

УДК: 61:004.45

## АНАЛІЗ ЕКСПЛУАТАЦІЙНИХ ХАРАКТЕРИСТИК БАЗИ ДАНИХ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПЕРЕВІРКИ ЗНАНЬ У МЕДИЧНІЙ ОСВІТІ

**А. В. Семенець***Тернопільський державний медичний університет імені І.Я.Горбачевського*

Метою даної роботи є проведення аналізу експлуатаційних характеристик розробленої БД ІСПЗМО використовуючи резервні копії файлів БД ІСПЗМО, збережені при проведенні пробних тестувань. Зокрема, оцінено вплив застосування BLOB-полів даних та RTF-тексту в якості ІХ вмісту на фрагментованість таблиць БД. Також проаналізовано вплив рівня фрагментованості таблиць на швидкодію БД шляхом вимірювання часу виконання найбільш важливих для роботи ІСПЗМО запитів.

**Ключові слова:** медична освіта, база даних, BLOB-поле, фрагментація таблиць, час виконання запиту.

## АНАЛИЗ ЭКСПЛУАТАЦИОННЫХ ХАРАКТЕРИСТИК БАЗЫ ДАННЫХ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРОВЕРКИ ЗНАНИЙ В МЕДИЦИНСКОМ ОБРАЗОВАНИИ

**А . В. Семенец***Тернопольский государственный университет имени. И.Я.Горбачевского*

Целью данной работы является, используя резервные копии файлов БД ИСПЗМО, сохраненные при проведении пробного тестирования, проведение анализа эксплуатационных характеристик разработанной БД ИСПЗМО. Произведена оценка влияния BLOB-полей данных и RTF-текста в качестве их содержимого на фрагментированность таблиц БД. Также проанализировано влияние уровня фрагментации таблиц на быстродействие БД путём измерения времени выполнения наиболее важных для работы ИСПЗМО запросов.

**Ключевые слова:** медицинское образование, база данных, BLOB-поле, фрагментация таблиц, время выполнения запроса.

## THE DATABASE PERFORMANCE BENCHMARK OF THE TEST KNOWLEDGE INFORMATION SYSTEM IN THE MEDICAL EDUCATION

**A. V. Semenets***Ternopil State Medical University by I.Ya. Horbachevsky*

The purpose of this work is the analysis of the database performance parameters of the test knowledge information system in the medical education. The BLOB-fields usage follow the database tables fragmentation. The ASKS and ASK\_ANSW tables average and real fragmentation levels are presented. The impact of the tables fragmentation level to the database performance was examined too. For this purpose the execution time was measured to the most important database query.

**Key words:** medical education, database, application development, BLOB-field, table fragmentation, query execution time.

**Вступ.** Сучасні інформаційно-аналітичні системи (ІАС), у тому числі і освітні, являють собою конкретні програмні додатки, кожен з яких працює з певним набором інформації - **базою даних** [1]. База даних (БД) - це єдине сховище даних, що одночасно визначається, а потім використовується одночасно багатьма користувачами - представниками різних підрозділів [2].

Як систему керування базою даних для технічної реалізації проекту інформаційної системи перевірки знань в медичній освіті (ІСПЗМО) в ТДМУ було обрано СУБД **Firebird** [3]. Основні причини такого вибору:

- Firebird - вільно розповсюджене, безкоштовне програмне забезпечення, що доволі актуально для навчального закладу;

© А. В. Семенець

- дана СУБД використовується в АСУ "Контингент", що призначена для автоматизації управлінської діяльності в медичному навчальному закладі;

- невеликі системні вимоги до комп'ютера-сервера бази даних при високій продуктивності й надійності [3];
- гарна підтримка в сучасних інструментальних системах.

