

МЕТОДИКА РОЗРОБКИ ТЕРМІНОЛОГІЇ ПРЕДМЕТНОЇ ОБЛАСТІ ЯК БАЗИСУ ДЛЯ ФОРМУВАННЯ ОНТОЛОГІЙ ТА ТЕЗАУРУСІВ

¹Міжнародний науково-навчальний центр інформаційних
технологій та систем НАН України і МОН України
²Інститут програмних систем НАН України

Розглянуто засоби та методи створення онтологій та тезаурусів – інтероперабельних моделей подання знань предметної області, що використовуються розподіленими Web-застосуваннями. Щоб коректно визначити відношення між різними термінами тезаурусу, пропонується використовувати елементи онтологічного аналізу та основи мереологічного підходу. Крім того, в роботі пропонуються застосовні правила створення визначень термінів тезаурусу, які дозволяють формувати більш якісні знання про предметну область

Вступ

Сучасні напрями розвитку інформаційних технологій (ІТ) пов'язані зі створенням інформаційних систем, що базуються на знаннях. У сучасних дослідженнях в сфері розподіленого керування знаннями застосовують термін "онтологія" для явної 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 [1]. У найбільш загальному випадку вона являє собою угоду про спільне використання понять, що містить засоби представлення предметних знань і домовленості про методи розуміння.

Термін онтологія, запозичений з філософії, де він позначав частину метафізики – навчання про всім суццюму, про його найбільш загальні філософські категорії (буття, субстанція, причина, дія, явище тощо), зараз активно застосовується в інформатиці і штучному інтелекті для формального і декларативного визначення концептуалізації певного розділу знань.

Часто онтологією називають базу знань спеціального вигляду, яку можна розділяти, відчувувати і самостійно використовувати. Онтології дозволяють представити поняття так, що вони стають

придатними для машинної обробки. Незалежно від виду онтології, у неї необхідно включити словник термінів і деякі специфікації їхніх значень, що дозволяє обмежити інтерпретацію цих термінів і відбити їхню взаємодію. Помітимо, що при такому підході поняття онтології перетинається з уже давно прийнятим в інформатиці і лінгвістиці поняттям тезауруса. Онтологія – це база знань, що описує факти, що передбачаються завжди істинними в рамках певного співтовариства на основі загальноприйнятого значення тезауруса.

Онтологія може використовуватися як посередник: між користувачем і інформаційною системою або між членами співтовариства, наприклад, між користувачами деякого корпоративного сховища даних.

Наявність онтології певного ІР дозволяє автоматизувати обробку семантики такого ІР (приміром, шукати в Інтернеті саме ті ІР, що допомагають користувачеві розв'язати певну задачу). Побудова онтології є складною задачею, яка не може виконуватися повністю автоматично, але, використовуючи певні правила та технологічні прийоми, можна полегшити та пришвидшити цей процес.

Онтологічний підхід до подання знань предметної області

Онтологія – угода про спільне використання понять (термінів), що містить засоби подання предметних знань і домовленості про методи логічного висновку (міркувань) [2]. Це 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 галузі [3]. Знання в онтологіях формалізують, використовуючи п'ять видів компонентів: класи, відношення, функції, аксіоми та екземпляри [4]. Формальна модель онтології – упорядкована трійка $O = \langle T, R, F \rangle$,

де T – скінчена множина термінів ПрО, що описує онтологія O ;

R – скінчена множина відношень між термінами заданої ПрО;

F – скінчена множина функцій інтерпретації, заданих на термінах і/чи відношеннях онтології O .

При створенні онтологій найбільшу складність становить формування множини F , тому що цей процес вимагає застосування спеціальних навичок з галузі інженерії знань і формальної логіки. У той же час стосовно трудомісткості основна робота з формування онтологій припадає на формування множини X , але слід зазначити, що ця робота доступна більшості фахівців довільної галузі. Складніше визначити множину відношень R , які треба використовувати для моделювання знань. У [5] виділені найбільш загальні онтологічні відношення в реальних доменах – зв'язки еквівалентності, таксономічний, структурний, залежності, топологічний, причинно-наслідковий, функціональний, хронологічний, подоби, умовний і цільовий. Таксономія – це окремих випадок онтології, в якій присутні тільки ієрархічні зв'язки одного типу.

