

3. *Лысенко И.Е.* Проектирование сенсорных и актюаторных элементов микросистемной техники. – Таганрог: Изд-во ТРТУ. 2005. – 103 с.
4. Норенков И.П. Основы теории и проектирования САПР. – М.: Высш. Шк., 1990. – 334 с.
5. Системы автоматизированного проектирования: Учеб. пособие для вузов: В 9 кн. / *И.П.Норенков*. Кн.1. принципы построения и структура. – М.: Высш.шк., 1986. – 127 с.
6. Microsoft Corporation. Проектирование и реализация баз данных Microsoft SQL Server 2000. / Пер. с англ – 3-е изд. – Издательско-торговый дом “Русская Редакция”; СПб.: Питер, 2006.- 512стр.: ил.
7. *Троелсен, Эндрю.* Язык программирования С# и платформа .NET 2.0, 3-е издание. : Пер.с англ. – М.: ООО “И.Д.Вильямс”, 2007 – 1168с.: ил.

Поступила 30.01.2009р.

УДК 004.652.4+004.827

Н.Б.Шаховська, Д.І.Угрин

ІНТЕГРАЦІЯ РОЗРІЗНЕНИХ ДАНИХ ПРО КЛІЄНТА З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АГЕНТА ВИЗНАЧЕННЯ СТРУКТУРИ ДЖЕРЕЛА

An order, methods and facilities of getting, concordance, integration of client information, creation of operative depositories of information and load of information, is in-process worked out in a central depository.

ВСТУП

Туристичний бізнес – це сучасна галузь, що динамічно розвивається. Впровадження спеціалізованих програмних продуктів, розроблених винятково для потреб туристичної галузі, компанії-розроблювачі почали вже понад п'ятнадцять років тому. Проте наявність великої кількості програмних продуктів не дає змоги здійснювати обмін даними між туроператорами, що б дозволило значно підвищити їхні прибутки, оскільки дані зберігаються у різних моделях, керуються різними операційними системами тощо. Технологія Customer Data Integration (CDI) дозволяє збільшити швидкість і точність ідентифікації клієнтів і не потребує при цьому створення єдиного сховища даних про клієнтів. Ця технологія інтеграції відрізняється від попередніх тим, що в ній використовується так званий "заснований на знаннях" підхід. В основі CDI є каталог даних про клієнтів та опис методів доступу до інформації про них.

1. АКТУАЛЬНІСТЬ РОБОТИ

Розроблені на сьогодні підходи інтеграції даних за своєю функціональністю поділяються на два типи: інтеграції веб-застосувань та інтеграція на основі сховищ даних (з утворенням локального сховища даних) [1 – 3]. Проте специфіка роботи з даними про клієнтів, а саме:

- наявність великої кількості джерел даних, інформація у яких є різної структури, не виключаються протиріччя та суперечливість,
- наявність великої кількості моделей зберігання джерел даних (реляційні бази даних (РБД), сховища даних (СД), структуровані текстові файли, електронні таблиці, статичні та динамічні веб-сайти тощо),
- відсутність стандартів назв характеристик клієнтів,
- ієрархічне впорядкування об'єктів туристичної сфери та агрегування інформації у ході передачі її до верхніх рівнів ієрархії,

вказує на те, що для отримання інформації від усіх об'єктів туристичної сфери необхідно поєднати обидва типи інтеграції.

Отже, задача інтеграції розрізненої інформації, зокрема клієнтської, є актуальною, оскільки:

- на сьогодні немає консолідованого розв'язання задачі інтеграції даних,
- зростання кількості виробників та споживачів інформації призводить до лавинного зростання кількості інформації,
- відсутність стандартів назв об'єктів не дозволяє розробникам на рівні створення програмного забезпечення узгодити єдиний каталог назв, забезпечивши цим узгодженість схем даних.

