

УДК 004.658.2

Д. Б. БУЙ, А. В. ПУЗІКОВА

Київський національний університет ім. Тараса Шевченка, Київ, Україна

## ТЕОРІЯ НОРМАЛІЗАЦІЇ В РЕЛЯЦІЙНИХ БАЗАХ ДАНИХ (ОГЛЯД)

У даній роботі здійснена спроба охарактеризувати сучасний стан теорії нормалізації в реляційних базах даних. Розглядається еволюція класичних нормальних форм (а саме: першої-третьої нормальних форм, нормальної форми Бойса-Кодда, четвертої нормальної форми та проєктивно-з'єднувальної або п'ятої нормальної форми) та історичні спроби їх покращення. Наводяться різноманітні варіанти не-класичних нормальних форм, зокрема, обговорюється концепція доменно-ключової нормальної форми.

**Ключові слова:** реляційні бази даних, нормалізація, нормальні форми.

## Вступ

Теорія нормалізації в реляційних базах даних (БД) почала формуватись у 70-х роках минулого століття і на даний час є достатньо розвиненим розділом теорії реляційних БД. Однак, накопичених теоретичних досліджень недостатньо для задоволення потреб розробників БД; про це свідчать роботи, присвячені шляхам вирішення існуючих проблем проектування схем баз даних (наприклад, [1-3]) та вдосконаленню алгоритмічного апарату (наприклад, [4-7]).

## Логіко-історичний огляд розвитку теорії нормалізації

Вперше термін «нормалізація» був застосований у 1970 р. Е. Коддом для назви процедури усунення непростих доменів [8]. Протягом наступних декількох років серед професіоналів розгорнулася полеміка щодо поняття «непростого домену», що впливало на інтерпретацію першої нормальної форми (1НФ) [9-14].

У 1971 р. Е. Кодд у роботі [15] вказує на надлишковість даних та аномалії, які виникають при здійсненні операцій над відношеннями, вперше представляє концепцію функціональної залежності (ФЗ), демонструє можливість її використання для розв'язання проблем проектування БД, наводить означення другої нормальної форми (2НФ), транзитивної ФЗ та третьої нормальної форми (3НФ). Щодо означень 2НФ див. [16, 17, 11].

Означення транзитивної ФЗ, сформульоване Коддом у вигляді: «Нехай  $A, B, C$  різні атрибути (або різні множини атрибутів) у відношенні і нехай вони задовольняють наступним умовам: виконуються ФЗ  $A \rightarrow B$ ,  $B \rightarrow C$ , та не виконується ФЗ  $B \rightarrow A$ . Тоді ФЗ  $A \rightarrow C$  називається

транзитивною залежністю», на даний час має два уточнення:

1) атрибут  $C \subseteq A \cup B$ , де  $A$  і  $B$  – множини атрибутів [17]. Дана умова виключає з розгляду тривіальні ФЗ;

2)  $B \subseteq A$ , де  $A$  і  $B$  – множини атрибутів [16].

Цим також виключаються тривіальні ФЗ в посилці  $A \rightarrow B$ .

В залежності від використовуваних уточнень означення 3НФ сформульоване Коддом у вигляді «Кожний непервинний атрибут не є транзитивно залежним від будь-якого потенційного ключа у відношенні», на даний час має два варіанти.

1) Відношення знаходиться у 3НФ, якщо воно знаходиться у 1НФ і не містить транзитивних залежностей (з вище наведеним уточненням 1) непервинних атрибутів від потенційних ключів [17]. Додатково доводиться теорема, яка стверджує, що з знаходження відношення у 3НФ випливає знаходження цього відношення у 2НФ [17].

2) Відношення знаходиться у 3НФ, якщо воно знаходиться у 2НФ і не містить транзитивних залежностей (з вище наведеними уточненнями 1 і 2) непервинних атрибутів від потенційних ключів [11, 16].