Одним з найпоширеніших відношень в онтології є відношення іменування. Воно є фундаментальним, тобто на його базі може бути побудована формальна система, що дозволяє виражати основні математичні поняття. Існує чотири фундаментальних відношення: приналеж-

ності (теорія множин ZF і NF), між функцією, її аргументом і результатом (теорія множин фон Неймана), іменування (онтологія Лесьневського) і "частина-ціле" (мереологія) [6].

Мереологія – це формальна теорія про частини і зв'язані з ними поняття, розроблена С.Лесьневським [7]. Об'єктом мереології є дослідження відношення "частина-ціле". Мереологія – частина тріади дедуктивних теорій, що включає також прототетику й онтологію (у Лесьневського онтологія розглядається тільки як система з єдиним відношенням «є» - "is_a"). Відношення "частина-ціле" є винятково важливим тому, що воно утворює основу поняття *системи*, яке є центральним у сучасному науковому пізнанні. Система являє собою структурне з'єднання своїх елементів. Її базовою формальною характеристикою є те, що елементи не просто входять у систему, а входять у неї в результаті взаємодії з іншими елементами.

Інші аксіоми мереології описують взаємозв'язки між системою і її елементами. Приклад – аксіома системи і частин: якщо елемент зв'язаний в одну сторону, то він зв'язаний і в іншу. Мереологія виходить за межі вивчення часткових відносин між елементами спільних систем. Вона також займається тими об'єктами, частини яких релевантні цілому. Такі об'єкти ідентифікуються як екземпляри. Серед мереологічних відношень можна виділити сім різних класів, і взагалі, транзитивність не прийнята серед екземплярів різних класів:

- 1) компонент-об'єкт: сторінка-книга;
- 2) член-колекція (наприклад, дерево-ліс);
- 3) частина-маса (наприклад, шматок-хліб);
- 4) матеріал-об'єкт (наприклад, алюміній-літак);
- 5) властивість-діяльність (наприклад, бачити-читати);
- 6) стадія-процес (наприклад, заварювання-готування чаю);

7) місцевість-область (наприклад, Закарпаття-Україна).

Більшість досліджень відношення "частина-ціле" присвячені вивченню частин, але можна також ідентифікувати різні типи цілого відповідно до таких властивостей: 1. чи віддільна частина від цілого (мелодія-пісня або вагон-потяг); 2. чи є частини просторовими або часовими (кімната-квартира або зима-рік); 3. чи відіграє частина певну функціональну роль відносно цілого (двигун-автомобіль); 4. чи є частини неподільними (атом-молекула).

Знання цих теоретичних принципів допомагає більш точно визначити мереологічні відношення, що вводяться до онтології. Визначивши тип відношення за такою класифікацією, можна більш чітко визначити, чи можна використовувати для визначення зв'язків між поняттями одне або різні відношення.

Постановка задачі

Встановивши певні застосовні правила, за якими потрібно створювати описи понять онтологій та тезаурусів, можна підвищити якість онтологій, що створюються, та забезпечити їх більшу інтероперабельність. Щоб коректно визначити відношення між різними термінами тезаурусу, пропонується використовувати основи мереологічного підходу для більш чіткого вибору відношень типу "частина-ціле".

Формування тезауруса предметної області, що цікавить користувача

Окремим випадком онтології, який простіше формувати та обробляти, є тезаурус – повний систематизований набір даних про будь-яку галузь знань, що дає змогу людині чи комп'ютеру в ній орієнтуватися. Можна досліджувати як тезауруси окремих фахівців, так і тезауруси галузей знань.

Слово тезаурус грецького походження і означає – скарбниця, запас, скарб; у широкому розумінні використовується для позначення словника, що міс-

ить синоніми, антоніми, родинні слова, а також ієрархічне дерево понять із відношеннями „рід – вид”, „від цілого до частини”.

Термін "тезаурус" є досить древнього походження. Уперше його застосував у значенні, близькому сьогоdnішнього, ще в XIII столітті Брунетто Датіні, автор систематизованої енциклопедії. Свою працю він назвав "Книга про скарб". Про труднощі складання тезауруса говорить той факт, що на сьогodнішній день існує всього два повноцінних універсальних тезауруси: тезаурус Роджета для англійської мови й Робера для французької мови [8].

Першою, хто вжив термін тезаурус у зв'язку з обчислювальними машинами, була Анетта Мастерман [9]). В 1954 році вона звернула увагу на те, що тезаурус – досить загальний засіб для опису семантичних структур природної мови й запропонувала використати цей засіб в області машинного перекладу. У сімдесятих роках почався цілий тезаурусний бум, пов'язаний з розробкою інформаційно-пошукових систем. Саме тоді термін тезаурус міцно узвичаївся як у лінгвістів, так і у фахівців з інформаційних технологій.

