

АРХІТЕКТУРА ТА КОМПОНЕНТИ КОМП'ЮТЕРНИХ СИСТЕМ

УДК 004.652

А. Берко

Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

МОДЕЛІ ТА МЕТОДИ ФОРМУВАННЯ ІНТЕГРОВАНОГО КОНТЕНТУ СИСТЕМ ЕЛЕКТРОННОГО БІЗНЕСУ

О Берко А., 2008

Розглянуто способи і можливості створення інтегрованих структур даних для зберігання інформаційного наповнення в гетерогенних системах електронного контент-бізнесу. Запропоновано формальні моделі та методи інтеграції структури, синтаксису та семантики різнорідних даних в системах електронного бізнесу.

Some ways and possibilities of integrated data model development for information storage of heterogeneous electronic content– business systems are considered in this paper. Formal models and methods of structure, syntax and semantic integration of data were proposed.

Вступ

Системи електронного бізнесу – це Web-системи поширення різноманітних товарів та послуг на основі мережових технологій. Сьогодні цей напрям є одним з найперспективніших підходів до обслуговування клієнтів як у мережі Internet, так і в інших інформаційних мережах. Значна кількість сучасних систем електронного бізнесу орієнтована на забезпечення користувачів інформаційними продуктами і сервісами. Особливістю таких систем є те, що об'єктом їх застосування, кінцевим результатом діяльності та основним елементом функціонування є інформаційний ресурс (контент). Такий вид підприємництва у глобальній мережі називають електронним контент-бізнесом [2]. Вимоги до обсягів та якості інформаційних послуг в середовищі Internet невідомо зростають, а їх реалізація стає все складнішою внаслідок неможливості простого механічного розширення обсягів продуктів та сервісів, які отримує споживач. Методологія й засоби реінжинірингу інформаційних процесів у галузі електронного контент-бізнесу має низку особливостей. Одним з напрямів реінжинірингу таких систем є інтеграція корпоративних ресурсів. У загальному випадку предметом інтеграції є бізнес-процеси, платформи, прикладні задачі та інформаційні ресурси систем електронного бізнесу. В цій статті розглянуто принципові аспекти інтеграції інформаційного наповнення систем електронного контент-бізнесу.

Як правило, інформаційний ресурс систем контент-бізнесу має гетерогенний характер, що обумовлено різноманітністю потреб і пропозицій у цій сфері діяльності. Ця особливість потребує поєднання в одному середовищі елементів баз даних, документів, web-сторінок, мультимедійних даних тощо. Отже, очевидною стає проблема інтеграції різнорідного контенту в єдиному середовищі зберігання, опрацювання та застосування. Інтеграція інформаційних ресурсів передбачає їх об'єднання, при якому спільне використання є простішим і ефективнішим, ніж локальне застосування кожної складової.

Аналіз останніх публікацій і досліджень

Системи та засоби електронного бізнесу часто ототожнюють з системами електронної комерції. Насправді електронна комерція є лише одним з напрямів електронного бізнесу, в якому застосовано класичну схему торгівлі – "товар-гроші" [8]. Електронний бізнес загалом охоплює значно ширше коло видів підприємницької діяльності в глобальних мережах і передбачає складніші схеми отримання прибутку (реклама, оплата трафіка, агентські відсотки тощо). Основні економічні принципи визначення продуктів інформаційних технологій як товару викладено в [6]. В роботі [6] подано основні характеристики інформаційного товару, а також визначено роль та місце інформації у системі виробничих відносин. Технологічні принципи формування інформаційних товарів та послуг, а також концепцію "*information on demand*", яку підтримує корпорація IBM, викладено в [7].

Основні принципи та напрями вирішення проблем кількості та якості інформаційних ресурсів було визначено у підсумкових матеріалах Лоуельської Міжнародної конференції з проблем баз даних та опрацювання інформації [7]. Одним з основних способів вирішення проблем опрацювання великих обсягів різнорідних даних було визначено їх інтеграцію. Концептуальні засади вирішення проблем інтеграції даних розглянуто зокрема в [3, 5]. Інтеграція даних охоплює широке коло питань, методів та засобів, що стосуються як технологічних, так і методологічних аспектів, які викладено у [3, 7, 8]. Зокрема, монографія Д. Ланде [8] присвячена проблемам інтеграції інформаційних потоків в Internet-системах електронної комерції. Особливі можливості застосування інтегрованих інформаційних ресурсів у електронному бізнесі відкриває напрям Internet-систем, побудованих за сервіс-орієнтованою архітектурою [7].