Строге математичне викладення концепції ФЗ та теорії нормальних форм станом на початок 80-х років виконано у монографії В. П. Дрібаса [16].

Відкриття аксіом та правил виведення ФЗ із заданої множини ФЗ [18] та побудова аксіоматики Армстронга<sup>1</sup> для ФЗ [19], дали можливість розробити алгоритми обчислення так званого канонічного покриття для заданої множини ФЗ та

<sup>1</sup> Строге математичне доведення коректності і повноти аксіоматики Армстронга, виконане в традиціях математичної логіки через встановлення збіжності семантичного та синтаксичного слідувань, наводиться у роботі [33], а критерій повноти аксіоматики Армстронга – у роботі [34].

замикання для множини атрибутів [20, 17, 21, 7].

Одним із способів приведення відношення до 3НФ є декомпозиція, обґрунтуванням якої стала теорема Хеза (Heath) [22] про виконувальність декомпозиції без втрат відношення  $R(A, B, C)$  на проєкції  $R_1 = \pi_R(A, B)$  та  $R_2 = \pi_R(A, C)$  за умови виконання ФЗ  $A \rightarrow B$ . До «підводних каменів» декомпозиції належить залежність проєкцій (за термінологією Ріссанена (Rissanen) [23]), яка може стати причиною аномалій. Дана проблема розв'язана у статті А. Филиповича [2], який розглянув взаємні ФЗ (ВФЗ), дослідив їх властивості, запропонував алгоритм виявлення ВФЗ, поняття взаємно-незалежної нормальної форми (ВННФ) та спосіб зведення до неї.

Інший спосіб запропонував Бернштейн (Bernstein), який сформулював альтернативне до запропонованого у [15] означення 3НФ: «Відношення знаходиться у 3НФ, якщо кожний атрибут, який транзитивно залежить від ключа, є первинним атрибутом», та використав його для побудови алгоритму синтезу повної схеми бази даних у 3НФ для заданої множини функціональних залежностей [24].

Специфіка різних підходів до задачі проектування схеми реляційної БД викликана відмінностями у формальних визначеннях еквівалентності та критеріїв якості схеми [25].

Відомими класичними алгоритмами зведення схеми відношення до 3НФ є алгоритми Ульмана [21], Делобеля-Гейсі (Delobel-Gasey) [18], Берштейна [24], Іслура (Isloor) [26], Неклюдової-Цаленка [27], Мейера [17].

Недоліки 3НФ були розглянуті і враховані у роботі Хеза [22] при формулюванні означення посиленої 3НФ та, пізніше, у роботі Кодда [28] (інша назва означення – нормальна форма Бойса-Кодда (НФБК)). Одним з перших відомих алгоритмів зведення «майже» до НФБК є алгоритм Берштейна [24]. Більш пізні алгоритми наводяться у [29, 4].

З введенням у розгляд багатозначних залежностей (БЗЗ) [30] Р. Фагін (Fagin) визначає нову четверту нормальну форму (4НФ) [31]. Зокрема, у зазначеній роботі показано, що якщо схема відношення знаходиться у 4НФ, то вона знаходиться у НФБК. Строгий та повний набір правил виведення для БЗЗ, а також правила, які пов'язують ФЗ та БЗЗ, представлені у [32], незалежність побудованої системи аксіом обговорюється у роботі Мендельсона (Mendelzon) [35], а її повнота – у статті Біскапа (Biskap) [36].

ЗЗ та аномалії, які вони викликають, були розглянуті у статті Ріссанена (Rissanen) [37] та досліджені у роботах [38, 39].

Концепцію п'ятої нормальної форми (5НФ) Р. Фагін розглянув у статті [40]. Зокрема, у даній роботі вперше обговорюється поняття залежності ключа, даються два означення НФБК в термінах множини залежностей ключа та доводиться теорема про їх еквівалентність з класичним означенням НФБК [28]; аналогічні означення формулюються для 4НФ; наводиться алгоритм перевірки того, чи ЗЗ є наслідком з множини залежностей ключів; формулюється означення 5НФ, доводяться теореми про включення НФБК у 4НФ, та 4НФ у 5НФ. На прикладах, наведених Дж. Ніколасом (Nicolas) у [41], обговорюється можливість здійснення подальшої нормалізації з використанням операторів *union* (об'єднання) та *split* (розщеплення).