В процесі розробки БД експериментального проєкту ІСПЗМО було здійснено:

- проаналізовано зовнішні представлення користувачів ІСПЗМО;
- визначено основні об'єкти інформаційної моделі БД;
- розроблено структуру бази завдань, як основи БД ІСПЗМО;
- визначення інформаційних атрибутів основних об'єктів бази даних і їхніх типів даних;
- нормалізацію для оптимізації структури бази даних [2];
- визначення додаткових об'єктів інформаційної моделі за результатами нормалізації, їхніх атрибутів і типів даних;
- уточнення зв'язків між основними й додатковими об'єктами;

Загальний вид отриманого в результаті зазначених операцій внутрішнього представлення бази даних ІСПЗМО показаний в роботі [4].

### 1. Аналіз фрагментованості таблиць БД ІСПЗМО

БД ІСПЗМО вказаної структури [4] проходить дослідну експлуатацію від початку 2007/2008 навчального року. З кожним новим навчальним роком в експерименті в ТДМУ система семестрового тестового іспиту поширювалася на студентів наступного курсу. Станом на 2010/2011 н.р. семестровий тестовий іспит складають студенти всіх курсів.

**Таблиця 1.** Загальна інформація про о

| № | Дата створення резервної копії БД | Розмір файлу БД |
|---|-----------------------------------|-----------------|
| 1 | 28/02/2008                        | 401 Мб          |
| 2 | 17/05/2008                        | 624 Мб          |
| 3 | 29/04/2009                        | 828 Мб          |
| 4 | 21/05/2009                        | 1,48 Гб         |
| 5 | 29/10/2009                        | 1,49 Гб         |

В якості тестового стенду використовувався домашній ПК автора роботи, що має наступну конфігурацію:

- Процесор - AMD Athlon XP, 1733 MHz 2100+ (ядро Thoroughbred-B);
- Материнська плата - ECS K7S6A;
- ОЗП - 1 Gb PC3200 DDR SDRAM
- Жорсткий диск - 320 Gb WDC WD3200AAKS-00VYA

Отже, розмір БД зростає через збільшення кількості предметів, з яких проводиться тестування. Слід також зазначити, що дана БД призначена для зберігання, насамперед, тестових питань, що містять мультимедійну інформацію. Тому, як текст питання, так і текст відповіді зберігаються в полях з типом даних BLOB [5]. Виникає необхідність в проведенні аналізу експлуатаційних характеристик БД, таких, як швидкодія, фрагментованість.

Особливість СУБД Firebird полягає в тому, дані в фізичному файлі зберігаються на окремих сторінках (Data Pages), розмір яких задається при створенні БД в межах 1024-8192 байт (найчастіше використовуються сторінки розміром 1024 або 4096 байт). При цьому ядро сервера може зберігати BLOB-поля трьома способами: [5]

- Якщо вміст BLOB-поля менший розміру вільного місця на сторінці даних, то дані BLOB-поля будуть збережені на тій же сторінці даних, що й запис (або його версія).

- Якщо вміст BLOB-поля більший розміру вільного місця на сторінці даних, то створюється нова сторінка даних, куди поміщаються лише дані BLOB-поля.

- Якщо в попередньому випадку вміст BLOB-поля більший розміру сторінки даних, то створюється сторінка вказівок для зберігання посилань на всі сторінки даних цього BLOB-поля.

Спершу оцінимо, як впливає застосування BLOB-полів даних та RTF-тексту в якості їх вмісту на фрагментованість таблиць БД, оскільки даний параметр має значний вплив на швидкодію виконання запитів до БД, а, отже, і ІАС в цілому. Для аналізу використано окремі резервні копії БД ІСПЗМО, зроблені в різні періоди дослідної експлуатації (табл. 1).

