

УДК 681.518

В. І. Зацерковний, канд. техн. наук

Чернігівський державний інститут економіки і управління, м. Чернігів, Україна

АНАЛІЗ МОЖЛИВОСТІ ІНТЕГРАЦІЇ ПРОСТОРОВИХ ДАНИХ У РЕГІОНАЛЬНІЙ ГІС

Розглянута можливість інтеграції просторових даних у регіональній ГІС. Розроблена математична модель інтеграції даних, яка адекватно описує міжсистемні особливості ІС і може слугувати основою для інтеграції розрізнених просторових даних.

Постановка проблеми

Сучасний етап розвитку суспільства характеризується зростаючою роллю інформаційної сфери, що є сукупністю інформації, інформаційної інфраструктури, суб'єктів, які здійснюють збір, формування, поширення і використання інформації, а також системи регулювання суспільних відносин, які виникають при цьому.

Широке застосування інформаційних технологій протягом останніх десятиліть в країні призвело до накопичення колосальних за об'ємом масивів різного роду інформації, організованої у вигляді баз даних, створених на базі сучасних програмних технологій, які об'єднують всебічну інформацію, що зібрана в різних масштабах, форматах і проєкціях. Базы просторових даних територіально рознесені, різномірні за структурою й форматами зберігання інформації, не пов'язані одна з одною каналами інформаційного обміну, що робить їх недоступними широкому колу користувачів з числа органів державної влади регіонального і міського рівнів, а також господарюючим суб'єктам регіонального і міського значення, і в багатьох випадках її навіть неможливо порівняти. Це також ускладнює процес пошуку необхідної інформації та її подальшого опрацювання, успішне завершення якого не гарантовано для кожного окремо взятого випадку.

Аналіз останніх досліджень і публікацій

Міжнародний досвід свідчить, що інтеграція розрізнених просторових даних на державному рівні в єдину інфраструктуру просторових даних за допомогою геоінформаційних технологій дозволяє здійснити кардинальний прорив у своєчасному забезпеченні органів влади, наукових досліджень, потреб промисловості й населення достовірними й несуперечливими просторовими даними про свою територію.

Україна також приступила до створення інфраструктури просторових даних (ІПД) на основі геоінформаційних технологій, яка покликана об'єднати й забезпечити колективний доступ до просторових даних, реалізований через мережу геопорталів на трьох рівнях: державному, територіальному (рівень суб'єктів держави) і муніципальному.

Питанням інтеграції просторової інформації підприємствами й органами державної влади присвячена низка робіт вітчизняних і закордонних авторів, зокрема, роботи Р.Н. Бахтизіна, О.В. Бакланова, Ш. Джонса, О.В. Кошкарьова, Я. Массера, Х. Мураками, Д. О'нілла, Д. Пітерса, Р. Томлінсона та ін., однак у них завданням створення математичної моделі інтеграції даних, яка адекватно описує міжсистемні особливості ІС і може слугувати основою для інтеграції розрізнених просторових даних, приділялося недостатньо уваги. У зв'язку з цим завдання розробки математичної моделі інтеграції даних, яка адекватно описує міжсистемні особливості ІС і може слугувати основою для інтеграції розрізнених регіональних просторових даних, є актуальним як у теоретичному, так і в практичному плані.

Постановка завдання

Метою роботи є розробка математичної моделі інтеграції даних, яка адекватно описує міжсистемні особливості ІС і може слугувати основою для інтеграції розрізнених

просторових даних. Складність і громіздкість системи управління територіальними процесами змушує здійснювати декомпозицію задач управління на підзадачі і для вирішення кожної з них створювати спеціалізовані організації (підрозділи), які уповноважені регулювати інформаційні процеси. Інтеграція даних в єдині інформаційні ресурси й узгоджена робота з ними сьогодні є однією з актуальніших проблем для регіонального управління, оскільки технологічні процеси мають чітко виражений міжвідомчий характер.

Виклад основного матеріалу дослідження

Сьогодні в Україні не сформульована єдина політика в галузі забезпечення користувачів достовірними геопросторовими даними, що суттєво ускладнює доступ до накопиченої інформації, її опрацювання і використання при оперативному вирішенні екологічних, соціальних і економічних проблем регіонів. Якщо ситуацію не змінити, то таке становище неминуче призведе до того, що цей процес стане все більш некерованим і позначиться на їх ефективному розвитку. У зв'язку з відсутністю єдиної методологічної і нормативно-правової бази, у відповідності з якою здійснюється збір даних, має місце дублювання збору однієї і тієї ж інформації різними органами державної влади і місцевого самоврядування.

