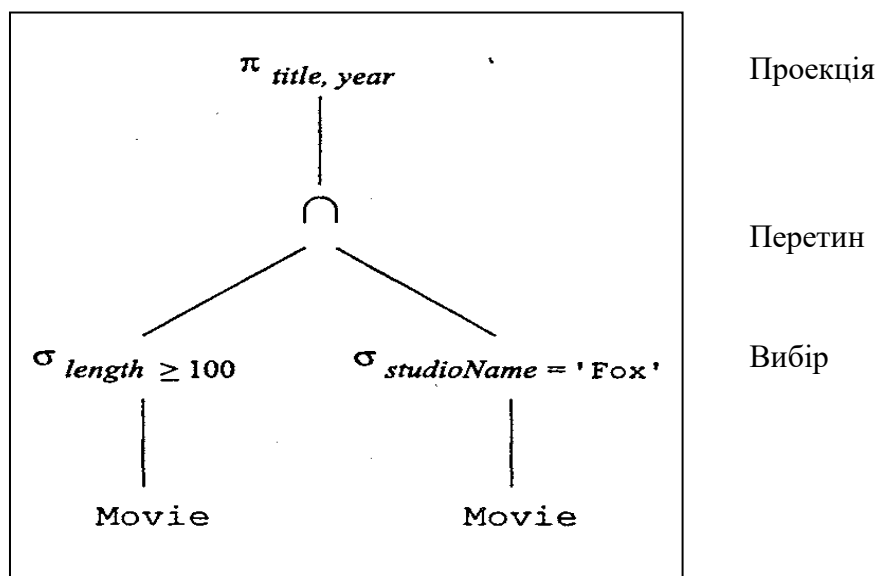


Рисунок 6 - Дерево алгебраїчних операторів

Цей вираз також можна надати у вигляді традиційної лінійної формули з дужками:

$$\pi_{title, year} \left(\sigma_{length \geq 100} (Movie) \cap \sigma_{studioName = 'Fox'} (Movie) \right).$$

Цей вираз можна спростити:

$$\pi_{title, year} \left(\sigma_{length \geq 100 \text{ AND } studioName = 'Fox'} (Movie) \right).$$

Еквівалентні перетворення запитів з метою їх спрощення і створення *Оптимізованого плану запити* виконує *Оптимізатор запитів* – компонента *Компілятору запитів* СУБД.

ДОДАТКОВІ ОПЕРАЦІЇ РЕЛЯЦІЙНОЇ АЛГЕБРИ

Конкретні мови, подібні SQL, пропонують ряд додаткових реляційних операцій, які довели свою корисність в реальних застосуваннях.

Таблиця 2 - Додаткові операції реляційної алгебри

№ пп	Назва	Позначення	Визначення	Примітка
9	Видалення кортежів- дублікатів	δ	Перетворює відношення-мультимножину у множину шляхом відкидання усіх копій кортежу, крім одної	
10	Оператори агрегування	SUM, AVG, MIN і т.п.	Призначені для обчислення сумарних, середніх, мінімальних і т.п. значень у стовпчику відношення, тобто обчислення одного значення на основі множини значень	Строго кажучи, не відносяться до реляційної алгебри
11	Групування	γ	Множина рядків розбивається на групи, після чого до певних атрибутів застосовується агрегування. В результаті група в одному рядку.	У групі значення певних атрибутів однакові
12	Сортування	τ	Перетворює відношення у список кортежів, упорядкований по змісту одного чи кількох компонентів.	Застосовується на завершення, оскільки інші операції не зберігають сортування.
13	Розширена проекція	π	Крім звичайної проєкції виконує додаткові функції над атрибутами для отримання нових атрибутів.	
14	Зовнішнє з'єднання	$R \bowtie S$	Варіант операції з'єднання, що зберігає „висячі”	„Висячі” кортежі відношення

			кортежі, тобто ті, для яких нема відповідника в кореспондуючому відношенні	включаються в результат, і атрибути з відсутнього кореспондуючого відношення отримують значення <i>null</i> .
--	--	--	----------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------

ГРУПУВАННЯ

Відношення, яке повертає оператор $\gamma_L(R)$, будується так:

1. Розбиваємо всю множину кортежів на групи. Кожна група включає всі кортежі, які містять певні значення групуючих атрибутів зі списку L. Якщо групуючі атрибути не задані, R розглядається як єдина група.
2. Для кожної групи створюємо по одному кортежу, що містить:
 - Значення групуючих атрибутів, що відповідають групі,
 - Результати агрегування, обчислені по агрегуючих атрибутах усіх кортежів групи, заданих у списку L.

