

ЧАСОВІ ВИМІРИ АНАЛІЗУ УСПІШНОСТІ СТУДЕНТІВ

© Ришковець Ю.В., Жежнич П.І., 2008

Розглянуто поняття часу в різних інтерпретаціях. Проаналізовано представлення часу в базах даних і висвітлено зв'язок ліній часу між ними, також введено чітку класифікацію часових вимірів аналізу успішності студентів.

In this article are considered concept of time in different interpretations. Analysed presentation of time in databases and reflected connection lines of time between them, also entered crisp classification of the temporal dimantions of students progress analysis.

Постановка проблеми

Найпоширенішою компонентою навколишнього світу є час. Майже неможливо знайти об'єкт (поняття) чи сукупність об'єктів (понять), що не мають жодного відношення до часу. Дослідження будь-якого процесу проводиться в контексті часу.

Контекст часу є надзвичайно багатограним. Реально він є не одним поняттям, а терміном, який позначає цілу сукупність властивостей нашого Світу. Він може бути і епохою нашого інформаційного суспільства, і розпорядком дня людини або періодом її життя, і найважливішим параметром усіх процесів у природі, суспільстві і техніці. Особливо це стосується комп'ютерних інформаційних систем та глобальної мережі Інтернет. В кожному з цих процесів час має свої властивості, свій ритм і масштаб.

Оскільки час сам по собі нічого не вартий, то він прив'язується до різних дій, подій, інформації, і саме такий підхід вкладає у зв'язку *час-дані* певний сенс.

Що ж таке часові дані? У дуже широкому сенсі – це довільні дані, які явно або неявно пов'язані з певними датами або проміжками часу. Під таке визначення потрапляють майже будь-які дані та інформація. Наприклад, навіть якщо немає явної залежності від часу для якої-небудь аксіоми, то все одно для неї є неявна залежність від часу, оскільки колись нам (або системі) стало відомо, що така аксіома існує. Крім того, є вірогідність, що в майбутньому аксіома буде спростована або на умови її застосування будуть накладені певні обмеження. Тому вже не можна буде розглядати її як деяку абсолютну істину, правильну у всіх ситуаціях і у будь-який момент часу.

Часові бази даних – це бази даних, що зберігають часові дані. Проте ці бази даних і дані, що в них містяться, можуть розглядатися як часові тільки в тому випадку, якщо відоме правило інтерпретації тимчасових міток та інтервалів для конкретної системи керування базами даних (СКБД). Щоб визначити, чи є дана СКБД часовою в повному розумінні цього слова, необхідно зрозуміти, чи можна окремо виділити і спеціально інтерпретувати дані атрибуту “час”. До категорії часових СКБД не потраплятимуть звичайні реляційні СКБД, в яких підтримуються пов'язані з часом типи даних. У “справжній” часовій СКБД враховуються специфічна природа часу і мінливість даних з часом.

У статті запропоновано різні погляди стосовно інтерпретації часу, з'ясовано у часові виміри на прикладі аналізу успішності студентів.

Зв'язок висвітленої проблеми із науковими завданнями

Метою статті є дослідження вимірів часу.

Наукова новизна статті полягає у введенні чіткої класифікації вимірів часу.

Практична цінність статті визначається можливістю представлення у базах даних основних часових характеристик даних.

Аналіз останніх досліджень

Розглянемо поняття часу з математичного погляду.

Час – це безмежна щільна множина елементів, на якій визначено відношення порядку. Ця множина є ізоморфною до множини дійсних чисел, тобто часова вісь подається як вісь дійсних чисел. Таке визначення відображає усі аспекти часу як одного з елементів дійсності.

В області баз даних **часову вісь** визначають як зліченну дискретну множину, на якій задано відношення порядку \leq (“менше рівне”). Елементи цієї множини називаються **часовими моментами**.

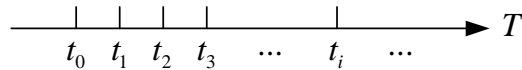


Рис. 1. Часова вісь

На рис. 1 зображено часову вісь з дискретними моментами часу. Причому:

$$T = \{t_0, t_1, \dots, t_i, \dots\}, \text{ де } t_0 < t_1 < \dots < t_i < \dots$$

Дискретна часова вісь ізоморфна до множини натуральних чисел.