2 ПОСТАНОВКА ЗАДАЧІ

На сьогодні більшість туристичних організацій використовують інформаційні технології для зберігання, опрацювання та аналізу інформації. Сфера застосування інформаційних технологій в туризмі подана на рис. 1.

Для керування туристичною сферою певного регіону (області, країни тощо) необхідно забезпечити інтеграцію інформації з туристичних організацій (агенцій).

Центри управління туризмом (Міністерство курортів і туризму, Міністерство у справі культури та молоді тощо) опрацьовують та аналізують узагальнену інформацію, отриману від об'єктів туризму у вигляді стандартних статистичних звітів [1]. Це призводить до того, що керівництво сферою здійснюється тільки за подіями, що настали (наприклад, при виявленні проблеми на рівні району), і нема можливості попереджати настання певної негативної події. Отримання деталізованої інформації, визначення споживачів певних туристичних послуг (клієнтів) за запитом дозволило б значно покращити якість прийнятих рішень з керування туристичною сферою.

Основна складність полягає у тому, що більшість компаній не може скласти клієнтський сервісний каталог, оскільки інформаційні системи

ізолювані й незалежні. Щоб ефективно вирішити проблему інтеграції розрізненої інформації з розрізнених інформаційних систем, CDI-система повинна забезпечити єдине й консолідоване представлення клієнтів. Побудова клієнтського пакета даних вимагає як перенесення значного обсягу інформації з оперативних систем, так і перевірки транзакційних даних і керування клієнтськими відомостями з точок зв'язку із клієнтами [3].

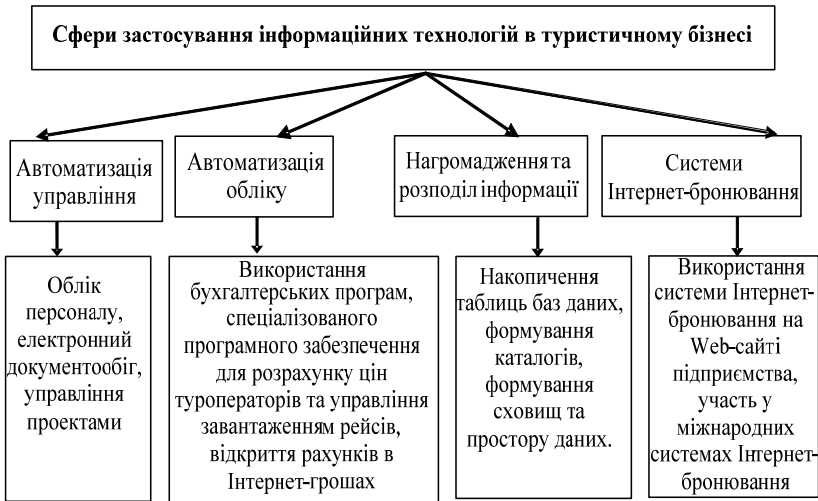


Рис 1. Сфери застосування інформаційних технологій в туристичному бізнесі

Окрім розрізненості назв та ідентифікаторів виникають проблеми некоректності та суперечливості даних у різних системах, а також застарілих даних, які не варто враховувати у подальшому аналізі.

Метою роботи є забезпечення інтеграції розрізненої інформації про клієнтів з метою її подальшого опрацювання та аналізу. Для цього використаємо модифіковану CDI-технологію, що дозволить:

- вирішити проблему дублювання і двозначності даних;
- збільшити кількість інформації про клієнтів, оскільки можливе доповнення інформації з різних джерел;
- забезпечити витягання клієнтських даних і створення інтегрованої клієнтської бази.

3. ОСНОВНИЙ МАТЕРІАЛ

Інтеграція клієнтських даних забезпечує інфраструктуру для перетворення вихідних даних у корпоративні інформаційні ресурси. Мета – уніфікований репозитарій (каталог) даних, або клієнтський концентратор даних.

Основою CDI є компоненти керування клієнтськими даними (рис. 2).