Мета та завдання досліджень

Цілі статті. Основними цілями, які поставлено в цій статті, є такі:

1. Узагальнення та класифікацій підходів до інтеграції інформації на різних рівнях її подання та сприйняття;
2. Розроблення формальних моделей і методів інтеграції синтаксису, структури та семантики даних у Web-системах електронного контент-бізнесу;
3. Визначення порядку та місця застосування методів і моделей інтеграції у процесах проектування структури інтегрованого контенту та управління ним.

Невирішені проблеми. Основні проблеми виникають насамперед через неоднорідність структури, форми та змісту елементів інформаційного ресурсу систем електронного контент-бізнесу. За оцінками експертів [7], лише близько 20 відсотків інформаційного ресурсу в сучасних Інтернет-системах зберігають у базах даних як структуровану інформацію, решту становлять дані, подані у різноманітних форматах без попередньо визначеної чіткої структури – так звані слабкоструктуровані дані [9]. Різнорідність форм та структури контенту, своєю чергою, викликає необхідність розроблення методів і технологій, які забезпечують спільне, узгоджене опрацювання і застосування таких неоднорідних даних.

Особливостями слабкоструктурованих даних є зокрема такі [9]:

- ◆ відсутність фіксованої схеми;
- ◆ відсутність чіткого розмежування між даними та схемою;
- ◆ дані не є строго типізованими;
- ◆ зміна схеми не викликає обов'язкової зміни даних;
- ◆ обсяг даних є сумірним з обсягом схеми;
- ◆ схема є описовим елементом, а не директивним і може бути отримана безпосередньо з даних;
- ◆ при формулювання запитів застосування схеми не є обов'язковим.

У загальному випадку, до категорії слабкоструктурованих даних відносять напівструктуровані дані, структура яких є нерегулярною або неоднозначною; самоструктуровані дані, структуру яких визначають за допомогою них самих та дані, структуру яких попередньо не встановлено, а її визначають в процесі застосування самих даних. Зберігають такі дані у формі текстових файлів, файлів документів, XML-документів, Web-сторінок, графічних і мультимедійних файлів тощо.

Неоднорідність інформаційного наповнення є характерною особливістю Web-систем. Оскільки Web-технології та Web-засоби сьогодні є невід'ємною частиною будь-якого масштабного інформаційного проекту, то проблеми інтеграції різноманітного контенту, в якому є як структуровані реляційні дані, так і дані слабкоструктурованого типу, є важливими та актуальними [7].

Основні результати досліджень

Метою інтеграції даних є отримання єдиної і цілісної картини корпоративної бізнес-інформації на основі різноманітних за формою та походженням вхідних наборів даних, отриманих з різних джерел. Концепція інтеграції даних є відомою достатньо давно і в різні часи була реалізована у формі вирішень, актуальних у свій час – обчислювальних ресурсів загального користування, корпоративних мереж, розподілених баз даних, сховищ даних тощо. Інтеграція даних є складним, багатогранним та об'ємним процесом, який передбачає зокрема реалізацію процедур видобування, перетворення і завантаження (extraction, transformation, loading, скорочено ETL) даних з різних систем та джерел в єдиний інтегрований набір даних, призначений для опрацювання і аналізу (підготовки звітності), а також узгодження методів та принципів подання, опрацювання та інтерпретації даних. Сховища, вітрини даних, розподілені та інтегровані корпоративні бази даних, інформаційне наповнення Web-систем є найтиповішими прикладами таких наборів, а інструменти ETL – це компоненти процесу інтеграції даних.

Згідно з [5] основними об'єктами у процесах інтеграції даних є глобальна схема інтегрованих даних, вхідна схема джерела даних і деяке відображення, що дає змогу перейти від способу зображення даних у вхідній множині джерел до способу їх глобального зображення в інтегрованому наборі даних. Загальну схему процесів інтеграції даних в гетерогенних системах зображено на рис 1.

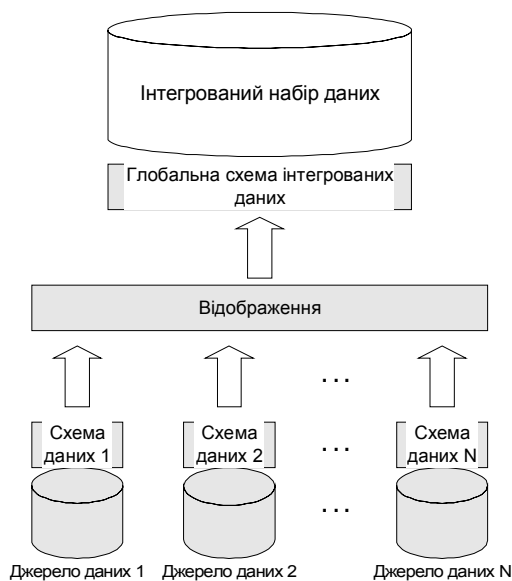


Рис 1. Загальна схема процесів інтеграції даних

Формально модель інтеграції даних I можна подати як трійку

$$I = \langle \Gamma, \Sigma, M \rangle, \quad (1)$$

де Γ – глобальна схема інтегрованих даних, описана в термінах деякої мови L_Γ над алфавітом A_Γ , Σ – вхідна схема джерела даних, описана в термінах мови L_Σ над алфавітом A_Σ , M – відображення, задане у вигляді відповідностей

