

10. Леффингуэлл Д., Уидриг Д. Принципы работы с требованиями к программному обеспечению. Унифицированный подход: Пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 448 с. 11. Козаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 800 с. 12. Критерии оценки и выбора, <http://www.citforum.ru/database>. 13. Ralph Kimball. The importance of the time dimension in data marts and data warehouses // DBMS – July 1997 <http://www.dbmsmag.com/9707d05.html>. 14. Андрей Прохоров, Informix Временной ряд как объект хранения в СУБД // CIT Forum [http://www.citforum.ru/seminars/cbd2001/day2\\_5\\_informix.shtml](http://www.citforum.ru/seminars/cbd2001/day2_5_informix.shtml).

УДК 683.1

А.Я. Вовчина

Національний університет “Львівська політехніка”,  
кафедра “Інформаційні системи та мережі”

## ВИДІЛЕННЯ ЧАСОВИХ ІНТЕРВАЛІВ ПРИ ПОБУДОВІ КАЛЕНДАРНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

© Вовчина А.Я., 2006

*In article approaches to division of a timebase into intervals structure of representation of time space in relational databases with formation optimum behind some attributes are considered. The primary goals of construction of time structures are allocated. The suggested campaign for definition of time dependences by division of time lines of values behind some criteria.*

*Розглянуто підходи до поділу часової осі на інтервали з формуванням оптимальної за деякими ознаками структури подання часового універсуму в реляційних базах даних. Виділено основні задачі побудови часових структур. Запропоновано підхід для визначення часових залежностей шляхом розбиття часового ряду значень за певними критеріями.*

### ВСТУП. ПОДАННЯ ЧАСУ В ІС

У більшості сучасних програмних засобів, зокрема СКБД, часові характеристики предметної області опрацьовуються з використанням загальноприйнятого структурування часу. Так, для подання деякої часової мітки на часовій осі, як правило, використовується структура типу <число:місяць:рік>, яка може бути доповнена структурою <година:хвилина:секунда>. Але таке загальноприйняте відображення деякої часової мітки не завжди є прийнятним при відображенні інформації про деяку предметну область в інформаційній системі її бази даних. Наприклад, в системі обліку робочого часу присутні такі часові виміри, як зміна, робочі /не робочі дні; в музейній справі – ера, епоха. Відображення та опрацювання таких часових інтервалів в реляційних базах даних породжує деякі проблеми, пов'язані з природою реляційної моделі даних. Також слід зазначити, що інформація, накопичена в

сховищах даних, є “сировиною” для отримання (видобування) певних знань (часових залежностей) про предметну область. Зокрема, структурування часу може виявити часові залежності деяких інформаційних показників предметної області, які набагато легше подати з використанням структури часових міток, відмінних від згальноприйнятих. Підходи до розв’язання таких задач подано нижче. Спочатку розглянемо суть задачі структурування часу.

### АНАЛІЗ ДОСЛІДЖЕНЬ. СТРУКТУРНЕ ПОДАННЯ ЧАСУ

Структурний спосіб подання часу ґрунтується на понятті структурування.

- Нехай задано скінченну множину  $S_0 = \{s_0^{(0)}, s_1^{(0)}, \dots, s_{n_0}^{(0)}\}$ , на якій задане відношення порядку. Тобто  $s_0^{(0)} < s_1^{(0)} < \dots < s_{n_0}^{(0)}$ .

Структуруванням множини  $S_0$  називається пара  $\langle S_1, f_0^1 \rangle$ , в якій  $S_1 = \{s_0^{(1)}, s_1^{(1)}, \dots, s_{n_1}^{(1)}\}$ ,  $n_1 \leq n_0$ , і  $f_0^1$  – бієктивне відображення, яке кожному елементу множини  $s_i^{(1)} \in S_1$  ставить у відповідність скінченний закритий інтервал  $f_0^1(s_i^{(1)})$  множини  $S_0$ , причому:

1.  $\bigcup_i f_0^1(s_i^{(1)}) = S_0$ .
2.  $\forall i \neq j \in [0, n_1]: f_0^1(s_i^{(1)}) \cap f_0^1(s_j^{(1)}) = \emptyset$ .