мі файли резервних копій БД ІСПЗМО

- Операційна система - Windows XP Professional SP2

Для оцінки фрагментованості таблиць БД застосуємо утиліту IBAAnalyst (автори Кузьменко Д. В. Ковязін А. Н.) [1]. Дана утиліта аналізує статистику сервера БД і формує загальний звіт по файлу БД в цілому та детальний звіт по таблицях та індексах. Автоматично формується і список проблем та рекомендації щодо можливості їх усунення. Через вели-

кий загальний обсяг далі подано лише найважливішу частину звітів утиліти IBAlyst що включає інфор-

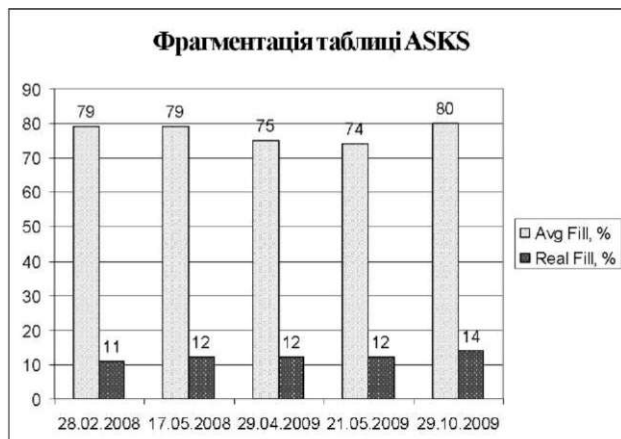


Рис. 1. Фрагментація таблиці ASKS БД ІСПЗМО.

Як видно з діаграм (див. рис. 1, 2), найбільш фрагментованими виявилися таблиці ASKS та ASK\_ANSW відповідних файлів БД, в яких зберігаються, відповідно, текст тестових питань та текст відповідей на них. На цьому етапі аналізу можна зробити наступні висновки:

- невисоке реальне заповнення сторінок даних свідчить, що дані BLOB-полів розміщуються безпосередньо на сторінках даних, а, отже, ступінь фрагментації таблиць ASKS та ASK\_ANSW є високим;
- вище реальне заповнення сторінок даних таблиці ASK\_ANSW вказує на те, що реальний розмір даних її BLOB-полів менший за розмір даних BLOB-полів таблиці ASKS;
- ступінь фрагментації таблиць ASKS та ASK\_ANSW БД із зростанням розміру файлу БД та числа записів змінюється мало;

## 2. Вплив фрагментованості таблиць на швидкість виконання запитів

У більшості БД високий ступінь фрагментації таблиць шкідливо впливає на швидкодію програмного додатку. Оцінка впливу фрагментації сторінок даних на швидкодію БД проводиться шляхом визначення часу виконання певного запиту [4]. Ця ж методика застосовується і для дослідження можливостей масштабування БД ІКС, зокрема, появи нелінійних ефектів в БД.

В першу чергу слід відмітити невеликі мінімальні системні вимоги СУРБД Firebird [3]:

- Процесор - Intel Pentium II або новіший;
- ОЗП - мінімально 16, рекомендовано 64 і більше;
- Операційна система Windows NT 4.0 SP5.

В ранніх версіях СУРБД Firebird підтримувалися лише 32-розрядні операційні системи Windows, Solaris

та Linux. Починаючи з версії 2.0 (2007 рік) введено повноцінну підтримку 64-розрядних версій Windows та Linux.

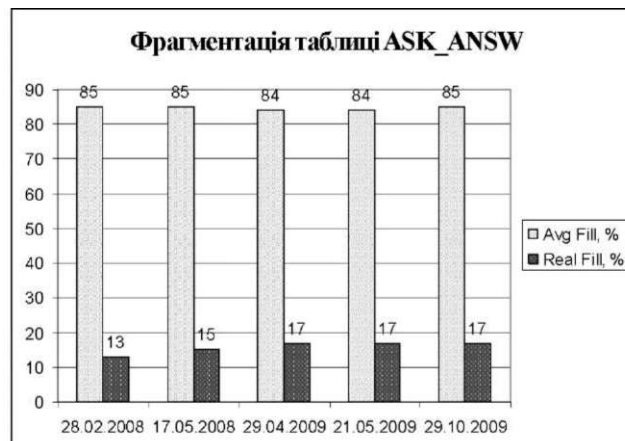


Рис. 2. Фрагментація таблиці ASK\_ANSW БД ІСПЗМО.