Все більше організацій, що представляють різні регіони і галузі господарства, потребують оновлення, інтеграції й аналізу достовірних і різноманітних геопросторових даних та намагаються налагодити використання для цього сучасних геоінформаційних технологій (ГІТ), що оперують геопросторовими даними у цифровому форматі [1]. У свою чергу ГІТ, пропонуючи все більш досконалі апаратно-програмні рішення і нові підходи до аналізу територіальних проблем, у все зростаючих масштабах потребують достовірних і узгоджених наборів цифрових геопросторових даних. Однак ГІТ, які є найважливішим засобом інтеграції інформаційних ресурсів, використовуються головним чином для розв'язання різноманітних галузевих завдань (створення земельних, лісових, водних та інших кадастрів, організації й проведення екологічного моніторингу, оцінки й прогнозування надзвичайних ситуацій, управління муніципальними службами і міським господарством тощо). Виконання комплексних за змістом і призначенням геоінформаційних систем вкрай обмежене. При цьому, в процесі проектування і створення галузевих ГІС на одній території сьогодні використовуються різні базові картографічні основи, які зазвичай погано співставляються між собою; застосовуються неузгоджені, часом вузькоспеціалізовані і несертифіковані програмні засоби, дублюється значний об'єм інформації.

У створюваних ГІС практично не використовується аналітичний апарат, який дозволяє досліджувати структурні взаємозв'язки між соціально-економічними, виробничими і ресурсно-природними факторами, які у сукупності визначають стан і мінливість елементів просторової організації території, що важливо для забезпечення завдань сталого розвитку. Необхідність інтеграції інформаційних ресурсів регіону в рамках єдиної регіональної ГІС (РГІС) сьогодні не викликає сумнів ні у кого. Це розуміння ґрунтується на цільності об'єкта й органу управління (адміністрація регіону) при чималій кількості розрізнених поглядів на напрямки розвитку території (рис. 1).

Інтеграція даних у єдині інформаційні ресурси та узгоджена робота з ними сьогодні є однією з актуальніших проблем для регіонального управління, оскільки технологічні процеси мають чітко окреслений міжвідомчий характер. У розв'язанні проблеми інтеграції інформаційних ресурсів зацікавлений кожний підрозділ, що входить до складу адміністрації регіону: управління економіки, управління архітектури і містобудування, податкова адміністрація, департаменти з управління муніципальним майном, землевпорядні організації, житлово-комунальні господарства, інженерні й дорожні служби тощо,

оскільки у повсякденній діяльності їм часто необхідно узгоджувати власні інформаційні ресурси з даними інших інформаційних систем (ІС).



Рис. 1. Функціональна схема системи регіонального управління

Незважаючи на це, кількість інтеграційних проєктів у регіоні, при всій їх необхідності, вкрай мала. Причина одна – це надзвичайно складна проблема.

Розглянемо деякі аспекти цієї проблеми. При цьому визначимо інтегровану ІС як систему, яка забезпечує взаємодію як мінімум двох відносно незалежних ІС.

Головними цілями РГІС є:

- інтеграція різних існуючих ІС з метою організації міжвідомчих технологічних (інформаційних) і документарних процесів. Все це – міжвідомчі процеси, в яких на різних стадіях оформлення документів беруть участь багато організацій. Терміни оформлення затягуються у часі, організації часто не знають, які дані про об'єкт змінилися (наприклад статус об'єктів нерухомості) у суміжній організації;

- інтеграція даних з метою запобігання повторного введення інформації у різних регіональних структурах й усунення помилок невідповідності даних у різних ІС. Наприклад, загальнорегіональні дані – адресний реєстр, реєстри фізичних і юридичних осіб, картографічна основа – дозволяють у інтегрованій системі брати за основу одні еталонні дані і не займатися багаторазовим введенням даних у свої ІС при роботі з одними й тими ж документами. Такі регіональні довідники і реєстри задають еталонні масиви – єдині для різних організацій, що створює принципову можливість використовувати у своїй роботі інформацію суміжних організацій;

- підготовка інтегрованої звітності, при формуванні якої потрібно обробити різномірні, але взаємозв'язані дані різних організацій, що їх використовують для ведення власних ІС. До важливих завдань також відноситься прагнення регіональних організацій скоротити витрати на супровід ІС: проєктні, інформаційні, технічні й інфраструктурні. Інтеграційний проєкт можна визначити як створення і впровадження ІС для забезпечення взаємодії більше, ніж однієї відносно автономної ІС.

Серед найбільш істотних проблем інтеграції виділяються наступні.