**Часовим універсумом** називається скінченна множина

$$TU = \{T_0, \langle T_1, c_0^1 \rangle, \dots, \langle T_n, c_{n-1}^n \rangle\},$$

де  $T_i, i = \overline{1, n}$  – часові домени,  $c_{i-1}^i$  – відображення, яке ділить (структурує) множину  $T_{i-1}$  на блоки (інтервали) елементів, що відповідають елементам множини  $T_i$ .

Відображення  $c_{i-1}^i$  часового універсуму задають **календарну систему часу**.

#### Приклад 1.1

Нехай елементи часової осі відображають секунди в годині: при порядковому записі кожна секунда матиме свій номер від 0 до 3599. При структурному записі виділимо два домени:

1.  $T_0 = \{S_0, \dots, S_{3599}\}$  – секунди.
2.  $T_1 = \{M_0, \dots, M_{59}\}$  – хвилини.

Можна ввести традиційну календарну систему:  $c_0^1(M_0)=[S_0, S_{59}]$ ,  $c_1^1(M_1)=[S_{60}, S_{119}]$ ,  $c_2^1(M_2)=[S_{120}, S_{179}]$  і т.д. Тоді елемент 25 порядкового запису матиме структурний вигляд 00:24 (0 хвилини 24 секунди), а елемент 125 запишеться 02:04 (2 хв 4 с).

Існує велика кількість календарних систем, які характерні для різноманітних предметних областей. Для відображення предметної області в базі даних необхідно забезпечити збереження та опрацювання часових структур, котрі притаманні конкретній предметній області. Це, відповідно, дасть змогу опрацювати часові залежності різноманітних інформаційних характеристик предметної області. Для створення повноцінного відображення предметної області в базі даних необхідно створити відповідні календарні системи, які відображають структури часу в предметній області. При створенні календарної

системи завжди виникає задача вибору часового домену. В більшості випадків ця задача зводиться до задачі поділу або розділення деякого часового інтервалу на дрібніші. Критерії і підходи до вибору способів розділення часових інтервалів подано нижче.

### ПОСТАНОВКА ЗАДАЧІ ПОДІЛУ ЧАСОВОГО ІНТЕРВАЛУ

Виділимо фактори, котрі визначають вибір часових доменів чи поділ часових інтервалів у базі даних на рис. 1.