У роботі [10] використане наступне визначення терміна тезаурус: "Тезаурус являє собою словник понять та словник класифікаційних зв'язків між ними. Основним призначенням тезауруса є переклад на інформаційно-пошукову мову смислового змісту документів і запитів, що ставляться до різних галузей науки й техніки, які охоплені класифікаційною схемою тезауруса. Тезаурус покликаний забезпечити єдине уніфіковане й формалізоване подання інформації в інформаційних системах."

Формальна модель тезауруса –

$$Ts = \langle T, R \rangle,$$

де T – скінчена множина термінів, R – скінчена множина відношень між цими термінами.

Тезаурус можна розглядати як семантичну мережу, у вузлах якої знаходяться терміни, пов'язані відношеннями з обмеженого набору *R*. Основні технологічні фази створення тезауруса [11]:

- виділення лексичних одиниць, тобто формування словника (глосарія) *T*;
- розробка набору семантичних зв'язків;
- актуалізація зв'язків – установлення зв'язків між термінами.

При цьому дуже важливо сформулювати принципи, за якими буде здійснюватися кожна процедура. Для першого пункту визначальними є два аспекти – джерело лексичних одиниць та критерій їх добування. При розробці набору семантичних відношень можна знаходити їх у тексті, що описує дану галузь (намагатися вичленувати й уніфікувати ті відношення, що існують в текстах між термінами) або безпосередньо аналізувати знання. На практиці звичайно використовують поєднання обох методик. Для актуалізації семантичних зв'язків між термінами тезауруса можна використовувати знання експертів, а також документи, призначені як для фіксації структури знань (словники, класифікатори тощо), так і самі знання, що відображають ПрО (реферати, статті, монографії тощо).

Термін – це слово або словесний комплекс, що співвідноситься з поняттям визначеної організованої галузі пізнання (науки, техніки) та вступає в системні відношення з іншими словами і словесними комплексами і утворює разом з ними в будь-якому окремому випадку чи у певний час замкнуту систему, що відрізняється високою інформативністю, однозначністю, точністю й експресивною нейтральністю.

Слово "*термін*" походить від латинського "*terminus*" – "*границя*". Множина термінів тезауруса *T* відповідає множині концептів *X* онтології *O*. Такі властивості термінів і терміностаггів (ТС), як системність, стійкість і регулярність зв'язків, відсутність експресії, установка на об'єктивність опису, дають можливість моде-

лювати ТС за допомогою тезаурусів. Класифікація понять ПрО через набір слів, умовно синонімічних і утворюючих клас умовної еквівалентності, лежить в основі тезаурусів, використовуваних для інформаційного пошуку (рис. 1).

На наш погляд, тезаурус може стати ефективним інструментом формування запитів до універсальних пошукових машин *Internet* й істотно підвищити якість інформаційного пошуку в спеціалізованій тематичній області при виконанні наступних умов:

- тезаурус відображує термінологію досить вузької наукової/предметної області;
- у тезаурусі використовується набір сильно диференційованих семантичних відношень;
- тезаурус незалежний від машини пошуку.

Завдання побудови універсального тезауруса є дуже трудомісткою. Побудову тезауруса вузької предметної або наукової області здійснити набагато простіше. З одного боку, завдання є набагато меншим за обсягом; з іншого боку – у цьому випадку ми маємо справу не із загальноживаною лексикою, а з термінами. Властивості термінів – системність, стійкість та регулярність зв'язків, відсутність експресії, спрямованість на об'єктивність опису, – уможливають їхній адекватний опис за допомогою тезаурусів.

Особливо точно описати термінологію можна за допомогою тезауруса з набором сильно диференційованих семантичних відношень. Основна ідея такого опису – використання не тільки універсальних (наприклад, "рід-вид", "частина-ціле" тощо), але й специфічних для конкретної предметної області відношень. Таким чином, кожен тип відношень сам по собі несе значне смислове навантаження, визначає різні аспекти семантики терміна.