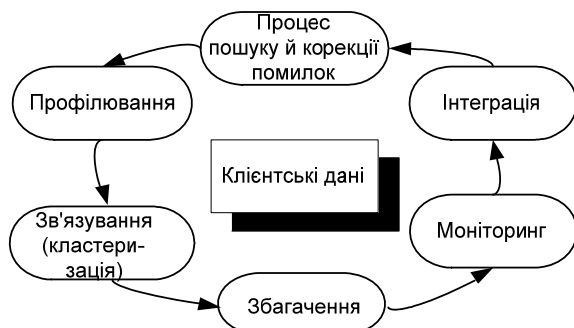


Рис. 2. Схема компонентів інтеграції клієнтських даних

Розглянемо ці компоненти детальніше.

1. *Профілювання даних* — перший крок для будь-якого проекту, що дозволяє виявити, проаналізувати та задокументувати всі джерела, що містять клієнтську інформацію. Цей етап також включає звіти по частоті оновлення, описують характеристики даних, зв'язки між таблицями, виявлення бізнес-правил.

Для ідентифікації та роботи з неоднорідними колекціями можна використовувати глобальну схему імен (Uniform Resource Identifiers – *URI*) як механізм посилань на глобальні константи, щодо яких є деяка угода між декількома постачальниками даних. Проте, якщо інформаційні системи вже створені, ініфікацію даних забезпечити важко. Тому необхідно розробити засоби, які б аналізували схеми даних, визначали характеристики-синоніми та вказували б на розбіжності схем даних. Результатом профілювання є створення каталогу клієнтів, інформація яких зберігається у локальному сховищі даних. Каталог містить не тільки описову інформацію (тобто виконує роль метаданих), але й зберігає для кожного учасника схему даних, статистичні дані, швидкість зміни, точність, можливості відповідей на запити, інформацію про власника і дані про політику доступу і підтримку конфіденційності.

Зв'язок між каталогом клієнтів *CG*, середовищем керування моделями *EM* та локальним сховищем *ODW* можна подати як функцію інтелектуального агента:

$$EM(CG) \xrightarrow{Agent} ODW.$$

Інтелектуальний агент запам'ятовує стан навколишнього середовища (структури даних опису клієнтів) та вміє розпізнати його стан на основі накопиченої інформації.

Агент *Agent* подається сімкою об'єктів:

$$Agent = \langle CG, EM, Dic, Experience_Base, Solver, Effector \rangle,$$

де *CG* – ідентифікатор внутрішнього стану агента (інформація про джерела, що вже є у локальному сховищі даних);

EM – компонента агента, що відповідає за сприйняття середовища (сенсор), тобто середовище керування моделями;

Dic – база знань, що містить знання агента про власні можливості (словник назв-синонімів, який дозволяє встановлювати подібності з різними назвами, але однаковим змістом);

Experience_Base – база накопиченого досвіду агента, що містить “історію” впливів на агента з боку середовища й відповідної їм реакції агента

$$(Experience_Base = \sigma_{evdate=Date()}(Dic));$$

Solver – компонента, що відповідає за навчання агента (подає список розбіжностей схем даних, які виявив агент);

Effector – компонента, яка відповідає за дії агента (формування запиту по декількох джерелах, приведення результатів запитів по джерелах до єдиної структури, відмова у запиті).

В основі інтелектуального агента лежить інформація про джерела даних про клієнтів, які вже є у просторі. Його задачею є порівняння структур даних джерела даних, що входить у простір клієнтських данх, з структурами даних джерел, що вже є у просторі, та визначення різниці. Це дозволить автоматизувати формування запитів, що виконуватимуться у просторі даних. Чим більше джерел здатний «розрізнити» інтелектуальний агент, тим точніше буде інформація в *ODW* і тим ефективніше можна буде проводити процедури інтеграції, пошуку та опрацювання даних про клієнтів. Для автоматизації процедури профілювання даних використано схему даних, подані на рис. 3. Відношення *tbl_First* та *tbl_Second* містять, відповідно, інформацію про структуру першого джерела даних та другого джерела.