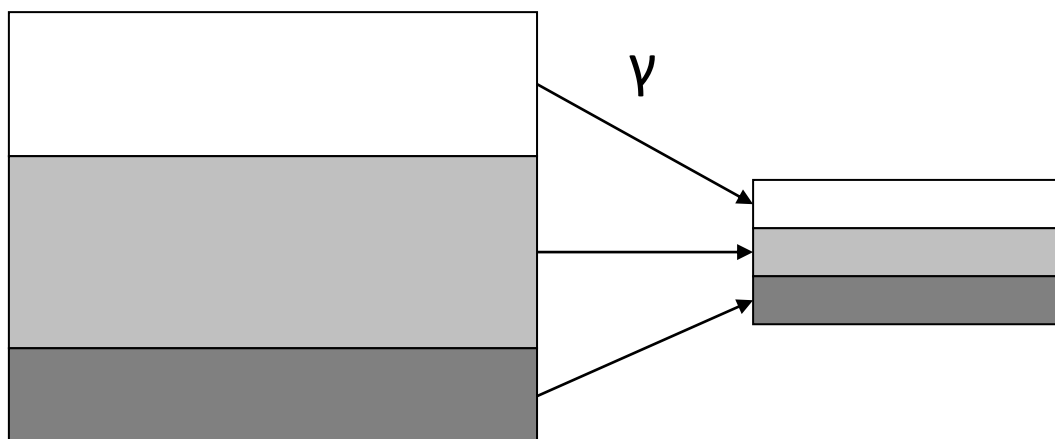


Рисунок 7 - Операція групування

Приклад 5. Групування

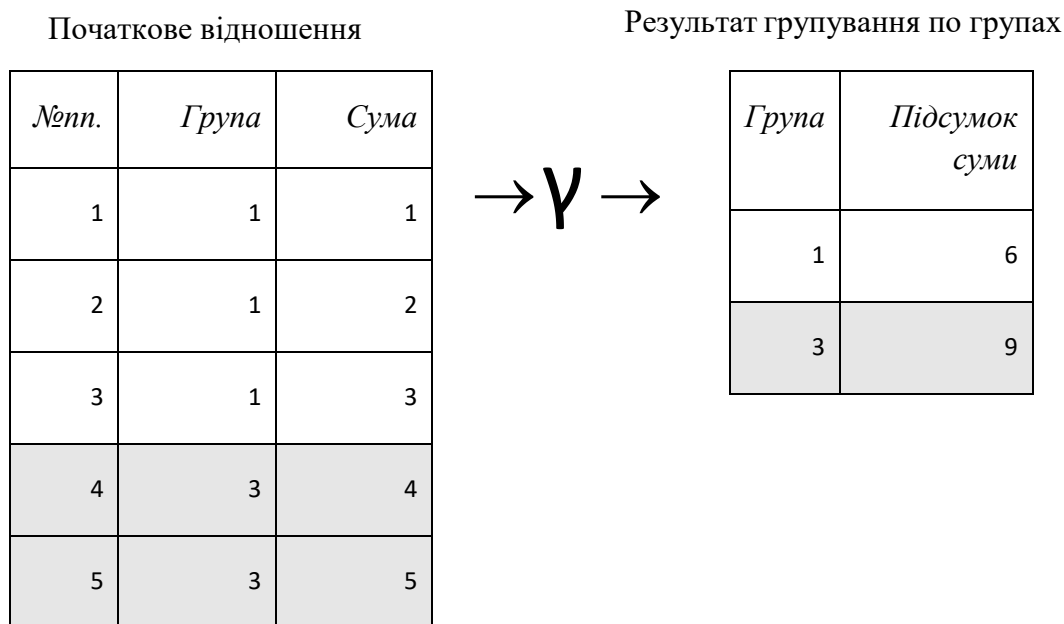


Рисунок 8 – Групування по групах

ЗАЛЕЖНІ ТА НЕЗАЛЕЖНІ ОПЕРАЦІЇ

Нагадаємо п'ять основних операцій реляційної алгебри

Група I: Три стандартних теоретико-множинних бінарних операцій:

- об'єднання
- різниця
- декартовий добуток

Група II. Дві спеціальні унарні операції над відношеннями:

- проєкція
- селекція

Реляційна алгебра складається з усіх виразів, отриманих шляхом комбінування цих п'яти основних операцій у синтаксично правильній комбінації.

