

МОДЕЛЮВАННЯ БАГАТОРІВНЕВОЇ РЕКУРЕНТНОЇ БАЗИ ДАНИХ У БАЗИСІ ГАЛУА

© Шаряк В.В., 2009

Викладено принципи побудови рекурентної бази даних у базисі Галуа та проаналізовано ефективність кодування ідентифікаційних даних на прикладі ієрархічно-реляційної бази даних та системного бізнесу просторів даних Global Space of Innovation and Development Inmarket та IntWay World Corporation USA. Результати аналізу організаційної структури бази даних, якими користуються великі міжнародні корпорації, показують, що найпоширенішою для кодування ідентифікаційних даних є багаторівнева ієрархічно-реляційна структура. При цьому очевидно є велика надлишковість кодування та їх низька завадозахищеність, що аналізувалося в дослідних роботах, де на основі ієрархічних структур обґрунтовані принципи та показана ефективність кодування ідентифікаційних даних на основі багатомірної рекурентної організації бази даних у базисі Галуа.

In the article the expounded principles of construction of recurrent database in the hierarchically relyaciynoy base of Galois and conducted analysis of efficiency of code of identification information are on an example hierarchically- relyaciynoy databases system business specious database Global Space of Innovation and Development Inmarket end IntWay World Corporation USA . An Analyses of organizational structure is databases which are used by large international coporation shows naubils shiponobraubem for the code of identification information are multilevel hierarchicalle-relation structured. Thus obviously there is large surplus of code but them low hindrance protection that analised in experimental works, where on the of hierarchical structures the grounded principles and efficiency of code of identification information is rotined on the basis of multidimensional recurrent organization of database in the base of Galua.

Вступ. Простори даних – нова абстракція керування даними. Бази даних та сховища даних дають змогу опрацювати деталізовані та інтегровані дані, побудовані на основі гарантовано доступних моделей даних. У випадку роботи із Всесвітньою мережею з величезною кількістю ресурсів (прикладом таких задач є Системний бізнес – державний, приватний, корпоративний, мережевий та інший бізнес, який об’єднується в світову базу даних та знань сфери товарів та послуг. Туристичний бізнес: збирання інформації про місця відпочинку, її інтеграція та зберігання у внутрішніх базах даних та геоінформаційні системи).

Сучасні інформаційні системи потребують відповідного потужного розвитку теоретичних засад реалізації баз даних та баз знань. Особливо великі масиви даних формуються, зберігаються та оперативно циркулюють у каналах зв’язку міжнародних корпорацій, до яких належить: GSIDI, або скорочено Inmarket. Це результат спільної діяльності групи компаній України, США та Росії на ринку України з 21 лютого 2005 року, штаб-квартири яких знаходяться в Києві та Панамі. Сьогодні інмаркет працює в режимі підготовки, тестування та розроблення системного бізнесу. Відділ юридичного супроводу проекту проводить реєстрацію на всіх рівнях, необхідних для міжнародної діяльності. Ведеться робота із створення партнерських груп і відкриття офіційних представництв компанії в різних країнах. Відпрацьовану і відтестовану методику навчання використано в „Інтерактивних навчальних комп’ютерних курсах Інмаркет”. Клієнтами компанії стають не тільки

фізичні особи: учні, студенти, викладачі, науковці, але й комерційні і громадські організації, школи, комп'ютерні курси, ПТУ, коледжі, інститути, університети та інші навчальні заклади (табл. 1). У 2009 році GSIDI планує запуск свого власного супутника Землі, до функцій якого входитимуть: телекомунікаційні системи – структури національної системи зв'язку, удосконалення електрозв'язку, перехід від SMS до E-mail у стільниковому зв'язку до 2012 року, відео-IP- телебачення, термінали AUTOPAY та поповнення електронних гаманців платіжних Інтернет-сиситем, інтернет-бізнес для всіх. Оскільки сфери впливу і світові ринки-бізнесу вже поділені, то бізнес в Інтернеті має велике майбутнє, оскільки є неосвоєним і доступний для кожного бажаючого [1].

