

ПІДХОДИ ДО ІНТЕГРАЦІЇ НЕОДНОРІДНИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ В РОЗВИНУТИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

В даній роботі досліджені методи і засоби забезпечення структурної і семантичної інтероперабельності інформаційних систем.

In this work were investigated methods and approaches and means of structural and semantic interoperability of information systems.

1. ВСТУП

Забезпечення інтероперабельності інформаційних систем (ІС) являє собою актуальну і разом з тим досить складну проблему, що виникає при створенні та інтеграції інформаційних систем. Ця проблема є особливо актуальною в даний час, оскільки обсяг даних, що обробляються інформаційними системами, постійно збільшується і необхідно забезпечувати обмін інформацією між різними ІС. Актуальність цієї проблеми визначається також необхідністю підвищення ефективності та спрощення формування правил зазначеного обміну.

2. ІНТЕГРАЦІЇ ІНФОРМАЦІЙНИХ СИСТЕМ

Ще починаючи з 80-х років існувала проблема спільного використання декількох баз даних (БД). Їх інтеграція вимагала багато зусиль зі сторони розробників програмного забезпечення. Це було спричинено перш за все через неоднорідність розподілених БД, їх об'єднання було великою проблемою, перш за все через недостатню ефективність інтегрованих систем.

На сьогоднішній день задача інтеграції інформаційних систем (ІС) набула ще більш важливого значення, у зв'язку зі зростанням кількості різноманітних за змістом, структурою, обсягом інформаційних ресурсів (баз даних, баз знань, програмних компонентів і т.д.), створених на різних програмно-апаратних платформах. В останні роки з'являються більш масштабні інформаційні системи, мова йде про системи дуже великих баз даних, які підтримують петабайти даних, системи текстового пошуку з масивними колекціями документів.

Сьогодні розвинуті ІС використовують різноманітні способи зберігання інформації, такі як бази та сховища даних, бази знань, ресурси

¹ Національний університет «Львівська політехніка»

на файлової системі тощо. Інформація, представлена цими ресурсами, багаторазово дублюється, втрачається цілісність даних, виникає проблема їх несумісності та ізольованості [1]. Виникає необхідність забезпечити тісний взаємозв'язок інформаційних систем, що забезпечує вищий рівень цілісності даних і швидкість знаходження необхідної інформації. Тому розробляються спеціальні механізми передавання даних між системами, розробляються спеціальні алгоритми здійснення взаємопослання окремих інформаційних ресурсів із забезпеченням необхідного рівня безпеки та максимальної швидкості їх взаємного використання.

При інтеграції інформаційних систем основну роль відіграє властивість інтероперабельності ІС. Під інтероперабельністю розуміють властивість ІС взаємодіяти з іншими ІС. Така взаємодія може виражатись у виді обміну даними, розподіленого виконання пошукових запитів, узгодженого оновлення записів в базах даних, тощо. Необхідність забезпечення інтероперабельності виникає при зв'язуванні бізнес-процесів підприємств, узгодженні роботи існуючої ІС з прийнятими стандартними рішеннями.

Також властивість інтероперабельності використовується при інтеграції декількох ІС, включення в створювану систему БД раніше використовуваних сховищ даних, розробці комплексних автоматизованих систем управління, побудові мереж інформаційних сховищ, а також в багатьох інших випадках. Проблема забезпечення інтероперабельності ІС має фундаментальний характер. Вона актуальна як для унаслідованих систем, які потрібно зв'язати з знову створюваними (або, як мінімум, отримати можливість використовувати їх БД), так і для проєктованих сховищ даних, в які необхідно передбачити можливість реалізації взаємодії з іншими ІС в перспективі, при зміні вимог до них.

При розробці інтероперабельної системи необхідно забезпечити однозначне відображення сутностей інформаційних систем і відносин між ними, тобто інтероперабельність компонентів ІС. Розрізняють два види інтероперабельності: структурну і семантичну.