1. Кожна інтегрована ІС є унікальною внаслідок великої складності і «замкненості» щодо місцевих умов. У кожному населеному пункті регіону існує своя насичена подіями історія розвитку систем, свої власні рішення, свої фахівці, особливості нормативної бази тощо. Тому не існує такої моделі інтегрованої системи, яку було б легко встановити у певному регіоні і швидко отримати результат. Чітких методик «як треба будувати ІС» також не існує: в кожному регіоні впровадженню передують обстеження, під час якого фахівці знайомляться з існуючим станом справ і потім розробляють концепцію створення інтегрованої ІС саме в регіоні N, яка пропонує діяти тільки так, як можна діяти саме тут, з цією структурою управління, з цим складом фахівців.

2. Інтегрована РГІС є великою, складною і унікальною системою. Отже, створюється протягом тривалого періоду часу, за який творчість депутатських і чиновницьких мас породжує велику кількість змін нормативно-правової бази. Тому створення інтегрованої РГІС доводиться вести в умовах вимог, які постійно змінюються. Зміни нормативної і директивної бази відбуваються на усіх рівнях управління, і законотворчість у нашій країні перешкоджає інтеграції. Якщо РГІС складається з багатьох підсистем, то об'єм постійних і неочікувано виникаючих змін є надзвичайно великим. Іноді, починаючи з певного масштабу складності інтегрованої РГІС, група розробників уже не встигає розвивати систему, а змушена займатися виправленням вже створених підсистем РГІС з урахуванням нових вимог. Робота з розвитку системи гальмується, а запланований ефект від інтеграції не вдається досягти. Для запобігання таким ситуаціям сформульована найважливіша вимога до інструментальних програмних засобів, які використовуються для створення РГІС, а також до інтеграційних ІС – можливість їхньої модифікації, тобто властивість, яка дозволяє вносити зміни швидко й без істотних витрат. Проблему збільшують помилки вибору архітектури системи, оскільки їх облік і «обхід» також вимагає немало зусиль. А якщо хтось з розробників за тими або іншими причинами звільняється, взагалі відбувається процес дезінтеграції.

3. Існує доволі складна проблема узгодження вже наявних у системах даних, які треба об'єднати в рамках РГІС. Окремі ІС, що підлягають інтеграції, містять великий об'єм накопичених даних. Наприклад, ДЗК і БТІ мають реєстри фізичних і юридичних осіб, що складаються з десятків і тисяч записів. Завдання інтеграції цих систем у РГІС передбачає, що інформація про всі документи повинна бути пов'язана між собою і віднесена до одного запису, наприклад, юридичної особи. Однак тут виникають певні проблеми. Може виявитися, що у кожній організації найменування однієї і тієї ж юридичної особи записані різним способом. Для прикладу розглянемо гіпотетичну фірму «Вікторія». У кожній БД, що інтегрується, вона може бути записана трьома різними способами: ТОВ «Вікторія», Товариство з обмеженою відповідальністю «ВІКТОРІЯ», ТОВ Вікторія (без лапок). Як наслідок, програма узгодження даних у РГІС створить запис не про одну юридичну особу, а про три різні. Зрозуміло, що є додаткові можливості, які полегшують інтеграцію: можливість порівнювати юридичні особи за ПІН, за іншими ознаками, виділяти організаційно-правову форму «ТОВ» трьома способами (інтелектуалізація конвертора) тощо. Однак деякі організації не завжди вводили до своєї бази ПІН для юридичної особи, а іноді й робили помилку в такому ПІН. Тоді інтеграцію даних не вдається виконати автоматично і з'являються дублікати однієї й тієї ж юридичної особи. А це означає, що й інтеграція даних повною не виходить, а інформація про юридичну особу буде «розірваною»: або земельпорядна справа, або договір оренди будуть пов'язані з різними «формами запису», тобто фактично з двома різними копіями системи. І таких різночитань в «об'єднаних» базах даних тисячі. В результаті після об'єднання декількох баз множини даних залишаються розірваними і хтось з провідних фахівців повинен протягом декількох тижнів, а то й місяців перевіряти всі неспівпадіння і виправляти їх. Часто для цього доводиться використовувати паперові документи. Оскільки і правильні, і помилкові дані занесені до БД і вже знайшли відображення в юридичних документах, то документи стають невідповідними. А за усіма нормативними документами «якщо дані в електронному вигляді і в паперовому не співпадають, то пріоритет мають дані в паперовому документі». Ось і утворюється нульовий варіант такої інтеграції.

