

## ПОРІВНЯЛЬНИЙ АНАЛІЗ ПІДХОДІВ ДО ПОБУДОВИ ОНТОЛОГІЧНИХ МОДЕЛЕЙ ТА БАЗ ДАНИХ

**Коваленко В. Ю. Новогрудська Р.Л.**

*Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна*

*E-mail: vladimir.kovalenko.y@gmail.com*

### COMPARATIVE ANALYSIS OF ONTOLOGIES AND RELATIONAL DATABASES

The article gives a comparative analysis of information representation using ontologies and relational databases. The analysis was conducted on four levels: conceptualization of information, data representation, data modeling, and efficiency.

Стаття наводить порівняльний аналіз способів представлення інформації з використанням онтологій та баз даних. Аналіз проведено на наступних умовних рівнях: концептуалізація інформації, представлення даних, моделювання даних, ефективність.

Незважаючи на те, що область моделювання знань налічує значну кількість моделей представлення, жодна з них не була популяризована так само, як онтології протягом останнього десятиліття. Онтології почали активно розвиватися в області обчислювальної техніки та науки, зокрема, в області штучного інтелекту та баз даних (БД). Саме БД найчастіше використовуються для представлення інформації про реальний світ, але вони накладають кілька обмежень для забезпечення ефективного доступу до даних та управління ними.

У даний час відбувається дискусія про те, чи можна вважати онтологією базу даних. Наукова спільнота, в більшості, вважає онтології найкращим способом відображення реальності завдяки їх властивості моделювати семантичні терміни. Онтологія використовує для представлення даних специфічні елементи: класи, властивості, екземпляри, агреговані відносини, узагальнені відносини, а також аксіоми, представлені логічними мовами (дескриптивна логіка або логіка першого порядку) щоб додати семантику до моделей. Але онтологія не описує конкретні представлення комп'ютерних даних і, отже, не залежить від реалізації. Варто відзначити, що досить легко знайти прямі зв'язки між онтологією та базою даних. Клас може відповідати таблиці бази даних, властивість реляційним атрибутам, узагальнюючі відносини або обмеження аксіомам тощо. Проте таке порівняння не таке тривіальне, як це здається на перший погляд.

Пропонується провести порівняльний аналіз підходів до побудови онтологічних моделей та баз даних на чотирьох рівнях: Концептуалізація інформації; Представлення даних; Моделювання даних; Ефективність.

**Рівень концептуалізація інформації.** Бази даних можна розглядати як спрощенні онтології [2], оскільки в БД відсутні аксіоми, що описують реальність. Схема бази даних являє собою механізм представлення, який розроблений в основному для задоволення вимог конкретного застосування або користувача, тому, коли ці вимоги змінюються, необхідно також змінити точку огляду і схему БД. На противагу, онтології вважаються незалежними від тексту і реалізації і, як такі, працюють на більш високому рівні абстракції. Що стосується визначення онтології, то згідно з [3] онтології представляють лише узагальнені знання, а їх

предметна область зазвичай узгоджена і залежить від рівня та мети використання. Таким чином, ті сукупності концептів які використовуються (або розробляються) тільки для конкретного додатка, не розглядаються як онтології. Крім того, при використанні онтологічної моделі немає необхідності розрізняти базові або складні типи даних і, більш того, їх властивості мають набагато більше семантики, ніж типи даних БД, оскільки онтології не потрібно нормалізувати.

**Рівень представлення даних.** Інформація, представлена в онтології, змішує поняття специфікації схеми з даними. На відмінно від БД, яка чітко диференціює схему та дані. При використанні онтології, екземпляри класів можуть повторюватися. Проте така функція не завжди може вважатися недоліком, оскільки, за рахунок наявності в онтології семантичних зв'язків, вона дозволяє підвищити ефективність при пошуку по онтології. Якщо розглядати процес ініціалізації даних, варто зауважити, що для онтологічної моделі немає специфічних вимог які регламентують цей процес. При додаванні нового екземпляру не накладаються жодні обмеження. В протиположності для БД кортеж не може бути вставлений в базу даних, якщо він не задовольняє всі семантичні обмеження схеми, такі як первинні ключі або зовнішні ключі, нульові перевірки, тощо. Отже, для реляційної бази даних такий факт може бути обмеженням на внесення даних в порівнянні з онтологією.

