УДК 004

Н.Б. Шаховська, В.Я. Крайовський, Є. Могильський
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІНТЕГРАЦІЯ НЕОДНОРІДНИХ ДЖЕРЕЛ НА ОСНОВІ АНАЛІЗУ МЕТАДАНИХ

Описано методику інтеграції інформації з неоднорідних джерел на прикладі предметної області туризму. Розроблено схеми метаданих та методи їх опрацювання.

Ключові слова – інтеграція, корпоративні дані, простір даних, сховище даних.

In the article the method of information integration from heterogeneous sources based on the domain of tourismis is described. The schemas of metadatas and methods of their working are developed.

Keywords – integration, corporation content, dataspase, dataware.

Вступ

Опрацювання інформації, що потрапляє з неоднорідних джерел (бази даних, сховища даних) тощо сьогодні є надзвичайно актуальним завданням, оскільки лавиноподібне зростання інформації не завжди означає таке саме підвищення якості її аналізу. Сьогодні на ринку туристичних послуг є безліч інформації з приводу того, куди можна поїхати, до яких туристичних операторів звернутися по допомогу, яка вартість тієї чи іншої послуги та прогнози вражень від здійснення поїздки. Але така інформація зберігається у різних системах, у яких зазвичай не погоджено назв структурних елементів та немає розвинених засобів інтеграції.

Для зберігання та опрацювання консолідованої інформації використовують сховища даних [1–4]. Вони найкраще пристосовані до роботи з великими обсягами інформації, що потрапляє періодично з джерел даних.

Тому у статті розглянуто розроблення засобів інтеграції неоднорідної інформації з можливістю її подальшого опрацювання та підтримки прийняття рішень щодо керування певною галуззю.

Постановка задачі дослідження

На різних етапах роботи з інформацією в туристичній галузі виникає необхідність застосування сховищ даних, щоб ефективно й оптимально вирішувати проблеми аналітичного характеру. За допомогою сховищ даних можна легше реалізувати надскладні задачі, наприклад, визначення сфер покриття туристичними фірмами регіонів і сфер покриття відпочиваючими та відшукати кореляцію між цими сферами.

У роботах [5–7] обгрунтовано необхідність використання сховища даних для підтримки прийняття рішень у галузі туризму.

Щоб опрацювати аналітичну задачу сховищ даних туристичного бізнесу, необхідно звернути увагу на наявність факторів привабливості туристичних ресурсів:

– рекреаційні ресурси – це сукупність природних, природно-технічних, соціально-економічних комплексів та їх елементів, що сприяють відновленню та розвитку фізичних та духовних сил людини, її працездатності;

– природні рекреаційні ресурси – це особливості природи, природні та природно-технічні геосистеми, тіла, явища природи, їх компоненти й властивості, природоохоронні об'єкти;

– соціально-економічні рекреаційні ресурси – культурні об'єкти, пам'ятки історії, архітектури, етнографічні особливості території.

Подамо основні умови та задачі, які повинен розв’язувати пакет програм для комплексного інформаційного забезпечення, формування аналітичних та прогнозних даних та підтримки прийняття рішень стосовно гармонійного розвитку території:

1) обліку

- рекреаційно-оздоровчої сфери та забезпеченості її послугами населення;
- нормативної та законодавчої бази;
- структури адміністративно-територіальних одиниць;
- обліку та аналізу картографічної інформації;
- кліматичних умов;
- структури населення;
- культурно-освітнього забезпечення населення;
- транспортної мережі та сполучення;

2) аналізу

- структури населення;
- зайнятості та структури зайнятості;
- впливу господарської діяльності на навколишнє середовище, прогнозування та запобігання надзвичайним ситуаціям;
- розвитку господарської діяльності;
- забезпеченості житлово-комунальними послугами;

3) планування та прогнозування

- обсягів забезпечення та розвитку туристичної діяльності;
- забезпеченості житлово-комунальними послугами;
- рекреаційно-оздоровчої сфери та забезпеченості її послугами населення;
- культурно-освітнього забезпечення населення;
- в галузі капітальних ремонтів та будівництва;
- природних умов проживання населення;
- демографічного стану території;
- споживання мінеральних, паливно-енергетичних, водних, біологічних, земельних ресурсів у туристичній діяльності;
- розвитку туристично-рекреаційної сфери;
- аналіз та планування функціональної структури території;
- аналіз та планування умов життєдіяльності;
- аналіз ефективності та планування економічної діяльності;
- типове проектування.