Час визначається як “кількість руху” відносно попереднього та майбутнього. Парадокс цього твердження Аристотеля полягає в тому, що минулого вже немає, майбутнє ще не настало, а “тепер” є не частиною часу, а скоріше границею між минулим і майбутнім. Хоча вважається, що існують тільки події “теперішнього”. Події минулого вже не існують, а події майбутнього ще не існують. Також важливо, що “тепер” – це не просто деяка точка часової осі, а момент усвідомлення теперішнього.

Для того, щоб встановити цей момент “усвідомлення” теперішнього в інформаційних системах (ІС), скористаємося поняттям транзакції в архітектурі клієнт/сервер. Транзакція – це набір операцій, які змінюють вміст бази даних і виконуються як єдине ціле. Якщо транзакція виконується успішно, то усі зміни вмісту бази даних стають загальнодоступними, у протилежному випадку усі зміни відміняються, і вміст бази даних залишається таким як до початку виконання транзакції.

Отже, теперішнє для ІС – це період виконання транзакцій. “Усвідомлення” теперішнього – це момент завершення транзакцій. Якщо інформація про подію потрапляє до бази даних після завершення транзакції, то така подія з погляду ІС вже є минулою. Тобто база даних зберігає інформацію лише про минуле. Події майбутнього не зберігаються у базі даних.

Кожному теперішньому відповідає своє минуле і майбутнє (множина можливих майбутніх). Усю цю складну структуру називають метатеперішнім, або метамоментом. З цього погляду, плин часу полягає в переході від одного метамоменту до іншого.

Отже, будь-яка подія має певний часовий зв'язок з минулим і з майбутнім. З погляду бази даних, це означає, що події можуть характеризуватися трьома групами часових вимірів, кожна з яких відповідає минулому, теперішньому і майбутньому. Зобразимо метамоментну структуру часу в базах даних так [1]:

$$(PAP_1, \dots, PAP_m, T_1, \dots, T_n, FAP_1, \dots, FAP_k),$$

де PAP_1, \dots, PAP_m – періоди активації минулого (Past Activation Period), T_1, \dots, T_n – часові виміри теперішнього, FAP_1, \dots, FAP_k – періоди активації майбутнього (Future Activation Period).

Періоди активації минулого відображають ті часові проміжки, які вплинули на появу події. Аналогічно періоди активації майбутнього – це ті часові проміжки, протягом яких відчуватиметься вплив події. Часові виміри теперішнього відображають момент часу, в який трапилася подія.

Дослідники часу розрізняють два способи його подання: **порядковий** та **структурний** [1]. Порядковий спосіб передбачає подання часу за допомогою натуральних чисел (див. вище). Структурний спосіб ґрунтується на понятті структурування.

Розглянемо скінченну множину $S_0 = \{s_0^{(0)}, s_1^{(0)}, \dots, s_{n_0}^{(0)}\}$, на якій задане відношення порядку. Тобто $s_0^{(0)} < s_1^{(0)} < \dots < s_{n_0}^{(0)}$. Визначимо оператори $\text{first}(S_0) = s_0^{(0)}$ та $\text{last}(S_0) = s_{n_0}^{(0)}$.

Означення 0.1. Структуруванням множини S_0 називається пара $\langle S_1, f_0^1 \rangle$, де $S_1 = \{s_0^{(1)}, s_1^{(1)}, \dots, s_{n_1}^{(1)}\}$, $n_1 \leq n_0$, 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. $\forall i \neq j \in [0, n_1]: f_0^1(s_i^{(1)}) \cap f_0^1(s_j^{(1)}) = \emptyset$.

Означення 0.2. 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 фактично задають календарну систему часу. Кожна предметна область визначає часовий універсум.

За допомогою структурування визначається структурний підхід до подання часу. На відміну від порядкового підходу, де кожен момент часової осі записується у вигляді порядкового номера (натурального числа), за структурним підходом момент часу задається як послідовність номерів блоків кожного з часових доменів T_i .

У базах даних не достатньо лише формального визначення часу, оскільки з даними пов'язують певну семантику, саме тому потрібно розглянути час з погляду бази даних. При цьому треба зауважити, що семантичне представлення часу в різних базах даних може істотно відрізнитись.

