

Роль онтології у поданні знань

Львівський національний університет імені
Івана Франка

Кафедра оптоелектроніки та інформаційних технологій

Підготувала
студентка групи ФЕІ-41
Литвин Віра

Вступ

Тенденції розвитку інформаційних технологій сьогодні тісно пов'язані зі створенням інформаційних систем, що базуються на знаннях.

У сучасних дослідженнях у сфері розподіленого керування знаннями застосовують термін онтологія для явного опису системи знань певної галузі або інформаційного ресурсу. Онтології забезпечують загальний словник певної сфери діяльності та визначають (з різними рівнями формалізації) значення термінів і відношення між ними.

Наявність онтології певного інформаційного ресурсу(IP) дозволяє автоматизувати обробку семантики такого IP (наприклад, шукати в Інтернеті саме ті IP, що допомагають користувачеві розв'язати певну задачу). Побудова онтології є складним завданням, яке не може виконуватися повністю автоматично, але, використовуючи певні правила та технологічні прийоми, можна полегшити та пришвидшити цей процес.

Знання. Властивості знань.

Знаннями називають формалізовану інформацію, яку використовують у процесі рішення задачі для отримання нової інформації. Перехід від обробки даних до обробки знань – результат розвитку й ускладнення структур, що є об'єктом ІТ. Знання відображають закономірності певної предметної області (принципи, зв'язки, закони), отримані в результаті практичної діяльності та досвіду розв'язання її задач.

Властивості знань:

- інтерпретованість – знання мають семантичне наповнення;
- структурованість – знання про складні об'єкти можуть бути подані як декомпозиція знань на простіші об'єкти та зв'язки між ними;
- зв'язність – знання відображають відношення між фактами та явищами (причинно–наслідкові, родо–видові тощо);
- ситуативна сумісність знань – знання відображають ситуації, припустимі при взаємодії об'єктів певної ПрО;
- активність – знання забезпечують генерування нової інформації.

Представлення знань.

У штучному інтелекті основна мета представлення знань — навчитися зберігати знання так, щоб програми могли опрацьовувати їх і досягати подібності з людським інтелектом.

Проблема представлення знань є однією з найважливіших серед тих, які характерні для систем, що базуються на знаннях. Це пояснюється тим, що форма представлення знань впливає на характеристики і властивості системи. Центральним завданням побудови систем, що базуються на знаннях, є вибір форми (моделі, мови, методу) представлення знань.

Представлення знань - це сукупність домовленостей стосовно синтаксичних і семантичних правил, що роблять можливим вираження знань про предметну галузь у комп'ютерно-інтерпретованій формі, яка вимагає їх структурування.

Моделі представлення знань.

У основі **логічних моделей** представлення знань лежить поняття формальної теорії, кортеж $S = \langle V, F, A, R \rangle$, де V — зліченна множина базових символів (алфавіт); F — множина, елементи якої називають формулами; A — виділена підмножина апіорі справедливих формул (аксіом); R — зліченна множина відношень між формулами, яку називають правилами висновку.

Продукційна модель — найпростіший засіб подання знань. Вони складаються з продукційних правил типу "Якщо A , тоді B ". A називають посилкою (умовою, антецедентом), а B — дією (консеквентом). Це означає, що "якщо всі умови A є істинними, тоді B — також істинне" або "якщо всі умови A виконуються, тоді потрібно виконати дію B ".

Мережні моделі є найбільш адекватним способом формалізації подання знань у природномовних текстах. Знання описуються сукупностями трійок (a, r, b) , де a і b — об'єкти або поняття, а r — бінарне відношення між ними. Якщо в мережній моделі допускаються зв'язки різних типів, то її називають семантичною мережею. Вона формалізує знання у вигляді орієнтованого графа з розміченими вершинами (поняттями) і дугами (відношеннями).

Поняття онтології.

Онтологія – угода про спільне використання понять (термінів), що містить засоби подання предметних знань і домовленості про методи логічного виведення. Вона може розглядатися як формалізований опис погляду на світ у конкретній сфері інтересів, що складається з набору термінів і правил використання цих термінів, що обмежують їхнє значення в рамках конкретної ПрО. Онтології дозволяють формалізувати знання користувачів про ту ПрО, яка їх цікавить. При цьому такі знання стають доступними іншим користувачам і можуть застосовуватися в інших ІС.

Складові частини онтологій.

Екземпляри (англ. *instances*) або індивіди (англ. *Individuals*) — це основні, низькорівневі компоненти онтології. Екземпляри можуть бути як фізичними об'єктами (люди, будинки, планети), так і абстрактними (числа, слова).