Залежно від типу об’єкта інформація може зберігатися у різних моделях та надходити з різних джерел.

- туристичне агентство – база даних, динамічний Веб-сайт з базою даних, розміщеною на Веб-сервері;
- адміністративна одиниця – сховище даних;
- особа – Веб-сайт, база даних, текстові дані тощо;
- відпочинковий ресурс – база даних, Веб-сайт.

Оскільки специфікою галузі туризму є подання інформації в Інтернеті у вигляді реклами, замовлень тощо, то на найвищому рівні ієрархії моделей даних – колекції іменованих ресурсів з базовими властивостями – розмір, дата створення і тип (наприклад, зображення JPEG, база даних MYSQL).

Одним з основних видів збирання даних туристичного бізнесу є каталогізація елементів даних учасників. **Туристичний каталог** – це реєстр ресурсів туристичних даних, що містить базову інформацію про кожний з них: туристична одиниця, ім'я, місцезнаходження в туристичній одиниці, розмір, дата створення і власник тощо. Туристичний каталог містить не тільки описову інформацію (тобто виконує роль метаданих), але й зберігає для кожного учасника схему туристичної одиниці,

статистичні дані, швидкість зміни, точність, можливості відповідей на запити, інформацію про власника і дані про політику доступу і підтримку конфіденційності. Оскільки туристичні одиниці простору даних фізично не переносять у нього інформацію та можуть обмінюватись між собою інформацією, то у туристичному каталозі необхідно зберігати дані і про зв'язки між туристичними одиницями [6].

Подання джерел даних до каталогу схематично зображено на рис. 1. Тут джерелами даних сховища даних є бази даних (реляційні, об'єктно-реляційні або багатовимірні). Тому у каталозі необхідно також вказувати тип джерела та засоби його опрацювання (програмні продукти, стандарти передавання тощо).

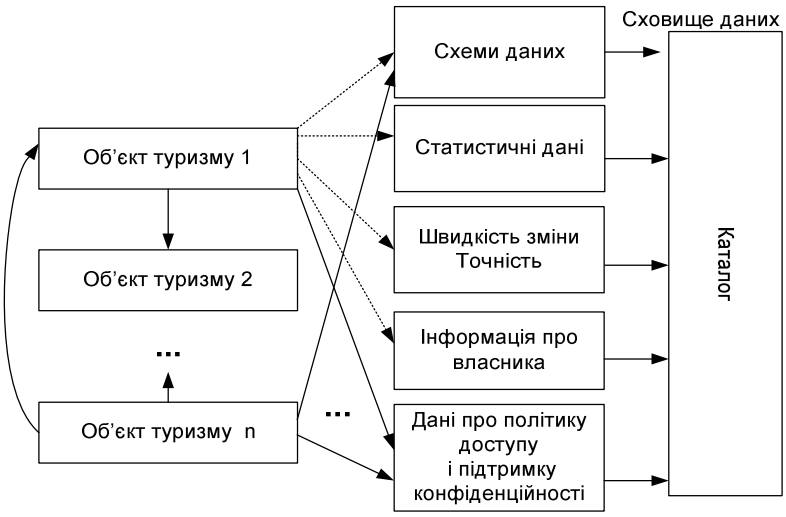

Рис. 1. Надходження даних до каталогу

Метою роботи є розроблення концептуальної моделі туристичної сфери та засобів інтеграції та опрацювання інформації з різних джерел.

