

# МОДЕЛЮВАННЯ ПРОЦЕСІВ І СИСТЕМ

УДК 004.652

А. Берко

Національний університет “Львівська політехніка”,  
кафедра інформаційних систем та мереж

## СТРУКТУРНО-СЕМАНТИЧНА ІНТЕГРАЦІЯ ДАНИХ НА ОСНОВІ ФАКТОЛОГІЧНОЇ РЕЛЯЦІЙНОЇ МОДЕЛІ

© Берко А., 2010

**Описано деякі можливості подання неоднорідних та слабкоструктурованих даних. Запропоновано структурно-семантичний підхід, що ґрунтується на фактологічній реляційній моделі для розв’язання поставленої задачі для зовнішнього подання даних на рівні користувача.**

**Any possibilities of heterogeneous and semi-structured data presentation in relation form are considered in proposed paper. Structural-semantic approach based on factological relation model for external presentation of data on end-user level has been proposed for solution of this problem.**

### Вступ

Сьогодні основним напрямом вирішення проблем функціонування інформаційних систем корпоративного та соціального типу в напрямі опрацювання великих обсягів різномірних і різноформатних даних є інтеграція. Інтеграцію визначають як перетворення множини локальних ресурсів на єдиний узгоджений глобальний ресурс, ефективність застосування якого перевищує сумарну ефективність локальних [3]. Загалом принципи інтеграції застосовують до різних складових інформаційних систем. Розрізняють такі напрями інтеграції, як інтеграція бізнес-процесів, інтеграція застосувань, інтеграція платформ та інтеграція даних [3].

Актуальність проблеми інтеграції ресурсів інформаційних систем обумовлена насамперед такими факторами. По-перше – постійне фізичне та змістовне зростання обсягів даних, які застосовують інформаційні системи у різноманітних сферах людської діяльності. По-друге – велика кількість і різноманіття способів, форм і форматів подання даних, а також методів та засобів їх опрацювання. По-третє – активний розвиток інформаційних ресурсів суспільного використання, таких як сховища даних, корпоративні системи, електронні бібліотеки, інформаційні web-системи, системи типу "cloud computing" тощо [4].

Запропоновано структурно-семантичний підхід до інтегрованого подання різномірних даних, який ґрунтується на застосуванні принципів реляційної моделі баз даних, слабкоструктурованих даних та поняття факту як одиниці зображення та сприйняття даних.

### Сучасні методи та засоби інтеграції даних

**Основні проблеми інтеграції даних.** Основною метою інтеграції даних є формування глобального інформаційного ресурсу на основі множини локальних ресурсів для спільного узгодженого застосування. Інтеграція не є простим механічним об’єднанням даних, отриманих з різних джерел. Процедури інтеграції передбачають вирішення цілої низки завдань відбору, перетворення, узгодження, об’єднання, контролю якості даних та багатьох інших [3].

Завдання спільного опрацювання даних різної природи та формату сьогодні вирішують у багатьох сферах застосування інформаційних технологій: корпоративних системах, системах

комп'ютерного моніторингу, в електронному бізнесі, у системах прийняття рішень та бізнес-аналітики тощо.

Сьогодні проблемам інтеграції даних приділяється особлива увага як з боку провідних виробників систем і засобів управління базами даних (Microsoft, Oracle, IBM, SAS, SAP, Informatica), так і в середовищі міжнародних некорпоративних структур – таких, як W3C, OASIS, Integration Consortium та інших.

Дуже багато проблем у галузі інтеграції даних викликано великою кількістю різноманітних підходів і технологій, які часто є несумісними між собою, та відсутність єдиної теоретичної моделі і методики інтеграції даних, незалежної від їх змісту, формату, засобів реалізації та призначення. Загалом нинішня ситуація в сфері інтеграції даних є подібною до ситуації в галузі баз даних до запровадження реляційної моделі [4].

Окремою проблемою є інтеграція структурованих (баз даних) з так званими слабкоструктурованими чи напівструктурованими інформаційними ресурсами. Якщо проблема створення єдиного середовища опрацювання баз даних є загалом достатньо дослідженою і забезпеченою відповідними методиками та технологіями [4], то спільне застосування та опрацювання даних неоднорідної структури сьогодні продовжує залишатися проблемним. Особливістю неоднорідних структур є різноманітність їх форми та змісту, способів і засобів подання та опрацювання, а також дуже часто неповнота, неточність і часткова невизначеність. Використання традиційних технологій баз даних у таких застосуваннях не завжди є ефективним, а часто неприйнятним, а опрацювання структурованих баз даних за принципами слабкоструктурованих призводить до втрати значної частки їхніх властивостей і переваг.