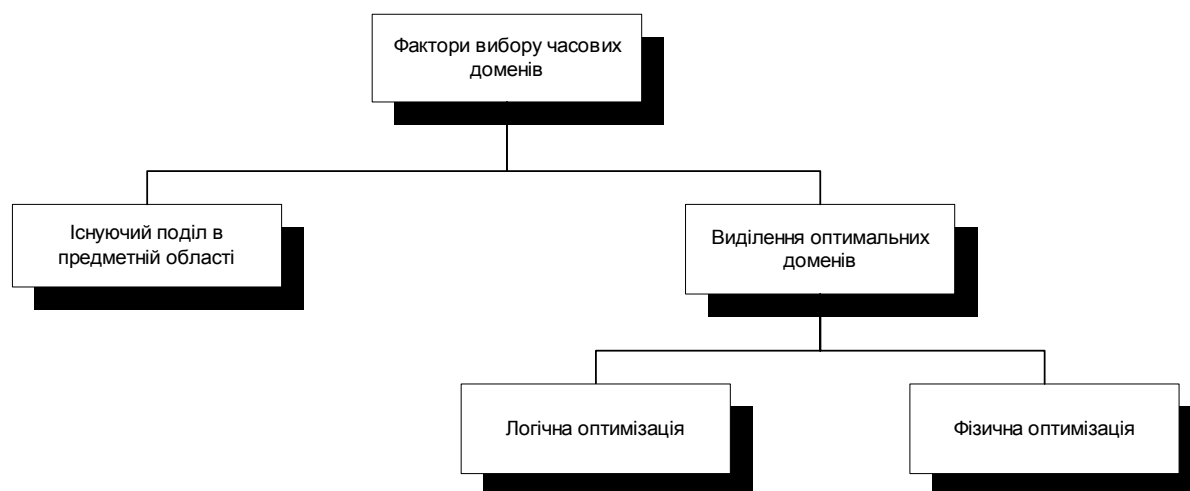


Рис. 1. Фактори вибору часових доменів для реалізації в БД

Фактори вибору часових доменів, які будуть представлятися і оброблятися в БД, можна розділити на дві категорії:

- *Існуючий поділ в предметній області* – обумовлює вибір для представлення в БД таких часових доменів(інтервалів), які безпосередньо присутні в предметній області.
- *Виділення оптимальних доменів* – вибір таких часових доменів(інтервалів) для представлення чи опрацювання в БД, які задовольняють деякі критерії оптимальності.

Загалом вибір оптимальних інтервалів часу можна розділити на *логічну* та *фізичну* оптимізацію (рис. 1).

Під *фізичною* оптимізацією розуміємо виділення часових проміжків, які дають вигоду на фізичному рівні опрацювання даних в БД (напр., побудова індексів, зменшення надлишковості, зменшення часу виконання часового з'єднання відношень і т.д.).

Під *логічною* оптимізацією розуміємо виділення часових проміжків, які дають можливість виявляти деякі узагальнені часові залежності показників предметної області (наприклад, у системі обліку доступу до Інтернет ресурсів можна виділити проміжки часу (домени), які можуть яскраво демонструвати залежність трафіка: робочі – неробочі дні, ранок-день-вечір-ніч).

Як бачимо, вибір оптимальних часових структур передбачає попередню наявність в БД самих даних, на основі яких можна проводити деяку оптимізацію. Також можна стверджувати, що часові домени за першим фактором (“предм. обл.”) вибираються на етапі проектування схеми БД, тоді як за другим фактором (“оптимізація”) – вже при функціонуванні (наповненій даними) БД.

Прикладом фізичної оптимізації є вирішення так званої задачі “інтервального поділу”.

### Задача інтервального поділу

Задача полягає в оптимальному розподілі множини деяких часових інтервалів, які перекриваються. Для прикладу розглянемо графічне представлення деякої множини часових інтервалів, зображене на рис. 2.

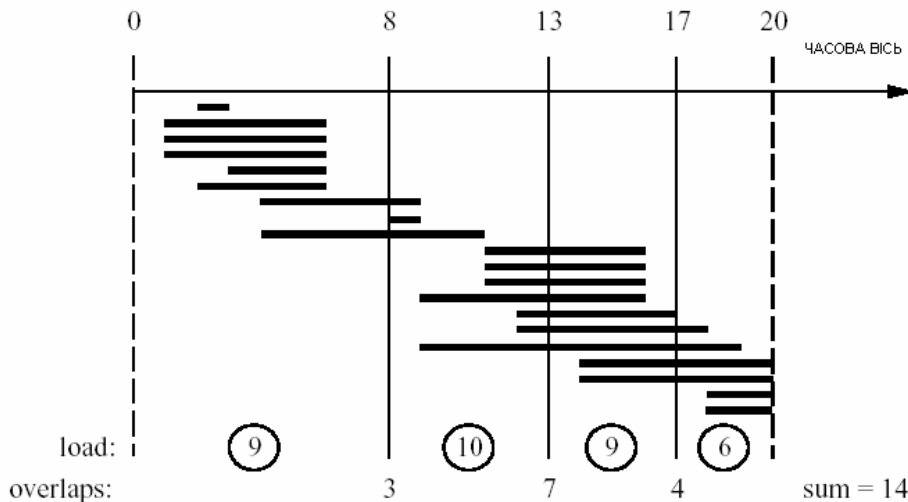


Рис. 2. Приклад характеристик перекриття деяких часових інтервалів

На часовій осі задано домен з часовими мітками  $p_j$ , для  $j=0..m$ . Визначимо такі функції:

- $load(p_{j-1}, p_j) = m$  – кількість інтервалів  $(u_1, u_2, \dots, u_m)$ , для яких виконується умова:  $begin(u_i) \in [p_{j-1}, p_j]$  Or  $end(u_i) \in [p_{j-1}, p_j]$ , для  $i = 1..m$ ; тобто кількість інтервалів, які перекриваються з інтервалом  $[p_{j-1}, p_j]$ .
- $overlaps(p_j) = n$  – кількість інтервалів  $(u_1, u_2, \dots, u_n)$ , для яких виконується умова:  $p_j \in u_i$ , для  $i = 1..n$ ; тобто кількість інтервалів, в які потрапила часова мітка  $p_j$ .

#### Приклад задачі інтервального поділу

Нехай задано множину  $U$  як колекцію інтервалів  $\langle u_1, u_2, \dots, u_N \rangle$ ,

і деяке натуральне число  $X$ .

Потрібно знайти таке розбиття  $P = \{p_1, p_2, \dots, p_{m-1}\}$  множини  $U$ , яке мінімізує

$$\sum_{p \in P} overlaps(p)$$

при умові:  $load(p_{j-1}, p_j) \leq X$  для  $j = 1..m$  (примітка:  $p_0 = -\infty$ ;  $p_m = +\infty$ ).

Так, для прикладу, поданого вище, оптимальним розподілом буде розподіл  $\{6, 9, 17\}$ , відображений на рис. 3.

Розв'язання цієї задачі має такі застосування:

- часові з'єднання;
- структури часових індексів;
- фізичний розподіл.

Для розв'язання цієї задачі існують алгоритми, які ґрунтуються на динамічному програмуванні.