та Linux. Починаючи з версії 2.0 (2007 рік) введено повноцінну підтримку 64-розрядних версій Windows та Linux.

Вимоги до дискового простору повністю визначаються розміром файлу БД. Однак слід враховувати обмеження, що накладаються різними версіями операційних систем та різними файловими системами, наприклад, при використанні файлової системи FAT32 розмір файлу не повинен перевищувати 4 Gb. Всі версії СУРБД Firebird підтримують розбиття БД на окремі файли. Станом на кінець 2009 р. (версія СУРБД Firebird 2.1.3) теоретична межа розміру одного файлу БД становить 64 Tb (терабайта), при використанні багатобайтового режиму - 131 Tb.

Слід відмітити, що практично було проведено експеримент по створенню і тестуванню роботоздатності БД розміром 1 Tb (1 терабайт) з використанням СУРБД Firebird. В якості сервера БД використовувався звичайний офісний комп'ютер. З деталями даного експерименту можна ознайомитися в [1]. В результаті авторами роботи підтверджено відмінну масштабованість СУРБД при мінімальних затратах. Експеримент підтвердив можливість застосування навіть типового офісного ПК в якості сервера БД і його роботоздатність при великому розмірі файлу БД.

В офіційній документації до СУРБД Firebird [3] немає вказаних обмежень на кількість користувачів, що можуть обслуговуватися одним сервером. Отже, ця кількість може бути обмежена лише апаратною та програмною конфігураціями комп'ютера-сервера СУРБД.

Порівняємо час виконання деяких типів запитів, що застосовуються в ІСПЗМО. Текст запитів та їх призначення вказані в таблиці 2.

Таблиця 2. Призначення окремих запитів в ІСПЗМО

| № | Текст запиту  | Призначення запиту в ІСПЗМО   |
|---|---|---|
| 1 | select *<br>from asks   | Вибирає всі записи з таблиці включно з вмістом BLOB-полів.  |
| 2 | select *<br>from asks<br>where contentmod id =56  | Вибирає записи з таблиці включно з вмістом BLOB-полів для окремого змістового модуля.   |
| 3 | select id_ask, contentmod_id, points, clinical_sign, frequent,<br>ask time, asktype id, answwcount, datechange, archive, ki, di,<br>from asks where contentmod id =56   | Вибирає записи з таблиці для окремого змістового модуля без вмісту BLOB-полів.  |
| 4 | select<br>d modules.cnt module theme,<br>d_modules.cnt_semestr,<br>d contentmodules.id contentmod,<br>d contentmodules.contmod theme,<br>disciplines.chair_id,<br>disciplines.cnt_discipline_name,<br>disciplines.faculty id,<br>disciplines.speciality_id,<br>d_modules.cnt_edu_year_id,<br>d_modules.cnt_module_num,<br>count(asks.id_ask) count_of_id_ask<br>from asks<br>inner join d contentmodules on (asks.contentmod id =<br>d contentmodules.id contentmod)<br>inner join d_modules on (d_contentmodules.module_id =<br>d_modules.id_module)<br>inner join disciplines on (d_modules.discipline_id =<br>disciplines.id_discipline)<br>group by d_modules.cnt_module_theme,<br>d_modules.cnt_semestr, d contentmodules.id contentmod,<br>d_contentmodules.contmod_theme, disciplines.chair_id,<br>disciplines.cnt_discipline_name, disciplines.faculty id,<br>disciplines.speciality id, d_modules.cnt_edu_year_id,<br>d_modules.cnt_module_num | Використовується для підрахунку кількості питань по кожному змістовому модулю з усіх навчальних дисциплін для всіх кафедр. Практично запит застосовується при формуванні структури тесту та при формуванні управлінської звітності. |

Проаналізуємо час виконання вищевказаних запитів Для моніторингу за виконанням запитів застосовано до всіх наявних файлів резервних копій БД (табл. 3). програму IVExpert [1 ]. При проведенні тестування