Якщо розглядати дані, представлені в базі даних, як деяке віддзеркалення поточного стану дійсності для модельованого світу, то кожен запис може сприйматися як деякий факт, який є істинним в певний інтервал часу [1]. При переході до часової бази даних для кожного факту не завжди можна вказати той проміжок часу, коли цей факт був істинним в модельованому світі, представленою в базі даних, а також не завжди можна однозначно стверджувати, що цей факт був актуальним в минулому чи буде істинним в конкретний проміжок часу у майбутньому. Подібне представлення часу, коли з даними зв'язується проміжок часу їх актуальності (з погляду модельованого світу), називається **дійсним (valid)**, або **модельним часом**. Причому в самій базі даних дійсний час задається як конкретний момент часу. Оскільки досить часто в базі даних відображається саме реальний світ, можуть бути задані співвідношення між значеннями часу реального світу і представленою в базі даних моделлю. Значеннями цього типу часу можуть бути моменти часу як у минулому, так і в майбутньому. Крім того, ці значення можуть змінюватися, тобто істинність факту в певні моменти часу може прийматися або відхилятися.

Іншим типом часу, який розглядається дослідниками часових баз даних, є **час транзакцій**. У будь-якій СКБД з кожним записом бази даних можна зіставити той проміжок часу, коли даний запис був представлений в базі даних, тобто проміжок часу між моментами додавання запису і його видалення з бази даних. Зазначимо, що при цьому операція оновлення, яка дійсно вносить зміни до запису, розуміється як складена операція видалення старого запису і додавання нового. Очевидно, що значення часу транзакцій не можуть належати до майбутнього. У переважній кількості СКБД час транзакцій використовується для роботи з блокуваннями та журналом для відновлення системи. У деяких системах навіть можна використовувати спеціальні розширення мови SQL, що дають змогу отримати доступ до часу транзакцій та історії змін записів у базі даних.

Виділення невирішених частин проблеми

У повсякденному житті людина майже не замислюється, що використовує тільки один вид часу. Події могли вже відбутися у минулому або тільки плануються в майбутньому, але час завжди вимірюється однаково. З іншого боку, в базі даних може зберігатися інформація про події та інтервали часу, які відповідають різним уявленням і зв'язкам. Якщо обробляє ці дані сам користувач, то тип часу, який він використовує, можна назвати часом, що визначає користувач. Його особливістю є відсутність інтерпретації з боку СКБД, оскільки обробка даних, пов'язаних з часом, повністю покладається на користувача. Фактично всі сучасні СКБД забезпечують підтримку подібного різновиду часу, наприклад, за допомогою введення спеціальних типів даних DATA або TIMESTAMP. Хоч він і відображається в зручній для користувача формі, але насправді час зберігається як число. Отже, час – це число, представлене у деякій формі.

Отже, сьогодні традиційні системи баз даних не розрізняють контексти часу (навіть дійсний час та час транзакції, без яких практично не може належним чином існувати жодна інформаційна система), і в найближчій перспективі цього не передбачається. Тобто вся “часова” обробка даних в сучасних інформаційних системах проводиться на рівні функцій ІС, що вимагає певних “домовленостей” інтерпретації часу.

Ми розглянемо класифікацію часових вимірів та їх інтерпретацію на прикладі предметної області (ПО), пов'язаної з успішністю студентів. Центральним об'єктом в цій предметній області є студент, який навчається в певній групі і протягом кожного семестру отримує бали на модулях, лабораторних та практичних заняттях. У результаті заліку чи іспиту з навчального предмету студент отримує бали, які спочатку фіксуються у відомості, а потім заносяться до бази даних. Після закінчення сесії формується один або декілька наказів. Накази бувають різні: на зарахування, відрахування, переведення, академічну відпустку. Кожна зміна оцінки у студента фіксується як у відомості, так і в базі даних.

Основні дослідники часових баз даних (Snodgrass R.T., Jensen C.S., Gadia S.K., Clifford J. та ін.) вважають, що реальний час в базі даних відображається за допомогою часових інтервалів [3–6]. З цими інтервалами вони пов'язують період актуальності факту. Однак аналіз ПО про успішність студентів показує, що не завжди можна однозначно стверджувати, що протягом всього цього періоду часу факт буде актуальним. Більше того, один і той самий факт в різних ситуаціях може одночасно вважатися як актуальним, так і неактуальним.

Наведемо приклад таких двозначних ситуацій. Нехай існує база даних, в якій зберігається інформація про рейтинги оцінок студентів за весь період їх навчання. Інформація про рейтинг оцінок студентів буде актуальною, наприклад, тоді, коли потрібно визначити, чи буде студенту призначена стипендія на наступний семестр, або який рейтинг оцінок він має за декілька періодів навчання – семестрів. Факт вступу студента на перший курс не гарантує, що студент протягом всього свого навчання постійно отримуватиме стипендію, матиме найвищий рейтинг оцінок чи взагалі навчатиметься в цьому навчальному закладі й надалі.