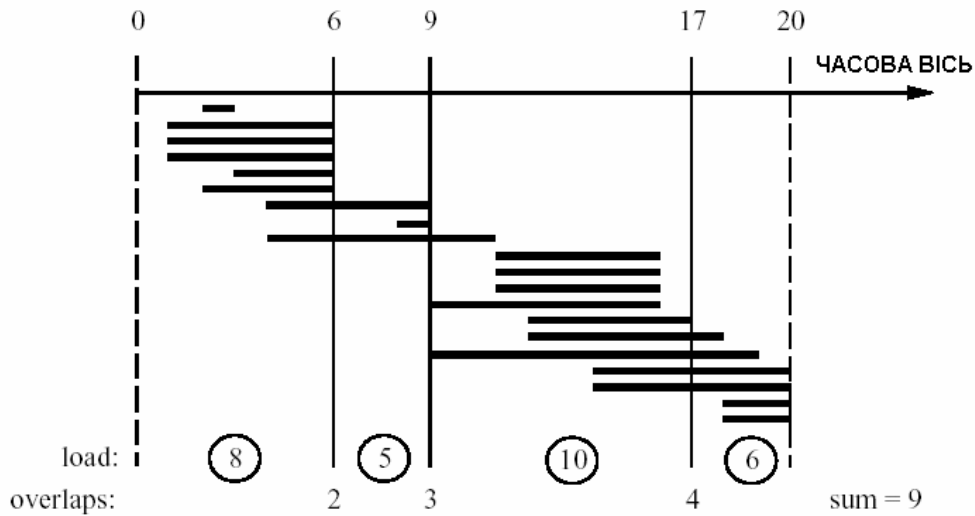


Рис. 3. Приклад характеристик перекриття деяких часових інтервалів

### 3. СТРУКТУРУВАННЯ ЧАСУ ДЛЯ ІДЕНТИФІКАЦІЇ ЧАСОВИХ ЗАЛЕЖНОСТЕЙ

Для прикладу розглянемо розвиток стану предметної області, зображений на рис. 4. Нехай задано історію значення обсягів продажу  $Q$  деякого товару протягом року. У випадку відображення такої інформації в реляційних базах даних часова вісь являє собою дискретну множину часових міток. Кожній мітці відповідає календарний день. Рік заданий інтервалом  $[t_s, t_e]$ . Візуально множину всіх значень характеристики за середнім значенням можна поділити на три інтервали:  $U_1 = [t_s, t_i]$ ,  $U_2 = [t_{i+1}, t_j]$ ,  $U_3 = [t_{j+1}, t_e]$ . Про значення характеристик на цих інтервалах можна стверджувати: для  $U_1$  – низьке значення ( $Q_l$ ), для  $U_2$  – високе значення ( $Q_h$ ), для  $U_3$  – середнє значення ( $Q_m$ ). Такий поділ часового інтервалу (у даному випадку року) дає можливість з'ясувати залежність об'ємів продажу протягом року від запропонованих інтервалів.

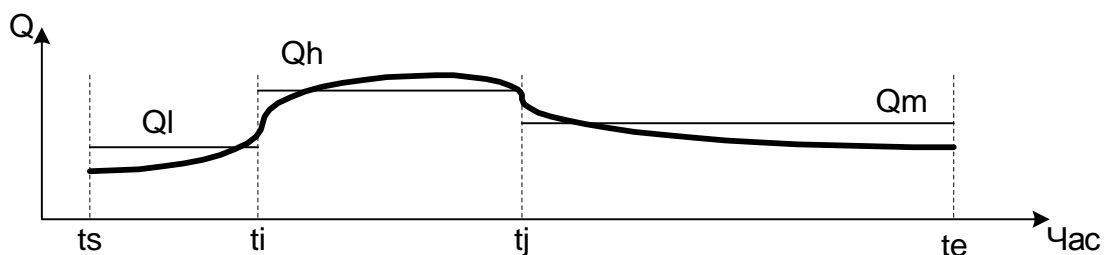


Рис. 4. Приклад історії зміни деякого показника предметної області

Такий підхід до поділу часового інтервалу дає змогу зручніше подати узагальнену залежність значення показника  $Q$ . Іншими словами, доцільно в цій предметній області виділити нову структуру часу, яка розбиває рік на вказані проміжки. Відповідно в реляційній базі даних, яка відображає стан предметної області, потрібно забезпечити подання вказаних часових інтервалів. До того ж, може виявитися, що це відомі проміжки предметної області, які просто не були відображені в базі даних. Так, для даного прикладу можна припустити, що  $U_1$  – весна,  $U_2$  – літо,  $U_3$  – осінь – зима. Не виключено, що така

залежність обсягів продажу буде спостерігатись і надалі, що обумовлює відображення виділених інтервалів в базі даних, яка описує вказану предметну область, з подальшим опрацюванням.

Як бачимо з прикладу, виділення нових часових інтервалів в деякій предметній області, з відображенням їх у відповідній базі даних, може забезпечити якісно новий рівень для проведення аналізу часових залежностей. Такий підхід до опрацювання часових залежностей в базах даних може також бути інструментом прогнозування розвитку предметної області.

Розв'язання сформованої вище задачі являє собою розбиття (за деякими ознаками) множини значень деякого інформаційного показника предметної області на окремі класи в розрізі часу. Розв'язання таких задач межує з кластерним аналізом. У цій роботі запропоновано уніфікований підхід для розв'язання таких задач, котрий сформований у вигляді алгоритму.

### Уніфікований алгоритм розбиття часового інтервалу

Нехай на часовій осі задано послідовність часових моментів (хрононів) у вигляді інтервалу  $\{t_0, t_1, \dots, t_n\}$  (розглядаємо дискретну лінійну модель часу, в якій між будь-якими двома моментами однаковий проміжок часу). Кожному моменту часу відповідає значення деякого інформаційного показника ( $Q_i, i=0\dots n$ ) предметної області (рис. 5).