Одним з найскладніших аспектів інтеграції даних є семантична інтеграція [3], яка передбачає формування єдиного змістового простору інтерпретації значень, отриманих з різноманітних джерел. Відомі сьогодні підходи, зокрема [3, 6], відокремлюють проблеми інтеграції семантики від проблем інтеграції структур даних. Однак, згідно з [5], саме структура є формалізованим зображенням певних змістовних понять та відповідностей між ними і значною мірою впливає на порядок сприйняття та інтерпретації як окремих одиниць даних, так і їх поєднань.

У цьому напрямі, на думку автора, найпродуктивнішим є підхід, який поєднує функціональні і технологічні можливості та теоретичний апарат баз даних з вільним форматом та широтою спектра подання і застосування слабкоструктурованих даних, а також структурні і семантичні аспекти подання та опрацювання результатів інтеграції даних.

**Сучасні підходи до реляційної моделі даних.** Серед найістотніших проблем реляційних баз даних, які на початку 1990-х років дали поштовх розвиткові альтернативних підходів до організації інформаційних ресурсів, називають зокрема такі: по-перше, звуження методів і способів опрацювання даних до поняття таблиці, яке не є загалом еквівалентним до початкового поняття відношення і не завжди адекватно відображає логіку та семантику даних, по-друге, недостатньо коректно, з погляду змісту і застосування, подання невизначеностей за допомогою трізначної логіки та псевдоконстанти Null, по-третє, невідповідність між формалізованою табличною структурою даних та різноманітними нетабличними підходами їх зовнішнього подання та опрацювання [1].

Незважаючи на активний розвиток таких новітніх підходів до організації та опрацювання інформаційних ресурсів, як об'єктно-орієнтовані бази даних, слабкоструктуровані і напівструктуровані дані, web-ресурси, графічні та мультимедійні зображення тощо, реляційні бази даних продовжують залишатися основним засобом зберігання і опрацювання даних в інформаційних системах і технологіях різноманітного спрямування. Основними чинниками незмінної популярності реляційної моделі можна назвати такі:

– ґрунтовні теоретичні положення та прогресивні інформаційні технології роботи з реляційними базами даних забезпечують високу ефективність їх опрацювання;

– сьогодні не існує методів і засобів опрацювання нереляційних даних, які б забезпечували однаково ефективність роботи з різними інформаційними ресурсами, співмірну з ефективністю застосування баз даних;

– функціональні можливості реляційної моделі не є вичерпаними, зокрема, вони не обмежені опрацюванням табличних структур, їх можна бути поширити на такі сфери, як опрацювання слабкоструктурованих і неоднорідних даних, застосування об'єктно-орієнтованих та інших новітніх технологій [1, 6].

Принципові положення щодо додаткових можливостей реляційної моделі було викладено і обґрунтовано у Третньому маніфесті К. Дейта та Х. Дарвена [1]. Основною тезою цього документа є твердження, що реляційна модель у класичному трактуванні (не у версії SQL) має достатньо функціональних можливостей для вирішення проблем роботи з різномірними, зокрема, нереляційними слабкоструктурованими даними та застосування об'єктно-орієнтованих принципів у межах реляційної структури. Модель даних, яку пропонують у третьому маніфесті, автори називають "істиною реляційною моделлю". Особливістю реляційної моделі даних у поданні К. Дейта та Х. Дарвена є те, що її може бути застосовано для спільного подання, зберігання та опрацювання як реляційних, так і нереляційних даних.

Базовими положеннями істиною реляційної моделі, викладеними в [1], є такі.

*Значення і змінні.* Різницю між поняттями "значення" і "змінна" автори вважають принциповою і фундаментальною у процесах зображення та застосування даних. Змінною позначають певне поняття чи концепт, з яким може бути пов'язано деяку константу – значення.

*Скалярний тип даних.* Поняття скалярного (або, точніше, інкапсульованого) типу подається як певне узагальнення домену і передбачає зображення одиничних елементарних значень даних у такий спосіб, який не потребує втручання користувача у їх внутрішню структуру при сприйнятті та застосуванні даних. Згідно з таким поданням, значенням скалярного типу можуть бути як число чи символічний рядок, так і бінарний файл, текст, XML-документ, web-сторінка або будь-яка одиниця, над якою визначено дії, що виконують без втручання до її внутрішньої структури. Атрибут визначають як поіменовану визначену множину значень одного скалярного типу [1].

*Генерований тип "кортеж".* Такий тип [1] застосовують як засіб утворення основної одиниці даних – реляційної, призначенням якої є подання певних фактів. Розрізняють змінні типу кортеж та значення типу кортеж. Значенням типу кортеж є послідовність триплетів виду  $\langle A, T, v \rangle$ , де

$A$  – ім'я атрибута,

$T$  – певний скалярний інкапсульований тип,

$v$  – константа відповідного типу.

