

УДК 681.3.016

Фісун М.Т. д-р техн. наук, професор, завідувачий кафедрою комп'ютерних технологій МФ НаУКМА – відповідальний редактор за напрямом “Технічні науки”;



Фісун Микола Тихонович,

1940 р.н. Зав. кафедри комп'ютерних технологій

МФ НаУКМА. Доктор технічних наук, професор. 83 наукові праці. Напрямки наукової діяльності: системи автоматизованого проектування; бази даних і бази знань; інформаційно-пошукові системи.

Про співвідношення між нормалізацією таблиць та ER-схемами при проектуванні структури бази даних

Показано, як поєднати роботу із створення структур баз даних методами побудови діаграм “Сутності – зв’язки” (ERD – Entity Relationship Diagram) з нормалізацією реляцій. Зроблено висновок, що коректна побудова ERD предметної сфери забезпечує для системи реляцій створюваної бази даних третю нормальну форму. Такий висновок продемонстровано на прикладі побудови ERD для інформаційної системи щодо приміщень будівлі та предметів і технічних систем, що знаходяться в них.

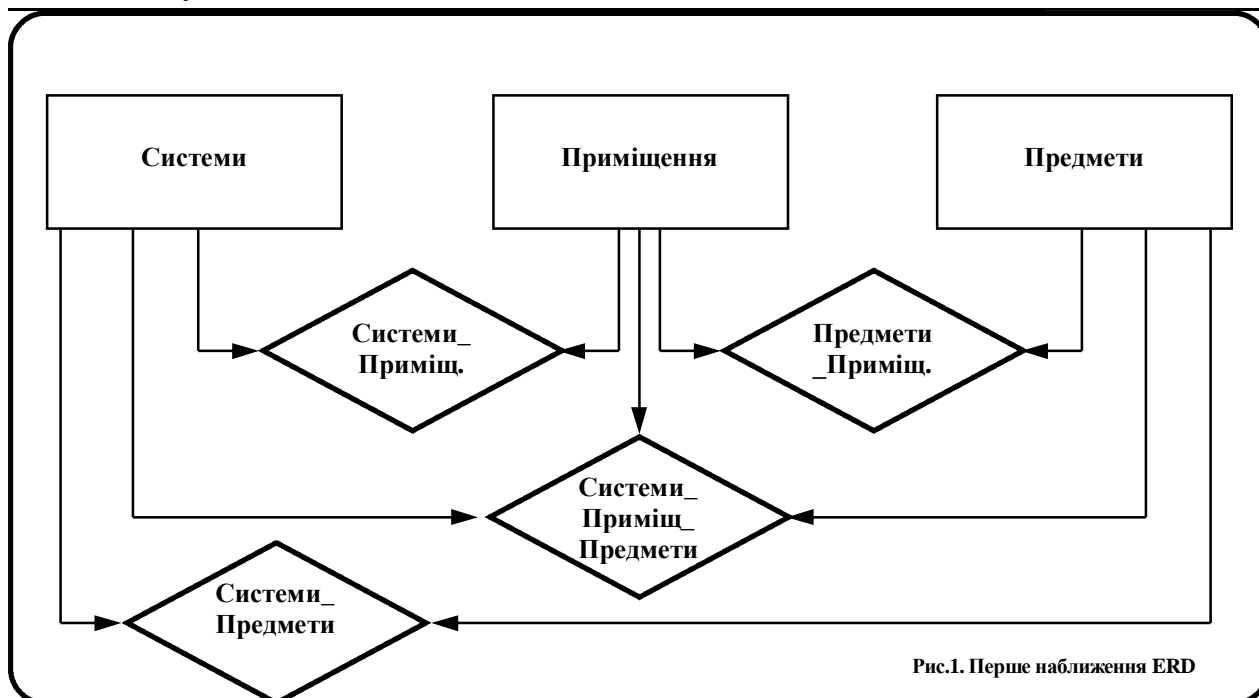
Is shown, how to connect such work on creation structures of database structures, as the construction ERD (Entity Relationship Diagram) and normalization of the relations. A conclusion is made, that at the correct construction ERD of the subject area the system of the relations of a created database will be in the third normal form. The conclusions are illustrated by an example of construction ERD for an information retrieval system for premises and subjects, which are in them.

