

## АДАПТАЦІЯ EAV МОДЕЛІ ДЛЯ МЕДИЧНИХ БАЗ ДАНИХ

Застосування традиційних реляційних моделей для реалізації клінічних баз даних (БД) не є ефективним внаслідок великих вимог останніх щодо масштабованості, оперативності доступу та специфіки самої структури БД – розрідженості таблиць та необхідності зміни схеми БД за нарощування кількості і типів досліджуваних показників з появою нових можливостей та інструментів дослідження. Подолання цих недоліків застосуванням noSQL моделей, в першу чергу документоорієнтованих, розв'язує проблеми масштабованості, але не забезпечує повноцінної підтримки цілісності і вимог ACID, що, на наш погляд, є неприпустимим для БД медичного призначення.

Як наслідок, найбільшого розповсюдження у клінічних базах даних набула модель EAV (Entity-Attribute-Value) [1], що є різновидом реляційної моделі, в якій обмеження жорсткості схеми реляційної БД долається збереженням сутностей (Entity), атрибутів (Attribute), склад яких при описі сутності може змінюватися, та їхніх значень (Value) в окремих таблицях. Опис конкретного екземпляру сутності через набір значень динамічних атрибутів здійснюється через асоціативну таблицю «Значення» («EAV»), що містить ідентифікатори сутності, атрибуту та його значення. Зміна складу атрибутів не вимагає зміни схеми БД, оскільки новий атрибут просто додається як кортеж у таблицю «Атрибути». Зникає також і розрідженість таблиць, оскільки в таблиці «Значення» присутні лише ненульові рядки. Модель EAV уперше була застосована в базі даних електронних медичних карт TMR (The Medical Record) і сховищі клінічних даних HELP Clinical Data Repository. Наявність одного-єдиного стовпця «Значення» в відповідній таблиці моделі EAV передбачає однаковість типів даних значень всіх атрибутів, що для медичних баз даних, у яких результати досліджень можуть відбиватися чисельними значеннями, графіками, знімками тощо, є неприйнятним. Для забезпечення можливості введення різноманітних типів даних у БД клінічних досліджень TrialDB використали кілька таблиць EAV (по одній для кожного типу даних) [2]. Такий підхід забезпечує динамічність складу атрибутів (досліджуваних характеристик) та довільність типів їх даних за незмінної схеми БД, але не відбиває зв'язку між експериментально досліджуваним показником і фізіологічним параметром, який ним характеризується, що ускладнює застосування даних для діагностики захворювань. Так, тиск крові оцінюється за значеннями діастолічного, систолічного і пульсового тисків; норма його залежить від віку, ваги, статі, анамнезу пацієнта. Відповідно окрему групу вимірюваних / виявлених показників можуть складати стандартні показники стану пацієнта (вага, зріст, стать, вік, паління, перенесені захворювання тощо), які впливають на оцінку результатів досліджень. Для усунення цих недоліків ми пропонуємо розділити останню групу показників на статичні (стать, рік народження тощо), які зберігати як значення атрибутів сутності «Пацієнт», і динамічні (вага, паління тощо), які виділити в окрему таблицю «Показники стану» зі внесенням у неї значень цих атрибутів на момент дослідження, а також використати ієрархічну структуру досліджуваних показників, виділивши підтипи за типами даних, та відобразити склад показників, за якими оцінюється той чи інший параметр, введенням асоціативної таблиці «Характеристики параметра» (рис. 1).

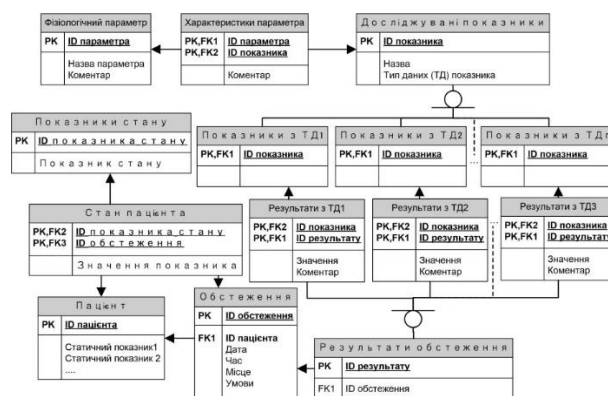


Рис. 1. Фрагмент схеми БД за модифікованою EAV моделлю

Як наслідок, за запитом щодо значення параметру за результатами обстеження виводимуться значення всіх відповідних показників, а зміна складу показників при оцінці параметру за результатами нових досліджень також не вимагатиме зміни схеми БД.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Batra S., Sachdeva, S., Bhalla S. Entity Attribute Value Style Modeling Approach for Archetype Based Data. *Information*. 2018. URL: <https://www.mdpi.com/2078-2489/9/1/2> (дата звернення 12.03.2021).
2. Brandt C., Nadkarni, P., Marengo L., Karras B., Lu C., Schacter L., Fisk J.M., Miller P.L. Reengineering a database for clinical trials management. *Control. Clin. Trials*. 2000. № 21. P. 440–461.