Основний матеріал

Передусім опишемо зовнішні сутності, які водночас є і джерелами даних у задачі інтеграції. Концептуальна схема подана на рис. 2.

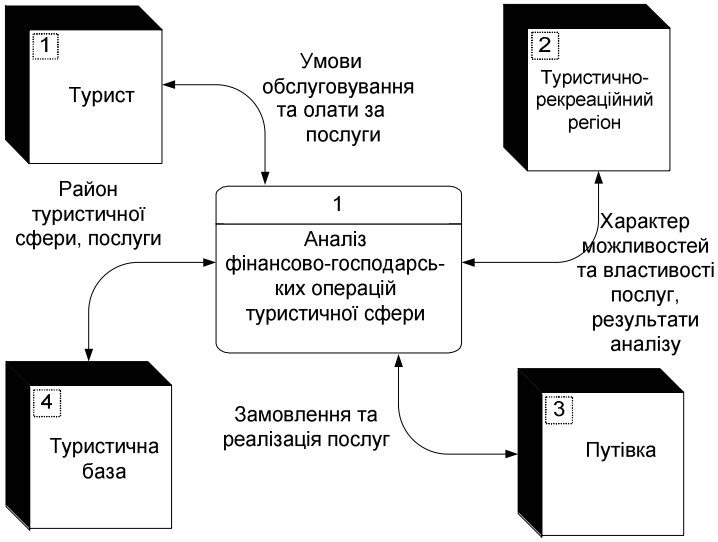

Рис. 2. Концептуальна схема

Зовнішніми сутностями є *Турист*, *Туристично-рекреаційний регіон*, *Путівка*, *Туристична база*. *Турист* – споживач туристичної сфери, що розглядає можливі умови обслуговування та оплату за послуги. *Туристично-рекреаційний регіон* – область, ландшафт та туристичні умови якої можуть привабити туриста. Сутність *Путівки* містить інформацію про готель, де відпочиватиме клієнт.

Туристична база – це одиниця туристичної сфери, послуги якої можуть характеризуватися залежно від її напряму: база відпочинку, оздоровча, спортивна тощо.

Основна робота Аналізу фінансово-господарських операцій туристичної сфери охоплює чотири процеси: Аналіз путівок від туроператорів, Аналіз замовлень на турбази, Аналіз виконання замовлень та Пропозиції для туриста (рис. 3).

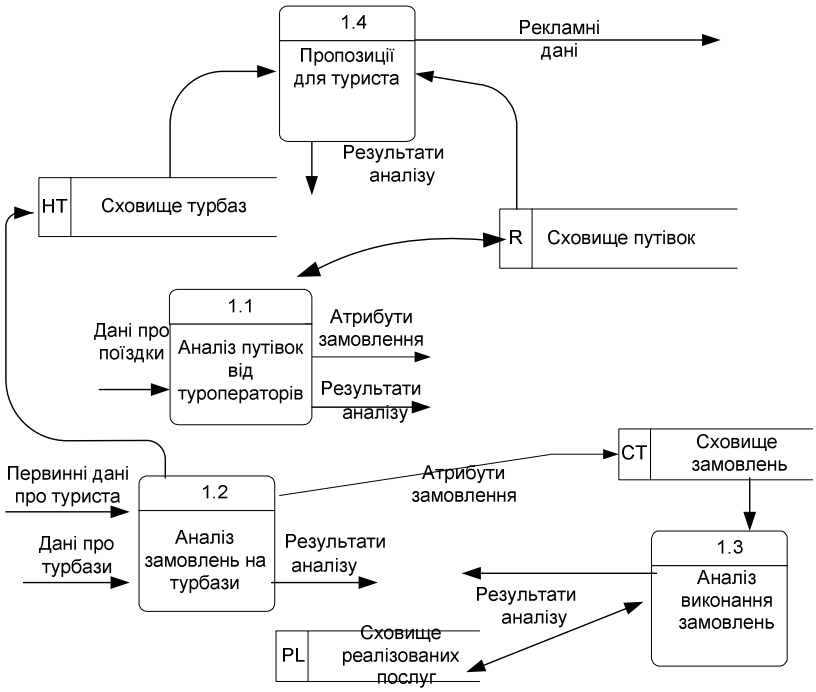

Рис. 3. Діаграма розгортання потоків даних ***Аналізу та обліку фінансових процесів туристичної сфери***