$$Q_\Sigma \rightarrow Q_\Gamma; Q_\Gamma \rightarrow Q_\Sigma, \quad (2)$$

у яких – Q_Σ і Q_Γ вирази однакової розмірності, визначені відповідно над вхідною схемою та глобальною схемою [3].

У результаті процес інтеграції даних повинен забезпечити релевантність виразів, запитів та операцій над даними в інтегрованому наборі до відповідних понять у вхідних джерелах даних. Основні шляхи розв'язання таких проблем на формальному рівні опубліковано зокрема в [1, 3].

Концептуально дані можна розглядати як деяку формальну мову, яку застосовують для позначення множини понять у середовищі інформаційної системи. Обов'язковими і невід'ємними властивостями даних є синтаксис, семантика та структура. В загальному випадку засоби визначення довільного набору даних D утворюють деяку формальну систему [10] вигляду

$$D = \langle V, G, S, H \rangle, \quad (3)$$

де V – множина значень, якими зображають множину понять деякої предметної області, G – синтаксис даних, S – структура даних, H – семантика даних.

Процес інтеграції даних – це послідовність дій, що передбачає з їх узгодження, перетворення, об'єднання, фільтрування і має на меті утворення кінцевого набору даних D на основі множини початкових наборів:

$$D = I(D_1, D_2, \dots, D_N), \quad (4)$$

де I – оператор інтеграції даних, $D_1, D_2, \dots, D_N$ – множина вхідних початкових наборів даних, які у загальному випадку можуть містити значення, що повторюються, тобто

$$D_1 \cap D_2 \cap \dots \cap D_N \neq \emptyset. \quad (5)$$

Інтеграція множини різнорідних наборів даних в єдине ціле не є їх простим механічним об'єднанням. Цей процес передбачає низку дій, пов'язаних з їх перетворенням і утворенням нових значень на основі початкових. Враховуючи модель даних, яка ґрунтується на визначенні їх синтаксису, семантики та структури [2]

$$D = \langle V, G, S, H \rangle, \quad (6)$$

формальне визначення процесу інтеграції можна звести до дії над цими компонентами, замінивши опис

$$D = I(D_1, D_2, \dots, D_N) \quad (7)$$

на деталізований опис всіх складових визначення даних

$$\langle V, G, S, H \rangle = I(\langle V_1, G_1, S_1, H_1 \rangle, \langle V_2, G_2, S_2, H_2 \rangle, \dots, \langle V_N, G_N, S_N, H_N \rangle), \quad (8)$$

де $\langle V_i, G_i, S_i, H_i \rangle, i=1,2, \dots, N$ – деталізоване формальне зображення i -го набору даних.

У такий спосіб проблему інтеграції даних можна декомпонувати на окремі проблеми інтеграції значень даних, інтеграції синтаксису, інтеграції структури та інтеграції семантики. Загальний оператор інтеграції даних I при цьому може бути подано через комбінацію

$$I = \langle I^V, I^G, I^S, I^H \rangle, \quad (9)$$

де I^V – оператор інтеграції значень, I^G – оператор інтеграції синтаксису, I^S – оператор інтеграції структури даних, I^H – оператор інтеграції семантики. Процес інтеграції при цьому буде декомпоновано на відповідні підпроцеси, які можна описати формальною схемою вигляду

$$\langle V, G, S, H \rangle = \langle I^V(V_1, V_2, \dots, V_N), I^G(G_1, G_2, \dots, G_N), I^S(S_1, S_2, \dots, S_N), I^H(H_1, H_2, \dots, H_N) \rangle. \quad (10)$$

Взаємне співвідношення цих процесів та їхня класифікація за рівнями зображені на рис. 2.



Рис. 2. Рівні інтеграції даних

Згідно з такою схемою, кожен наступний рівень інтеграції ґрунтується на результатах попереднього і не може бути реалізованим без них. Так, семантична інтеграція даних є можливою

лише після інтеграції їх структури, котра, своєю чергою, потребує побудови інтегрованого синтаксису, який визначає способи зображення даних в інтегрованому наборі.