Таблиця 3. Інформація про файли резервних копій БД ІСПЗМО

| Дата створення резервної копії файлу БД | Розмір файлу БД, Мб | Кількість змістових модулів (таблиця D CONTENTMODULES) | Кількість питань (таблиця ASKS) |
|---|---------------------|--|---------------------------------|
| 16.10.2007                              | 256                 | 193  | 69215                           |
| 28.02.2008                              | 401                 | 264  | 94678                           |
| 17.05.2008                              | 624                 | 424  | 153573                          |
| 21.11.2008                              | 754                 | 540  | 210017                          |
| 13.01.2009                              | 726                 | 601  | 204253                          |
| 04.03.2009                              | 726                 | 602  | 204395                          |
| 03.04.2009                              | 731                 | 604  | 206448                          |
| 29.04.2009                              | 828                 | 665  | 248461                          |
| 21.05.2009                              | 1521                | 765  | 283518                          |
| 30.05.2009                              | 1539                | 798  | 293806                          |
| 23.10.2009                              | 1551                | 832  | 304093                          |
| 29.10.2009                              | 1526                | 840  | 306620                          |
| 12.11.2009                              | 1538                | 894  | 317361                          |
| 24.11.2009                              | 1553                | 969  | 328974                          |
| 11.12.2009                              | 1582                | 1065   | 356374                          |

часу всі файли резервних копій БД були розміщені на ПК автора на одному логічному диску одночасно. На кожному етапі експерименту здійснювалося підключення лише до одного файлу резервної копії

БД. В кожному з файлів резервних копій БД кожен запит виконувався 4 рази, після чого розраховувалося середнє значення. Результати виконання запитів наведені в таблиці 4.

**Таблиця 4.** Середній час виконання питів до файлів резервних копій БД

| Дата створення резервної копії файлу БД | Середній час виконання запиту, с |         |          |          |
|---|----------------------------------|---------|----------|----------|
|   | 1                                | 2       | 3        | 4        |
| 16.10.2007                              | 4,4445                           | 0,097   | 0,062125 | 3,40425  |
| 28.02.2008                              | 5,99225                          | 0,10175 | 0,0665   | 4,6565   |
| 17.05.2008                              | 8,84425                          | 0,1015  | 0,06625  | 7,83975  |
| 21.11.2008                              | 10,2185                          | 0,0995  | 0,068625 | 9,1525   |
| 13.01.2009                              | 11,59275                         | 0,09775 | 0,07075  | 10,46475 |
| 04.03.2009                              | 11,997                           | 0,1015  | 0,0665   | 10,44525 |
| 03.04.2009                              | 12,25375                         | 0,10925 | 0,06275  | 10,473   |
| 29.04.2009                              | 15,28525                         | 0,1015  | 0,0665   | 12,844   |
| 21.05.2009                              | 17,44175                         | 0,10125 | 0,06625  | 14,11875 |
| 30.05.2009                              | 20,371                           | 0,0975  | 0,07825  | 15,04    |
| 23.10.2009                              | 39,47675                         | 0,10575 | 0,074    | 16,75025 |
| 29.10.2009                              | 17,841                           | 0,09375 | 0,06275  | 15,39475 |
| 12.11.2009                              | 25,22725                         | 0,10975 | 0,08175  | 15,9415  |
| 24.11.2009                              | 19,418                           | 0,10125 | 0,0705   | 16,301   |
| 11.12.2009                              | 19,67975                         | 0,09775 | 0,062    | 18,49575 |

Розглянемо час виконання запитів більш детально. Запит № 1 (рядок 1 в таблиці 2) вибирає всі записи з таблиці включно з вмістом BLOB-полів. Хоча час виконання цього запиту великий - (до 55 секунд в окремих випадках), це не впливає на загальну швидкість БД ІСПЗМО, оскільки практично такий запит ніколи не виконується (таблиця 4).