На діаграмі подано такі сховища даних: сховище турбаз, сховище путівок, сховище замовлень та сховище реалізованих послуг (рис. 3). Крім того, є ще сховище даних, яке використовується для опису об’єктів туристичного бізнесу. Це репозиторій метаданих. Цей репозиторій уможливлює ще й автоматичне завантаження та узгодження даних з різних джерел. Сховище реалізованих послуг містить інформацію про путівки, які вже надані туристу та оплачені ним. Кожен з розглянутих процесів передбачає виконання таких основних кроків (рис. 4):

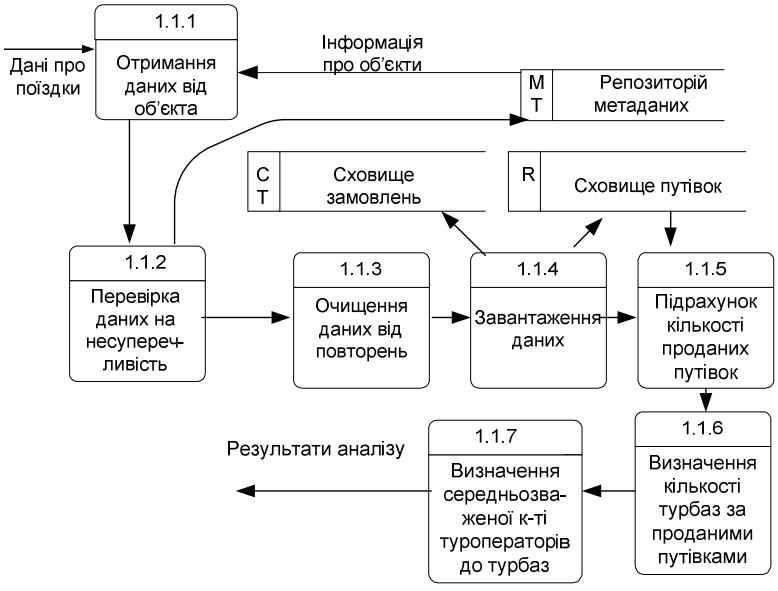

Рис. 4. Діаграма розгортання потоків даних ***Аналізу путівок від туроператорів***

- отримання даних від об’єкта;
- перевірка даних на несуперечливість;
- очищення даних від повторень;
- завантаження даних;
- власне сам аналіз.

Побудуємо основні об’єкти та їх характеристики, що використовуються для подання туристичної інформації.

Вхідною інформацією для комплексної туристичної інформаційної системи є:

- рекреаційні дані;
- природно-кліматичні умови;
- історичні, культурні та оздоровчі дані;
- дані соціальної сфери;
- дані людської діяльності.

Після подання діаграм розгортання потоків даних наведемо схему бази даних серверної частини. База даних, реалізована в SQL Server, під’єднується через ODBC до клієнта корпоративного сховища Access, у якому розроблений інтерфейс для отримання даних. Схема даних подана на рис. 5.