Інтероперабельність ІС на структурному рівні означає здатність до структурного погодження їх сутностей систем [2]. Даний аспект пов'язаний із забезпеченням єдиного розуміння структур даних різними взаємодіючими ІС.

Семантична інтероперабельність, тобто забезпечення спільного використання різнорідних інформаційних ресурсів на семантичному рівні, характеризується здатністю до поведінкового погодження. Для досягнення семантичної інтероперабельності необхідно вирішувати проблеми порівняння вмісту цих ресурсів, відшукування відповідностей і

вирозв'язування конфліктів між ними, а також проблему сполучення різнорідних ресурсів. Ця сторона інтероперабельності вимагає використання онтологічного підходу і метаданих, розроблених на його основі.

Сучасні системи інтеграції неоднорідних інформаційних ресурсів використовують концепцію медіатора [3], тобто посередника між розподіленими інформаційними ресурсами у межах інтероперабельної системи та користувачами цих ресурсів.

Структура і поведінка конкретного інформаційного ресурсу завжди визначені семантикою предметної області, що її відображає цей ресурс. Специфікація даних може бути виконана у вигляді інформаційної схеми (як, наприклад, у реляційних моделях даних). Однак на більш загальному рівні, будь-яку специфікацію семантики даних можна записати деякою формалізованою мовою, наприклад, численням предикатів першого порядку, або багатосортною логікою.

Для складних проблемних областей потрібні загальні та гнучкі способи представлення знань. Способи представлення абстрактних понять інколи називають онтологічною інженерією. Онтологія являє собою інструмент створення узагальнених специфікацій.

Онтологія включає в себе сукупність термінів та правил, згідно з якими ці терміни можуть бути скомбіновані для побудови достовірних тверджень про стан аналізованої системи в певний момент часу. Крім того, на основі цих тверджень, можуть бути зроблені відповідні висновки, що дозволяють вносити зміни в систему, для підвищення ефективності її функціонування [4].

В будь-якій системі існує дві основні категорії предметів сприйняття, такі як самі об'єкти, що становлять систему і взаємозв'язки між цими об'єктами, що характеризують стан системи. У термінах онтології, поняття взаємозв'язку однозначно описує залежності між об'єктами системи в реальному світі, а терміни, відповідно, описують самі реальні об'єкти. Онтологічна модель представляє найбільш важливі твердження в предметній області. Додатково, ця модель допомагає описувати поведінку об'єктів і відповідну зміну взаємозв'язків між ними; тобто поведінку системи. Таким чином, онтологія являє собою словник даних, що включає в себе і термінологію, і модель поведінки системи. Оскільки кожна концептуальна модель предметної області є підмножиною онтології, завдання інтеграції інформаційних систем зводиться до задачі об'єднання метамоделей інформаційних систем, тобто побудови відображень між цими метамоделями, в термінах онтології.

Для забезпечення поширення наукових знань у світі розробляються засоби взаємопоеднання наукових архівів. Ініціатива "відкритих архівів" OAI (Open Archives Initiative) [5], започаткована розробника-

ми системи EPrints, яка створює і поширює стандарти взаємопоеднання та інтероперабельності бібліотечних архівів для більш ефективного використання наукових знань. Ця ініціатива включає в себе протокол збирання метаданих, оформлених за визначеними стандартами. Системи, в яких підтримуються OAI, можуть бути об'єднані в загальну мережу архівів, що взаємопоеднані через задані інтерфейси.

Для того, щоб користувач міг швидко і з однієї точки доступу знайти потрібну йому інформацію була розроблена система OCLC (Online Computer Library Center) [6]. OCLC забезпечує всіх бібліографічною, реферативною та повнотекстовою інформацією. На основі метаданих система проводить пошук по інформаційним ресурсам різних організацій. Більше 60000 бібліотек у 112 країнах і територіях використовують послуги OCLC для пошуку, придбання, каталогізації, позичання і збереження бібліотечних матеріалів [7].