З результату про відсутність повної скінченної множини правил виведення для ЗЗ, який наводиться у роботі С. Петрова [42], впливає неможливість побудови повної та коректної аксіоматики, звідси – неможливість побудови канонічного покриття.

Грунтовне математичне викладення теорії залежностей та окремі питання теорії нормалізації схеми реляційної БД здійснено у главах 5-6 роботи М. Цаленка [43].

Окрім вказаних 1-5НФ пропонувались різноманітні варіанти покращення класичних нормальних форм, наприклад, покращена 3НФ (An Improved Third Normal Form) [44], а також – введення інших видів нормальних форм, наприклад, нормальної форми з елементарним ключем (Elementary Key Normal Form – EKNF), яка займає проміжне положення між 3НФ та НФБК [45], нормальної форми з повним ключем (Key-Complete Normal Form) [46], доменно-ключової нормальної форми (ДКНФ) [47].

Останню роботу прокоментуємо детальніше. Концепція ДКНФ базується на поняттях залежності ключа та залежності домена: «Відношення знаходиться у ДКНФ тоді і тільки тоді, коли кожне обмеження з його схеми є логічним наслідком з об'єднання множин залежності ключа та залежності домена». Доводиться, що реляційна схема знаходиться у ДКНФ тоді і тільки тоді, коли немає аномалій вставки та знищення (означення яких формулюються у роботі); проводиться аналогія між класичними [28, 31, 40] та альтернативними означеннями НФБК, 4НФ та 5НФ, розглянутими у роботі [40], пропонуються їх модифікації з урахуванням комбінаторного ефекту розмірів обмежувачого домену та показується, що ДКНФ включає в себе ці модифікації; частковим є випадок, коли потужність усіх доменів є нескінченною, тоді ДКНФ включає в себе НФБК, 4НФ та 5НФ у класичному виді. Але під

обмеженнями в означенні ДКНФ розуміються не тільки ФЗ, БЗЗ та ЗЗ, означення яких використовуються у попередніх нормальних формах, а й усі, які можна записати у вигляді висловлювання логіки предикатів 1-го порядку [48].

Оскільки задача перевірки імплікації для таких видів обмежень, як ВБЗЗ, є нерозв'язною [48, 43], то неможливо побудувати алгоритм синтезу реляційної схеми у ДКНФ. До такого ж висновку можна прийти, врахувавши включення 5НФ у ДКНФ за умови нескінченності доменів та нерозв'язності задачі синтезу оптимальної схеми у 5НФ.

Окремим видом є шоста нормальна форма (6НФ), запропонована для хронологічних БД у 2002 р. в [49]. З означення 6НФ випливає, що 6НФ є посиленням 5НФ.

### Сучасний стан і напрямки подальшої роботи

На сьогодні розвиток класичної теорії нормалізації реляційних БД відбувається у декількох напрямках, одним з яких є поширення її принципів на нечіткі реляційні бази даних (НРБД) [53, 54]. Для НРБД розглядаються концепції нечіткої ФЗ та неповної нечіткої ФЗ, які використовуються для визначення понять нечіткого ключа, транзитивного замикання та нечітких нормальних форм, починаючи з 1НФ (Fuzzy Fit NF) і закінчуючи нечіткою НФБК (FBCNF). Зокрема, у роботі [54] наводяться алгоритми для побудови транзитивного замикання множини атрибутів, визначення нечітких потенційних ключів, зведення до нечіткої 2НФ, зведення до нечіткої 3НФ шляхом декомпозиції без втрат із збереженням множини нечітких ФЗ; в кінці роботи наводиться список літератури з даного напрямку.

