

ІНТЕГРАЦІЯ ДАНИХ З ВИКОРИСТАННЯМ ГЕООНТОЛОГІЙ

О. Дишлик
КДП “Київгеоінформатика”

Постановка проблеми

До складу органів територіального управління входить велика кількість структурних підрозділів, які часто використовують різні системи управління базами даних. Технологічний розвиток, географічний розподіл та постійне зростання децентралізації сприяють цьому різноманіттю. Зазвичай ці бази даних можуть мати різні інтерфейси управління. Пошук інформації через загальний, зручний для розуміння інтерфейс, при повному розумінні наявних та доступних даних, був би надзвичайно корисним в цій ситуації. При визначенні техніки інтегрованого доступу до даних, маючи набір розподілених, неоднорідних та автономних баз даних найчастіше пропонується створення **федеративних баз даних**. Федерація, крім баз даних, може містити інформаційні системи.

Невирішені проблеми

Досі проблемною нішею при розробленні систем управління територіями залишається задача інтеграції даних і представлення інформації. Потоки текстової і числової інформації поступово осідають в сховищах даних. Наскільки повно на практиці використовуються приховані в цих даних закономірності, які, цілком можливо, є надзвичайно цінними? Можна припустити, що відсоток опрацювання “сирих” даних для отримання практично значущих знань поки що вельми скромний.

При інтегруванні даних потрібно враховувати декілька аспектів, зокрема, проблеми семантичної неоднорідності. Це проблеми, пов’язані з семантично еквівалентними поняттями або семантично пов’язаними/непов’язаними поняттями.

Неоднорідність може поділятися на чотири категорії: *структура, синтаксис, система та семантика*. Неоднорідність *структури* містить різні моделі даних; неоднорідність *синтаксису* – різні мови та відображення даних і неоднорідність *системи* передбачає технічні засоби та операційні системи. Семантична неоднорідність може класифікуватись як:

- *семантично еквівалентні поняття (концепції)* (моделі використовують різні терміни для підтвердження подібного поняття, наприклад, синоніми; властивості, змодельовані окремими системами по-різному тощо);
- *семантично непов’язані поняття* (один і той самий термін може використовуватись різними системами для визначення абсолютно різних понять);
- *семантично пов’язані поняття* (узагальнення/деталізація, різні класифікації тощо).

Інші подібні класифікації неоднорідності можна знайти в роботах [1–4].

Аналіз досліджень та публікацій

Останнім часом активно розвиваються та поширюються нові методи аналізу даних і видобування знань, які ґрунтуються на інших, ніж традиційна інтегро-диференційна парадигма, підходах. Мова йде про застосування онтологій та геоонтологій, доцільно використати для розв’язання проблем, які виникають при інтеграції геопросторових та непросторових даних.

У галузі географічних інформаційних систем дослідниками Фонсекою (Fonseca 2002), Френком (Frenk, 2001), Агарвалом (Agarwal 2005), Кавурасом (Kavouras, 2005), Лутцом (Lutz, 2006, 2007), Кліном (Klien, 2006), Агустіном (Agustina, 2007) та іншими зроблено чимало напрацювань щодо геоонтологій та додатків з застосуванням геоонтологій і досягнуто суттєвих результатів. Проте геоонтологія – це надто комплексне і неоднозначне поняття. Більшість експертів, досліджуючи геоонтологію, переносять поняття онтології з інформаційної науки в ГІС, причому рідко зосереджуються на особливостях географічної інформації.

Виклад матеріалу

У цій статті зроблено спробу прояснити зміст поняття геоонтологія та принципи створення застосувань (додатків) на основі геоонтологій.

Онтологія – це філософська концепція, яка відображає сутність об'єктів в реальному світі, і має на увазі, передусім, знання і розуміння.

Онтологія – це точна специфікація деякої предметної галузі. Вона забезпечує словник для уявлення і обміну знаннями про цю предметну галузь і безліч зв'язків, встановлених між термінами в цьому словнику. У найпростішому випадку побудова онтології зводиться до виділення концептів – базових понять цієї предметної галузі та побудови зв'язків між концептами – визначенню співвідношень і взаємодій базових понять.