Питання: Чи є всі п'ять основних операцій реляційної алгебри дійсно потрібні? Може, одна з них виражається через інші чотири?

Теорема: Кожна з п'яти основних операцій реляційної алгебри незалежна від інших чотирьох, тобто, вона не може бути виражена за допомогою виразу реляційної алгебри, що включає в себе тільки чотири інші операції.

Ідея доказу: для кожної операції реляційної алгебри ми повинні показати властивість, якими володіє ця операція, але не володіє будь-який вираз реляційної алгебри, що включає в себе тільки інші чотири операції.

Наведемо доказ для проєкції та декартового добутку.

Властивість проєкції: це єдина операція, результат якої може мати арність (кількість колонок) менше, ніж його аргумент.

Для доведення потрібно показати по індукції, що арність результату кожного виразу реляційної алгебри в інших чотирьох основних операцій принаймні така ж як арність аргументів.

Властивість декартового добутку: це єдина операція, результат якої має арність більше, ніж на вході.

Для доведення потрібно показати по індукції, що арність результату кожного виразу реляційної алгебри з інших чотирьох основних операцій не більше як найбільша арність своїх аргументів.

Деякі з додаткових операцій реляційної алгебри можуть бути виражені в термінах інших операцій. Наприклад, операція перетину допускає заміну парою операцій різниці:

$$R \cap S = R - (R - S)$$

Операцію тета-з'єднання легко представити у вигляді поєднання операцій декартового добутку та вибірки

$$R \bowtie S = \sigma_c(R \times S)$$

ОБМЕЖЕННЯ РЕЛЯЦІЙНОЇ АЛГЕБРИ

Хоча реляційна алгебра, здається, достатньо потужна для більшості практичних цілей, є кілька простих і природних операторів, які не можуть бути виражені операціями реляційної алгебри. Одним з них є **транзитивне замикання бінарного відношення**. Нехай D є домен D та бінарне відношення R , що є підмножиною $D \times D$. Транзитивне замикання R^+ над R є найменшою підмножиною $D \times D$, що містить R , яке задовольняє наступна умова:

$$\forall x \forall y \forall z ((x, y) \in R^+ \wedge (y, z) \in R^+ \Rightarrow (x, z) \in R^+),$$

тобто всі транзитивні бінарні відношення також входять у транзитивне замикання.

Не існує виразу реляційної алгебри $E(R)$ з R в якості аргументу, яка генерує R^+ .

Наступний приклад ілюструє клас запитів, невимовних засобами реляційної алгебри або реляційного числення з причини невимовності засобами реляційної алгебри транзитивного замикання.

Приклад 6. Якщо ми побудуємо транзитивне замикання для n ієрархічних рівнів, завжди може існувати елемент $n+1$ -го рівня ієрархії. Розглянемо відношення, яке описує співробітників якогось підприємства. Відношення містить дані про табельний номері співробітника, прізвища, посади та табельний номер керівника співробітника - СПІВРОБІТНИКИ (ТАБ_НОМ, ПРІЗВИЩЕ, ПОСАДА, ТАБ_НОМ_КЕР):

ТАБ_НОМ	ПРІЗВИЩЕ	ПОСАДА	ТАБ_НОМ_КЕР
1	Іванов	директор	1
2	Петров	Глав.бух.	1
3	Сидоров	бухгалтер	2
4	Васильєв	начальник цеху	1
5	Сухов	майстер	4

6	Шаріпов	робітник	5
...

Рисунок 9 – Для відношення *Співробітники* можливо побудувати транзитивне замикання для чотирьох рівнів ієрархії

Розглянемо запит "Перерахувати всіх керівників (прямих і непрямих) даного співробітника».

Відповіддю може бути запит, отриманий за допомогою поняття транзитивного замикання. Однак транзитивне замикання не може бути виражене операторами реляційної алгебри.

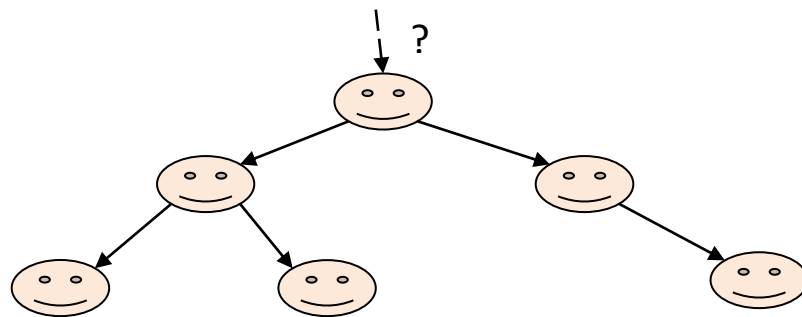


Рисунок 10 – Засобами реляційної алгебри не можна побудувати перелік керівників для кожного підлеглого в ієрархії з довільним числом рівнів

Ця задача потребує не алгебраїчного, а алгоритмічного підходу з використанням циклу до виконання умови. Тому виникає необхідність розширень, як-от PL-SQL та Transact SQL.