Наведений приклад показує, що інтервали часу не доцільно пов'язувати з дійсним часом. Також це підтверджує аналіз паперового документообігу в ПО, пов'язаною з успішністю студентів.

У ПО, пов'язаною з успішністю студентів, оперують такими документами, як відомості, накази і т.д. Відомості характеризуються датою, що в базі даних інтерпретується як дійсний час, причому кожна відомість характеризується своїм кодом, номером, датами виписки та закриття відомості, датами проведення основного, першого та К-го талонів. Кожний наказ характеризується своїм кодом, номером і датою його підписання. В наказах дійсний час задається датою, починаючи з якої цей наказ набуває чинності. Коди документів також можна вважати часовими вимірами, оскільки вони відображають послідовність заведення цих документів у базу даних, тобто за кодом документа можна визначити, який документ був заведений раніше, а який – пізніше. Семантично код документа близький до часу транзакції, однак на відміну від часу транзакції, який може бути однаковим для декількох документів (якщо вони формуються в межах однієї транзакції), коди документів не можуть повторюватися.

Наведений приклад показує, що використання лише моментів часу для позначення дійсного часу адекватно відображатиме стан ПО про успішність студентів у базі даних.

Основний матеріал

У ПО, пов'язаною з успішністю студентів, виділяються дві складові масивів даних: довідникова і документальна. Довідникова частина містить списки студентів, груп та ін., документальна – відомості, накази, довідки, нарахування стипендії.

Довідникова частина не передбачає аналізу даних в розрізі часу навіть за наявності атрибутів типу Дата/Час (наприклад, дата народження), оскільки з погляду аналізу успішності студентів не доцільно ставити питання раніше/пізніше (хто зі студентів народився раніше). Іншими словами, довідникову частину можна вважати статичною і не враховувати.

Власне дані в часі аналізують саме за документальною частиною. Кожний документ ПО про успішність студентів з погляду їх структури складається з двох частин: заголовка і тіла. Основними атрибутами, що вказуються у заголовку, є код документа, дата, з якої цей документ набуває чинності, семестр, період дійсності, період академічної відпустки, час транзакції, а в тілі – прізвище, ім'я та по батькові студента, дія, що над ним виконується (зарахування до числа студентів, відрахування, переведення на наступний курс чи в іншу групу, надання академічної відпустки) і час транзакції.

Щоб відповісти на питання, як саме співвідносяться між собою дійсний час і час транзакцій, розглянемо такий приклад.

Нехай задано таблицю, в якій зберігається інформація про поточну оцінку студента (табл.1).

Таблиця 1

Таблиця без часових параметрів

Студент	Оцінка
Андрушко	5
Колінко	4

За підтримки дійсного часу можна було б у будь-який момент часу сказати, яка у студента була оцінка в певний момент часу (табл. 2). Тобто можна контролювати, яку оцінку отримав студент в конкретний момент часу і до якого моменту вона була істинною. Отже, дані про оцінку можуть бути представлені як послідовність значень, що змінюються.

Таблиця 2

Таблиця з підтримкою дійсного часу

Студент	Оцінка	Дійсний час
Андрушко	5	25 грудня 2006
Колінко	3	1 березня 2006
Колінко	5	25 січня 2007
Колінко	4	14 лютого 2007

За підтримки часу транзакцій можна було б сказати, в який момент в таблицю були внесені зміни (табл. 3).

Таблиця 3

Таблиця з підтримкою часу транзакцій

Студент	Оцінка	Час транзакцій
Андрушко	5	10 січня 2007
Колінко	3	3 березня 2006
Колінко	5	12 лютого 2007
Колінко	4	18 лютого 2007

Тепер припустимо, що для таблиці підтримується як дійсний час, так і час транзакцій (табл. 4). Тоді у випадку, якщо неправильно введені дані були згодом виправлені, можна буде точно сказати, коли це було зроблено. Крім того, інформація про подібні зміни необхідна, оскільки некоректні дані могли бути вже використані в яких-небудь звітах. Тому в цьому випадку потрібна підтримка часу транзакцій. При оновленні значень в системі (навіть у разі виправлення помилки в даних) інтервал часу транзакцій також оновлюється, тому можна переглянути список змін в базі даних.