IntWay World Corporation USA-утворилася 29 вересня 2005 року внаслідок злиття трьох компаній: ринок Інтернет-технологій; ринок інвестицій та цінних паперів; ринок нерухомості; ринок товарів; міжнародний туризм, які працюють на ринку України [1].

Загальні принципи обробки лінійних списків. Лінійні структури поділяються на два типи: масиви і зв'язані лінійні списки.

Масиви – статичні структури однотипних елементів, доступ до яких здійснюється за номером елемента. Переваги: швидкий доступ до будь-якого елемента. Недолік: неможливість зміни розміру масивів і виконання програми.

Зв'язані лінійні списки за перевагами і недоліками протилежні до масивів. Список являє собою послідовність вузлів, кожен з яких – це структура, що містить, крім самого значення вузла, вказівник – наступний вузол. Зі списком працюють, маючи вказівник на його перший вузол. Для доступу до потрібного значення по чергово перебирають усі вузли, поки не зустрінеться шукане значення в однонаправлених списках. Рухаються з початку списку до його кінця.

У двонаправленому списку переміщення між вузлами здійснюється у довільному напрямку, кожен вузол, крім вказівника на наступний, містить і вказівник на попередній вузол. Останній вузол списку містить нульовий вказівник. Всі операції зведені до введення змін за сортуванням, вставкою і видаленням.

Загальні принципи обробки нелінійних списків. Нелінійний список – це дерево – структура даних, що являє собою множину вузлів з одним виділеним вузлом, який називається корінь, а інші вузли мають розділену скінченну кількість множин, які не перетинаються і кожна з яких теж є деревом (рис.4) Множина – це набір елементів, які не мають спільних елементів, а дерево має.

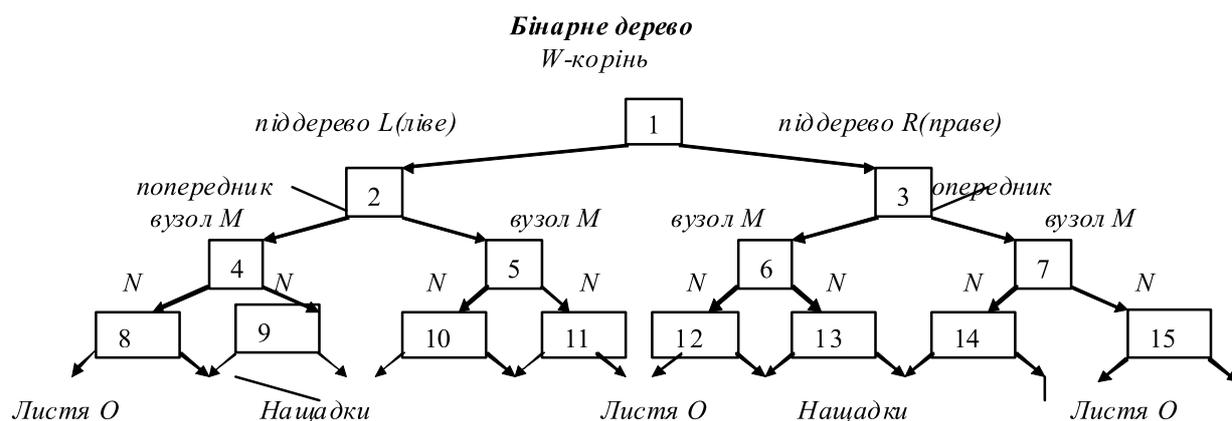


Рис. 1. Ієрархічна структура бінарного дерева

Властивості дерев:

1. Існує один вузол, який не має попередників – це корінь *W* дерева.
2. Решта вузлів мають тільки одного попередника.

3. Для будь-якого вузла K існує послідовність вузлів від K до W -кореня, яка називається віткою дерева.