На нашу думку, теорія нормалізації в реляційних БД, незважаючи на вагомні результати, носить фрагментарний характер і далека ще до задовільного завершення.

### Література

1. Darwen, H. *A Normal Form for Preventing Redundant Tuples in Relational Databases [Text]* / H. Darwen, C. Date, R. Fagin // *Proceedings of the 15th International Conference on Database Theory – ICDT'2012, March 26–30, 2012, Berlin, Germany.* – P. 114-126.

2. Филипович, А. *Взаимные функциональные зависимости [Текст]* / А. Филипович // *Системный администратор.* – 2002. – № 1. – С. 84-89.

3. Fotache, M. *Why Normalization Failed to Become the Ultimate Guide for Database Designers? [Електронний ресурс]* / М. Fotache. – Режим доступу: <http://ssrn.com/abstract=905060> or <http://dx.doi.org/10.2139/ssrn.905060>. – May 1, 2006.

4. Bahmani, A. *Automatic database normalization and primary key generation [Text]* / A. Bahmani, M. Naghibzadeh, B. Bahmani // *CCECE/CCGEI May 5-7 2008, Niagara Falls, Canada.* – P. 11-16.

5. Зорин, И. *Теоретико-графовое приведение реляционной базы данных к третьей нормальной форме Э. Кодда [Електронний ресурс]* / И. Зорин // *Электронное научное издание «Устойчивое инновационное развитие: проектирование и управление», 2009.* – Т. 5. – С. 50-59. – Режим доступа: [www.gypravlenie.ru](http://www.gypravlenie.ru).

6. Виноградова, М. В. *Конструктор баз данных на основе сущностей и их реквизитов с возможностью нормализации [Електронний ресурс]* / М. В. Виноградова, Э. Г. Изушев // *Электронное научно-техническое издание «Наука и образование», 2011.* – № 10. – Режим доступа: <http://technomag.edu.ru/doc/242645.html>

7. Григорьев, Ю. А. *Алгоритм синтеза частично оптимальной схемы реляционной базы данных [Електронний ресурс]* / Ю. А. Григорьев // *Электронное научно-техническое издание «Наука и образование», 2012.* – № 1. – Режим доступа: <http://technomag.edu.ru/doc/294486.html>.

8. Codd, E. F. *A Relational Model of Data for Large Shared Data Banks [Text]* / E. F. Codd // *Communications of the ACM.* – 1970. – Vol. 13, № 6. – P. 377-387 (русский перевод: Кодд, Е. Ф. *Реляционная модель данных для больших совместно используемых банков данных [Текст]* / Е. Ф. Кодд // *СУБД.* – 1995. – № 1. – С. 145-160).

9. Дейт, К. Дж. *Введение в системы баз данных [Текст]* / К. Дж. Дейт. – М.: Наука, 1980. – 464 с.

10. Elmasri, R. *Fundamentals of Database Systems [Text]* / R. Elmasri, S. R. Navathe. – Massachusetts: Addison-Wesley, 2000. – 893 p.

11. Дейт, К. Дж. *Введение в системы баз данных [Текст]* / К. Дж. Дейт. – 8-е издание. – М.: Вильямс, 2005. – 1328 с.

12. Date, C. J. *Database in Depth: Relational Theory for Practitioners [Text]* / C. J. Date. – California: O'Reilly, 2005. – 208 p.

13. Jaeschke, G. *Remarks on the Algebra of Non First Normal Form Relations [Text]* / G. Jaeschke, H. J. Schek // *Proceedings of the 1st ACM SIGACT-SIGMOD symposium on Principles of database systems, Los Angeles, California., 1982.* – P. 124-138.

14. Makinouchi, A. *A Consideration of Normal Form on Not-necessarily Normalized Relations in the Relational Data Model [Text]* / A. Makinouchi // *Proceedings of the Third International Conference on Very Large Data Bases,*

October 6-8, 1977, Tokyo, Japan. IEEE Computer Society. – 1977. – P. 447-453.