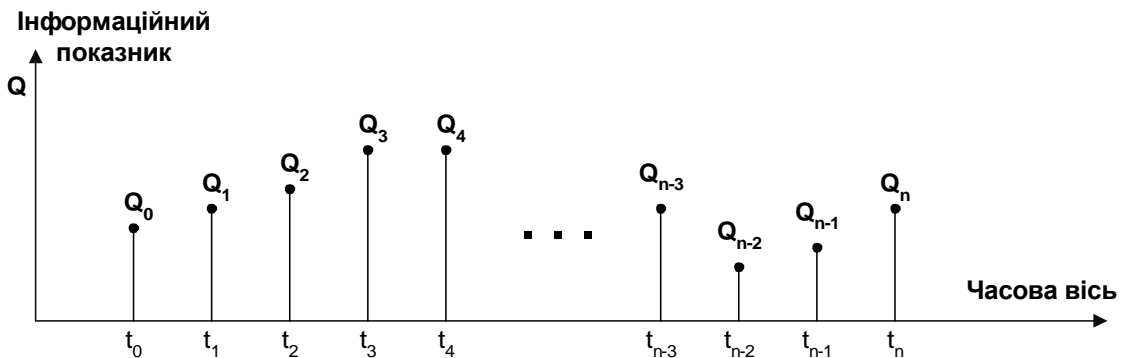


Рис. 5. Зміна деякого показника предметної області

Задача полягає в розподілі інтервалу  $[t_0, t_n]$  на множину інтервалів  $\{U_1, U_2, \dots, U_m\}$ , де  $U_j = [t_k, t_l)$ ,  $j = 1..m$  та  $k < l$  (умова відкритості інтервалу не є обов'язковою). Відповідно  $\text{begin}(U_j) = t_k$ ,  $\text{end}(U_j) = t_l$ . Нехай  $\text{QOND}(U_j)$  – деяка умова (критерій) формування інтервалу.

Сформулюємо послідовний алгоритм.

#### Алгоритм формування інтервалів

На початку роботи алгоритму:  $i := 0, j=0, m := 1$  – кількість інтервалів (рис. 6).

Алгоритм послідовно перебирає часові моменти і додає їх до поточного виділеного інтервалу, поки виконується (задовольняється) умова формування інтервалу  $\text{QOND}(U_m)$ . Якщо при переході до наступного часового моменту порушується умова формування інтервалу, то з цієї часової мітки починаємо формувати наступний інтервал. Такий послідовний ітераційний процес продовжуємо, поки не досягнемо кінцевого часового моменту. Після закінчення роботи алгоритму буде сформовано множину з  $m$  інтервалів, які розділяють вихідний інтервал  $[t_0, t_n]$  згідно з деяким критерієм декомпозиції.