Запит № 2 (рядок 2 в таблиці 2) вибирає записи з таблиці включно з вмістом BLOB-полів для окремого змістового модуля. При експлуатації ІСПЗМО такий запит виконується при редагуванні тестових питань викладачами (переходи між змістовими модулями навчальної дисципліни). Час виконання запиту менший 0,2 с достатній для забезпечення комфортної роботи користувача (табл. 4).

Запит № 3 (рядок 3 в таблиці 2) вибирає записи з таблиці для окремого змістового модуля без вмісту BLOB-полів. При експлуатації ІСПЗМО такий запит виконується при автоматизованому формуванні тест-білетів. Кожного семестру закінчується вивчення 10-12 модулів з різних дисциплін, при цьому кожен з них включає 2-4 змістових модулів. Таким чином, в процесі формування тест-білету запити, аналогічні запиту № 3 будуть виконуватися 20-50 разів. Враховуючи середній час виконання одного запиту, для вибірки даних для кожного варіанта тест-білету може знадобитися від 2,2 до 7,05 секунд (табл. 4).

Запит № 4 (рядок 4 в таблиці 2) використовується для підрахунку кількості питань по кожному змісто-

вому модулю з усіх навчальних дисциплін для всіх кафедр. Практично запит застосовується при формуванні структури тесту та при формуванні управлінської звітності (табл. 4).

В процесі розробки тест-білетів найбільш важливим є виконання запиту № 4. Представимо отримані дані в вигляді графіка (рис. 3). Очевидно, що на даному етапі експлуатації БД ІСПЗМО між кількістю змістових модулів та часом виконання запиту, що підраховує в них кількість тестових питань, спостерігається залежність, близька до лінійної, що підтверджується коефіцієнтом апроксимації лінійного тренду (рис. 3). Ознаки появи нелінійних ефектів відсутні.

Наявні незначні відхилення реального часу виконання запиту від ідеального лінійного тренду (рис. 3) зумовлені цілим рядом факторів, зокрема:

1. Інтервалом часу, що минув від останнього "збирання сміття" до моменту створення окремого файлу резервної копії.

2. Інтенсивністю виконання запитів на вставку чи оновлення записів, яка також впливає на дочасний запуск процедури "збирання сміття" в СУРБД Firebird.

3. Наявністю або відсутністю інших операцій резервного копіювання та наступного відновлення файлу БД з заміною оригінального файлу, що передували останній операції резервного копіювання.

4. Особливостями налаштування ПК автора роботи, ОС якого оптимізована для виконання активних

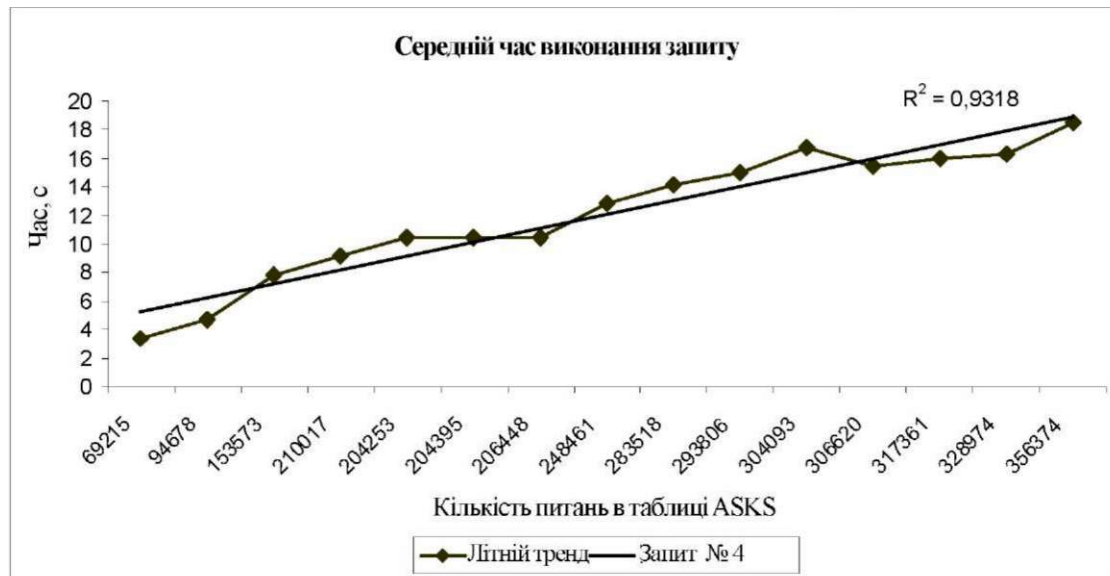


Рис. 3. Швидкодія виконання запиту № 4 до файлів резервних копій БД ІСПЗМО.