15. Codd, E. F. *Further Normalization of the Data Base Relational Model [Text]* / E. F. Codd // (Presented at Courant Computer Science Symposia Series 6, "Data Base Systems", New York City, May 24-25, 1971.) IBM Research Report RJ909 (August 31, 1971). Republished in Randall J. Rustin (ed.), *Data Base Systems: Courant Computer Science Symposia Series 6*. Prentice-Hall, 1972.

16. Дрибас, В. П. Реляционные модели баз данных [Текст] / В. П. Дрибас. – Минск : Изд-во БГУ им. В. И. Ленина, 1982. – 192 с.

17. Мейер, Д. Теория реляционных баз данных [Текст] / Д. Мейер. – Москва : Мир, 1987. – 608 с.

18. Delobel, C. *Decomposition of a database and the theory of Boolean switching function [Text]* / C. Delobel, R. Gasey // *IBM Journal of Research and Development*. – 1973. – Vol. 17, № 5. – P. 374-386.

19. Armstrong, W. W. *Dependency structures of data base relationships [Text]* / W. W. Armstrong // *Proc. IFIP '74*, North Holland Pub. Co., Amsterdam, 1974. – P. 580-583.

20. Beeri, C. *Computational problems related to the design of normal form relation schemas [Text]* / C. Beeri, P. A. Bernstein // *ACM Transactions on Database Systems*. – 1979. – Vol. 4, № 1. – P. 30-59.

21. Ульман, Дж. Основы систем баз данных [Текст] / Дж. Ульман. – М. : Финансы и статистика, 1983. – 334 с.

22. Heath, I. J. *Unacceptable File Operations in Relational Database [Text]* / I. J. Heath // *ACM SIGFIDET Workshop on Data Description, Access, and Control*. – San Diego, Calif. – 1971. – P. 19-33.

23. Rissanen, J. *Independent components of relations [Text]* / J. Rissanen // *ACM Transactions on Database Systems*. – 1977. – Vol. 2, № 4. – P. 317-325.

24. Bernstein, P. A. *Synthesizing Third Normal Form relations from functional dependencies [Text]* / P. A. Bernstein // *ACM Transactions on Database Systems*. – 1976. – Vol. 1, № 4. – P. 277-298.

25. Beeri, C. *A sophisticate's introduction to database normalization theory [Text]* / C. Beeri, P. Bernstein, N. Goodman // *Proceedings of 4th International Conference on Very Large Data Bases*, West Berlin, 1978. – P. 113-124.

26. Isloor, S. S. *An algorithm with logical simplicity for designing third normal form relations data base schema for functional dependencies [Text]* / S. S. Isloor // *Proceedings of International Conference on DBMS (ICMOD 78)*, Fast Milano, Italy, 1978. – P. 31-50.

27. Неклюдова, Е. А. Синтез логической схемы реляционной базы данных [Текст] / Е. А. Неклюдова, М. Ш. Цаленко // *Программирование*. – 1979. – № 6. – С. 58-68.

28. Codd, E. F. *Recent Investigations into Relational Data Base Systems [Text]* / E. F. Codd // *Proceedings of IFIP Congress 74*, Stockholm, Sweden, August 5-10, 1974. North-Holland, 1974. – P. 1017-1021.

29. Lin, W.-Y. *Efficient algorithm for BCNF-decomposition [Text]* / W.-Y. Lin // *Information and*

*Software Technology*. – 1992. – Vol. 34, № 5. – P. 308-312.

30. Zaniolo, C. *Analysis and design of relational schemata for database systems [Text]* : Ph.D. dissertation, Tech. Rep. UCLA-Eng-7769, Dep. Computer Science, Univ. California at Los Angeles, July 1976.

31. Fagin, R. *Multivalued Dependencies and a New Normal Form for Relational Databases [Text]* / R. Fagin // *ACM Transactions on Database Systems*. – 1977. – Vol. 2, № 1. – P. 262-278.