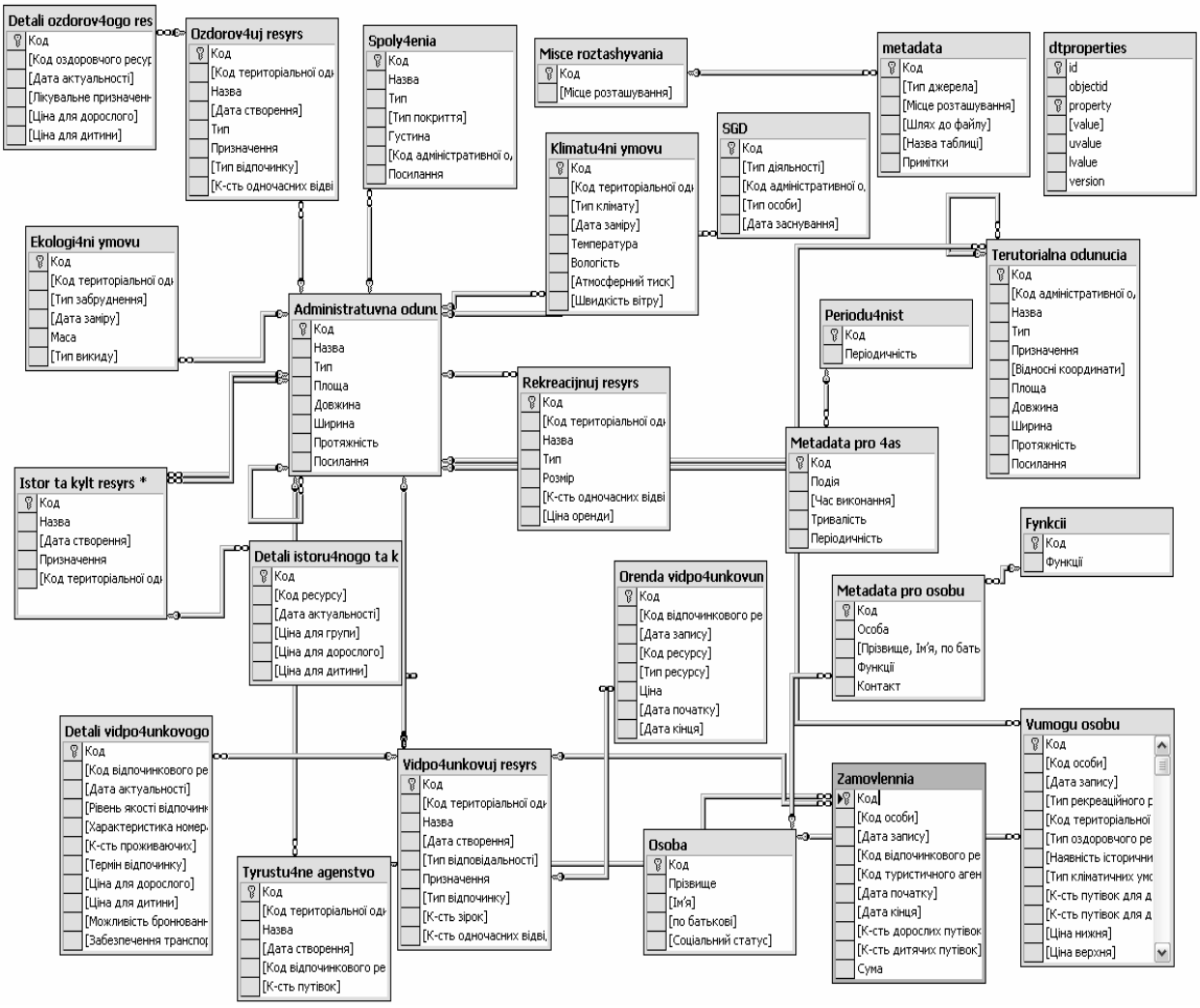

Рис. 5. Схема серверної бази даних

Структура сховища даних побудована за схемою “подвійна сніжинка”, оскільки планується одержувати інформацію з реляційних баз даних – від адміністративної одиниці, туристичної бази, СГД тощо. Надалі у дослідженні планується об’єднувати вищезгадану інформацію у сховище – за допомогою методів інтеграції (модифікований метод ETL) та агрегування.

Відношення фактів першої сніжинки – відпочинковий ресурс, який агрегує інформацію з вимірів адміністративної одиниці, кліматичних умов, екологічних умов, оздоровчих ресурсів, урбанізаційних умов та сполучень. Відношення вимірів другої зірки – замовлення відпочиваючого, який агрегує інформацію з вимірів туриста, туристичне агентство, оренда відпочинкового ресурсу.