tbl_First		tbl_Second		tbl_Paths	
PK	<u>fld_number</u>	PK	<u>fld_number</u>	PK	<u>id</u>
	tbl_name		tbl_name		
	fld_name		fld_name		
	fld_tp		fld_tp		
	fld_sz		fld_sz		
	fld_def_value		fld_def_value		
	fld_reqv		fld_reqv		
	fld_empty		fld_empty		
	fld_ind		fld_ind		
	fld_prim		fld_prim		
	fld_uniq		fld_uniq		
	fld_source		fld_source		pth

Рис. 3. Схема відношень бази знань інтелектуального агента

Tbl_Path зберігає шляхи або сторінки усіх джерел даних туристичного бізнесу. Принцип роботи агента полягає у попарному переборі усіх джерел, завантаженні інформації про них у відношення бази знань та їх порівнянні. Покажемо результат роботи агента на прикладі. Нехай є дві бази даних туристичних організацій, схеми даних яких подані на рис 4.

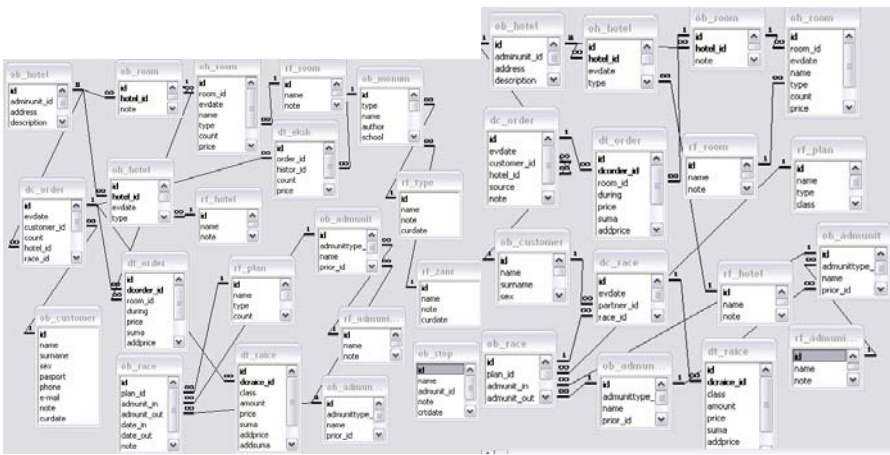


Рис. 4. Схема бази даних а) туроператора; б) туристичної бази

Результат порівняння двох джерел подано на рис. 5. Тут показано таблиці, які відсутні у першій або другій базі даних, а також відмінності у таблицях з однаковими назвами.

2. *Процес пошуку й корекції помилок.* Часто відомості бувають невірними, виходять зі заданого діапазону, несумісні і не погоджені з існуючими бізнес-правилами. Тому для попередньо отриманих і збережених у оперативному сховищі даних необхідно розробити систему норм і критеріїв перетворення їх до форми, придатної для збереження у сховищі даних. Приклади критеріїв та вимог:

ТАБЛИЦІ, яких нема в одній з баз:					
dc_race		нема у першій базі			
dt_eksk		нема у другій базі			
ob_monum		нема у другій базі			
ob_stop		нема у другій базі			
oh_race		нема у другій базі			
rf_type		нема у другій базі			
rf_zanr		нема у другій базі			
ПОЛЯ, яких нема в одній з баз:					
dc_order	count	відсутнє у другій базі			
dc_order	race_id	відсутнє у другій базі			
dc_order	realiz	відсутнє у другій базі			
dc_order	recr_name	відсутнє у другій базі			
dc_order	type_recr	відсутнє у другій базі			
rf_plan	count	відсутнє у другій базі			

Таблиця	Поле	База	Властивості	Розмір поля
dt_race			Тип поля	
	dcrace_id	1		
		2		

Рис. 5. Результат роботи інтелектуального агента

- обмеження часового періоду фактів "від ... – до ...", наприклад, вибір даних за останній місяць;
- обов'язкове значення виміру чи показника факту;

- входження значень виміру у попередньо визначену множину чи область визначення.