Ефективність роботи комп'ютерних інформаційних систем (ІС), які обробляють структуровану у вигляді таблиць інформацію, суттєво залежить від структури бази даних (БД). На жаль, розробники таких ІС мало приділяють уваги етапу проектування логічної моделі ІС. Розробка структури БД більшості ІС проводяться стихійно: створюються таблиці, на базі яких можна реалізувати першочергові запити, потім з появою нових запитів поширюються існуючі таблиці або створюються нові, і цей процес триває довго, поки не звільниться головний розробник або керівництво фірми не запросить нового конструктора БД. Тому, коли приходиться розбиратися зі структурою наявних баз даних, в більшості випадків приходиш до висновку, що вона розроблялася “за натхненням”, а не за правилами, які визначені теорією баз даних. Частково це можна пояснити тим, що методичні нароби стосовно проектування структур БД не дуже пов’язані між собою. Так, зокрема, у всіх підручниках з баз даних [1,2,3 та ін.] не вказується взаємозв’язок між методологією побудови ERD (Entity Relationship Diagram) та нормалізацією відношень при застосуванні реляційної моделі даних.

Побудова ERD пов’язана з встановленням об’єктів (сутностей) предметної галузі та їх властивостей і зв’язків між об’єктами та

властивостями цих зв’язків. При переході на логічний рівень бази даних в реляційних моделях даних об’єкти і зв’язки перетворюються в таблиці, а їх властивості – в атрибути цих таблиць. При застосуванні ієрархічної та мережної моделей даних об’єкти та їх властивості також перетворюються в таблиці та атрибути цих таблиць, а зв’язки перетворюються в структури типу “link”. Якщо зв’язки мають свої властивості, то вони також перетворюються в таблиці з відповідними атрибутами.

При нормалізації відношень розглядаються або існуючі “паперові” таблиці, або таблиці існуючої бази даних. В підручниках, як правило, подається приклад ненормалізованої “великої” таблиці (за кількістю та різноманітністю атрибутів), яка послідовно переводиться в першу, другу, третю і навіть в четверту нормальну форму. Але за начебто простими з алгоритмічної точки зору діями губиться сутність перетворень (з концептуальної точки зору). Тому слід ці два комплекси робіт (побудова ERD предметної галузі та нормалізація відношень) проводити паралельно. Якщо проектування структури бази даних починається з нуля, то можна взагалі обійтися побудовою ERD. При правильній їх побудові відновлені таблиці будуть знаходитися у третій нормальній формі. Тому більш



фундаментальним методологічним підґрунтям для розуміння предметної галузі і проектування “гарної” структури бази даних є побудова ERD.

Продемонструємо це на прикладі проектування структури бази даних стосовно приміщень та предметів, що там знаходяться. Сутність задачі полягає в наступному. Треба розробити інформаційно-пошукову систему (ІПС) стосовно приміщень комплексу будівель та насичення цих кімнат предметами як автономними, так і в складі яких-небудь систем (опалення, водопостачання, комп’ютерної мережі тощо). Таким чином, основними інформаційними об’єктами (сутностями) такої ІПС є **приміщення**, **предмети** та **системи**. Ці об’єкти мають свої властивості. Так, кожен екземпляр об’єкта **приміщення** має такі властивості, як площа, функціональне призначення та ін., а кожен екземпляр об’єкта **предмети** має такі, властивості як назва предмета, вартість і т.д. Тип взаємозв’язків між об’єктами **приміщення** і **предмети** – **1:n**, тобто в одному приміщенні може бути декілька предметів, а один предмет може бути в даний час тільки в одному приміщенні, а між об’єктами **предмети** та **системи** – **1:n**, між об’єктами **приміщення** та **системи** – **m:n**. Так на першому кроці побудови структуру БД можна представити схемою, яка представлена на рис.1.

На рисунку прямокутниками відображені об’єкти, а ромбами – зв’язки між об’єктами. Одночасно в прямокутниках і в ромбах попередньо записуються властивості об’єктів, в тому числі такі, що роблять ці об’єкти унікальними (тут це не зроблено через обмеженість обсягу).