Моделі та методи інтеграції синтаксису даних. Проблема інтеграції синтаксису даних є базовою відносно інтеграції інших складових їх загального опису. Вирішення проблем побудови узагальненої структури та семантики даних є можливим лише на основі єдиної системи їх позначення. Поняття синтаксису даних саме по собі є комплексним і враховує різні аспекти їх зображення у документах, базах даних, сховищах та джерелах даних. Враховуючи це, синтаксис даних можна подати як деяку формальну систему

$$G = \langle A, T, R \rangle, \quad (11)$$

де A – алфавіт, T – множина типів даних, R – множина синтаксичних обмежень.

Алфавіт визначає множину символів, які застосовують для зображення значень даних у визначеному середовищі. Як правило, алфавіт складається з літер, цифр, спеціальних та службових символів. Однак на визначення алфавіту впливають зокрема такі фактори, як локалізація середовища опрацювання даних до мови користувачів, характер завдань, для вирішення яких застосовують дані, особливості процесів їх збереження, передавання та опрацювання, специфіка змісту різноманітних значень даних. Поряд із традиційними засобами позначення даних в сучасних системах широко застосовують графічні, звукові, мультимедійні та інші елементи для їх зображення і опрацювання, а також дані складних і комплексних типів, потокові та активні дані.

Поняття типу даних може бути визначено як результат класифікації їх значень за способами зображення та опрацювання. Сьогодні поряд із такими класичними типами, як числові, символні, логічні, дата-час тощо широко застосовують специфічні типи даних, які відображають особливості їх змісту, опрацювання та застосування. Це зокрема такі скалярні типи, як "гіперпосилання", "валюта", "об'єкт", "локатор" та інші, комплексні (агрегатні) типи – "масив", "запис", "множина", "XML-документ" тощо та типи даних, що визначає користувач. Така різноманітність типів даних, з одного боку, створює додаткові можливості щодо зображення та опрацювання інформаційного ресурсу, з іншого – значно ускладнює засоби підтримки середовища зберігання даних, процедури їх сумісного застосування, перетворення та об'єднання [1].

Обмеження як елемент синтаксису даних застосовують з метою створення значень максимально адекватних до понять та величин, які вони зображають. Обмеження синтаксису задають у вигляді кількісних показників, розмірності, форматів, шаблонів, правил формування значень, визначення підмножини допустимих символів тощо. Такі обмеження може бути визначено як на рівні систем і технологій підтримки даних, так і на рівні користувачів.

Отже, проблему інтеграції синтаксису даних можна декомпонувати на проблеми інтеграції алфавітів, інтеграції типів та інтеграції обмежень. Співвідношення цих проблем та результатів їх вирішення подано на рис. 3.

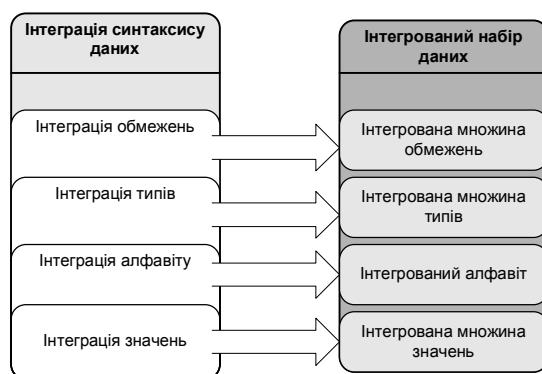


Рис. 3. Схема процесу інтеграції синтаксису даних

Згідно з такою схемою, синтаксис зображення значень інтегрованого набору G^I даних може бути подано як поєднання трьох складових

$$G^I = \langle A^I, T^I, R^I \rangle, \quad (12)$$

де $A^I = I^A(A_1, A_2, \dots, A_N)$ – алфавіт інтегрованого набору даних, утворений шляхом інтеграції алфавітів вхідних наборів даних $A_1, A_2, \dots, A_N$; $T^I = I^T(T_1, T_2, \dots, T_N)$ – множина типів даних, які застосовують в інтегрованому наборі, отримана як результат інтеграції типів даних, визначених для вхідних даних; $R^I = I^R(R_1, R_2, \dots, R_N)$ – множина обмежень інтегрованого набору даних, сформованих інтеграцією обмежень, які застосовано для вхідних даних; I^A, I^T, I^R – оператори інтеграції, відповідно, алфавітів, типів даних та обмежень.