4. Степінь вузла – це кількість напрямів вузла, з найбільшим степенем визначає порядок дерева.

5. Дерево степені M називається повним, якщо будь-який вузол має степінь N або O .

6. Вузли дерева степенем O називаються листями, а решта – внутрішніми вузлами.

Пряме проходження дерев: спочатку відвідують корінь, а потім, у прямій послідовності, його піддерева – 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15.

Обернене проходження дерева: в оберненому порядку проходять піддерева, а потім проходять його корінь–2,4,8,9,5,10,11,6,12,13,14,15,7,3,1 [2–5].

Алгоритми обробки бінарних дерев. Бінарні дерева – особлива категорія збалансованих або незбалансованих деревоподібних структур, в яких допускається не більше двох гілок (ліва гілка – L, права гілка –R). Це дерева другого порядку, використовуються для запису арифметичних виразів. У корені дерева розміщується знак операції, а операнти – в двох гілках, в результаті всі числа і зміни подаються в листи, а внутрішні вузли міститимуть знаки операції. Операція, яка виконується останньою, буде у корені дерева (рис.2). Бінарні дерева поділяються на два типи: збалансовані і незбалансовані дерева.

У збалансованих деревах кожен вузол має однакове число гілок, причому процес включення нових гілок у вузли дерева йде згори донизу, а кожен рівень дерева – зліва направо.

У незбалансованих деревах кожен вузол не має однакового числа гілок, вони з перемінним числом гілок, причому процес включення нових гілок у вузли дерева йде згори донизу, а на кожному рівні дерева – зліва направо.

Алгоритм вставки вузла на дереві ґрунтується на його означенні і є рекурсивним.

У вигляді структури бінарного дерева може бути представлений, наприклад, мережевий маркетинг-план корпорації, „GSIDI та Intway World Corporation” [2–5].

Аналіз характеристики кодування ідентифікаційних даних в ієрархічно-реляційній БДструктурі багаторівневого маркетингу корпорації, „GSIDI та Intway World Corporation”.

Функції корпорації реалізуються ієрархічними рівнями на основі бінарного дерева, що забезпечує структуру дистриб’юторської організації. Приклад ієрархічної структури системного бізнесу корпорації показано у табл.1.

Таблиця 1

Інтерактивні мультимедійні курси INMARKET

Пакет повного доступу	Інтерактивні Мультимедійні курси INMARKET	Кількість навчальних уроків	Ціна пакета повного доступу до ресурсів в євро	Кількість Партнерів	
				бали	осіб
D – 2847	Директор	12*450	2847	1200	12
B – 219	Партнер, „Бізнес”	450	219	100	1
S – 89	Партнер, „Старт”	220	89	40	1
M – 47	Партнер, „Міні”	80	47	20	1
C – 25	Партнер, „Клієнт”	40	25	10	1
T – 0	Партнер, „Тест”	0	0	0	1
Всього:	—	—	—	—	—

Згідно з табл. 2 на рис. 2 відображено кодування бінарного дерева ієрархічної структури дистриб’юторської організаційної, а на рис. 3 – схема кодування діяльності інфраструктури корпорації, „GSIDI та Intway World Corporation”.

Система ідентифікації партнерської організації системного бізнесу (ПОСБ), яка належить до класу дерев зі зв'язаними списками [2, 3], де у вигляді табл. 2 подана структура ієрархічно-реляційної організації ІД корпорації, з такими даними: НП – номер підприємства, ІНП – ідентифікаційний номер партнера, ПБП – партнера, ІНС – ідентифікаційний номер спонсора, ПБС – спонсора, СТ – ступінь, ВВС – відстань до верхнього спонсора, ОЗ – особисті закупки, ТОМЗ – товарообіг місячних закупок, СО – структурний об'єм, ОНО особистий нагромаджувальний об'єм, ОНОСБ – основний нагромаджувальний об'єм системного бізнесу.