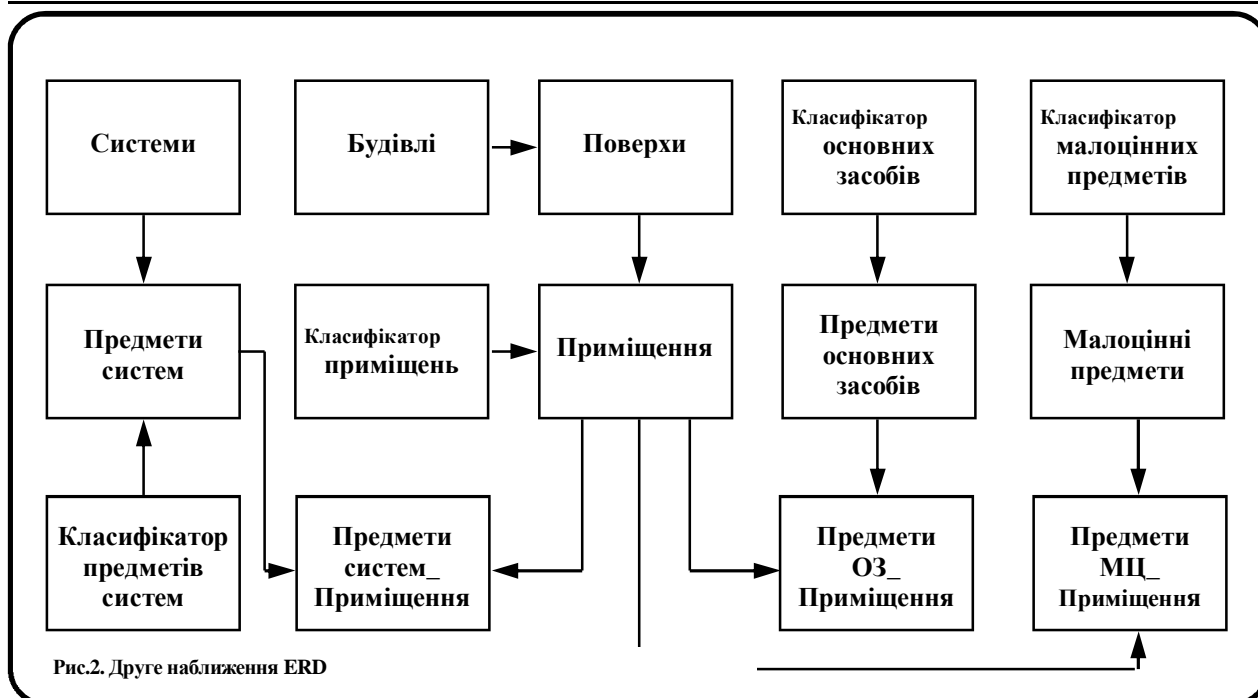
Для робітників адміністративних, господарчих та

фінансових підрозділів важливою інформацією є узагальнена інформація і про предмети, і про приміщення. Якщо фізично приміщення розташовані в будівлях на певних поверхах, то треба ввести природні угруповання приміщень – **поверхи** і **будівлі**, взаємозв’язки між якими та з об’єктом **приміщення** ієрархічні. У відповідності до [3] об’єкт **будівлі** відноситься до правильних, а об’єкти **поверхи** і **приміщення** – до слабких.

Для предметів нема природнього групування, тому треба ввести штучну класифікацію, виходячи з потреб користувачів інформації, здорового глузду та правил класифікації та кодування інформації, наприклад “столи”, “шафи” і т.д. Але на самому верхньому рівні, виходячи з різних підходів до бухгалтерського обліку предметів, розподілимо їх на такі три класи:

- предмети, що за вартістю відносяться до основних засобів (вони мають інвентарні номери);
- предмети, що за вартістю відносяться до так званих малоцінних предметів (вони не мають інвентарних номерів);
- предмети, що входять до складу систем.

Ці три класи предметів слід тепер розглядати як окремі об’єкти (за [3] підтипи супертипа – **предмети**): **основні засоби**, **малоцінні предмети** та **предмети систем**. Для кожного з них повинна бути розроблена або застосована якась з існуючих систем класифікації. При застосуванні ієрархічної системи класифікації взаємозв’язки кожного з цих трьох класів об’єктів та їх класифікаційними угрупованнями будуть **n:1**. Відповідно втричі збільшується і кількість взаємозв’язків нових



об'єктів з об'єктом **приміщення**.

Виходячи також з потреб користувачів інформації, введемо ще класифікацію самих приміщень за їх функціональним призначенням: аудиторія, лабораторія і т.д. Після проведених доповнень ERD розглядуваної інформаційної системи буде виглядати так, як зображено на рис. 2 (виходячи з того, що вже зрозуміла різниця між об'єктами та зв'язками, на рисунку як перші, так і другі показані прямокутниками).

Як бачимо, питання класифікації та кодування техніко-економічної інформації має велике значення

при побудові баз даних, але обговорення цієї проблеми виходить за межі статті. При конкретній реалізації вирішення цього питання треба покласти на відповідні структурні підрозділи установи (організації): адміністративно-господарчий відділ та бухгалтерію.

Далі треба визначити, виходячи з поточних та майбутніх потреб в інформації, перелік властивостей (атрибутів) об'єктів та встановити ключові та неключові атрибути у визначених об'єктах, а також зовнішні ключі. Одночасно встановлюються ключові та неключові атрибути у зв'язках (відношеннях), де

Склад та зміст таблиць бази даних

BLDNGS – будівлі	
KD_BD	Код будівлі
NM_BD	Назва будівлі
FL_SQ	Загальна площа, кв.м
PR_SQ	Виробнича площа
ET_SQ	Допоміжна площа
QT_FL	Кількість поверхів
PR_BD	Балансова вартість, грн.