Поняття (англ. *concepts*) (або класи (англ. *Classes*)) — абстрактні групи, колекції або набори об'єктів. Вони можуть містити в собі екземпляри, інші класи, або ж сполучення й того, і іншого.

Об'єкти в онтології можуть мати *атрибути*. Кожен атрибут має принаймні ім'я й значення, і використовується для зберігання інформації, що специфічна для об'єкта й прив'язана до нього. Значення атрибута може бути складеним типом даних.

Важлива роль атрибутів полягає в тому, щоб визначати залежності (*відношення*) між об'єктами онтології. Завичай відношенням є атрибут, значенням якого є інший об'єкт.

Класифікація онтологій.

Спеціалізовані (предметно-орієнтовані) онтології — це представлення якої-небудь галузі знань або частини реального світу. У такій онтології містяться спеціальні для цієї галузі значення термінів.

Загальні онтології використовуються для подання понять, спільних для великої кількості галузей. Такі онтології містять базовий набір термінів, глосарій або тезаурус, використовуваний для опису термінів предметних галузей.

Тезаурус.

Тезаурус – це окремий випадок онтології. Зазвичай тезаурус T визначають як словник, що містить лексичні одиниці з явною вказівкою семантичних зв'язків між ними. Згідно "Сучасному словнику іноземних слів": тезаурус -

1) словник, у якому максимально повно подані усі слова мови з вичерпним переліком прикладів їхнього вживання в текстах; у повному обсязі це може бути реалізовано лише для мертвих мов;

2) ідеографічний словник, у якому показані семантичні відносини (синонімічні, родо-видові тощо) між лексичними одиницями;

3) в інформатиці - повний систематизований набір даних про будь-яку область знань, що дозволяє людині чи комп'ютеру в ній орієнтуватися.

Формальна модель тезаурусу – $T_s = \langle T, R \rangle$, де T - множина термінів, а R – множина відносин між цими термінами. Множини T та R скінчені.

Інтеграція 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.

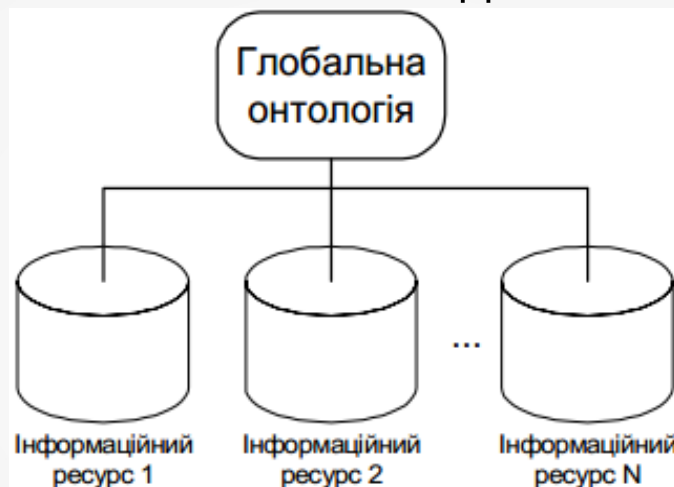
Інтеграція семантики даних передбачає формування єдиного змістового простору для сприйняття, інтерпретації та застосування даних незалежно від формату їх подання та структури.

Основною метою таких процесів є однакова інтерпретація всіх елементів даних, отриманих з різноманітних джерел, та складових деякої об'єднаної інформаційної структури.

Використовуючи онтології для явного визначення неявних чи прихованих знань про дані та їх зміст, можна вирішити значну частину проблем семантичної неоднорідності даних.

Інтеграція на основі єдиної онтології.

Особливістю такої семантичної інтеграції даних є спільне використання її ресурсів для опису семантики кожного вхідного набору даних.