32. Beeri, C. *A complete axiomatization for functional and multivalued dependencies [Text]* / C. Beeri, R. Fagin, J. Howard // *Proc. ACM-SIGMOD Conf. (Toronto, Canada, Aug. 3-5) ACM*, New York, 1977. – P. 47-61.

33. Буй, Д. Б. Повнота аксіоматики Армстронга [Текст] / Д. Б. Буй, А. В. Пузікова // *Вісник Київського національного університету імені Тараса Шевченка, Серія: Фізико-математичні науки*. – 2011. – Вип. № 3. – С. 103-108.

34. Буй, Д. Б. Критерій повноти аксіоматики Армстронга [Текст] / Д. Б. Буй, А. В. Пузікова // *Матеріали міжнародної конференції «Теоретичні та прикладні аспекти побудови програмних систем» – ТААПСД'2011 (Україна, Ялта, 19-23 вересня 2011 року)*. – С. 30-34.

35. Mendelzon, A. O. *On axiomatizing multivalued dependencies in relational databases [Text]* / A. O. Mendelzon // *ACM Transactions on Database Systems*. – 1979. – Vol. 26, № 1. – P. 37-44.

36. Biskup, J. *Inferences of multivalued dependencies in fixed and undetermined universes [Text]* / J. Biskup // *Theoretical Computer Science*. – 1980. – Vol. 10, № 1. – P. 93-105.

37. Rissanen, J. *Independent components of relations [Text]* / J. Rissanen // *ACM Transactions on Database Systems*. – 1977. – Vol. 2, № 4. – P. 317-325.

38. Aho, A. V. *The theory of joins in relational databases [Text]* / A. V. Aho, C. Beeri, J. D. Ullman // *Proc. 18th Symp. on Foundations of Computer Science*, Providence, R.I., 1977. – P. 107-113.

39. Dayal, U. *The fragmentation problem: lossless decomposition of relations into files [Text]* / U. Dayal, P. A. Bernstein // *Proceedings of the ACM SIGMOD international conference on Management of data*, 1979. – P. 143-151.

40. Fagin, R. *Normal Forms and Relational Database Operators [Text]* / R. Fagin // *Proceedings of the ACM SIGMOD International Conference on Management of Data (Boston, Mass., May 30-June 1)*, ACM, New York, 1979. – P. 153-160.

41. Nicolas, J. M. *Mutual dependencies and same results on indecomposable relations [Text]* / J. M. Nicolas // *Proceedings of the fourth international conference on Very Large Data Bases*, 1978. – Vol. 4. – P. 360-367.

42. Петров, С. В. Об аксиоматизации зависимостей по соединению [Текст] / С. В. Петров // *Применение методов математической логики: Тезисы докл. IV Всес. конф. «Предоставление знаний и синтез программ»*. – Таллин: АН ЭССР, 1986. – С. 151-152.