Таблиця 2

**Середньостатистичні параметри даних
реляційно-ієрархічної структури дванадцятирівневої системи СБ**

<i>І І Н П</i>	ІНД	ПБД	ІНС	ПБС	СТ	ВВС	ОЗ	ТОМЗ	СО	ОНО	ОНОСБ
6·X	8·X	24·X	8·X	24·X	1·X	2·X	3·X	5·X	5·X	9·X	12·X
...
...
6·X	8·X	24·X	8·X	24·X	1·X	2·X	3·X	5·X	5·X	9·X	12·X

ІДП

ОДП

де X – байт; ІДП – ідентифікаційні дані партнера, ОДП – організаційні дані партнера

У табл. 2 символ n·X означає число алфавітно-цифрових даних у кожній позиції двомірного файлу (n = 1,2,...). Отже, при заданій кількості користувачів за табл.2 можна розрахувати об'єм алфавітно-цифрових даних одного кортежа БД корпорації, GSIDI та Intway World Corporation”. Для зберігання даних кортежа реляційної БД, враховуючи умову байт-орієнтованого кодування, загальний об'єм даних можна обчислити за формулою:

$$I_j = \sum_{j=1}^m \sum_{i=1}^k d_{ij}, \quad (1)$$

де d_{ij} – об'єм даних ij – го атрибута кортежа, ($i=1,2,\dots, k$) – число атрибутів, ($j = 1,2,\dots, m$) – число кортежів.

Аналіз табл. 2 показує, що об'єми оперативних даних (ОД) діяльності дистриб'юторів визначаються доменами : ОЗ, ТОМЗ, СО, ОНО та ОНОСБ, а ідентифікаційні дані описуються доменами : НП, ПБД, ІНС, ПБС, ІНС, СТ та ВВС. Отже, ефективність кодування БД ієрархічно-реляційної архітектури можна розрахувати за формулами відносного та приведенного оцінювання:

$$K_{e_1} = \frac{I_{ID}}{I_{OD}} ; K_{e_2} = \frac{I_{ID} - I_{OD}}{I_{OD} + I_{ID}} \cdot 100\%. \quad (2)$$

Ефективність кодування БД ієрархічно-реляційної архітектури, яка відповідає корпорації Intway World Corporation та GSIDI, ” можна розрахувати за формулою (2):

$$I_{OD} = 34 \cdot X; I_{ID} = 73 \cdot X. \quad (3)$$

Підставляючи дані у вирази формули (2), отримуємо значення коефіцієнта ефективності у вигляді

$$K_{e_1} = 2,15; K_{e_2} = 37 \%,$$

що відповідає більш ніж двократній надлишковості кодування (ІД) відносно (ОД) або 37 % відносній надлишковості кодування (ІД).