З іншого боку, онтології використовують семантичний механізм міркувань (СММ) для вирішення проблем ініціалізації даних. Саме він визначає, чи екземпляр належить онтології чи ні, в залежності від виконання усіх накладених обмежень. Наприклад, чи узгоджені усі визначені аксіоми, правила успадкування, тощо. У результаті кожного разу, коли потрібна перевірка, викликається СММ. Онтології мають власний механізм для вилучення інформації зі сховища онтології. Такий механізм працює аналогічно механізмам мов запитів, таких як SQL. Розглянемо основну відмінність такого механізму від аналогічних механізмів БД. СММ можуть вибрати нову інформацію, незалежно від того, чи містяться дані в онтології. Для запитів по простору даних онтології використовуються такі мови як SPARQL або SERQL. Ці мови повертають інформацію про відношення між термінами онтології з використанням предикатів, властивих структурі графів, тоді як мова SQL (мова запитів БД) представляє плоску та табличну структуру інформації, що зберігається в словнику.

**Рівень моделювання даних.** На відміну від онтологій, концептуальна модель, яка використовується для представлення даних в БД, є семантично багатшою, ніж дескриптивні логічні моделі, які використовуються для представлення більшості онтологій [1,4,5]. Проте концептуальні моделі пропонують занадто низький рівень виразності. Онтології не можуть розглядатися як концептуальні моделі, оскільки вони є багаторазовими, тоді як концептуальні моделі можуть бути багаторазово доступними лише на нижчому рівні.

Необхідно також враховувати відмінності між різними моделями баз даних (в контексті різниці між моделями на різних рівнях проектування: концептуальна модель, логічна модель, фізична модель). Логічне представлення бази даних не визначає той самий рівень семантики, що концептуальна модель тієї ж БД. Обмеження моделі даних призводять до семантичних втрат у представленні інформації. Наприклад, нехай розширена схема сутність-зв'язок або діаграма UML

представляє відношення IS-A. Воно може зникнути, коли схема не була належним чином перетворена в відповідну реляційну логічну модель, наприклад за допомогою SQL. Однак, більшість подібних перетворень не змінюють семантики інформації, представлені у сховищі.

Така ж проблема актуальна і для онтологій. Їх семантика залежить від мови представлення. Отже, онтологія, представлена за допомогою конкретної мови, такої як KIF, RDF, OWL або конкретного інструменту (Protege, WebOde, WebOnto, тощо) має власні семантичні обмеження. Тим не менше, переходи між цими мовами можуть призвести до семантичних втрат або навіть до деяких несумісностей. Такі недоліки не є актуальними для баз даних.

Мови опису онтології є набагато виразнішими в сенсі вираження семантичних концепцій, ніж бази даних, оскільки в БД містяться лише конструкції для визначення або вибору даних, а в онтологіях присутні також конструкції для опису семантичних відношень та зв'язків між даними. Крім того, онтологічні мови забезпечують більш коректну та точнішу концепцію домену.

**Рівень ефективності використання.** Бази даних є більш ефективними ніж онтології, коли мова йде про рівень підтримки даних. Для баз даних наявний широкий спектр інструментів для їх налаштування. При великих об'ємах баз даних не втрачається впорядкованість даних. Але вони однозначно програють при необхідності розвивати модель і вносити зміни.

**Висновки.** В роботі було проаналізовано підходи до побудови онтологічних моделей та баз даних. Порівняння БД та онтологічних моделей проведено по чотирьом умовним рівням: рівень концептуалізації інформації, рівень представлення даних, рівень моделювання даних, рівень ефективності використання. Подальші дослідження буде спрямовано на опис підходу до інтеграції базових понять БД та онтологічних моделей з метою спрощення та автоматизації процесу представлення даних з онтологічних моделях за допомогою БД.

## Література

1. Meersman R Ontologies and databases: More than a fleeting resemblance. In: Proceedings of inter-national workshop on open enterprise solutions: systems, experiences, and organizations (OES-SEO),Luiss Publications Rome (2001).
2. K Gómez-Pérez A, Fernández-López M, Corcho-García O (2003a) Metodologies, tools and languages for building ontologies. where is their meeting point? In: Data knowledge engineering, vol 46. Elsevier Science Publishers B.V. Amsterdam, The Netherlands, pp 41–64.
3. Dillon T, Chang E, Hadzic M, Wongthongtham P (2008) Differentiating conceptual modelling from data modelling, knowledge modelling and ontology modelling and a notation for ontology modelling. In: APCCM '08: proceedings of the fifth on Asia-Pacific conference on conceptual modelling. Australian Computer Society Inc., Darlinghurst, Australia, pp 7–17.
4. Cullot N, Parent C, Spaccapietra S, Vangenot C (2003) Ontologies: a contribution to the dl/db debate. In: Proceedings of the first international workshop on semantic web and databases (VLDB workshop).
5. Ruiz F, Hilera JR (2006) Using ontologies in software engineering and technology ontologies for software engineering and software technology. Springer, Berlin 49–102.