Множина впорядкованих пар виду  $\langle A, T \rangle$  утворює схему (опис складу та структури) кортежу.

*Генерований тип "відношення".* Автори [1] вводять поняття змінної типу "відношення" та значення типу "відношення". Значенням такого типу є множина значень типу "кортеж", які мають однакову схему. Схема кожного кортежу при цьому збігається зі схемою відношення. Значення типу "відношення" застосовують як характеристику стану певної множини однотипних фактів, визначених у заданій предметній області.

### Формулювання цілей статті

Основною метою роботи є подання концепції та обґрунтування принципів запропонованого автором структурно-семантичного підходу до подання інтегрованих даних на зовнішньому рівні, а саме:

- формулювання та опис концепції фактологічної реляційної моделі;
- визначення та обґрунтування поняття факту як структурно-семантичної одиниці зовнішнього подання даних та його властивостей;
- обґрунтування можливості подання реляційних, слабкоструктурованих даних та даних без попередньо визначеної структури за принципами фактологічної реляційної моделі;
- формулювання основних принципів маніпулювання фактами та визначення повної, несуперечливої і замкненої множини операцій над фактами;
- обґрунтування можливості та ефективності подання результатів інтеграції неоднорідних інформаційних ресурсів у формі фактологічних відношень.

## Основні результати досліджень. Фактологічна реляційна структура даних

**Факти та фактологічні відношення.** Поняття факту в реляційній моделі не є новим. Вперше факт як одиницю даних у відношеннях реляційної бази даних було визначено у [2], пізніше його почали застосовувати як структурну та змістовну одиницю у технологіях сховищ даних, аналітичних системах, системах і технологіях прийняття рішень тощо [4]. Згідно з концепцією факту, сформульованою у [2], поняття кортежу є занадто формалізованим і недостатньо відповідає суті та змісту даних. Це створює низку проблем при роботі з таблицями (відношеннями), зокрема, в операціях вибору та оновлення даних. Саме тому у [2] запропоновано замінити кортеж поняттям факту як логічно завершеної достовірної змістовної одиниці, яка має власну інтерпретацію у визначеній предметній області. Кортеж (чи підкортеж) відношення вважають константою, яка є зображенням деякого факту [2]. За таким визначенням один кортеж відношення може містити зображення множини фактів, кожен з яких має власну інтерпретацію. Така концепція цілком узгоджується з положеннями "істинно реляційної моделі" у поданні К. Дейта та Х. Дарвена [1].

Порівнянням концепцій організації даних, викладених у [1, 2], можна поєднати їх у принципово новій моделі зображення та опрацювання даних. В основу підходу, запропонованого автором, покладено поняття факту як множини значень, що подає певні достовірні відомості, релевантні щодо деякої предметної області. Множину атрибутів, значення яких застосовують для зображення деякого факту  $f$ , називатимемо його схемою і позначимо як  $sch(f) = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$ . Схема факту визначає його зміст та склад значень, за допомогою яких цей факт задано. Реалізацією факту  $f$  назвемо деякий кортеж загального вигляду

$$r^f = \langle a_{i1}, a_{i2}, \dots, a_{ik} \rangle,$$

складений з припустимих значень атрибутів, що входять до схеми цього факту. Загалом факт може позначати такі поняття, як:

- існування деякого концепту (сутності),
- наявність властивостей (атрибутів) у певного концепту;
- зв'язок між атрибутами, що описують властивості концепту;
- відношення чи зв'язок між концептами тощо.

Наведемо приклад утворення фактів на основі відношення реляційної бази даних. Нехай кортеж  $r^*$  відношення *Студент* вигляду  $r^* = \langle 12345, \text{Петренко}, \text{Сергій}, \text{КН-41}, 4.72 \rangle$ , наведений на рис. 1, є загальним поданням зокрема таких фактів :

- $f_1$  : *Петренко Сергій* є студентом, поданого кортежем  $r_1 = \langle \text{Петренко}, \text{Сергій} \rangle \subseteq r^*$ ,
- $f_2$  : *Петренко* навчається в групі *КН-41*, реалізованого як  $r_2 = \langle \text{Петренко}, \text{КН-41} \rangle \subseteq r^*$ ,
- $f_3$  : *Петренко* має середній бал *4.72*, зображеного кортежем  $r_3 = \langle \text{Петренко}, 4.72 \rangle \subseteq r^*$ ,
- $f_4$  : *Петренко* має залікову книжку з номером *12345*, реалізованого кортежем  $r_4 = \langle 1234, \text{Петренко} \rangle \subseteq r^*$  тощо.