**Кодування бінарного дерева дистриб'юторської організації структури корпорації,
GSIDI та Intway World Corporation ”**

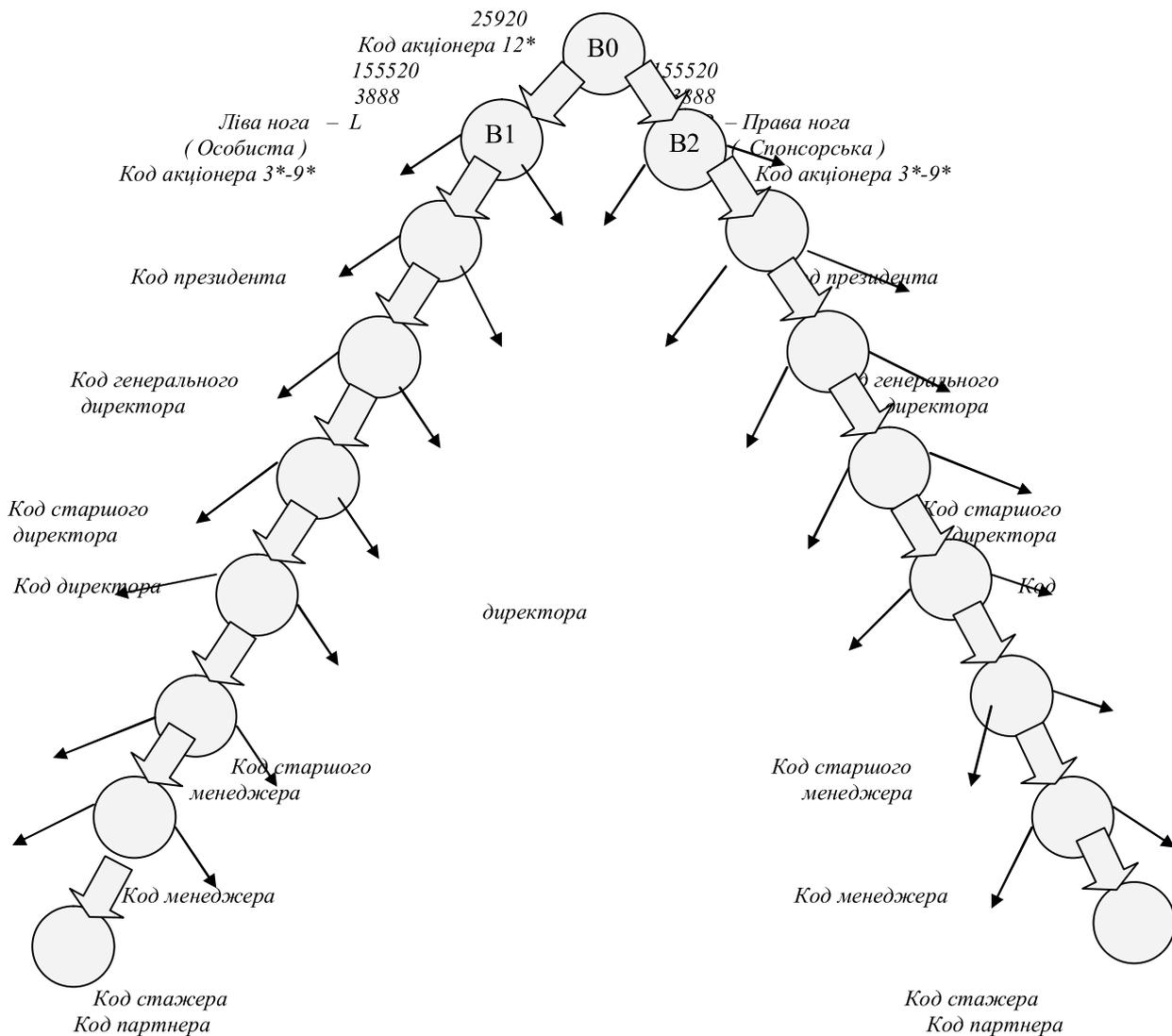


Рис. 2. Ациклічний граф. Велике бінарне дерево T-кодування організаційної структури корпорації

Принципи організації багаторівневої рекурентної БД у базисі Галуа

Теоретичні положення кодування ID у багаторівневій рекурентній організації БД викладені у [5–7]. Рекурентно-реляційна організація БД еквівалентна ієрархічно-реляційній організації, що використовується в корпорації “GSIDI та Intway World Corporation”.

Числовий аналіз кодування ідентифікаторів НП –11200·X показує, що число символів, які потрібно зберігати в БД корпорації „Intway World Corporation та GSIDI” про код країни та ідентифікаційний номер дистриб’ютора, дорівнює – 11200 · X. При кодуванні аналогічних даних на основі рекурентної БД в базисі Галуа маємо 208·X ·300.