KLS_FUN – класифікатор функціональних призначень	
KD_FN	Код функціонального призначення
NM_FN	Назва функціонального

KLS_OSF – класифікатор ОЗ	
KL_OF	Код угруповання основних засобів (ОЗ)
NM_KLOF	Назва угруповання основних засобів

FLOORS – поверхи	
KD_FL	Код поверху будівлі
NM_FL	Назва поверху
TS_F	Загальна площа поверху, кв.м
PS_FL	Виробнича площа поверху
ES_FL	Допоміжна площа поверху

SYSTEMS – системи (мережі)	
KD_SYS	Код системи (мережі)
NM_SYS	Назва системи (мережі)
HR_SYS	Характеристика системи

KLS_MCP – класифікатор МП	
KL_MC	Код угруповання малоцінних предметів (МП)
NM_KLMC	Назва угруповання малоцінних предметів

ROOMS – приміщення	
KD_RM	Код приміщення
KD_FN	Код функціонального призначення
LONG	Довжина
WIDE	Ширина
ALTD	Висота
SQWR	Площа
VOLM	Об'єм

KLS_PSS – класифікатор ПС	
KL_PS	Код угруповання предметів систем (ПС)
NM_KLPS	Назва угруповання предметів систем

NOM_OSF – номенклатор ОЗ		NOM_MCP – номенклатор МП		NOM_PSS – номенклатор ПС	
ID_OF	Системний номер предмету	ID_MC	Системний номер предмету ОЗ	ID_OF	Системний номер предмету ОЗ
KL_OF	Код угруповання основних	KL_MC	Код угруповання малоцінних предметів (МП)	KL_PS	Код угруповання предметів
NM_PROF	Назва предмету ОЗ	NM_PRMC	Назва предмету МП	NM_PRSS	Назва предмету системи
HR_PROF	Характеристика предмету ОЗ	HR_PRMC	Характеристика предмету МП	HR_PRSS	Характеристика предмету
PR_PROF	Вартість предмету ОЗ	PR_PRMC	Вартість предмету МП	PR_PRSS	Вартість предмету системи

RMS_OSF – розташування предметів ОЗ у приміщеннях		RMS_MCP – розташування		RMS_PSS – розташування	
KD_RM	Код приміщення	KD_RM	Код приміщення	KD_RM	Код приміщення
ID_OF	Системний номер предмету ОЗ	ID_MC	Системний номер предмету	ID_PS	Системний номер предмету
KLK_OF	Кількість предметів ОЗ у	KLK_MCP	Кількість предметів МП у	KLK_PSS	Кількість предметів ПС у
DTV_OSF	Остання дата встановлення	DTV_MCP	Остання дата встановлення предмета МП	DTV_PSS	Остання дата встановлення

ключі є складеними по меншій мірі з двох ключових атрибутів об'єктів, що зв'язуються.

Як відомо, проектування баз даних на основі нормалізації відношень є більш формалізованим у порівнянні з семантичним моделюванням [3], одним з підходів якого є побудова ERD. Але при використанні першого підходу припускається, що на початковому етапі якимось чином отримана невелика кількість великих (за кількістю атрибутів) відношень, які після операцій з нормалізації на завершальному її етапі перетворюються у велику кількість невеликих “правильних” відношень. Однак, при опису метода нормалізації ніде не згадується, яким чином отримані початкові “великі” і “неправильні” відношення. Об'єднання ж двох методів проектування структур баз даних дозволяє пройти етап отримання великих відношень, оскільки виконувати при побудові ERD процедури: виділення слабких об'єктів як ієрархічно підлеглих правильним, визначення супертипів та виділення його підтипів і визначення таких додаткових інформаційних об'єктів, як класифікатори, – відповідають певним процедурам нормалізації. При такій послідовності робіт база даних у більшості випадків буде вже знаходитися у третій нормальній формі, що гарантує від аномалій при вводі, коригуванні та видаленні інформації. Для підтвердження цього нижче наводиться повний склад та зміст вказаної бази даних, яка була створена для інформаційної системи адміністративно-господарчого відділу МФ НаУКМА (в наведених нижче таблицях ключові атрибути затемнені).

Література

1. Атре Ш. Структурный подход к организации баз данных. – Москва: Финансы и статистика, 1983.
2. Ахьян Р., Горев А., Макашарипов С. Эффективная работа

с СУБД. – Санкт-Петербург; Москва; Харьков; Минск: Юпитер, 1997.

3. Дейт К.Дж. Введение в системы баз данных. – 6-е изд. – Киев; Москва: Диалектика, 1998.