### Студент

№ залікової	Прізвище	Ім'я	Група	Середній_бал
...	...	...	...	...
12345	Петренко	Сергій	КН-41	4.72
...	...	...	...	...

кортеж  $r^*$

Рис. 1. Приклад зображення фактів у відношенні

На основі такого трактування факту пропонується визначити фактологічне відношення (або відношення фактів) як структурну та функціональну одиницю даних. Попередньо визначимо поняття схеми фактологічного відношення. Схемою фактологічного відношення  $R^F$  назвемо вираз вигляду

$$Sch(R^F) = R^F(A_1, A_2, \dots, A_n),$$

де  $\{A_1, A_2, \dots, A_n\}$  – множина атрибутів, значення яких застосовують для подання певної категорії фактів.

Фактологічним відношенням  $R^F$  зі схемою  $Sch(R^F)$  назвемо множину фактів, схема яких входить до складу схеми відношення

$$R^F = \{f / sch(f) \mid f \in Sch(R^F)\}.$$

У такий спосіб класичне поняття відношення реляційної бази даних значно розширюється і виходить за межі простого табличного зображення даних. Користувач отримує можливість за допомогою такого засобу в межах реляційної моделі баз даних оперувати структурами довільного вигляду, як показано, наприклад, на рис. 2.

Відношення  $R^F$

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$
факт $f_1$								
факт $f_2$								
факт $f_3$								
факт $f_4$								
факт $f_5$								
факт $f_6$								
факт $f_7$								
факт $f_8$								
факт $f_9$								

Рис. 2. Приклад фактологічного відношення з кортежами різної розмірності

З погляду структури фактологічне відношення можна розглядати як таблицю з рядками різної розмірності, що часто зустрічається на практиці. Більш формально таке відношення утворюється як множина локальних відношень, відображених на спільну схему

$$R^F = \{R_1^F[Sch(R^F)], R_2^F[Sch(R^F)], \dots, R_N^F[Sch(R^F)]\},$$

де  $R_i^F[Sch(R^F)]$  – відображення локального відношення  $R_i^F$  на схему  $Sch(R^F)$ . На відміну від схеми відношення у класичній реляційній моделі, схема фактологічного відношення є неформальним переліком властивостей деякого класу сутностей. З погляду змісту, фактологічне відношення утворює деякий простір даних зі системою семантичних координат (вимірів). Умовно назвемо його фактологічним семантичним простором. Кожен вимір цього простору утворюють на основі деякого атрибута, а значення атрибутів формують метрику цього виміру. Факт у такому сенсі розглядають як певну точку фактологічного семантичного простору, якій відповідає специфікований набір значень вимірів (атрибутів). Приклад семантичного простору та фактів, визначених у ньому, показано на рис. 3.

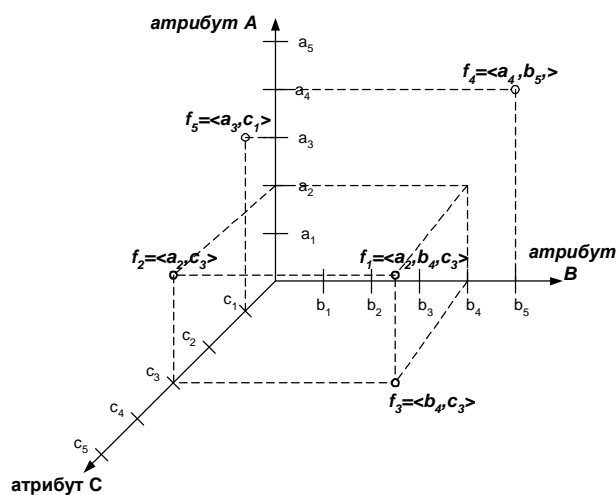


Рис. 3. Приклад зображення фактів у семантичному просторі

Подібну інтерпретацію фактів застосовують, наприклад, у сховищах даних, але на відміну від традиційних підходів, у фактологічній моделі кількість значень, які застосовують для опису факту,

може бути різною. Множина точок семантичного простору, зображених на рис. 3, відповідає фактологічному відношенню, наведеному на рис. 4.

**Відношення  $R_i^F$**

	A	B	C
факт $f_1$	$a_2$	$b_4$	$c_3$
факт $f_2$	$a_3$		$c_4$
факт $f_3$		$b_4$	$c_3$
факт $f_4$	$a_4$	$b_5$	
факт $f_5$	$a_3$		$c_1$

Рис. 4. Результат відображення фактологічного семантичного простору у табличну форму

Запропонована фактологічна реляційна модель даних не заперечує класичної реляційної моделі баз даних, оскільки стосується способів подання та опрацювання даних на зовнішньому рівні користувача. При цьому на концептуальному рівні бази даних зберігаються реляційні принципи роботи з даними, тобто утворюється комбінація "реляційна база даних – фактологічні засоби зовнішнього подання даних".