Отже, коефіцієнт надлишковості кодування ID в БД корпорації, Intway World Corporation та GSIDI ”згідно з виразом (2) становить :

$$K_{e_1} = \frac{11200}{208} = 54.$$

Відповідно коефіцієнт надлишковості кодування ідентифікаційних кодів дистриб’юторів згідно з рис. 2 у першому випадку становить: 24·X , а в другому випадку : 11·X. Результати розрахунків кодування ідентифікаційних даних за рівнями дистриб’юторської організації на основі

рекурентної БД у базисі Галуа відображено в табл. 3, де G_i – послідовність бітів коду Галуа згідно з певним ключем, а \bar{G}_i – інвертовані біти Галуа, які вказують початок ієрархічного підрівня БД.

Схема кодування діяльності інфраструктури корпорації GSIDI та Intway World Corporation

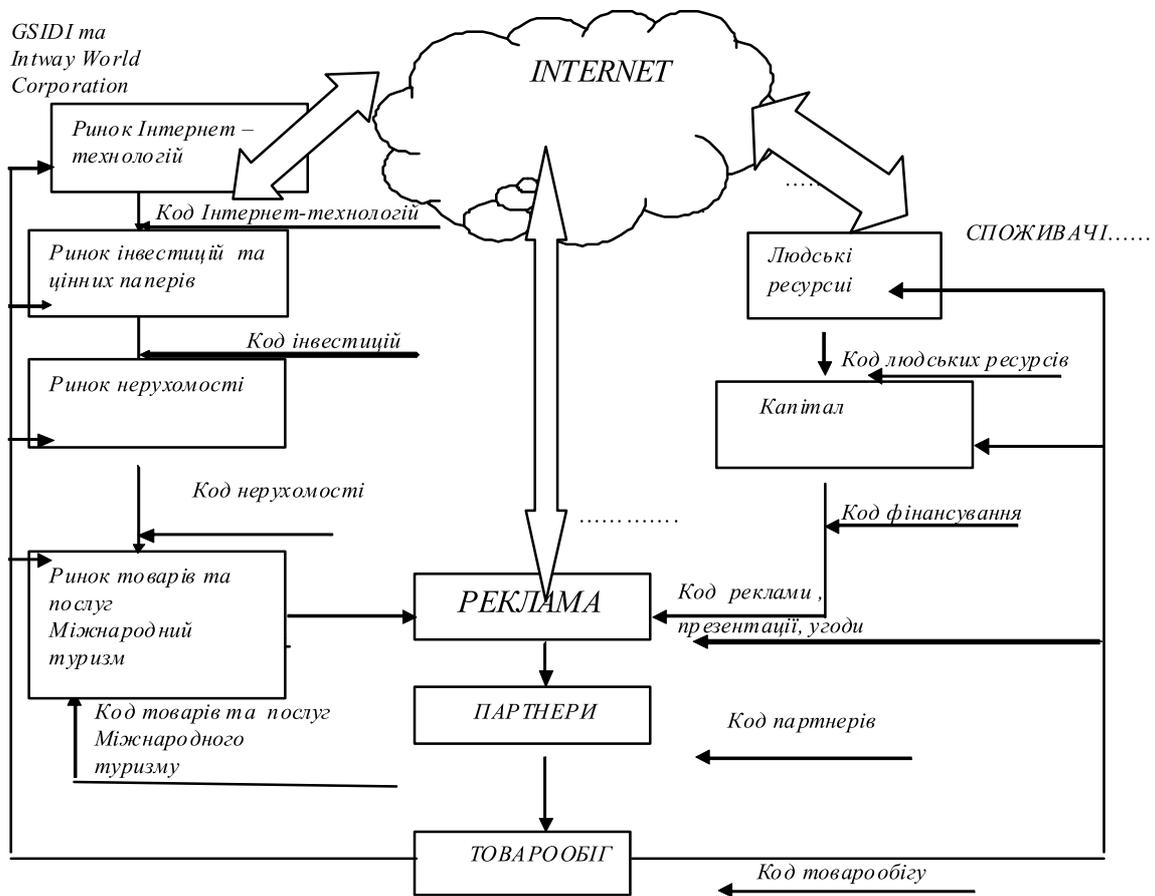


Рис. 3. Циклічний граф мереж міжнародної корпорації GSIDI та Intway World Corporation