программ, а не системних сервісів (яким являється СУРБД Firebird).

Примітки:

1. Особливості роботи вбудованого механізму "збирання сміття" СУРБД Firebird описані в роботі [1].

2. При виконанні резервного копіювання файлу БД також виконується операція очистки БД від "сміття" та неактуальних версій записів, тому наступне відновлення приводить до зменшення розміру файлу БД та підвищення його швидкодії [2].

3. СУРБД Firebird застосовує стиснення даних BLOB-полів на основі алгоритму ZIP, тому послідовність операцій "резервне копіювання - віднов-

лення" зазвичай додатково зменшує розмір файлу БД та підвищує його швидкодію.

**Висновки.** Використовуючи резервні копії файлів БД ІСПЗМО, збережені при проведенні пробних тестувань, проведено аналіз експлуатаційних характеристик розробленої БД ІСПЗМО. Зокрема, оцінено вплив застосування BLOB-полів даних та RTF-тексту в якості їх вмісту на фрагментованість таблиць БД. Також проаналізовано вплив рівня фрагментованості таблиць на швидкодію БД шляхом вимірювання часу виконання найбільш важливих запитів. Отримані результати показали, що попри високий ступінь фрагментації головних таблиць БД, швидкодія БД в цілому є достатньою для ефективної роботи ІСПЗМО.

#### Лтература

1. Урнов В. Базы данных - основа всего / В.Урнов // ИКТ в образовании. - 2007. - №> 4. - С. 5-6.
2. Коннолли Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика / Т. Коннолли, Б.Карелии; [пер. с англ.]. - М. : Издательский дом "Вильямс", 2003. -1440 с.
3. Firebird - The RDBMS that's going where you're going. [Електронний ресурс] / Режим доступу до сайту : <http://www.firebirdsql.org/>.
4. Марценюк В.П. О структуре базы данных информационной системы проверки знаний в медицинском образовании / В. П. Марценюк, А. В. Семенец // Искусственный интеллект. - 2009. - №> 1. - С. 267-278
5. Кузьменко Д. В. Фрагментация (или разрежение) таблиц блобами / Д.В.Кузьменко [Електронний ресурс] / 2006. - Режим доступу до документу : <http://ibase.ru/devinfo/tablefraghtm>

6. IBaseAnalyst. [Електронний ресурс] / Режим доступу до сайту : [http://www.ibase.ru/download/ibanalyst\\_r.zip](http://www.ibase.ru/download/ibanalyst_r.zip)
7. База размером 1 терабайт на Firebird. [Електронний ресурс] / 2009. - Режим доступу до документу : <http://www.ibase.ru/devinfo/fb1tb.htm>
8. IBaseExpert Developer Studio [Електронний ресурс] / Режим доступу до сайту : <http://ibexpert.net/ibe/index.php?n=Main.IBaseExpert>
9. Харрисон А. Состояния транзакций, сборка мусора, интересующиеся и активные транзакции, sweep, и примеры / А. Харрисон; [пер. з англ. Д. В. Кузьменко]. - 1997. - [Електронний ресурс] / Режим доступу до документу : <http://www.ibase.ru/devinfo/oitoat.htm>
10. Кузьменко Д. В. Утилита GBAK, Firebird и InterBase / Д. В. Кузьменко [Електронний ресурс] / 2009. - Режим доступу до документу : <http://ibase.ru/devinfo/gbak.htm>