4. Сильно впливає на інтеграцію систем відсутність тісного оперативного зв'язку між інтеграторами і фахівцями поєднаних систем. Іноді розробники інтегрованої РГІС у процесі її створення навіть не підозрюють, що в поєднаних, але ще не поєднаних ІС (процес розтягнутий у часі) щось вже змінилося і це створить несподівані нові

труднощі. Наприклад, в одній із систем раптом замінили довідники. А в інтегрованій РГІС цей факт ще не відобразили за різними причинами. Ситуація ускладнюється, коли у складі РГІС є державна система, наприклад, Державний земельний кадастр (ДЗК), Державний кадастр нерухомості (ДКН), система податкової служби тощо. На рівні регіону можна довго займатися розробкою конверторів даних для обміну інформацією між системами, але постійні зміни законодавства, бюджету можуть нанівець звести всі зусилля. У цьому випадку інтегрована РГІС опиняється в ситуації, коли один з каналів зв'язку перестає працювати. У інтеграторів починаються певні складності. Те ж саме може трапитись і з системами податкової служби, системами реєстрації й іншими державними ІС. На державному рівні часто не враховують той факт, що в регіонах хтось займається інформатизацією, і не задумуються про те, що в одну мить руйнується робота РГІС.

5. У результаті інтеграції передбачається певний перерозподіл частини функцій між організаціями-учасниками, особливо у випадку реінжинірингу міжвідомчих процесів. Але коли люди збираються в колектив, їм доводиться «притиратись» один до одного. Це означає, що частині учасників доведеться чимось жертвувати заради того, щоб колектив був працездатний, перерозподілити між собою функції і відповідальність. При інтеграції РГІС у певних організацій-учасників можуть виникнути побоювання, що перерозподіл функцій виявиться не такий, як вони собі це уявляли, і це призведе, на їх думку, до посилення позицій одних організацій, а їхня організація буде відсторонена від справедливого розподілу ресурсів тощо. Тому іноді затягуються терміни узгодження, з'являються невинуваті «капризи» певних організацій, а то й саботаж інтеграційних процесів. Власне, ці сумніви небезпідставні: частина функцій разом з відповідними робочими місцями може бути легко передана до іншої організації, якщо такий перерозподіл повноважень виявиться оптимальним. Наявність інтегрованої РГІС полегшує реорганізацію: системі байдуже, де працює співробітник, головне – наявність доступу до неї.

6. Сьогодні є фірми-розробники програмних систем, які намагаються не поставляти замовникам конвертори на експорт даних у визнані формати, щоб якомога довше змушувати замовників залишатися «прихильниками» їх обмежених продуктів.

7. Розробники «зникли» (поїхали в інші країни, тощо) і корегувати підсистеми раптом стало нікому. Ця ситуація досить поширена і пов'язана в цілому з низькою розвиненістю індустрії розробки програмного забезпечення у нашій країні, наслідком чого є відносно невисокі заробітки програмістів.

Крім того, роботи зі створення інтегрованих РГІС супроводжуються чисельними ілюзіями як замовників, так і розробників. Ілюзії керівництва замовника такі: 1) інтегрована система створюється легко, невимушено і швидко; 2) створення інтегрованої РГІС – процес, не пов'язаний з істотними витратами; 3) інтегровану РГІС можна створити, не занадто переймаючись впровадженням: «є договір – виконуйте».

Типовою помилкою є випадки, коли компанії, що називають себе системними інтеграторами, тільки постачають техніку, розгортають мережі і базове програмне забезпечення, а замовник вважає, що придбав систему «під ключ». Часто замовнику вселяють думку, що придбання РГІС вирішить майнову, містобудівну або будь-яку іншу проблему. Прозріння настає швидко – відразу після підпису актів здачі системи в експлуатацію. Часто компанії-інтегратори погано розуміють специфіку предметних галузей, ІС в яких вони намагаються інтегрувати. Як відомо, досвід є у багатьох інтеграторів, але тільки не у вирішенні питань територіального управління. На жаль, кадастрові, моніторингові, майнові, містобудівні та інші системи набагато складніше фінансових систем і систем документообігу. Компанії, які спеціалізуються на постачанні й впровадженні систем електронного документообігу впевнені, що документообіг є найбільш важливим

у завданні інтеграції систем. Компанії, які спеціалізуються на постачанні ГІС, вважають, що ГІС і цифрова карта вирішують майже всі основні проблеми інтеграції систем регіонального управління тощо. Компанії і замовники, які не мають достатнього досвіду впровадження великих систем, часто вважають, що головне – написати або купити програму. Однак це ілюзія. Головна проблема полягає у вирішенні питань та безлічі організаційних завдань під час інтеграції в РГІС.

Способів побудови інтеграційної РГІС можна назвати багато, однак всі вони є варіаціями двох – створення інтеграційного комплексу РГІС «з нуля» та інтеграція наявних ІС у РГІС.