Однією з переваг використання онтологій як інструмента пізнання є системний підхід до вивчення предметної галузі. При цьому досягають:

- систематичності – онтологія представляє цілісний погляд на предметну галузь;
- одноманітності – матеріал, представлений в єдиній формі, набагато краще сприймається і відтворюється;
- науковості – побудова онтології дає змогу відновити відсутні логічні зв'язки у всій їх повноті.

Термін “онтологія” довгий час використовувався в різних сферах діяльності по-різному. В інформатиці онтології були запропоновані Грубером [5] як “*визначена (явна) специфікація концептуалізації*”.

Концептуалізація стосується абстрактної моделі загального розуміння людьми реальних речей в світі. *Явна специфікація* означає, що поняттям та відношенням абстрактної моделі були дані – явні назви та визначення [6]. Онтологія дає назву та описує сутність специфічних сфер діяльності з використанням тверджень, які визначають відношення між цими сутностями. Вона забезпечує словник для представлення та передачі знань про сферу діяльності і набір взаємовідношень, що містить терміни словника на концептуальному рівні.

Тому онтологія може використовуватись для задач інтеграції даних через її здатність до опису семантики джерел інформації та вирішення інших різнопланових проблем [6].

Онтологія пов'язана з ієрархічною структурою даних. *Формальна модель онтології*. O – це упорядкована трійка

$$O = (C, R, F),$$

де C – скінченна множина термінів предметної галузі (теми), яку описує онтологія O ; R – скінченна множина відношень між термінами заданої предметної галузі; F – скінченна множина функцій інтерпретації, заданих на термінах і/або відношеннях онтології O .

Щоб створити онтологію, користувач має задати скінченну множину термінів предметної галузі, скінченну множину відношень між цими термінами R і скінченну множину функцій їх інтерпретації F , а потім вказати, які відношення існують між якими саме термінами. Онтологія предметної галузі може бути візуалізована у вигляді лісу орієнтованих графів з навантаженими дугами, у якому вершини відповідають термінам предметної галузі, а дуги – відношенням між ними.

Моделі концепцій в геоонтології – це просторова концептуалізація, яка сформована на розпізнаванні і добуванні просторових феноменів. Звичайно, ці просторові концепції мають бути формалізовані для машинного розуміння через обмеження області визначення.

Як зазначив Марк (Mark, 1999): “Онтологія в ГІС є свого роду додатком взаємного перетину багатьох наук, який має строге відношення до знань, опису, міжопераційності і сумнівності в географічній інформації. Найважливішим є вивчення феноменів на основі просторової семантики, або іншими словами, це – дослідження людського мислення для інформаційних систем і географічного світу.”

Геоонтологія є об'єднаною характеристикою географічної інформації і її можна визначити як:

$$O(\text{geo}) = \{C, R, A, X, I\},$$

де C (концепція) – набір концептів географічного об'єкта; R (відношення) – це набір відношень між концептами; A (атрибути) – набір атрибутів географічних об'єктів; X (аксіоми) – правила обмеження серед понять, відношень і атрибутів; I (зразки) – набір визначень зразка.

Інтеграція даних із застосуванням онтологій

Поняття *інтеграція даних*, *прикладна інтеграція* та *прикладна можливість до взаємодії* є подібними, проте необхідно їх розрізняти. **Інтеграція даних** стосується поєднання даних, які використовують спільну семантику, але походять від непов'язаних джерел. **Прикладна можливість до взаємодії** намагається стандартизувати інтерфейси окремих застосувань так, щоб дані, підготовлені в одному застосуванні, могли легко перетікати до іншого. **Прикладна інтеграція** містить аспекти інтеграції даних та прикладної можливості до взаємодії.

Існує декілька систем, які відповідають потребам інтеграції даних. Розробники кожної системи спробували різні підходи для пошуку якнайкращого шляху для надання потрібних сервісів. Деякі з найпопулярніших систем: Garlic, TSIMMIS, ObjectGlobe, SIMS тощо. Всі вони були створені для вирішення неоднорідностей будь-яких рівнів. Якщо ми обмежимося питаннями семантичної неоднорідності, то можна виділити два різні підходи: з *онтологіями та без онтологій*. У напрямі інтеграції даних без онтологій є чимало досліджень та напрацювань, але вони не розглядаються в цій роботі.