Приклад 7. Підготовка Відомості наявності матеріальних цінностей (МЦ) на довільну дату

Нехай у БД зберігаються:

- R - залишки МЦ на складах на початок дня кожного першого числа кожного місяця, а також іще 2 відношення, які містять:
- I – рядки первинних документів надходження МЦ на склади,
- і O - рядки ПД видатків МЦ зі складів.

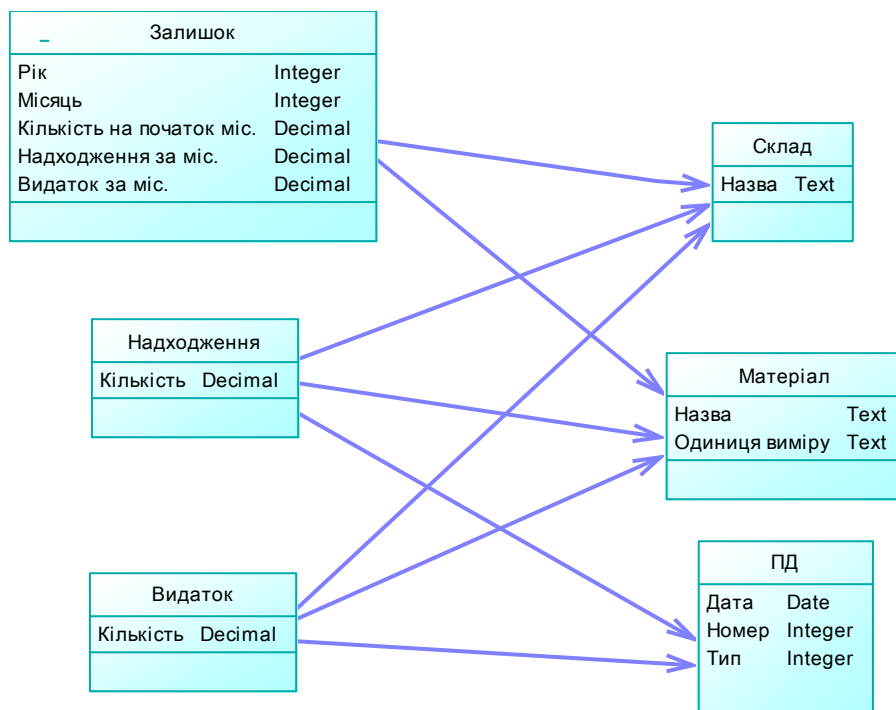


Рисунок 11 – ER-діаграма частини БД системи класу ERP з залишками і оборотами матеріальних цінностей

Обороти по регістру 01 - Залишки ТМЦ за рік - 2010

Балансова одиниця: Суб'єкт: Склад ПММ-зав. Глухенька
 Стаття: Процес:
 Підрозділ: Об'єкт: Бензин А76

Рік:	Початковий залишок	Надходження	Видаток
2010			
Обороти за попередній рік		0,00	0,00
Січень	7427,00	0,00	586,00
Лютий	6841,00	0,00	5595,00
Березень	1246,00	14429,00	2444,00
Квітень	13231,00	30,00	7397,00
Травень	5864,00	0,00	4206,00
Червень	1658,00	0,00	1230,00
Липень	428,00	0,00	61,00
Серпень	367,00	1389,00	1756,00
Вересень	0,00	11747,00	7429,00
Жовтень	4318,00	23659,00	9383,00
Листопад	18594,00	-9784,00	7992,00
Грудень	818,00	9365,00	2924,00
Кінець облікового періоду	7259,00		

Рисунок 12 - Подання руху залишків палива на АЗС підприємства

Для вирішення задачі :

- Об'єднаємо відфільтровані по складу, періоду відношення наявності на початок місяця, надходження і видатку (з мінусом) МЦ за період з початку місяця по дату, попередню даті відомості для складу.

$$\Psi = R_C \bigcup \pi(I_C) \bigcup \pi(O_C)$$

- Згрупуємо останнє відношення, тим самим додавши до залишку на початок місяця надходження за частину місяця і віднявши видаток.

$$\Omega = \gamma(\Psi)$$

Підготовка і видача Відомості наявності запускається з форми фільтру для цієї відомості. З форми беруться дані для фільтрів SQL-запиту відомості. Використовуються фільтри по даті, центрах відповідальності (тобто складах), номенклатурі МЦ (підгрупах або конкретних позиціях).