Моделі та методи інтеграції структури неоднорідних даних. Однією з найістотніших проблем інтеграції є поєднання реляційних структур, які описують значення, що зберігають у базах даних та слабко структурованих даних, структура яких є нерегулярною, непостійною або попередньо невизначеною. Загальну модель інтегрованого контенту систем електронного бізнесу можна подати як деякий кортеж вигляду

$$C = \langle R, S_1, S_2, \dots, S_k, J_R, J_S, J_{RS} \rangle, \quad (13)$$

де C – схема інтегрованого інформаційного ресурсу, R – схема реляційної складової, яку утворюють структуровані дані, подані у вигляді таблиць бази даних; $S_1, S_2, \dots, S_k$ – схеми слабкоструктурованих складових різного типу, J_R – множина зв'язків між реляційними елементами, J_S – множина зв'язків між слабкоструктурованими елементами, J_{RS} – множина зв'язків між реляційними та слабкоструктурованими елементами.

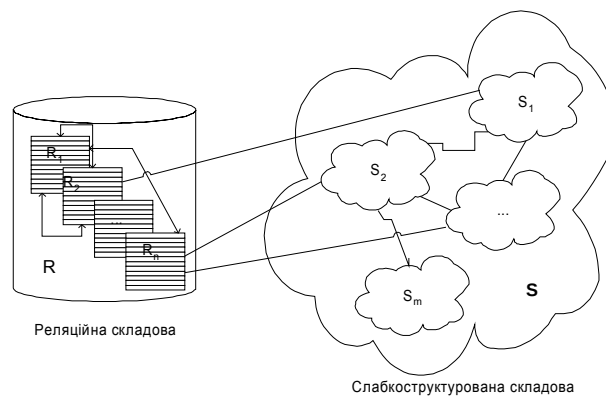


Рис. 4. Загальна структура інтегрованого контенту систем електронного бізнесу

Варіанти такої інтеграції реляційних та слабкоструктурованих даних структур може бути описано такими моделями.

1. Інтеграція структурованих даних з текстовими, просторовими, темпоральними, візуальними і мультимедійними даними. Сучасні системи управління базами даних, значною мірою забезпечують вирішення таких завдань за рахунок використання темпоральних (часових) типів даних, типів символьних та бінарних об'єктів та типу даних XML-документ. Значення цих типів інтегруються у таблиці і доповнюють перелік елементарних значень в описах сутностей та фактів. Структуру таблиць реляційної бази, в яких разом з реляційними зберігають слабкоструктуровані дані може бути описано як

$$R(A_1, A_2, \dots, A_k, X_1, X_2, \dots, X_m), \quad (14)$$

де $A_1, A_2, \dots, A_k$ – стовпчики таблиці, які зображають скалярні значення традиційних типів, $X_1, X_2, \dots, X_m$ – стовпчики, які зображають слабкоструктуровані значення.

2. Інтеграція баз даних та процедурних даних. Такий метод передбачає інтеграцію самих даних, що зберігаються у базах та множини об'єктних типів даних разом з методами, що інкапсулюють їх. У такому випадку кожен стовпчик таблиці може бути подано парою вигляду (A, M) , де A – стовпчик таблиці, M – набір методів, пов'язаних з цим стовпчиком. Схема такої таблиці може бути описана як вираз вигляду

$$R((A_1, M_1), (A_2, M_2), \dots, (A_k, M_k)). \quad (15)$$

3. Інтеграція в бази даних тригерів для опрацювання даних. Застосуванням тригерів можна реалізувати концепцію активної бази даних. Така база даних разом зі значеннями зберігає опис

певних правил та дій, які виконують при зміні стану бази даних. Таблиці, до складу яких входять традиційні дані та активні елементи, можна подати у вигляді такої моделі

$$R(A_1, A_2, \dots, A_k, T_1, T_2, \dots, T_m), \quad (16)$$

де $A_1, A_2, \dots, A_k$ – стовпчики таблиці, які зображають звичайні типізовані значення, $T_1, T_2, \dots, T_m$ – набір тригерів, які описують послідовність дій, пов'язаних з оновленнями даних таблиці.

4. Інтеграція статичних даних з потоками та чергами даних. Поточкові дані – це множина значень, яка не зберігається на носіях системи, а існує лише в момент застосування цих даних. Прикладом поточкових даних є моніторингова, біржова інформація, трансляція новин у встановленому форматі тощо. Потік даних формується в результаті виконання запитів, пересилання чи відбору даних. Черга – спеціальний тип потоку даних, в якому кожна одиниця має власну порядкову ознаку. Стан потоку даних S у певний момент часу t можна описати таким виразом

$$S_t = (s(R_t), s(X_t), s(S_t), s(W_t)), \quad (17)$$

де $s(R_t)$ – множина значень, отриманих як результат операцій вибору з таблиць реляційної бази даних, $s(X_t)$ – множина значень, отриманих як результат вибору з наборів слабкоструктурованих даних, $s(S_t)$ – множина значень, отриманих як результат вибору з інших потоків даних, $s(W_t)$ – множина значень отриманих з web-ресурсів.