Джерелом значень для формування фактологічного відношення у цьому випадку є таблиці (відношення) бази даних. Об'єкт, кортежі якого застосовують для утворення фактів, назвемо базовим відношенням. Загалом базове відношення може бути як таблицею, так і результатом перетворень однієї або більше таблиць бази даних. Загальну схему формування фактологічного відношення на основі реляційної бази даних показано на рис. 5.

Як видно з рисунка, зовнішнє фактологічне відношення можна розглядати як результат виконання послідовності операцій проєкції та селекції базового відношення.

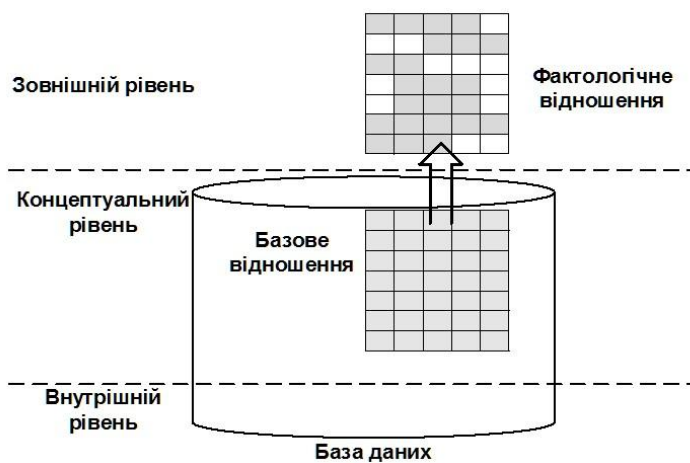


Рис. 5. Загальна схема формування фактологічного відношення на основі реляційної бази даних

Джерелом даних фактологічного відношення може бути не лише таблиця (відношення) реляційної бази даних. У загальному випадку факт як змістовно завершену одиницю можна утворювати на основі значень, поданих в інших форматах – тексту, XML-документів, web-ресурсів тощо. При цьому слід дотримуватися тих самих принципів, що і у випадку баз даних – кожен факт задається послідовністю значень, які можна однозначно інтерпретувати у визначеній предметній області.

На відміну від процедури формування фактів на основі кортежів базового відношення, утворення фактів з даних слабкоструктурованих форматів є значно складнішим. Порядок та методи видобування фактів значною мірою залежать від виду та формату джерела даних. Загалом цей процес можна розглядати як частковий випадок таких методик, як text mining, web mining, content mining тощо, які сьогодні є достатньо відомими і активно розвиваються. Ланцюжок значень, що

зображають факт, при цьому, утворюють способом, який обов'язково враховує семантику поєднання цих значень у цілісне поняття. Таку процедуру, яка виділяє з джерела даних множину взаємопов'язаних значень та поєднує їх у факти, назвемо видобуванням фактів (*fact mining*). Методи і засоби вирішення цієї проблеми є окремим предметом досліджень і виходять за межі цієї роботи.

Загальну схему утворення фактологічного відношення на основі джерел довільних слабкоструктурованих форматів показано на рис. 6.

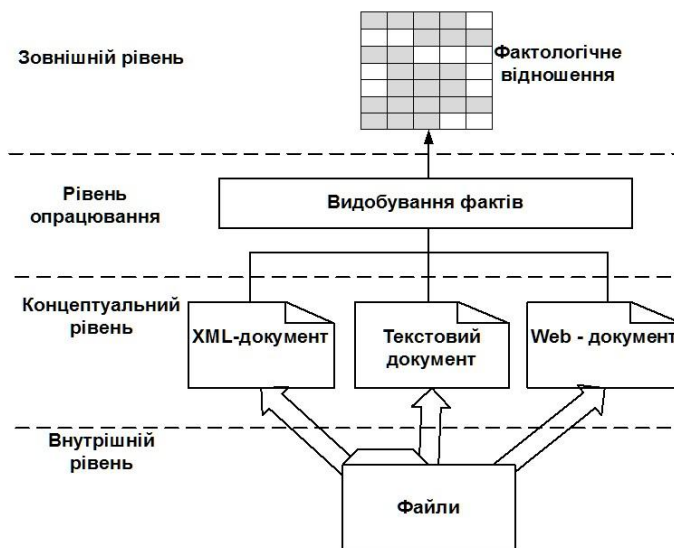


Рис. 6. Загальна схема формування фактологічного відношення на основі даних довільного формату