Створення інтеграційного комплексу РГІС «з нуля». Це самий простий і найкращий спосіб, який майже завжди гарантує успіх при умові правильної побудови роботи зі створення інтегрованої РГІС. Подібна ситуація реальна для малих і деяких середніх населених пунктів, в яких серйозні ГІС не впроваджувались, але є бажання керівництва адміністрації населеного пункту (району) впровадити відразу інтегровану систему, минаючи етап розробки самостійних ІС у різних організаціях. У цьому випадку відсутні дані, які потрібно вивіряти й узгоджувати; відсутні різнорідні системи, які потрібно поєднувати між собою; відсутнє застаріле обчислювальне і мережеве обладнання, яке для великої системи зазвичай доводиться змінювати, тобто відсутні обтяжуючі фактори. Система «з нуля» може відразу стати правильною, якщо буде розвиватись послідовно. Це означає, що ядро системи може будуватись як єдиний для всіх споживачів реєстр даних, і всі основні процеси обміну інформації закладаються системно. Якщо до роботи додатково долучиться компанія розробників, що має великий досвід, то результат може перевершити всі очікування. Розробка інтегрованої РГІС у цьому випадку навряд чи виявиться дешевшою, ніж в інших випадках, однак економія часу на створення вирається цілком реальною і, отже, результат від використання буде очевидним.



Рис. 2. Приклад інтегрованої регіональної системи у галузі нерухомості

Для створення інтегрованої РГІС «з нуля» найкращий варіант у всіх відношеннях – створення централізованої системи (рис. 2) з наступним розвитком її частин у спеціалізованих предметних галузях (ДЗК, БТІ, ЖКХ, моніторинг тощо) шляхом створення на центральному сервері предметних підсистем, уже пов’язаних з центральним ядром за допомогою єдиних регіональних довідників і реєстрів.

Не менш важливим фактором успіху є централізація всіх рішень у галузі інформатизації.

Інтеграція вже наявних систем у РГІС. Цей варіант має дуже багато версій реалізації і зустрічається найчастіше. Зазначений підхід є наслідком того, що в адміністраціях регіонів роботи зі створення інтегрованої системи не були організовані з самого початку або ідея інтеграції не підтримувалась, тому різні підрозділи розвивались своєрідно, і кожна організація будувала власну ІС, не турбуючись про те, що в майбутньому можлива інтеграція з іншими системами. Це «розподілене існування» породжує величезні масиви даних, які або недостатньо узгоджені між собою, або суперечать один одному. Дані подаються у різних форматах, по-різному ідентифікуються й адресуються. Відносна достовірність даних є загадкою для більшості учасників процесу, тому при інтеграції виникає проблема щодо узгодження («ув'язування») десятків тисяч записів у декількох базах даних, причому одночасно в різних підсистемах. Ніхто не готовий пожертвувати своїми даними, на накопичення яких пішли роки, та які вже «прописані» в тисячах юридичних документів (їх важко виправити). Ось і виходить, що з усіх наявних баз даних жодна не є еталонною. Більш того, дані всіх N баз вже містяться в юридично закріплених документах, і аніякі дані – (навіть не еталонні) – руйнувати не можна. Таким чином, інтеграція вже існуючих ІС в єдиний інформаційний комплекс (РГІС) пов'язана з вирішенням великої кількості проблем. Фактично у порівнянні зі створенням нової інтегрованої РГІС завдання ускладнюється на порядок, а то й взагалі не може мати вдалого вирішення за різними причинами. Аналіз проблемної області, цілей моделювання, а також дискретна природа об'єкта моделювання свідчать про ефективність застосування на високому рівні абстракції теоретико-множинного апарату для побудови математичної моделі інтеграції даних [2]. Базовим поняттям пропонованої моделі є поняття інформаційного об'єкта (ІО). Зазвичай ІО відповідають сутностям предметної області, кожний об'єкт характеризується значеннями заданого набору атрибутів. Тому ІО можна визначити як множину упорядкованих пар:

$$x = \langle \{a_1, d_1\}, \{a_2, d_2\}, \dots, \{a_n, d_n\} \rangle, \quad a_i \neq a_j, \quad i_i \neq j_j, \quad i, j \in \overline{1, n}, \quad (1)$$

де a – ім'я атрибуту, d – значення атрибуту.