Методи побудови геоонтологій

Дослідники пропонують різні методи побудови геоонтологій, більшість з яких схожі на методи побудови інших онтологій. Єдиного стандартного методу побудови геоонтологій сьогодні не існує, але накопичено велику кількість знань у певних галузях, як, наприклад, система картографічних умовних знаків для масштабного ряду національних карт, класифікації сутностей і кодування базової географічної інформації. Ці знання дуже важливі для побудови геоонтологій. Грунтуючись на цьому знанні, можна виділити основні процеси побудови геоонтологій:

- підтвердження змісту геоонтології;
- складання переліку властивостей онтології географічного поняття;
- встановлення взаємовідношень серед географічних понять;
- збирання визначень, властивостей, образів і винятків;
- прототипування системи геоонтологій.

Інтеграція даних з використанням онтологій має багато переваг. Найвагомішими з них є використання багатшого вбудованого словника, який слугує “розумним” інтерфейсом до баз даних і не залежить від схем бази даних; знання, представлені онтологією, краще підтримують перетворення всіх потрібних джерел інформації; онтологія підтримує послідовне управління і розпізнавання несумісних даних тощо.

Один з можливих методів інтеграції даних з використанням онтологій передбачає три основні етапи: *формування загальнодоступного словника, побудова локальних онтологій та визначення зображень*. Кожний етап складається з набору завдань, які необхідно виконати. Коротко зупинимось на кожному етапі.

Перший етап: *формування (побудова) загальнодоступного словника*: цей етап має три основні кроки: *аналіз інформаційних джерел, пошук елементів (або базових елементів) та визначення глобальної онтології*.

Перший крок передбачає повний аналіз інформаційних джерел: наприклад, яка інформація та як зберігається, значення цієї інформації (семантика) тощо. На цьому етапі локалізується проблема семантичної неоднорідності.

Другий крок – *пошук термінів (або примітивів)* – передбачає складання списку термінів або понять, погоджених з розподіленим словником.

Третій і останній крок – *визначення глобальної онтології* – передбачає використання термінів, вибраних на попередньому етапі для створення глобальної онтології.

Другий етап: Побудова локальних онтологій. Ця стадія містить два основні кроки: *аналіз джерел інформації і визначення локальних онтологій*. Перший крок подібний на першу стадію

попереднього етапу. Має бути виконаний повний аналіз джерел інформації. Цей аналіз виконується, незважаючи на інші джерела інформації. Після такого аналізу виконується другий крок.

Третій етап: визначення зображень (візуалізація схем): на цій стадії ми визначаємо схему (і відношення) між концептами, які визначені в глобальній і в локальній онтологіях. Ця стадія орієнтована на владнання проблеми семантичної неоднорідності через створення зв'язків між двома стадіями. Отже, коли користувачі виконують запити, для пошуку потрібної інформації використовуються глобальна онтологія і визначення зображення.

Визначення зображення не потрібне в випадку, коли в обох системах терміни мають однакове ім'я і означають аналогічні явища.

Геоонтології на основі технології семантичного вебу (Semantic Web) [8, 9] і, зокрема, OWL-онтології¹ відіграють ключову роль у забезпеченні інтероперабельного обміну інформацією. Можливості такого підходу, який дуже активно розвивається і використовується в зарубіжних країнах, зокрема в інтересах організації управління метаданими у межах робіт із створення Національної інфраструктури геопросторових даних. Створення OWL-онтології, які формально описують предметну галузь геопросторових даних і сервіси на основі стандартів ISO 19115:2003 і ISO 19119:2005, робить можливим створення систем інтеграції розподілених неоднорідних джерел просторових даних та сервісів.

Запропонована онтологічна модель може бути використана для генерування єдиної (базової) схеми метаданих, яка забезпечить семантичну інтероперабельність систем, які беруть участь у розподіленій взаємодії. При комбінуванні онтологій, які належать до різних галузей, з геоонтологіями з'являються нові можливості для створення семантичних метаданих для наборів даних. Методика орієнтована на надання постачальникам просторових даних і сервісів можливості вибирати рівні деталізації метаданих з онтологій, які якнайкраще описуватимуть метадані ресурсів. Ці метадані можуть бути використані кінцевими користувачами геопросторових даних з покращеними умовами пошуку за змістом і оцінкою придатності незалежно від того, де ці ресурси знаходяться і в якому форматі зберігаються.