Результатом інтеграції статичної та потокової складових є напівдинамічна структура, яка поєднує дані, збережені в базах даних та дані, сформовані у вигляді змінного в часі потоку.

Методи інтеграції семантики неоднорідних даних. Семантика є невід'ємною властивістю даних, що забезпечує їх змістовність та можливість застосування даних за їх призначенням. Загалом семантику визначає множина відповідностей між формальними позначеннями та реальними поняттями предметної області, що дає змогу однозначно інтерпретувати дані на різних стадіях роботи з ними. У сучасних інформаційних технологіях застосовують багато різноманітних засобів визначення семантики даних. Найпростішими засобами інтерпретації даних є, наприклад, описи стовпчиків таблиць, XML-тегів, розділів документа тощо. У великих системах, до яких належать системи електронного бізнесу, застосовують складніші засоби формування семантики даних, зокрема тезауруси (словники даних), метадані та онтології.

Інтеграція семантики даних передбачає формування єдиного змістового простору для сприйняття, інтерпретації та застосування даних незалежно від формату їх подання та структури. Найвідомішими способами семантичної інтеграції є такі [3]:

- контекстуальна семантична інтеграція;
- інтеграція на основі метаданих;
- інтеграція на основі онтологій.

Метод *контекстуальної семантичної інтеграції* ґрунтується на змістовому порівнянні тезаурусних термінів двох наборів даних. Тезаурусні терміни – це перелік ключових понять, пов'язаних з даними та внесеними до спеціального переліку – тезаурусу. Їх застосовують для опису семантичної відповідності між лексичними одиницями цього набору даних та конкретними значеннями з предметної області, наприклад, перелік стовпчиків таблиці бази даних, XML-теги, назви розділів та пунктів текстового документа тощо. Критерієм семантичної інтегрованості двох наборів даних у цьому випадку є функція контекстуальної семантичної віддалі між ними (CSD-функція) [4]. Вираховують значення CSD-функції так. Нехай D_i та D_j – множини тезаурусних термінів наборів даних D_i та D_j , відповідно, W_{ij} – множина тезаурусних термінів, які є семантично спільними для двох наборів, $|W_i|$, $|W_j|$, $|W_{ij}|$ – потужності відповідних множин. Тоді значення функції контекстуальної семантичної віддалі між наборами даних та вираховують як частку спільних значень у меншій за об'єм з множин тезаурусних термінів двох наборів даних

$$CSD(D_i, D_j) = \frac{|W_{ij}|}{\min(|W_i|, |W_j|)} \quad (18)$$

Схему семантичної інтеграції із застосуванням тезаурусних термінів та функції семантичної віддалі показано на рис. 5.

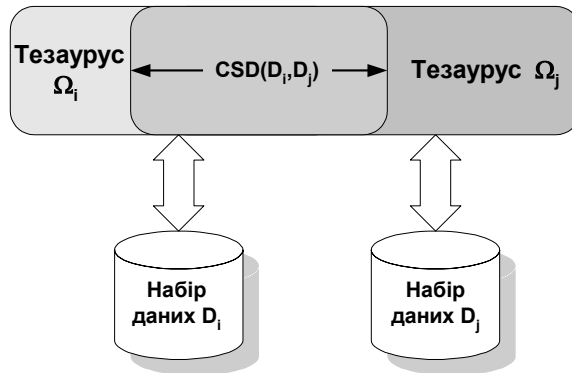


Рис. 5. Загальна схема контекстуальної семантичної інтеграції

Вважають[4], що семантична інтеграція двох наборів даних є можливою, якщо значення функції контекстуальної семантичної віддалі задовольняє умову (19)

$$CSD(D_i, D_j) \geq 0,8. \quad (19)$$

Семантична інтеграція на основі метаданих передбачає порівняння складу та змісту метаданих двох наборів. Найпоширенішою структурою метаданих є вимірна схема Захмана (рис. 6) [8], котра передбачає застосування шести категорій метаданих (вимірів), які описують об'єкти даних, суб'єкти, часові показники, способи розміщення, призначення та порядок застосування даних.