У цьому випадку під ІС можна вважати певну інформаційну схему, яка описує характеристики об'єктів, що входять до складу системи, і множину ІО, і які задовольняють даній схемі. Під інформаційною схемою будемо розуміти кортеж $S = \langle A, D, T, \varphi, \delta \rangle$, де $A = \{a_1, a_2, \dots, a_n\}$ – множина атрибутів інформаційних об'єктів; $D = \{D_1, D_2, \dots, D_m\}$ – родина множин можливих значень атрибутів; $T = \{t_1, t_2, \dots, t_l\}$ – множина типів об'єктів; $\varphi: A \rightarrow D$ – відображення, що ставить у відповідність кожному атрибуту множину його можливих значень; $\delta: T \rightarrow 2^A$ – відображення, що задає для кожного типу множину його елементів.

ІС, що побудована за схемою S , будемо називати кортеж $U^S = \langle S, U, \gamma \rangle$, де $S = \langle A, D, T, \varphi, \delta \rangle$ – інформаційна схема; $U = \{U_1, U_2, \dots, U_n\}$ – множина ІО; $\gamma: U \rightarrow T$ – відображення, що ставить у відповідність ІО його тип, і для будь-якого ІО виконані умови:

- множина атрибутів відповідає типу $\{a: \langle a, d \rangle \in x \} \subseteq \delta(\gamma(x))$;
- для будь-якої пари $\langle a, d \rangle \in x$ маємо $d \in \varphi(A)$.

Зміну ІС задамо відображенням $F: W^S \rightarrow W^S$, де W^S – множина всіх ІС, що задовольняють схемі S . Задамо множину ІС $U^{\bar{S}} = \{U_1^{s_1}, U_2^{s_2}, \dots, U_N^{s_N}\}$, де $U_i^s = \langle S_i, U_i, T, \gamma_i \rangle$ і $S_i = \langle A_i, D_i, T_i, \varphi_i, \delta_i \rangle$, і введемо позначення:

$$S = \langle \{S_1, S_2, \dots, S_N\}, \bar{A} \prod_{1 \leq i \leq N} A_i, \bar{D} \prod_{1 \leq i \leq N} D_i, \bar{T} \prod_{1 \leq i \leq N} T_i, \bar{U} \prod_{1 \leq i \leq N} U_i \rangle \quad (2)$$

Множина ІС $U^{\bar{S}}$ є несуперечливою, якщо існують $\bar{\varphi} : \bar{A} \rightarrow \bar{D}$, $\bar{\delta} : \bar{T} \rightarrow 2^{\bar{A}}$, $\bar{\gamma} : \bar{U} \rightarrow \bar{T}$, які є розширенням відповідних відображень $\varphi_i, \delta_i, \gamma_i (1 \leq i \leq N)$.

Якщо множина ІС $U^{\bar{S}}$ несуперечлива, то ІС $U' = \langle \bar{S}, \bar{U}, \bar{\gamma} \rangle$, де $S = \langle \bar{A}, \bar{D}, \bar{T}, \bar{\varphi}, \bar{\delta} \rangle$ будемо називати інтегрованою на $U^{\bar{S}}$ (рис. 3).

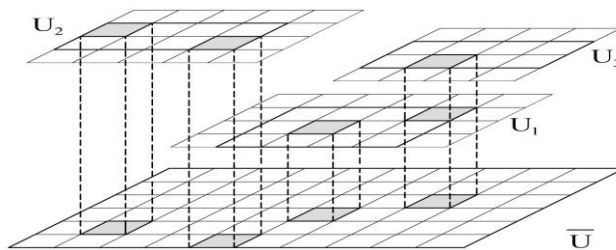


Рис. 3. Розподіл об'єктів в системі U

Зазвичай ІС повинні задовольняти більш суворим вимогам, ніж відповідність певній схемі. Такі вимоги зазвичай накладаються різними семантичними залежностями, що визначають, які стани ІС є припустимими, і використовують для узгодженої зміни даних в ІС. Таким чином, мета інтеграції полягає у збереженні відповідності множині ІС заданому набору семантичних залежностей. Семантичною залежністю, заданою на U^S , будемо називати s -предикат, заданий на W^S . Множина семантичних залежностей $C = \{c_1, c_2, \dots, c_p\}$ є несуперечливою, якщо існує ІС U^S така, що $c_i(U^S)$ для усіх $1 \leq i \leq p$. На практиці залежність між ІС необхідно зводити до залежностей між типами елементів, що входять до них. Нехай множина всіх можливих ІО X_i задовольняє схемі і має набір атрибутів, що відповідають типу t_i :

$$X_{t_i} = \{x = \{ \langle a_1, d_1 \rangle, \langle a_2, d_2 \rangle, \dots, \langle a_n, d_n \rangle \} \mid \&(A^x = \delta(t_i)) \& (\forall a_i \in A^x, a_i(x) \in \varphi(a_i)) \} \quad (3)$$