Напрями подальшого дослідження

Дослідження методу продовжується, і необхідно проаналізувати нові аспекти. Основною метою подальших досліджень є вирішення проблеми “контексту”, наприклад, проблему омонімів, різних представлень тощо. Продовжуються дослідження функцій схожості для пошуку умови схожості в межах різних локальних онтологій. Метод і його застосування необхідно протестувати на складних прикладах у справжніх практичних застосуваннях.

Висновки

Геоінформаційні системи сьогодні стали головною платформою для ефективного управління і опрацювання просторової інформації, починаючи від географічних сутностей та географічних даних до географічної інформації та географічних знань, які широко використовуються в практиці територіального управління. Значно зростає потреба в інтегруванні різних видів геопросторової інформації для того, щоб вона могла задовольнити потреби різних застосувань. Це, своєю чергою, викликає потребу внесення суттєвих змін в дослідження теорії та методології ГІС, зокрема застосування онтологій та геоонтологій.

Онтології через абстрагування/генералізацію в предметній галузі створюють серії концепцій і відносини між цими концепціями, що забезпечує їх придатність для більш “інтелектуального” комп'ютерного опрацювання інформації.

¹ OWL – Web Ontology Language, мова веб-онтологій: URL: <http://www.w3.org/2004/OWL>

Література

1. Agarwal P., 2005. Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5), pp.501–535.
2. Agustina Buccella, Alejandra Cechich., 2007. Towards Integration of Geographic Information Systems. *Electronic Notes in Theoretical Computer Science*, 168(8), pp. 45–59.
3. Fonseca F., Egenhofer M. and Agouris P., Camara G., 2002. Using Ontologies for Integrated Geographic Information Systems, *Transactions in GIS*, 6(3), pp.231–257.
4. Frank A. U., 2001. Tiers of ontology and consistency constraints in geographic information systems. *International Journal of Geographical Information Science*, 15(7), Special Issue on Ontology of Geographic Information, pp.667–678
5. Gruber T R. “Towards Principles for the Design of Ontologies used for Knowledge Sharing”. *International Journal of Human-Computer Studies*, 1995, 43:907-928.
6. Kavouras M., Kokla M. Tomai E., 2005. Comparing Categories among geographic ontologies. *Computers & Geosciences*, 31(2), pp.145–154.
7. Klien, E., Lutz, M., Kuhn, W., 2006. Ontology-Based Discovery of Geographic Information Services – An Application in Disaster Management. *Computers. Environment and Urban Systems*, 30(1), pp.102–123.
8. Lutz, M., Klien, E., 2006. Ontology-Based Retrieval of Geographic Information, *International Journal of Geographical Information Science*, 20(3), pp.233–260.
9. Lutz, M., 2007. Ontology-Based Descriptions for Semantic Discovery and Composition of Geoprocessing Services, *Geoinformatica*, 11(1), pp.1–36.

Інтеграція даних з використанням геоонтологій

О. Дишлик

Сьогодні під час розроблення систем управління територіями найбільшою проблемою залишається задача структуризації і представлення інформації. Запропоновано підхід до інтеграції даних з використанням геотехнологій як моделі описання знань в подібних системах. Крім цього, запропоновано метод для побудови онтології та проаналізовано його переваги і обмеження.

Интеграция данных с использованием геоонтологий

А. Дышлык

Сегодня при разработке систем управления территориями наибольшей проблемой остается задача структуризации и представления информации. Предложен подход к интеграции данных с использованием геоонтологий как модели описания знания в подобных системах. Кроме того, предложен метод для построения онтологии и проанализированно его преимущества и ограничения.

Data integration with the use of geontology

A. Dyshlyk

Nowadays the problem of structurization and information presentation remains the most problem area in territory control systems developing. In the article was offered an approach to the integration of data with the use of geontology as a model of knowledge description in similar systems. A method was offered for ontology creation with analyzing its advantages and limitations.