Алгоритм аналізу та верифікації вхідних даних із зовнішніх джерел подано на рис. 6.

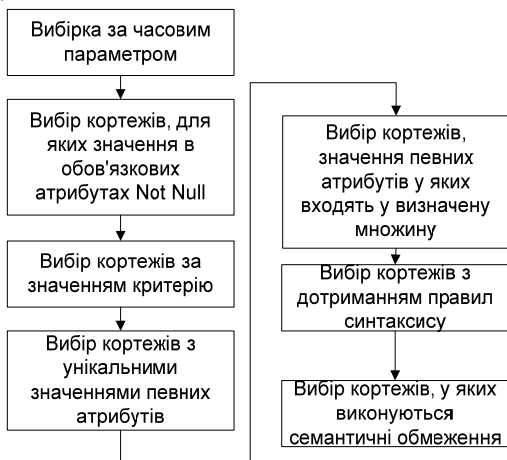


Рис. 6. Блок-схема аналізу та верифікації локального сховища

Процес пошуку та корекції помилок вирішує ряд завдань:

- у яких точках збору даних є клієнтська інформація;
- як зберігаються, оцінюються та перевіряються клієнтські відомості;
- які джерела містять кращі дані;
- яким чином можна інтегрувати дані з різних джерел;
- чином можна інтегрувати дані з різних джерел;
- яка інформація про клієнта необхідна та де її можна знайти.

Потім необхідно оцінити сформовану ситуацію та з'ясувати наступне:

- які критерії низької якості даних;
- яка інформація пов'язана з неповнотою даних;
- як можна забезпечити узгодженість та уніфікованість стандартів (поповнити словник Dic);
- чи вкладаються дані у встановлені бізнес-параметри;
- чи можна консолідувати інформацію з різних бізнес-підрозділів;
- які види консолідації необхідні.

3. *Інтеграція даних* – збір даних з територіально віддалених або різноплатформених джерел даних в єдине локальне сховище даних з метою їх подальшого опрацювання та аналізу після визначення попередньо їх структури та узгодження розрізних структур між собою:

$$ODW.Fact \xrightarrow{\text{consolid}} Agent(\text{Source}_1.Fact) \cup \dots \cup Agent(\text{Source}_n.Fact)$$

Пошук інформації про клієнта у різних джерелах вимагає, перш за все, знищення всіх записів-дублікатів у результуючому запиті. Крім того,

важливо забезпечити зв'язок між джерелами даних, щоб одержати агреговану інформацію про взаємозв'язки клієнтів [4].

З метою аналізу та верифікації отриманого сховища даних необхідно перевірити та обґрунтувати працездатність сховища даних. Для цього необхідно:

- перевірити наявність усіх необхідних значень у стовпчиках таблиці фактів за таким алгоритмом:

- 1) визначити кількість даних у джерелах,
- 2) обчислити суму отриманих кількостей записів,
- 3) визначити кількість введених у таблицю сховища записів,
- 4) якщо суми співпадають, то інтеграція пройшла без перешкод;

- перевірити таблицю фактів на наявність "підвішених" рядків, що виникають за наявності в ній значень кодів, яких немає у таблицях інших джерел;

- перевірити таблицю фактів на наявність повторюваних рядків:
 - визначаємо записи, що повторюються,
 - визначені записи знищуємо, залишивши одну з них;

- виконати тестові запити на отримання контрольних підсумкових значень, таких як підрахунок кількості фактів за певним виміром, сумарного значення за певним виміром, тощо і звірити отримані результати з фактичними;

- на підставі отриманих результатів зробити висновки про придатність чи непридатність побудованого сховища даних до застосування.