Таблиця 4

Таблиця з підтримкою обох типів часу

Студент	Оцінка	Дійсний час	Час транзакцій
Андрушко	5	25 грудня 2006	10 січня 2007
Колінко	3	1 березня 2006	3 березня 2006
Колінко	5	25 січня 2007	12 лютого 2007
Колінко	4	14 лютого 2007	18 лютого 2007

Отже, тимчасові мітки часу транзакцій надають інформацію про час зміни даних або виправлення помилок, а тимчасові мітки дійсного часу зберігають інформацію про зміну деяких параметрів реального світу. Тобто, дійсний час і час транзакцій виявляються ортогональними один одному (рис. 2).

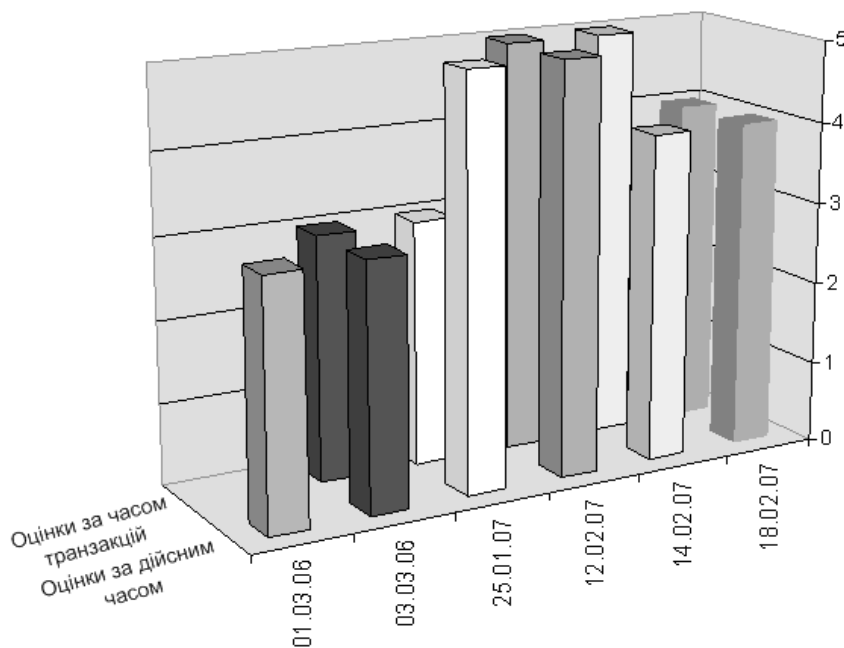


Рис. 2. Зв'язок ліній часу для студента Колінко

При розгляді дійсного часу йшлося про те, що існує деякий інтервал, в якому певний факт є істинним. Це так зване **інтервальне представлення**. Проте можна розглядати окремих момент часу і всі факти, які були істинні в цей конкретний момент (табл. 5). Тут йдеться про представлення часу з погляду користувача, тобто ті умовні моделі, у межах яких можуть формулюватися запити і повертатися їх результати. При використанні будь-якого з цих представлень істинність фактів не змінюється, але в разі **точкового**, або **моментного представлення** можна отримати зріз всіх фактів на якийсь конкретний момент часу, а для інтервального представлення нас цікавить певний факт і періоди його істинності.

Моментне представлення і зрізи на лінії часу

Час Дата	Дійсний час		Час транзакцій	
	Андрушко	Колінко	Андрушко	Колінко
28.02.2006				
01.03.2006		3		
02.03.2006		3		
03.03.2006		3		3
...		3		3
24.12.2006		3		3
25.12.2006	5	3		3
...	5	3		3
09.01.2007	5	3		3
10.01.2007	5	3	5	3
...	5	3	5	3
24.01.2007	5	3	5	3
25.01.2007	5	5	5	3
...	5	5	5	3
10.02.2007	5	5	5	3
11.02.2007	5	5	5	3
12.02.2007	5	5	5	5
...	5	5	5	5
13.02.2007	5	5	5	5
14.02.2007	5	4	5	5
18.02.2007	5	4	5	4
...	5	4	5	4
зараз	5	4	5	4

Для позначення часу можна використовувати не тільки представлення часу, до яких ми звикли, у таких форматах, як *день-місяць-рік* чи *година-хвилина-секунда*, а й у вигляді звичайних натуральних чисел.

Крім того, за допомогою точкового та інтервального часу можна представляти процес навчання студентів. Наприклад, в точковому представленні зручно представляти інформацію про рік навчання студента (курс), номер семестру чи номер модуля в семестрі. Термін навчання студентів складається з п'яти або шести років, кожний рік поділяється на семестри, а семестри, своєю чергою, поділяються на модулі. Все це є моментами часу, а інтервалами є проміжки часу від одного модуля одного семестру до другого модуля того самого чи іншого семестру, або ж проміжки часу між семестрами чи курсами (табл. 6). Тут умовно показано, що в кожному семестрі для предмета є лише два модулі, насправді їх може бути менше або більше.