Завдання підбору оптимальних умов розміщення відпочиваючих полягає у порівнянні даних з таблиць фактів “відпочинковий ресурс” та “замовлення”. Розв’язання цієї задачі можливе лише за наявності інтегрованої інформації з усіх відпочивальних комплексів.

Інтеграція даних – це об'єднання даних, які спочатку вводяться в різні системи. Самі ці системи можуть розташовуватися в одній локальній мережі, але мати різні платформи і внутрішню архітектуру. Зупинимося на деяких проблемах реалізації сховища даних, які приводять до виникнення задачі інтеграції даних. Серед них можна виділити [4]:

- неоднорідність програмного середовища;
- розподілений характер організації;
- підвищені вимоги до безпеки даних;
- необхідність наявності багаторівневих довідників метаданих;
- потреба в ефективному зберіганні й обробленні дуже великих об'ємів інформації.

Основою якісно виконаної інтеграції, як вже зазначалось, є туристичний каталог, який об’єднує інформацію про джерела даних та виконує роль метаданих.

Розроблено такі групи метаданих.

Метадані про користувачів

Metadata pro osoby	
Користувач	Дозволена таблиця сховища
Admin1	zamovlennya
Admin1	Vymogy osoby
Turist1	Tyrustu4na odunutsia
Turist1	Istor ta kylt resyrs
Turist1	Vidpo4unkovuj resyrs
Admin1	zamovlennja
Admin1	Metadata pro 4as
Admin1	Periody4nist

Метадані місцезнаходження джерел даних та самого сховища у локальній чи глобальній мережі (у тестовому варіанті використано локальну мережу):

Metadata					
Id	Тип джерела	Шлях до файлу	Ім'я файлу	Назва таблиці	Тип поповнення
1	Access	d:\	source1.mdb	documents	джерело даних
2	Access	d:\	База даних.mdb	documents	СД
3	Excel	d:\	avans.xls	documents	джерело даних

Метадані, що описують періодичність надходження даних:

Metadata pro 4as					
Код	Параметр	Опис	Запуск_процедури	Об'кт запуску	Періодичність
1	01.__.__	додавання даних з джерела	add_data	1	місяць
2	01.__.__	додавання даних з джерела	add_data	3	місяць
3	01.__.__	архівування даних	arch_data	2	місяць

Після опису усіх джерел даних та способу отримання інформації від них перейдемо до опису алгоритму інтеграції та розроблених для цього засобів.

Однією з найпоширеніших технологій інтеграції даних є технологія ETL. Проте вона працює ефективно тоді, коли структура джерел однорідна. У нашому випадку ми працюємо з неоднорідними джерелами інформації, користуючись їх описом, поданим у каталозі. Тому у роботі ми модифікуємо алгоритм технології ETL, встановивши залежності між значеннями зовнішніх ключів та їх описами.

Алгоритм інтеграції даних подано на рис. 6.

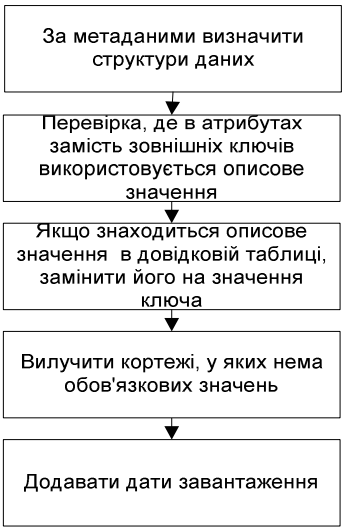

Рис. 6. Блок-схема аналізу, фільтрації та перетворення вхідних даних

Рис. 7. Блок-схема завантаження локального сховища.

Нехай дані надходять у відношення zamovlennia. На боці клієнтів є три джерела, які містять дані, що необхідно завантажити у сховище.