4. *Зв'язування (кластеризація)* полягає у пошуку залежностей між клієнтськими даними. Для цього використовуємо класифікаційні правила. *Класифікаційним правилом* назвемо залежність між підмножинами атрибутів X та Y , яка зустрічається у тестовому наборі відношення r із ступенем відповідності (довіри) s , при якій

$$(X = x) \rightarrow (Y = y) .$$

Будується класифікаційне правило на основі навчального набору даних у відношенні r' , де значення міток класу (значення підмножини атрибутів Y) відомі. Класифікаційне правило будується для схеми відношення і тому не буде залежати від нових кортежів, що надходять у відношення сховища даних.

4. *Збагачення даних* забезпечує повніше (порівняно з використанням неопрацьованих вхідних даних) використання консолідованої інформації. Метою збагачення даних є виявлення джерел даних, у яких відсутня частина інформації про клієнта та її доповнення [5].

$$Tir: (Source1.Fact = \overset{INSERT}{ODW.Fact}, \dots, Source n.Fact = \overset{INSERT}{ODWD.Fact}).$$

5. *Моніторинг даних* — завершальна стадія будь-якого CDI-проекту. Вона полягає у пошуку та коректуванні проблем у джерелах даних, виявленні процесів, що призвели до погіршення якості інформації. Результатом роботи

моніторинга даних є визначення:

- нових клієнтів у джерелі даних, інформація про яких ще не описана у каталозі та локальному сховищі,
- нових клієнтів у джерелі, інформація про яких є у каталозі та локальному сховищі, та повноту цієї інформації.

5. ВИСНОВКИ

Підхід інтеграції даних про клієнтів на основі CDI-технології підвищує здатність туристичних організацій до ідентифікації своїх клієнтів на основі вхідних даних. Опис клієнтів у єдиному каталозі даних дозволяє визначати власників інформації про клієнтів, а також використовувати цю інформацію для аналізу та прийняття рішень. *Науковою новизною* статті є визначення основних кроків інтеграції клієнтської інформації та розроблення інтелектуального агента визначення структури джерела. *Практичною цінністю* є розроблення схеми сервісу інтеграції клієнтських даних.

1. *Сенин В.С.* Организация международного туризма: Учебник. – М.: Финансы и статистика, 1999. – 400 с.
2. Интеграция данных и Хранилища. – 2005, Електронне джерело: <http://citcity.ru/12101/>
3. Интеграция корпоративной информации: новое направление. – 2005, Електронне джерело: <http://citcity.ru/11155/>
4. *Шаховська Н.Б.* Простір даних області наукових досліджень // Моделювання та інформаційні технології. – ІПМЕ НАН України ім. Пухова. – Київ, № 45, С.132-140
5. *Шаховська Н.Б.* Особливості моделювання просторів даних // Комп'ютерна інженерія та інформаційні технології. Вісник НУ "Львівська політехніка", № 608, с. 145-154, 2008.

Поступила 26.02.2009р.

УДК 621.3

Л.С. Сікора, д.т.н., НУ «ЛП», Львів, Ю.Г. Міюшкович, асп., НУ «ЛП», Львів, І.О. Малець, н.с., Н.К. Лиса, н.с., В.М. Сікора, н.с., Р.М. Владика, гол. інж. ЦСД «ЕБТЕС», Львів.

МЕТОДИ І МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ СТВОРЕННЯ МУЛЬТИМЕДІЙНИХ КОМПЛЕКСІВ ТЕРМІНАЛЬНОГО ДІАЛОГОВОГО УПРАВЛІННЯ

Анотація. Розглянута інформаційна технологія створення діалогових систем управління з використанням методів і моделей штучного інтелекту для прийняття оперативних рішень в режимі діалогу.

© Л.С.Сікора, Ю.Г.Міюшкович, І.О.Малець, Н.К.Лиса, В.М.Сікора, Р.М.Владика 173