Для практики, найбільш характерними є декілька типів залежностей для реальних зв'язків між інформаційними об'єктами, які будуть розглянуті нижче. Семантичною залежністю будемо називати T -залежність, якщо для об'єктів $x_i \in X_{t_i}$, $x_j \in X_{t_j}$ вона визначається предикатом $\alpha(x_i, x_j) \alpha \beta(x_i, x_j)$. Такого виду залежність дозволяє тільки перевіряти наявність або відсутність порушення залежності і не надає механізму корекції множини ІС у відповідності з даними обмеженнями. T -залежність між типами $t_i, t_j \in T$ будемо називати V -залежністю $(t_i \xrightarrow{V, \alpha, \beta} t_j)$, якщо існує відображення $V : X_{t_i} \times X_{t_j} \rightarrow X_{t_j}$ таке, що виконана умова: $\alpha(x_i, x_j) \alpha \beta(x_i, V(x_i, x_j))$.

Існує відображення V , що дозволяє отримувати, виходячи зі значень типу t_i і t_j , нове значення типу t_j , яке буде задовольняти заданому предикату β .

Основною проблемою, що виникає при відображенні V , є наявність циклічних послідовностей, які можуть привести до нескінченної рекурсії при корегуванні залежних об'єктів. У роботі [4] доведено, що використання V -залежностей не гарантує наявність автоматичної процедури узгодженого корегування множини ІС.

Множину атрибутів $K(t_i) \subseteq \delta(t_i)$ будемо називати ключем для типу $t_i \in T$, якщо $\forall x' \in U$ виконано: $\{x \in U, a(x) = a(x'), \forall a \in K(t_i)\} \neq \emptyset$. Ключові атрибути $a \in K(t_i)$ однозначно визначають будь-який інформаційний об'єкт з множини U .

A-залежністю $(t_i \xrightarrow{A^1 A^2} t_j)$ називають V-залежність спеціального виду: нехай $A^1 \subset K(t_i)$, $A^2 \cap K(t_j) = \emptyset$, тоді

$$\begin{aligned} \alpha(x, y) &\equiv \forall a \in A^1, a(x) = a(y), \\ \beta(x, y) &\equiv \forall a \in A^2, a(x) = a(y), \\ V(x, y) &= \langle a_1, d_1 \rangle, \langle a_2, d_2 \rangle, \dots, \langle a_p, d_p \rangle, \end{aligned} \quad (4)$$

де

$$d_i \equiv \begin{cases} a_i(x), & \text{якщо } a_i \in A^2 \\ a_i(y), & \text{якщо } a_i \notin A^2 \end{cases}$$

Для будь-якої несуперечливої множини A-залежностей $C = \langle c_1, c_2, \dots, c_p \rangle$ існує процедура, яка дозволяє скорегувати ІС U^S , використовуючи кінцеву послідовність відображень V_i , що відповідають T-залежностям з C таким чином, що U^S буде задовольняти C. Ключовий атрибут $a \in \delta(t_j)$ назовемо залежним (запозиченим) для t_j , якщо існують A^1, A^2 такі, що $a \in A^2$ і $t_i \xrightarrow{A^1 A^2} t_j$. У протилежному випадку атрибут $a \in \delta(t_j)$ будемо називати вільним (власним) атрибутом. Атрибути інформаційних об'єктів $x \in U^S$ можуть бути віднесені до однієї з 3 груп, що не перетинаються: ключові, вільні і залежні атрибути. Під активною інформаційною моделлю (АІМ) розуміють кортеж $\langle S, U^S, C, F \rangle$, де S – інформаційна схема, U^S – ІС зі схемою S. C – множина семантичних залежностей, $F: W^S \rightarrow W^S$ таке відображення, що $\forall U^S \in W^S, \forall c \in C$, виконано $c(F(U^S))$. У цьому випадку для будь-яких U, S, C , де C – несуперечлива множина A-залежностей, можна побудувати активну модель $\langle S, U^S, C, F \rangle$, де відображення F має вигляд:

$$\begin{aligned} F(\langle x_1, x_2, \dots, x_n \rangle) &= \langle y_1, y_2, \dots, y_n \rangle, \text{ де } y_i = \langle a_1^i, d_1^i \rangle, \langle a_2^i, d_2^i \rangle, \dots, \langle a_{k_i}^i, d_{k_i}^i \rangle, \\ d_j^i &= \begin{cases} a_j^i(x_k), & \text{якщо, } \exists x_k, c_1 \in C, \text{ виду } \gamma(x_k) \xrightarrow{A^1 A^2} \gamma(x_i), \\ & \text{такі, що } \alpha_1(x_k, x_i) \& a_j^i \in A^2; \\ a_j^i(x_k), & \text{інакше} \end{cases} \end{aligned} \quad (5)$$