Для зручної роботи з джерелами розроблено такі форми:

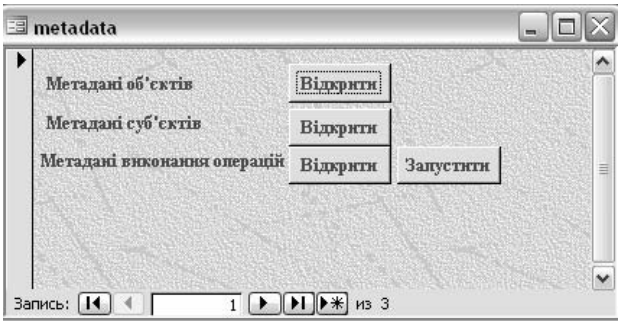

Рис. 8. Інтерфейсні елементи для завантаження даних з джерел

Для передавання у сховище подано такі дані:

– Джерело1 –таблиця MS Access

замовлення									
Код	Дата	Працівник	Тип документа	Підрозділ	Товар	Кількість	Ціна	Сума	Тип операції
7	12.02.2008	1	продаж	Леополіс		1	1 000,00 грн.	1 000,00 грн.	розхід
8	13.02.2008	2	продаж	Леополіс		1	2 000,00 грн.	2 000,00 грн.	розхід
9	14.02.2008	3	продаж	Леополіс		1	3 000,00 грн.	3 000,00 грн.	розхід
10	15.02.2008	4	продаж	Леополіс		1	4 000,00 грн.	4 000,00 грн.	розхід

– Джерело 2 – Лист Excel щодо продажів

Код	Дата	Працівник	Тип документа	Підрозділ	К-сть	Ціна	Сума	Тип операції
7	12.02.2008	Петренко	Замовлення	Леотур1	1	100,00 грн.	1 000,00 грн.	розхід
8	13.02.2008	Іваненко	Замовлення	Леотур1	1	200,00 грн.	2 000,00 грн.	розхід
9	14.02.2008	Вахненко	Замовлення	Леотур2	1	300,00 грн.	3 000,00 грн.	розхід
10	15.02.2008	Вахненко	Замовлення	Леотур2	1	400,00 грн.	4 000,00 грн.	розхід

– Джерело 3 – текстовий файл, сформований з html (файл з розділювачами)

60 12.02.08 3 1 4 10 3 3 9 -1

62 14.02.08 1 1 4 3 7 10 70 -1

Як видно з таблиці метаданих, джерела містять однотипну інформацію, яка має спільний характер, але відрізняється складом, способами подання і форматами. Також в одних джерелах значення атрибутів є рядками (джерело 2), у інших – зовнішні ключі відношень, які описують відповідні об’єкти (джерело 1, джерело 3 містять коди працівників). На основі метаданих та інформації з таблиць вимірів може бути утворена інтегрована таблиця, яка буде виконувати функції оперативного сховища даних.

Отримуємо таблицю фактів:

zamovlenia										
id	evdate	employee_id	Documents_type	delp_id	tovar_id	count	price	suma	type	real_date
1	12.02.2007	Вахненко	1	3	1	1	1 000,00 грн.	1 000,00 грн.	-1	
2	13.02.2007	Василенко	1	3	2	1	2 000,00 грн.	2 000,00 грн.	-1	
3	15.02.2007	Іваненко	1	3	3	1	3 000,00 грн.	3 000,00 грн.	-1	
4	15.02.2007	Петренко	4	3		1	4 000,00 грн.	4 000,00 грн.	-1	
9	15.02.2007	Іваненко	1	3	3	1	3 000,00 грн.	3 000,00 грн.	-1	
11	15.02.2007	Іваненко	1	3	3	1	3 000,00 грн.	3 000,00 грн.	-1	

Висновки

У статті описано предметну область сфери туризму та обґрунтовано необхідність інтеграції даних для прийняття рішень щодо управління туристичним бізнесом загалом. Побудовано концептуальну модель предметної області.