Особливе місце в системах інтеграції неоднорідних компонентів відіграє мережа Інтернет. Величезна кількість інформації розсіяна по всесвітній мережі, і тут важливу роль відіграє пошук релевантної інформації і відсікання непотрібної інформації.

В мережі виникає проблема досягнення високого рівня інтероперабельності, можливості взаємодії з іншими системами в Intranet/Internet мережах. Ця проблема вирішується за рахунок відкритості інтерфейсів доступу до своїх сервісів та шляхом використання єдиного формату для обміну даними, а саме XML і пов'язаної з ним об'єктної моделі подання документів DOM (Document Object Model), здійснюючи, якщо необхідно, XSL- перетворення. Такий підхід дозволяє вирішити задачу синтаксичної інтероперабельності.

Нажаль, для організації взаємодії між різними інформаційними системами в Інтранет/Інтернет мережах в більшості випадків недостатньо забезпечення тільки синтаксичної інтероперабельності. Це зумовлено насамперед тим що одну й ту ж інформацію можна синтаксично по-різному уявити, і, як наслідок може виникнути бар'єр між системами. Саме проблема відсутності чітких визначень семантичних заважає об'єднанню інформаційних систем різних виробників. Після складання такого словника для представлення даних предметної області можна з легкістю використовувати мову веб-онтологій OWL (Ontology Web Language) [8], яка є одною із самих останніх розробок консорціуму W3C в напрямку Semantic Web для організації високого рівня синтаксичної та семантичної здатності додатків до взаємодії.

Мова OWL дозволяє описувати класи і відносини між ними, властиві для веб-документів і додатків. OWL заснована на більш ранніх мовах OIL і DAML OIL і в даний час є рекомендованою консорціумом Всесвітньої павутини. В основі мови — уявлення дійсності в моделі

даних «об'єкт — властивість». OWL придатна для опису не тільки веб-сторінок, але і будь-яких об'єктів дійсності. Кожному елементу опису в цій мові (в тому числі властивостями, що зв'язує об'єкти) ставиться у відповідність URI.

3. ВИСНОВОК

Зроблено спробу конкретизації стану сучасних систем інтеграції неоднорідних баз даних і баз знань, і підходів, використаних у цих системах. Розглянуті вимоги до інтероперабельної системи, проблеми їх побудови. Розглянутий метод формалізації області знань за допомогою онтології. Розглянута організація взаємодії між різними інформаційними системами у Всесвітній павутині.

1. Малезжик М. П., Засоби і технології продукування навчальних інформаційних ресурсів / М. П. Малезжик, М. В. Закатнов, В. П. Сергієнко. // Наук. часопис НПУ ім. М.П. Драгоманова: зб. наук. Праць. – К.: НПУ ім. М.П. Драгоманова, 2010. - №8 (15). 2. Михайлов И.С. - Исследование и разработка методов и программных средств обеспечения структурной и семантической интероперабельности информационных систем на основе метамоделей. Московский энергетический институт. – Москва, 2008. 3. Н.Кеберле – Огляд сучасних систем інтеграції неоднорідних баз даних і знань. Вісник Львівського університету. Серія прикладна математика та інформатика – Львів, 2002. 4. W3C Recommendation 10 February 2004, RDF/XML Syntax Specification, <http://www.w3.org/>. 5. В.А. Резніченко, О.В. Новицкий, Г.Ю. Проскудіна - Інтеграція наукових електронних бібліотек на основі протоколу OAI-PMH. Пробл. програмув. — 2007. — N 2. — С. 97-112. 6. Офіційний сайт проекту OCLC <http://www.oclc.org>. 7. <http://www.bizjournals.com/columbus/stories/2007/10/22/daily2.html>, «OCLC renames European division». Business First of Columbus, 22.10.2007 року. 8. Специфікація мови OWL <http://www.w3.org/TR/owl-ref/>