43. Цаленко, М. Ш. Моделирование семантики в базах данных [Текст] / М. Ш. Цаленко. – М. : Наука, 1989. – 287 с.
44. Ling, T. W. An Improved Third Normal Form for Relational Databases [Text] / T. W. Ling, F. W. Tompa, T. Kameda // *ACM Transactions on Database Systems*. – 1981. – Vol. 6, № 2. – P. 329-346.
45. Zaniolo, C. A New Normal Form for the Design of Relational Database Schemata [Text] / C. Zaniolo // *ACM Transactions on Database Systems*. – 1982. – Vol. 7, № 3. – P. 489-499.
46. Vincent, M. W. Redundancy Elimination and a New Normal Form for Relational Database Design [Text] / M. W. Vincent // *In Semantics in Databases* (Libkin, L., Thalheim, B., eds.). – Vol. 1358 of LNCS. – 1998. – P. 247-264.
47. Fagin, R. A Normal Form for Relational Databases That Is Based on Domains and Keys [Text] / R. Fagin // *Communications of the ACM*. – 1981. – Vol. 6. – P. 387-415.
48. Fagin, R. The theory of data dependencies – a survey [Text] / R. Fagin, M. Y. Vardi // *In Mathematics of Information Processing, Proc. Symposia in Applied Mathematics*. – Vol. 34, American Mathematical Society. – 1986. – P. 19-71.
49. Date, C. J. *Temporal Data and the Relational Model* [Text] / C. J. Date, H. Darwen, N. Lorentzos. – Morgan Kaufmann, 2002. – 422 p.
50. Date, C. J. Simple Conditions for Guaranteeing Higher Normal Forms in Relational Databases [Text] / C. J. Date, R. Fagin // *ACM Transactions on Database Systems*. – 1992. – Vol. 17, № 3. – P. 465-476.
51. Ling, T. W. Logical Database Design with Inclusion Dependencies [Text] / T. W. Ling, C. H. Goh // *In Proceedings of the Eighth International Conference on Data Engineering, Tempe, Arizona, 1992*. – P. 642-649.
52. Levene, M. Justification for Inclusion Dependency Normal Form [Text] / M. Levene, M. W. Vincent // *IEEE Transactions on Knowledge and Data Engineering*. – 2000. – Vol. 12, № 2. – P. 281-291.
53. Chen, G. Normalization based on ffd in a fuzzy relational data model [Text] / G. Chen, E. E. Kerre, J. Vandenbulcke // *Inform Syst.* – 1996. – Vol. 21. – P. 299-310.
54. Bahar, Ö. Normalization and Lossless Join Decomposition of Similarity-Based Fuzzy Relational Databases [Text] / Ö. Bahar, A. Yazıcı // *International Journal of Intelligent Systems*. – 2004. – Vol. 19. – P. 885-917. Published online in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)).

Поступила в редакцію 24.02.2014, рассмотрена на редколлегии 24.03.2014

**Рецензент:** д-р техн. наук, проф. С. В. Листровой, Харьковская государственная академия железнодорожного транспорта, Харьков, Украина.

## ТЕОРИЯ НОРМАЛИЗАЦИИ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ (ОБЗОР)

*Д. Б. Буй, А. В. Пузикова*

В данной работе сделана попытка охарактеризовать современное состояние теории нормализации в реляционных базах данных. Рассматривается эволюция классических нормальных форм (а именно: первой-третьей нормальных форм, нормальной формы Бойса-Кодда и проективно-соединительной или пятой нормальной формы) и исторические попытки их улучшения. Приведены различные варианты неклассических нормальных форм, в частности, обговаривается концепция доменно-ключевой нормальной формы.

**Ключевые слова:** реляционные базы данных, нормализация, нормальные формы.

## THEORY OF NORMALIZATION IN RELATIONAL DATABASES (A SURVEY)

*D. B. Bui, A. V. Puzikova*

Given paper contains some attempt to characterize the current state of the theory of normalization in relational databases. The evolution of classical normal forms (first-third normal forms (1-3NF), Boyce-Codd normal form (BCNF), fourth normal form (4NF) and project-join normal form (PJ/NF or 5NF)) and historical attempts to improve them are considered. The various options no classical normal forms are reviewed, in particular, the concept of domain-key normal form (DKNF) are discussed.

**Keywords:** relational databases, normalization, normal forms.

**Буй Дмитро Борисович** – д-р фіз.-мат. наук, професор, професор кафедри теорії і технології програмування факультету кібернетики Київського національного університету ім. Тараса Шевченка, Київ, Україна, e-mail: [bui@unicyb.kiev.ua](mailto:bui@unicyb.kiev.ua).

**Пузикова Анна Валентинівна** – аспірантка кафедри теорії і технології програмування факультету кібернетики Київського національного університету ім. Тараса Шевченка, Київ, Україна, e-mail: [anna\\_inf@mail.ru](mailto:anna_inf@mail.ru).