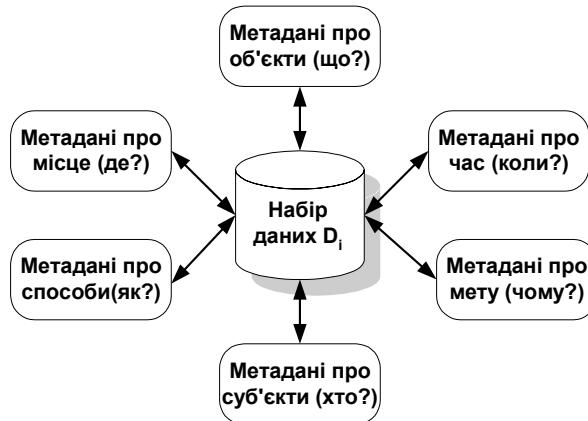


Рис. 6. Схема Захмана організації метаданих

Множину метаданих M_i деякого набору даних D_i можна подати як кортеж

$$M_i = \langle M_i^o, M_i^s, M_i^p, M_i^t, M_i^e, M_i^m \rangle, \quad (20)$$

де M_i^o – метадані про об'єкти, M_i^s – метадані про суб'єкти даних, M_i^p – метадані про розміщення даних, M_i^t – метадані про часові показники даних, M_i^e – метадані про мету застосування даних, M_i^m – метадані про порядок використання даних.

Збіг значень метаданих двох наборів за одним або більше вимірами застосовують як критерій семантичної інтеграції. Міра збігу залежить від конкретних випадків, але, як правило, її числовий вимір не повинен бути меншим за 80% [4]. Певним чином цей метод є схожий на метод контекстуальної семантичної інтеграції, але в цьому випадку замість тезаурусу застосовують інший інструментальний засіб – метадані.

Найперспективнішим на сьогодні підходом до інтеграції семантики даних є *інтеграція на основі онтологій*. Цей метод застосовує основні елементи двох попередніх – тезаурус та метадані, але є значно загальнішим за ці методи та враховує більше аспектів семантики даних [3]. Загалом онтологію розглядають як цілісну формалізовану специфікацію предметної області, яка має на меті забезпечити однакову інтерпретацію знань про цю предметну область на людському та комп'ютерному рівнях.

Формально онтологію подають як трійку вигляду

$$O = \langle X, R, F \rangle, \quad (20)$$

де X — скінченна множина понять предметної області (концептів), R — скінченна множина відношень між поняттями, F — скінченна множина функцій інтерпретації [4]. Оскільки об'єктом опису для такої онтології є дані, то її можна класифікувати як прикладну онтологію, засобом реалізації якої є метадані.

У процесі інтеграції даних для кожного вхідного набору даних D_i створюють власну онтологію $O_i = \langle X_i, R_i, F_i \rangle$, яка описує семантичний зв'язок елементів даних з поняттями предметної області, утворюючи цілісну структуру "дані–зміст". У такий спосіб проблему семантичної інтеграції даних можна звести до проблеми виявлення відповідностей та суперечностей між їх онтологіями.

Отже, два набори даних D_i та D_j вважають придатними до семантичної інтеграції, якщо для двох онтологій O_i та O_j , які відповідають цим наборам даних, виконуються правила:

- у множинах концептів X_i та X_j немає однакових понять, описаних різними способами;
- у множинах концептів X_i та X_j немає понять різного змісту, описаних однакою;
- з однотипними концептами різних онтологій не пов'язано обмежень, що не можуть бути виконані одночасно.
- у множинах зв'язків R_i та R_j немає зв'язків протилежного напрямку та змісту між однотипними концептами та зв'язків, що не можуть бути реалізованими одночасно;
- у множинах функцій інтерпретації F_i та F_j немає функцій, одночасна реалізація яких призведе до неоднозначності інтерпретацій.

Перевірити зазначену низку критеріїв семантичної інтеграції даних можна як на формальному, так і на інтуїтивному рівні, при цьому результат має бути однаковим. Виконання всієї множини вимог дає змогу зробити висновок про можливість інтеграції двох наборів даних на рівні їх змісту з отриманням семантично коректного результату.

Висновки

Процеси інтеграції даних мають достатньо широку сферу застосування. Це, зокрема, сховища даних різного типу та прямування, корпоративні ERP та CRM системи, інформаційні Web-системи, системи електронного бізнесу тощо. Інформаційні ресурси таких систем передбачають одночасне застосування значної кількості різноманітних за формою, структурою, змістом, способами подання і застосування даних. Однією з основних проблем інтеграції є створення та застосування єдиних правил і способів зображення таких різнорідних даних. Така проблема може бути вирішена за рахунок формування інтегрованого синтаксису даних, утвореного на основі синтаксичних методів і засобів вхідних даних.

У запропонованій роботі розглянуто низку питань, пов'язаних з одним із принципових аспектів інтеграції даних – інтеграції визначальних складових їх загального опису. В основу запропонованого вирішення покладено формальне подання даних як системи, властивостями якої є синтаксис, структура та семантика. Проаналізовано зокрема особливості створення інтегрованих способів зображення даних на етапі їх переміщення від джерел утворення до інтегрованого кінцевого набору. Синтаксис даних у роботі подається як деяка формальна система, що складається з окремих самостійних складових – алфавіту, набору типів даних, множини обмежень. Запропоновані способи поєднання гетерогенних за походженням даних, що ґрунтуються на гомогенізації, узгодженні або розширенні їх синтаксису. Такий підхід значно розширює можливості поєднання в єдиному середовищі значень різної природи та форми, зокрема числових, текстових, графічних, звукових чи мультимедійних даних.