На практиці під час інтеграції даних з ІС виникає питання: як мінімізувати залежності між ІС без втрати цілісності, коректності і доступності інформації. Для вирішення цієї задачі необхідно оцінити ступінь залежності між системами. Ступенем залежності типу $t_i \in \bar{T}$ від інформаційної системи U_j будемо вважати $R(t_i, U_j)$ – кількість залежностей $c \in C$ виду $t_i \rightarrow t$, де $t \in T_j$. Ступінь залежності між ІС U_i та U_j визначається наступним:

$$R(U_i, U_j) = \begin{cases} \sum_{t \in T_i} R(t, U_j), & \text{якщо } i \neq j \\ 0, & \text{якщо } i = j \end{cases} \quad (6)$$

Як показано в роботах [2, 4], при розподілі незв'язаних типів $T' = \langle t_1, t_2, \dots, t_k \rangle \in \bar{T}$ за ІС з \bar{U}^S мінімальне значення буде досягнуто при їх розподілі за $U_{t_1}^S, U_{t_2}^S, \dots, U_{t_k}^S$, таких, що максимальне значення k-мірної матриці G, де кожний елемент $G[i_1, i_2, \dots, i_k] = \sum_{j=1}^k \sum_{U_i \in U^S} R(t_j, U_i) \in G[i_1^i, i_2^i, \dots, i_k^i]$.

Побудована модель може слугувати основою для конкретизації семантичних залежностей і застосування технології інтеграції даних обраної предметної області (рис. 4).

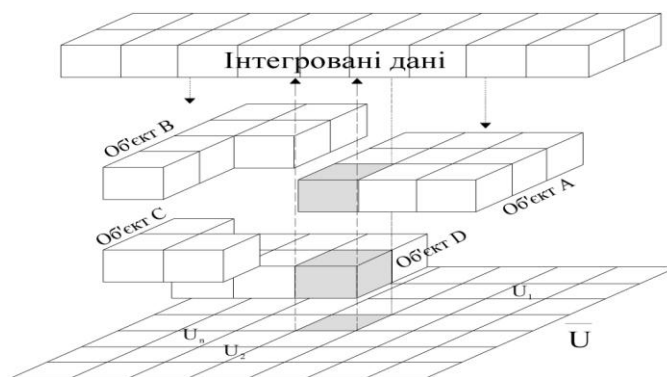


Рис. 4. Взаємодія інформаційних об'єктів

На наступному етапі моделювання знижується рівень абстракції моделі, виділяються основні класи об'єктів єдиного інформаційного простору, для яких визначається структура: властивості, поведінка і взаємне відношення. Будується багаторівнева об'єктна схема інтегрованої системи, в якій базовим класом об'єктів обліку є клас «ІО». Нащадками класу «ІО» є класи і підкласи («Реальний об'єкт», «Віртуальний об'єкт», «Класифікатор» тощо), що містять статичні елементи, які характеризують, головним чином, семантичні особливості класу і динамічні елементи для збереження історії зміни відповідних параметрів. Подальше розширення об'єктної схеми відбувається в залежності від специфікації обраної предметної області.

Для подання всіх ІО може використовуватись мова розмітки XML. Перетворення ІС, опис інформаційної схеми системи можуть задаватись відповідними XLST-перетвореннями, що дозволяє задавати складні перетворення документів на основі компактних декларативних програм. Відомі механізми корегування залежних атрибутів об'єкта інформаційної системи у відповідності з заданою множиною А-залежностей [3].

Висновки даного дослідження

Математична модель інтеграції даних, побудована таким чином, адекватно описує міжсистемні особливості ІС і може слугувати основою для інтеграції розрізаних просторових даних.

Список використаних джерел

1. Бурачек В. Г. Основи ГІС / В. Г. Бурачек, О. О. Железняк, В. І. Зацерковний. – Ніжин: ТОВ Видавництво «Аспект-Поліграф», 2011. – 512 с.
2. Воробьева М. С. Математическое моделирование и технологии интеграции данных в учетных информационных системах: автореферат диссертации на соискание ученой степени канд. техн. наук: спец. 05.13.18 «Математическое моделирование, численные методы и комплексы программ» / М. С. Воробьева. – Тюмень, 2006. – 20 с.
3. Грэхем Иан. Объектно-ориентированные методы: Принципы и практика = Object-Oriented Methods: Principles & Practice / Иан Грэхем. – 3-е изд. – М.: Вильямс, 2004. – 880 с.
4. Самарский А. А. Математическое моделирование: идеи, методы, примеры / А. А. Самарский, А. П. Михайлов. – М.: Физматлит, 2005. – 320 с.