Можливості формування фактологічного відношення на основі структурованих та слабкоструктурованих джерел дозволяють використати їх як уніфікований засіб інтеграції різномірних даних на зовнішньому рівні. Результатом такого процесу буде деяке фактологічне відношення  $R^f$

$$R^f = R_s^f \text{ UNION } R_{ss}^f,$$

де  $R_s^f$  – фактологічне відношення, утворене на основі базового відношення реляційної бази даних;  $R_{ss}^f$  – фактологічне відношення, утворене на основі слабкоструктурованих джерел даних;  $\text{UNION}^f$  – оператор об'єднання фактологічних відношень, який об'єднує схеми та інформаційне наповнення двох фактологічних відношень.

При цьому схему результуючого відношення утворюють об'єднанням схем відношень операндів, а наповнення – об'єднанням наповнення цих відношень, як показано на рис. 7.

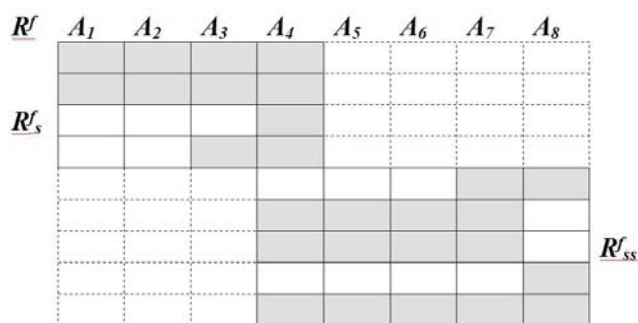


Рис. 7. Схема виконання операції об'єднання фактологічних відношень

У такий спосіб створюється можливість оперувати неоднорідними даними, зберігаючи їх як у структурованій реляційній формі, так і у слабкоструктурованих форматах, а також даними з високим рівнем невизначеності. Окрім того, фактологічна реляційна структура може бути легко перетворена до інших форматів (наприклад, XML, текстового тощо), виконуючи при цьому функції проміжної ланки між реляційними та слабкоструктурованими даними.

**Операції над фактами.** Для виконання перетворень фактів у фактологічних відношеннях та отримання нових фактів на основі заданих пропонується визначити множину операцій над фактами. Подібний підхід до визначення дій на концептуальному рівні було визначено у мові маніпулювання інтегрованими даними, описаній у [6]. Але на відміну від пропозицій, викладених у цій роботі, базовою елементарною одиницею маніпулювання інтегрованими даними у [6] визначено кортеж.

Основними вимогами до набору операцій над фактами є:

- замкненість – результатом перетворень фактів має бути лише факт;
- повнота – якщо деякий факт може бути отримано на основі інших, то у визначеному наборі існує послідовність операцій над фактами, за допомогою яких це може бути зроблено;
- несуперечність – не існує двох різних операцій, які утворюють однаковий результат і одна й та сама операція завжди дає однаковий результат;
- універсальність – результат операції не залежить від складу і змісту фактів, над якими її виконують.

Як концептуальну основу для множини операцій над фактами використано такий засіб маніпулювання відношеннями, як реляційна алгебра. Однак при цьому самим діям та їх змісту надано принципово іншу інтерпретацію, а деякі операції не мають прототипів у реляційній алгебрі і є визначеними винятково для фактів. Пропонується визначити таку множину операцій над фактами у фактологічній реляційній моделі.

1. **Логічний вивід фактів.** Деякий факт  $f''$  назвемо логічним наслідком факту  $f'$ , позначимо це як  $f' \models f''$ , якщо  $sh(f'') \bar{I} sh(f')$ , а для кортежів  $r'$  і  $r''$ , які є реалізацією, відповідно, фактів  $f'$  і  $f''$  виконується співвідношення  $r'' \bar{I} r'$ . Тобто, деякий факт можна отримати з іншого, якщо множина значень атрибутів, що описують факт входить до складу множини значень, які описують інший факт. Факт, на основі якого виконують операцію логічного виведення, назвемо фактом-прецедентом; факт, утворений як результат операції – фактом-консеквентом. Загальну схему виконання операції показано на рис. 8.

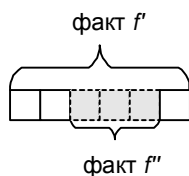


Рис. 8. Загальна схема виконання операції логічного виведення фактів

Зміст операції логічного виведення полягає у продукуванні нових фактів шляхом отримання часткових відомостей на основі повніших. Наприклад, логічним висновком факту, зображеного кортежем  $r^* = \langle 12345, \text{Петренко, Сергій, КН-41, 4.72} \rangle$  (див. рис. 1), є факт, реалізований кортежем  $r' = \langle \text{Петренко, КН-41} \rangle$ , який інтерпретується як "студент на прізвище Петренко навчається в групі КН-41". Технічно результат логічного виведення утворюють як проекцію деякого факту на частину атрибутів його схеми. Ця операція має властивість рефлексивності, тому будь-який факт можна вважати логічним наслідком самого себе.