Таблиця 3

Кодування ідентифікаційних даних на основі рекурентних кодів поля Галуа

XXX 0	1	2	3	4	...	12	Код Галуа
... 1	8·X						GGGGGG...G \bar{G} GGGGG
... 2	8·X	8·X					GGGGG...GG \bar{G} GGGGG
... 3		8·X	8·X				GGGG...GGG \bar{G} GGGGG
... 4			8·X	8·X			GGG...GGGG \bar{G} GGGGG
... ..				8·X	8·X		GG...GGGGG \bar{G} GGGGG
... 12					8·X	8·X	G...GGGGGG \bar{G} GGGGG
...
XXX							...

Практично у корпорації “GSIDI та Intway World Corporation” використовується універсальна система кодування ІД у вигляді двох 8-розрядних десяткових кодів, які ідентифікують код наставника і дистриб’ютора в будь-якій структурі і підрозділі в будь-якій країні світу, на ринку яких працює корпорація “GSIDI та Intway World Corporation” . Це становить 64 бітів двійково-десятькового коду, або символно-байтового стандартного коду. При цьому об’єм бази ідентифікаційних даних менеджерів нижчого рівня корпорації “GSIDI та Intway World Corporation” становить $11200 \times 128 = 1433600$, а загальна кількість можливих ідентифікованих членів становить 2^{64} осіб дистриб’юторів і споживачів населення планети Земля.

Висновки. Проаналізовано оцінку ефективності кодування ІДП та ОДП на прикладі корпорацій “Intway World Corporation та GSIDI”. Показано, що застосування ієрархічно-рекурентної організації БД в базисі Галуа є достатньо ефективними та може забезпечити зменшення надлишковості кодування даних, які формуються, зберігаються і передаються міжнародними каналами зв’язку на один–два порядки. При цьому, враховуючи відносно низьку вартість засобів зберігання даних на фізичних носіях (у діапазоні $10^6 - 10^{20}$ Гбайт), основного ефекту зниження собівартості руху даних у транснаціональних міжнародних корпораціях можуть досягти в каналах зв’язку локальні та глобальні комп’ютерні мережі та Інтернет, що при інтенсивному трафіку обміну даними 0 – 1 Гбайт/с може становити сотні тисяч у.о. на рік.

1. *Web: Global Space of Innovation and Development Inmarket (GSIDI) <http://www.inmarket.biz> ; IntWay World Corporation USA(IntWay) <https://www.intway.biz>.* 2. *Мартин Дж. Организация баз данных в вычислительных системах / Дж. Мартин-пер.с англ. М.: Мир.1980. – 662 с.* 3. *Системы баз данных. Гарсиа-Молина/[Гарсиа-Молина и др.]. –М.: Вильямс,2004. – 1088 с.* 4. *Структуры данных и алгоритмы/ [Ахо и др.]–пер.с англ.– М: Вильямс, 2001.– 384 с.* 5. *Шаряк В. Теоретико-числові базиси та їх застосування при організації бази даних / В.В. Шаряк* 6. *Николайчук Я.М. Оптико-електронні інформаційно-енергетичні технології. – Вінниця: ВНТУ, 2006. – № 2(12). – С. 59–66.* 7. *Шаряк В. Архітектура і кодування баз даних на основі теоретико-числових базисів / В.В. Шаряк // Вісник Тернопільського ДТУ ім. І. Пулюя. – 2007. – Том.12. – №1. – С.171–179.* 8. *Шаряк В. Методи дослідження системних характеристик моделей бази даних / В.В. Шаряк // Вісник Тернопільського ДТУ ім. І. Пулюя. – 2008.– Т. 13. – № 2. – С.116–121.*