Описано процес інтеграції даних та розроблено засоби автоматизованого завантаження даних із неоднорідних джерел інформації.

1. Inmon W., Building the Data Warehouse. John Willey & Sons, New York, 1992. 2. Ralph Kimball, Help for Hierarchies. DBMS, 1998 September. 3. Дрюэк К. (Katherine Drewek). Хранилища данных: сходство и различия подходов Билла Инмона и Ральфа Кимболла. – 2005. – [Електронний ресурс]: http://www.b-eye-network.com/view/743. 4. Кузнецов С. От баз данных к пространствам данных: новая абстракция управления информацией. – 2006. – [Електронний ресурс]: http://www.citforum.ru/ database/articles/from_db_to_ds. 5. Шаховська Н.Б. Простори даних туристичної сфери // Д.І. Угрин. –

Відбір і обробка інформації. Вісник ФМІ, Львів; № 13, 2008. – С.101–110. 6. Шаховська Н.Б. Інтеграція, консолідація та федералізація даних для інформаційних технологій туристичного бізнесу // Д.І. Угрин. – Відбір і обробка інформації. ФМІ, № 16, 2008. – С.98–108. 7. Шаховська Н.Б. Альтернативні рішення ситуацій опрацювання даних з розрізнених джерел в просторах даних туризму // Д.І. Угрин. – Інтернет-конференція РусНаука, [Електронний ресурс]: http://www. rusnauka. com/25_DN_2008/Matemathics/28431.doc.htm.

УДК 004.043

О.В. Шпортько
Рівненський державний гуманітарний університет,
кафедра інформатики та прикладної математики

ОПТИМІЗАЦІЯ БЛОКІВ СТИСНУТИХ ДАНИХ У ГРАФІЧНОМУ ФОРМАТІ PNG

Запропоновано алгоритм генерування альтернативних стиснутих блоків, вибору найкоротшого стиснутого блока з альтернативних та ітеративного зменшення його розмірів для покращання компресії зображень у форматі PNG. Як показують експерименти, реалізація цього алгоритму дає змогу підвищити коефіцієнт стиснення переважної більшості зображень на 2 – 6 %.

Ключові слова – компресія зображень, формат PNG, коефіцієнт стиснення.

This algorithm of the generation of alternative compressed blocks, choice of the shortest compressed block from alternative one and iterative diminishing of its size for the improvement of compression images in the format of PNG is offered in the following article. As the experiments show, the realization of this algorithm allows to raise the compression factor of the majority of images in 2 – 6 percent.

Keywords – compression images, format PNG, compression factor.

Вступ

Формат графічних файлів PNG був створений 1 жовтня 1996 року для ефективного збереження растрових зображень без втрат після того, як компанія Unisys почала вимагати плату за використання формату GIF [1]. Сьогодні дизайнери та розробники Web-сайтів найчастіше зберігають фотореалістичні зображення у форматі JPEG, а дискретно-тонові – у форматі PNG. Крім цього, формат PNG найчастіше використовується для зберігання зображень, де втрати недопустимі (наприклад, для рентгенівських знімків). Саме тому цей формат підтримує більшість сучасних програм для перегляду і створення зображень. Проблема підвищення ефективності стиснення зображень у форматі PNG є актуальною сьогодні і буде актуальною в найближчому майбутньому, оскільки навіть наближення до її вирішення дасть змогу зменшити розміри відповідних файлів, що, своєю чергою, сприяє підвищенню ефективності використання дискового простору та прискорення завантаження з мережі. У цій статті описується один із способів часткового вирішення вказаної проблеми.

Принципи та оптимізація стиснення зображень у форматі PNG. Аналіз останніх досліджень. Постановка задачі

Будь-яке стиснення даних можливе за рахунок зменшення надлишковостей. Чим більше видів надлишковостей виявляє й опрацьовує компресор і чим краще він ці надлишковості усуває – тим