Факт  $f^+$ , логічним наслідком якого є деяка скінченна множина фактів  $\{f_1, f_2, \dots, f^k\}$ , назвемо замиканням цієї множини.

2. **Доповнення фактів.** Якщо  $f$  – факт, реалізацією якого є кортеж  $r$ ,  $A^*$  – атрибут фактологічного відношення, то результатом операції доповнення факту  $f$  атрибутом  $A^*$  буде новий факт  $f^+ = f \text{ add } A^*$ , реалізацією якого є кортеж  $r^+ = \langle r, a_i^* \rangle$ , де  $\text{add}$  – оператор доповнення факту,  $a_i^*$  – значення атрибута  $A^*$ . Схему виконання операції показано на рис. 9.

Операція доповнення за змістом передбачає підвищення рівня деталізації факту, тобто збільшення кількості значень, які його описують. Наприклад, якщо факт  $f'$ , заданий як  $r' = \langle \text{Петренко, КН-41} \rangle$ , доповнити значенням атрибута "Середній\_бал=4.72", то у результаті



отримаємо новий факт  $f''=f' \text{ add } \textit{Середній\_бал}$ , реалізацією якого є кортеж  $r''=\langle r', 4.72 \rangle = \langle \textit{Петренко, КН-41, 4.72} \rangle$ .

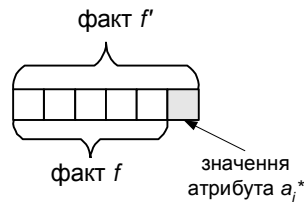


Рис. 9. Загальна схема виконання операції доповнення фактів

3. *Поглинання фактів.* Якщо у фактологічному відношенні одночасно присутні факти  $f'$  і  $f''$  такі, що  $f_i \models f_j$ , які реалізовано відповідно кортежами  $r'$  і  $r''$ , то результат операції поглинання фактів буде утворено як  $f' \text{ absorb } f'' \approx f'$ , де *absorb* – оператор поглинання фактів. Тобто, якщо деякий факт утворено на основі іншого шляхом логічного виведення, і множина значень, які його зображують, входить до складу кортежу, що його зображає цей інший факт, то у складі фактологічного відношення може бути залишено лише початковий факт. Наприклад, якщо у фактологічному відношенні мають місце факти  $f'$  і  $f''$ , задані відповідно кортежами  $r'=\langle \textit{Петренко, КН-41, 4.72} \rangle$  і  $r''=\langle \textit{Петренко, КН-41} \rangle$ , то для подання відомостей, зосереджених у них, достатньо лише одного факту  $f'$ , заданого кортежем  $r'$ . За своїм змістом операція поглинання фактів є оберненою до операції логічного виведення, тому її можна проілюструвати схемою, поданою на рис. 8.

4. *З'єднання фактів.* Якщо для двох фактів деякого фактологічного відношення  $f'$  і  $f''$  справедливе співвідношення схем  $sh(f') \zeta sh(f'') \in \mathcal{A}$  і співвідношення реалізацій  $r' \zeta r'' \in \mathcal{A}$ , то для них може бути визначено операцію з'єднання  $f' \text{ join } f''$ , де *join* – оператор з'єднання фактів. Результатом натурального з'єднання фактів буде новий факт  $f^*=f' \text{ join } f''$ , такий що схемою цього факту буде  $sh(f^*)=sh(f') \dot{\zeta} sh(f'')$ , а реалізацією кортеж  $r^*=r' \dot{\zeta} r''$ , що об'єднує значення атрибутів, які зображують факти  $f'$  і  $f''$ . Схему виконання операції з'єднання фактів наведено на рис 10.

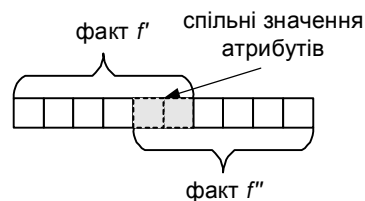


Рис. 10. Загальна схема виконання операції з'єднання фактів

За змістовною інтерпретацією, операція з'єднання фактів означає утворення розширеного опису одного поняття шляхом злиття двох окремих аспектів, які подані різними фактами і на зв'язок між якими вказує збіг значень деяких атрибутів.