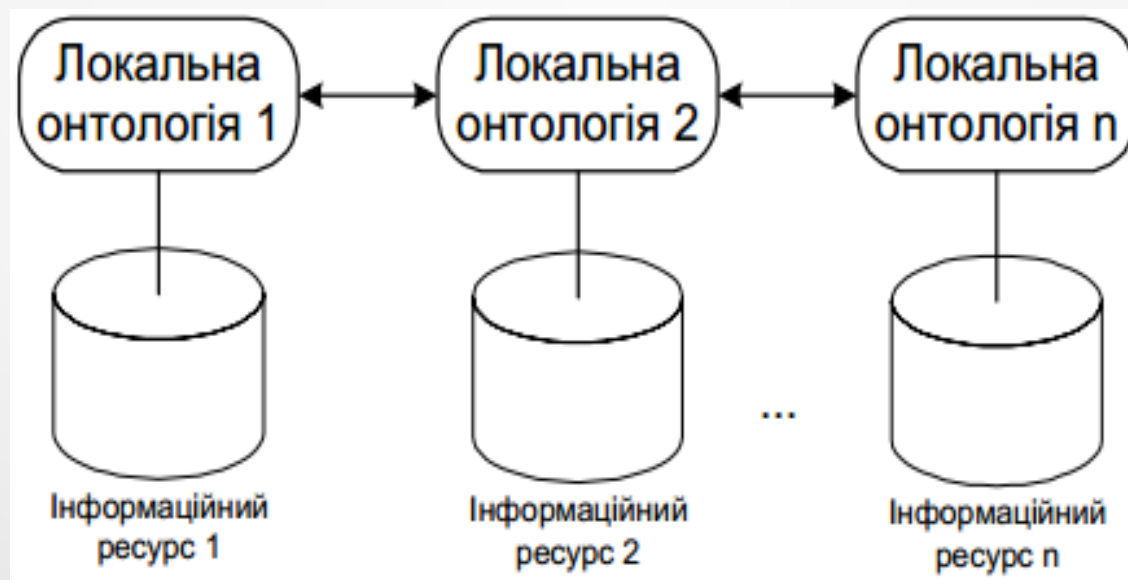


Перевагою є однотипність визначення та інтерпретації концептів в усіх вхідних наборах даних, відсутність неоднозначностей формулювання понять та суперечностей імен і метрик. Це, в свою чергу, спрощує процеси формування єдиного семантичного простору для інтегрованого набору даних, зменшує обсяги самих онтологій та додаткових метаданих.

Проблемою такого підходу є саме формування єдиної глобальної онтології, оскільки загалом не завжди можна забезпечити єдність концептуалізації усіх вхідних даних

Інтеграція на основі множини онтологій.

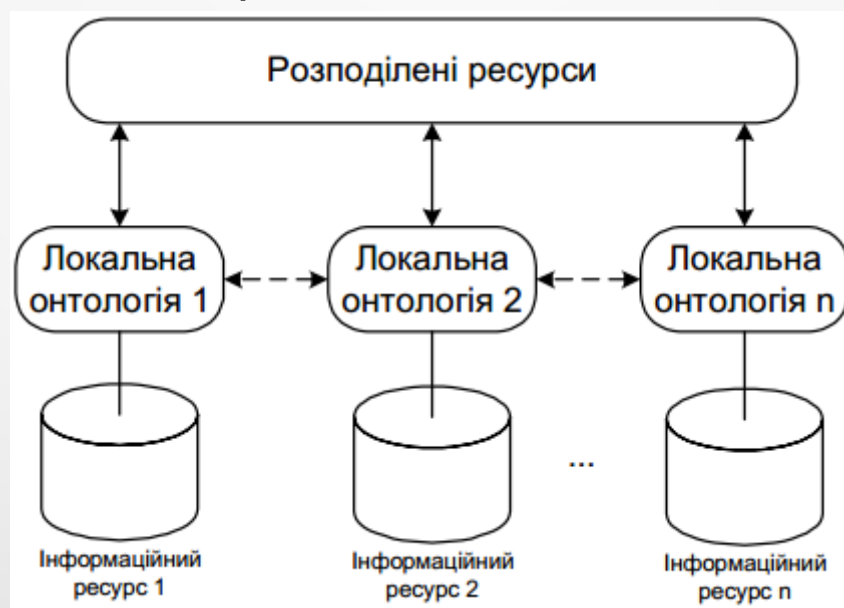
Цей підхід застосовують, коли побудова єдиної глобальної онтології для опису семантики всіх вхідних наборів є неможливою чи складною. У такому випадку кожен вхідний набір даних для семантичної інтеграції описують власною онтологією, яка не пов'язана з іншими і оперує власними нерозподіленими словниковими ресурсами. Процес семантичної інтеграції у цьому випадку ґрунтується на узгодженні, взаємодії та обміні ресурсами локальних онтологій.



Гібридний підхід.

За аналогією з єдиною онтологією, у цьому випадку створюють спільний, узгоджений розподілений ресурс, але використовують цей ресурс для специфікації семантики вхідних наборів даних через їх власні локальні онтології.