Таблиця 6

Часова вісь навчального процесу

1 курс		2 курс		3 курс		4 курс		5 курс		6 курс	
1 семестр		2 семестр		3 семестр		4 семестр		5 семестр		11 семестр	
Модуль 1	Модуль 2	Модуль 1	Модуль 2	Модуль 1	Модуль 2	Модуль 1	Модуль 2	Модуль 1	Модуль 2	Модуль 1	Модуль 2

Весь процес навчання, як і все наше життя загалом, складається з подій, появу яких спонукають дії. Дії та події пов'язані між собою, вони не можуть існувати одне без одного. Зрозуміло, що подія може відбутися внаслідок як однієї дії, так і їх сукупності. Таке явище – це неперервний процес.

Отже, передумовою факту виписки відомості є подія настання модульного або залікового тижня, передумовою факту допуску студента до складання сесії є подія отримання ним всіх заліків, передумовою факту переведення студента на наступний курс є подія складання ним всіх заліків та іспитів на попередньому курсі тощо.

Наприклад, знаючи, що студент навчається у певному семестрі, можна дізнатися, на якому він зараз курсі і скільки в цьому семестрі модулів чи знаючи, що студент навчається на першому курсі, можна стверджувати, що він успішно склав вступні іспити. Якщо студент навчається на останньому курсі, то це означає, що він є випускником і йому потрібно здати державні іспити та написати дипломну роботу. Тобто маючи лише певний момент або інтервал часу, можна отримати додаткову інформацію про факт, яка так чи інакше з ним пов'язана.

Висновки

Розглянуто поняття часу з математичного погляду та з погляду баз даних. У процесі досліджень встановлено, що час в базах даних не достатньо представляти лише як математичну величину, тому що з ним пов'язують своєрідну семантику.

Запропоновано пов'язувати з дійсним часом не часовий інтервал, а момент часу, оскільки не завжди можна однозначно стверджувати, що протягом всього періоду часу факт буде актуальним, причому один і той самий факт в різних ситуаціях може одночасно вважатися як актуальним, так і неактуальним.

Введено чітку класифікацію часових вимірів аналізу успішності, що дасть змогу не плутати і не підміняти поняття. Висвітлено зв'язок ліній часу між дійсним часом та часом транзакцій. Запропоновано модель часової осі для навчального процесу.

При аналізі інформації про успішність студентів різних груп та років навчання з деякого предмета, що накопичується в базі даних, можна з певною ймовірністю говорити, які рейтинги оцінок з цього предмета будуть у нових груп в майбутньому. Тобто можна прогнозувати рейтинг, спираючись на часові залежності.

Запропоновану класифікацію часу можна застосовувати не лише до процесу навчання студента, а й до будь-якої сфери людської діяльності. Зокрема, дійсний час у вигляді моментного (точкового) та інтервального представлень актуально використовувати і в археології, коли окремі ери розглядати як моменти часу, а під інтервалом часу розуміти час між двома, трьома і т.д. моментами часу – ерами.

1. Zhezhnych P., Peleschyshyn A. *Time Aspects of Information Systems. //Proceedings of the IXth International Conference CADSM, 2007. – P.530–533.* 2. Костенко Б.Б., Кузнецов С.Д. *История и актуальные проблемы темпоральных баз данных.* http://citforum.ru/database/articles/temporal/#_ftn2. 3. Clifford J., Rao A. *A simple, general structure for temporal domains. //Temporal aspects in information systems, IFIP, 1988.* 4. Jensen C.S., Clifford C., Elmasri R., Gadia S.K., Hayes P., Jajodia S. *A consensus glossary of temporal database concepts. //Technical Report R 93-2035, Dept. of Mathematics and Computer Science, Inst. for Electronic Systems, Denmark, Nov. 1993.* 5. Snodgrass R.T., Ahn I. *Temporal databases. //Computer, vol.19, no.9, pp.35–42, Sept. 1986.* 6. Snodgrass, R.T. *Addendum to Valid- and Transaction-time Proposals. ANSI X3H2-96-582, ISO/IEC JTC1/SC21/WG3 DBL MAD-203, November 1996, ftp://ftp.cs.arizona.edu/tsql/ysql2/ysql3/ansi-96-582.pdf.*