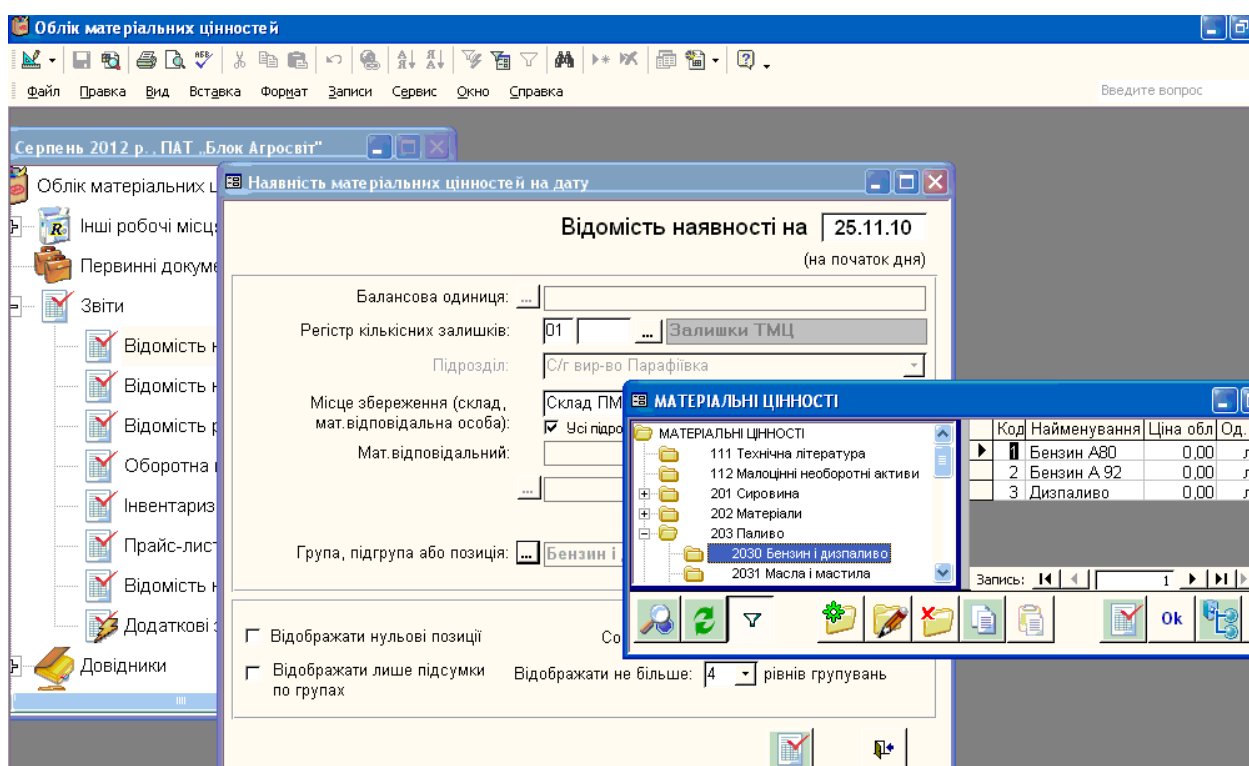


Рисунок 13 – Форма фільтру для Відомості наявності в задачі «Облік МЦ»

Відомість наявності матеріальних цінностей на 25.11.10

Код	Назва	Одиниця виміру	Кількість	Ціна	Сума
Місце збереження: 5 - Склад ПММ-зав. Глухенька					
203 - Паливо					
2030 - Бензин і дизпаливо					
1	Бензин А76	л	8584	5,923	50840,90
2	Бензин А 95	л	3697	5,493	20307,85
3	Дизпаливо	л	30131,291	6,288	189476,45
Разом по підгрупі:			42412,29		260625,20
Разом по групі:			42412,29		260625,20
Разом по місцю збереження:			42412,29		260625,20
Разом по підрозділу:			42412,29		260625,20
РАЗОМ:			42412,29		260625,20

Рисунок 14 - Відомість наявності палива на АЗС підприємства



Рисунок 15 - Запит до сервера для видачі відомості наявності МЦ на довільну дату