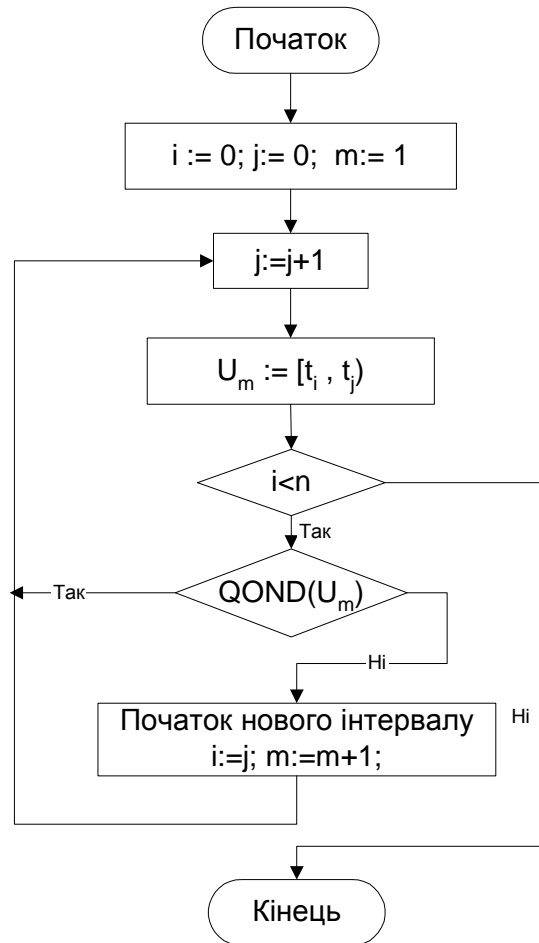


Рис. 6. Алгоритм формування часових інтервалів

Щодо критерію декомпозиції, то залежно від нього на множині значень інформаційного показника можна виділяти різноманітні проміжки. Наприклад, проміжки з різко відмінним значенням показника; проміжки зростання – спадання; проміжки різкої зміни значень показника та проміжки відносної стабільності тощо.

Можливі такі найпростіші варіанти вибору критерію  $QOND(U_j)$ :

- Обмеження на різницю між максимальним і мінімальним значеннями інформаційного показника на проміжку, який формується  $\Delta Q_{U_i} \leq \Delta Q_{\max}$ , де  $\Delta Q_{U_j} = (Q_{\max} - Q_{\min})_{t \in U_j}$ .

Такий критерій розбиває часовий ряд на проміжки з суттєво відмінними рівнями значення інформаційного показника.

- Можливий критерій – обмеження на різницю між максимальним і мінімальним значеннями швидкості зміни значень показника на проміжку. Такий критерій розбиває часовий інтервал на ділянки з відносно сталим значенням швидкості зміни значень інформаційного показника. Тобто виділяє ділянки різкого зростання – спадання чи відносної сталості значень показника.

Запропонований вище алгоритм допускає використання інших критеріїв формування часових інтервалів, що виходить за межі даної роботи. Запропонований уніфікований алгоритм дає змогу побудувати універсальний підхід до формування часових інтервалів на

часовій осі. Формування різноманітних критеріїв виділення часових інтервалів віднесене до перспективних напрямків подальших досліджень.

### ВИСНОВОК

У роботі було проведено огляд літератури за проблематикою структурування часу та виділення часових інтервалів. У результаті огляду проблематики в літературних джерелах та проведення досліджень отримано такі результати:

а) виділено фактори, котрі визначають підходи до структурування часу в деякій предметній області з подальшим відображенням в базі даних;

б) сформовано поділ задач оптимального вибору часових інтервалів на логічну та фізичну оптимізацію;

в) формалізовано підхід до вирішення задачі поділу часового інтервалу ряду значень інформаційного показника для визначення часових залежностей;

г) запропоновано уніфікований алгоритм виділення вищевказаних інтервалів і можливі критерії їх формування.

1. Claudio Bettini X. Sean Wang Sushil Jajodia, *A General Framework and Reasoning Models for Time Granularity* // *Time –96 workshop*. 2. Claudio Bettini, X. Sean Wang, Sushil Jajodia, *Solving multi-granularity temporal constraint* // *Artificial Intelligence 131 (2001) 135–170*. 3. Жежнич П.І. *Методи та засоби моделювання часового параметра в інформаційних системах на основі реляційних баз даних: Дис. ...канд. техн. наук. – 2000. – Машинопис*. 4. Прохоров А. *Informix Временной ряд как объект хранения в СУБД* // *CIT Forum* [http://www.citforum.ru/seminars/cbd2001/day\\_2\\_5\\_informix.shtml](http://www.citforum.ru/seminars/cbd2001/day_2_5_informix.shtml) 5. Guida G. and Mauri G. *Evaluating Performance and Quality of Knowledge-Based Systems: Foundation and methodology* // *IEEE Trans. Knowledge and Data Eng.* – 1993. – Vol. 5, № 2. P. 204–244.