Запропоновані вирішення можуть слугувати базисом для створення алгоритмів та методів організації процесів видобування, перетворення, завантаження даних та інших технологій інтеграції у сховищах, вітринах даних чи інтегрованих або розподілених базах даних.

1. Berko A. Consolidated data models for electronic business systems. / Andriy Berko // *Proceedings of IXth Internationale Conference CADSM 2007*. – Lviv, 2007. pp. 341 - 342. 2. Berko A. Models and methods of data integration for content-business systems. / Andriy Berko // *Proceedings of Internationale Conference*

ACSN 2007. – Lviv, 2007. – p. 112 – 113. 3. Lenzerini M. *Data Integration: A Theoretical Perspective*. / Marco Lenzerini // *Proc. of the ACM Symp. on Principles of Database. Systems (PODS)*, 2002. – pp. 233 – 246. 4. Tierney B. *Contextual Semantic Integration for Ontologies* / Brendan Tierney, Mike Jackson // www.macs.hw.ac.uk/BNCOD21/DC/Tierney.pdf, 2005. 5. White C. *Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise (Report Excerpt)*. / C. White // <http://www.tdwi.org/Publications/WhatWorks/display.aspx?id=7979>, 2007. 6. Берко А.Ю., Висоцька В.А. *Моделі та методи проектування інформаційних систем електронної комерції*. / Андрій Берко, Вікторія Висоцька // *Автоматизовані системи управління та прилади автоматики. Всеукраїнський міжвідомчий науково-технічний збірник № 138*. – Харків, 2007. – с. 55 – 66. 7. Кузнецов С.Д. *Ландшафт області управління даними: аналітичний обзор* / С.Д. Кузнецов, М.Н. Гринев // http://www.citforum.ru/database/data_management_overview/, 2008. 8. Ландэ Д.В. *Основы интеграции информационных потоков: Монография* / Дмитрий Ландэ – К.: Инжиниринг, 2006. – 240 с. 9. Литовский К.Ю. *Слабоструктурированные данные: некоторые методы их представления и обработки запросов* / К.Ю. Литовский, Г.С. Томусяк // *Московская Секция ACM SIGMOD* – <http://synthesis.ipi.ac.ru/sigmod/seminar/s20000224>, 2000. – с.1 – 2. 10. Смальян Р. . *Теория формальных систем* / Роберт Смальян. – М. : Наука, 1981. 11. Спирли Э. *Корпоративные хранилища данных. Планирование, разработка, развитие* / Эрик Спирли. – М. : Издательский дом "Вильямс", 2001. – 400 с.

УДК 681.325.5-181.4

В. Глухов

Національний університет “Львівська політехніка”,
кафедра електронних обчислювальних машин

ОЦІНКА АПАРАТНИХ ВИТРАТ НА РЕАЛІЗАЦІЮ БАГАТОРІВНЕВОЇ КОМП'ЮТЕРНОЇ СИСТЕМИ

© Глухов В., 2008

Сформульовано підхід до оцінювання апаратних витрат на реалізацію багаторівневої комп'ютерної системи.

In this article multilevel computer system complexity estimation is discussed.

Вступ. Постановка проблеми

Запропоновано метод оцінювання апаратних витрат на реалізацію багаторівневих комп'ютерних систем, який ґрунтується на використанні багаторівневої моделі взаємозв'язку відкритих систем. Також пропонується метод проектування багаторівневих систем, до якого метод оцінювання входить як складова частина.

Аналіз публікацій і окреслення проблеми

Однією з складових частин гарантоздатності є конфіденційність [1, 2]. Криптографічні методи забезпечення конфіденційності ґрунтуються на використанні шифропроцесорів (ШП). У роботах [3, 4] наведено багаторівневу структуру ШП, який здійснює криптографічні перетворення відповідно до стандартів [5–8]. Для побудови ШП використовується центральний протокольний процесор і криптографічний спецпроцесор (або декілька спецпроцесорів), при цьому спецпроцесори можуть бути виконані у вигляді ядер HBIC [9]. Загалом основою для побудови багаторівневих комп'ютерних систем є багаторівнева модель взаємозв'язку відкритих систем [11]. Найбільш доступним для проектування типом HBIC є програмовані логічні інтегральні схеми (ПЛІС), як відрізняються своїми технічними характеристиками (кількість і структура логічних блоків, блоків введення-виведення,