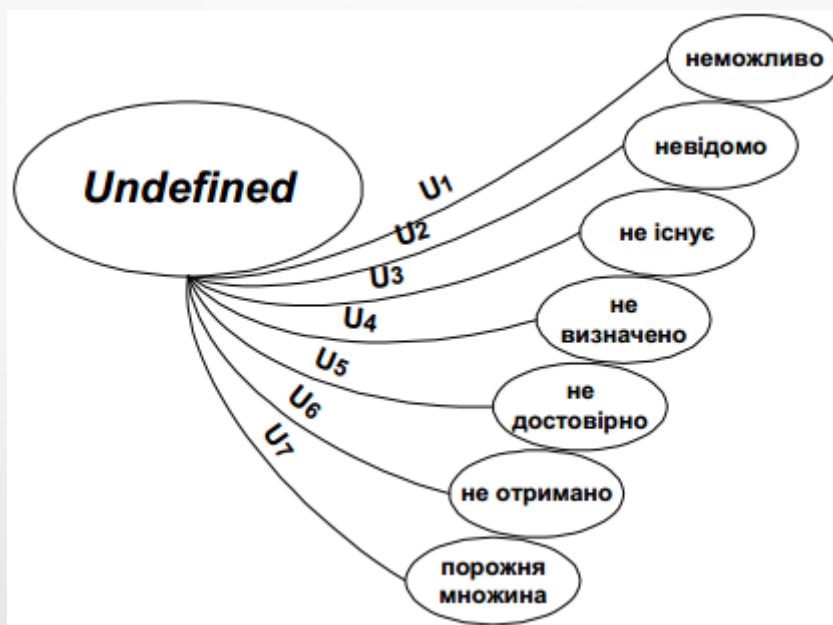
Аналогічно до попереднього випадку семантику інформаційного ресурсу, що підлягає інтеграції, описує окрема онтологія. Але для сумісності локальних онтологій створюють глобальні розподілені словникові ресурси, в яких зосереджено базові терміни і поняття, спільні для предметної області інтегрованих даних.



Інтеграція невизначеностей в даних.

Для опису невизначеностей як структурної та семантичної одиниці даних пропонується застосувати спеціальний концепт – клас "Невизначеність", елементами якого є різноманітні варіанти невизначеностей. Класифікація невизначеностей необхідна для уникнення суперечностей, які виникають.

Часткова інтерпретація невизначеностей за допомогою визначення спеціальних класів в онтологіях даних переводить проблему їх опрацювання на принципово інший рівень – від опрацювання невідомих, неповних, не достовірних понять до опрацювання явно визначених концептів.



Переваги використання онтологій

Переваги використання онтологій при вирішенні тієї чи іншої задачі:

- в обчислювальному плані (наприклад, для скорочення часу обчислень) ;
- в економічному плані (скорочення витрат на розробку програмного забезпечення, на інтеграцію даних);
- порівняно з вже існуючими рішеннями (заснованими на класичних підходах).

Найбільшою областю застосування семантичних технологій є:

- проект семантична Мережа (Semantic Web);
- інформаційний пошук (Information Retrieval).

Недоліки використання онтологій

Недоліки, що ускладнюють інформаційний пошук, зумовлені сприйняттям текстів як простого набору слів ("bag of words"):

- Надмірність – в послівному індексі використовуються слова-синоніми, що виражають одні й ті ж поняття ;
- слова тексту вважаються незалежними одно від одного, що не відповідає властивостям зв'язного тексту ;
- багатозначність слів – оскільки багатозначні слова можуть розглядатися як диз'юнкція двох або більше понять, що виражають різні значення багатозначного слова, то мало ймовірно, що всі елементи цієї диз'юнкції цікавлять користувача.

Висновки.

Онтологія визначає множину концептів та зв'язків між ними без специфікації предметної області. Загалом, онтології утворюють інфраструктуру, необхідну для визначення семантики даних, яка може бути сприйнятою і опрацьованою програмними засобами. Такий підхід дає змогу коректно вирішувати проблеми, пов'язані зі змістом даних на формальному рівні із застосуванням засобів інформаційних технологій.

На відміну від звичайного словника, для онтологічної системи характерні внутрішня єдність, логічний взаємозв'язок і несуперечність використовуваних понять.

Одним із напрямів у дослідженні методів і засобів опрацювання даних на основі онтологій, які сьогодні активно розвиваються, є напрям, пов'язаний із застосуванням Web-онтологій.

Дякую за увагу.