5. *Умовне з'єднання фактів.* Ця операція є подібною до попередньої, за винятком вимоги до складу фактів, над якими її можна виконати – для умовного з'єднання передбачається, що  $sh(f') \zeta sh(f'') = \mathcal{A}$ . У цьому випадку критерієм утворення нового факту на основі двох заданих є деяка умова, сформульована щодо значень атрибутів у кортежах, які зображають задані факти –  $Q(r', r'')$ . Найпростішим прикладом такої умови є порівняння значення одного чи кількох атрибутів одного факту зі значеннями атрибутів іншого факту – вираз вигляду  $A_i \theta A_j$  ( тут  $A_i$  і  $A_j$  – імена атрибутів,  $\theta$  – оператор порівняння). Результатом умовного з'єднання фактів  $f'$  і  $f''$  за умовою  $Q$  буде новий факт  $f^*=f' \text{ join}[Q(r', r'')] f''$  такий, що  $sh(f^*)=sh(f') \dot{\zeta} sh(f'')$ , реалізацією якого є кортеж, утворений шляхом конкатенації кортежів, що зображають початкові факти за умови виконання для значень атрибутів, у яких вимоги критерію  $Q$ , а саме  $r^*=\langle r', r'' | Q(r'_i, r''_i)=true \rangle$ , де  $\text{join}[Q(r', r'')]$  – оператор з'єднання фактів за умовою  $Q$ . Загальну схему виконання операції умовного з'єднання фактів показано на рис. 11. За своїм змістом операція дає змогу утворити новий факт шляхом злиття фактів, що стосуються

різних сутностей і не мають спільної ланки, зв'язок між якими реалізує задане співвідношення значень окремих атрибутів.

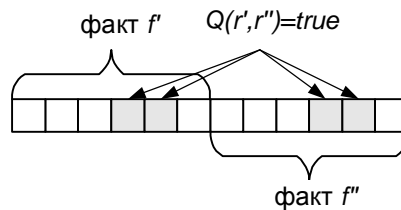


Рис. 11. Загальна схема виконання операції умовного з'єднання фактів

Запропонований набір операцій над фактами, який є, загалом, повним і несуперечливим, може бути означено як алгебру фактів. Алгебра фактів, своєю чергою, може слугувати теоретичною основою і підґрунтям для розроблення засобів опрацювання та перетворення результатів інтеграції даних на рівні користувача. Окрім операцій над фактами як окремими структурно-семантичними одиницями даних, можливим є визначення множини операцій над фактологічними відношеннями за схемою, подібною до реляційної алгебри.

### Висновок

Застосування фактологічних відношень як засобу зовнішнього зображення різнорідних даних, початково поданих і збережених у реляційних структурах, а також у вигляді напівструктурованих даних, має такі переваги:

- поєднати в єдиних процесах структурні та семантичні аспекти принципу подання та опрацювання даних;
- перейти від суто структурних, формалізованих одиниць у маніпулюванні даними до змістовних понять – фактів;
- застосовувати для зображення одиниць даних кортежі різної розмірності;
- відмовитись від застосування трізначної логіки та псевдоконстанти Null ;
- поєднати в єдиному середовищі структуровані та напівструктуровані форми подання даних.

Підхід, запропонований автором, може бути застосовано у розв'язанні зокрема таких проблем, як інтеграція інформаційних ресурсів, інтелектуальних аналіз даних, опрацювання неповних і неточних даних, перетворення форматів даних, створення гетерогенних структур тощо.

1. Date C.J. *Foundation for Future Database Systems: The Third Manifesto*, 2<sup>nd</sup> edn./ Date C.J., Darwen H.- Harlow: Addison Wesley Longman, 2000.
2. Desai B.C. *Fact structures and its application to updates in Relational Databases.*/ Desai B.C., Goyal P., Sadri F. // *Information Systems Vol. 12, No 2, 1987.*- p. 215-221.
3. Lenzerini M. *Data Integration: A Theoretical Perspective.* / Marco Lenzerini // *Proc. of the ACM Symp. on Principles of Database. Systems (PODS), 2002.* – pp. 233 – 246.
4. *The Lowell Database Research Self-Assessment Meeting, Lowell Massachusetts, 4-6 May 2003.*– [Електронний ресурс].–Режим доступу: <http://research.microsoft.com/~gray/lowell>. – June, 2003.
5. Wache H. *Ontology-Based Integration of Information – A Survey of Existing Approaches* / H.Wache, T. Voegelé, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hubner // *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 4-5, 2001*, pp.108-118.
6. Калиниченко Л.А. *Методы и средства интеграции неоднородных баз данных* / Л. Калиниченко. – М.: Наука, 1983. – 424 с.