Для створення тезауруса можна скористатися методологією розробки онтологічних моделей – стандарт *IDEF5* [12] сімейства *IDEF*, згідно з якою побудова

тезауруса ПрО складається з п'яти основних дій:

– вивчення і систематизація початкових умов – мети і контексту розробки

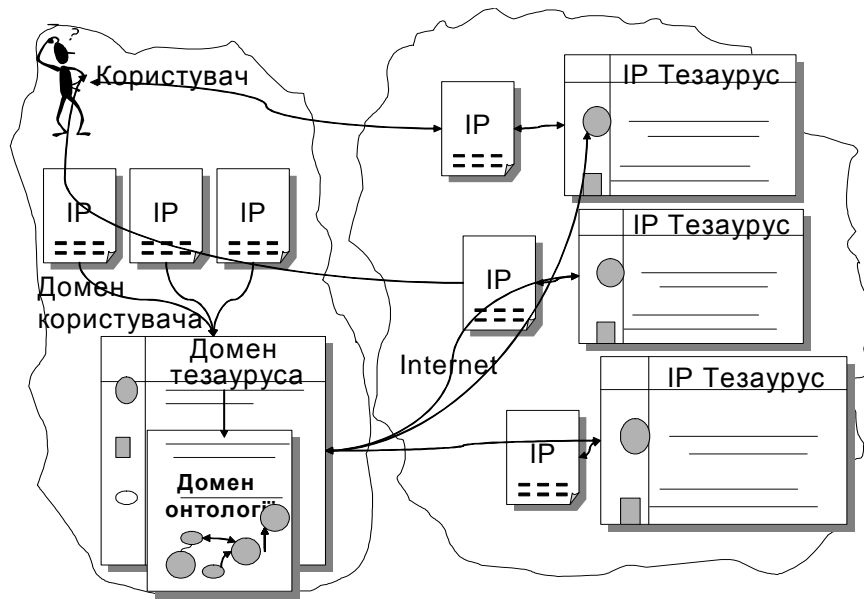


Рис. 1. Інформаційний пошук на основі тезаурусу

тезауруса, визначення меж ПрО, яка цікавить користувача;

– збирання і накопичення даних – підбір IP, що відносяться до цієї ПрО;

– аналіз даних – вивчення відібраних IP, формування словника термінів ПрО, що містяться у відібраних IP;

– початкова розробка тезауруса – установлення зв'язків між термінами ПрО, з якої потім витягуються базові терміни ПрО;

– уточнення і затвердження тезауруса – аналіз користувачем отриманого тезауруса та його коректування.

Під час формування тезауруса доцільно враховувати наступні рекомендації, які стосуються побудови визначень даних і метаданих та враховують вимоги, розроблені підкомітетом зі стандартизації ПК-6 (ТК-20) „Телекомунікації та обмін інформацією між системами” з урахуванням ISO/IEC 11179:

Абстрактний домен (conceptual domain) – набір змісту значущих значень.

Поняття (concept) – одиниця інформації, створена як унікальна комбінація характеристик.

Дані (data) – представлення інформації, яке може бути повторно інтерпре-

товане, подане у формалізованому вигляді, який дозволяє комунікацію, інтерпретацію або обробку.

Елемент даних (data element) – частина даних, для якої визначення, ідентифікація, представлення та припустимі значення визначаються через набір атрибутів.

Поняття елементу даних (data element concept) – поняття, яке може бути представлене у формі елементу даних, описаного незалежно від будь-якого конкретного представлення

Визначення (definition) – представлення поняття через описовий вираз, що дозволяє відрізнити його від інших пов'язаних з ним понять

Метадані (metadata) – дані, що визначають та описують інші дані.

Реєстр метаданих (metadata registry) – інформаційна система для реєстрації метаданих.

Елемент метаданих (metadata item) – екземпляр об'єкту метаданих. Елемент метаданих має пов'язані з ним атрибути, які відповідають екземплярам об'єкту метаданих.

Об'єкт метаданих (metadata object) – тип об'єкту, визначений у метамоделі

Ім'я (*name*) – позначення об'єкту через лінгвістичний вираз.

Домен значень (*value domain*) – набір припустимих значень. Домен значень забезпечує подання, але не надає висновки про те, з якими елементами даних ці значення можуть бути пов'язані, та який зміст мають ці значення.

Практичні рекомендації стосовно визначення термінів тезаурусу

Визначення термінів тезаурусу може здійснюватися в автоматичному режимі (шляхом аналізу повнотекстових документів та інших інформаційних джерел), шляхом вилучення з інших баз знань (тезаурусів, онтологій тощо) або надаватися безпосередньо експертом Про.

Формальні вимоги до визначень термінів тезаурусу:

а) визначення має бути **викладене в однині**. Виняток становлять поняття, які самі є множинними. Наприклад, “*номер статті*”: добре визначення – “номер посилання, що ідентифікує статтю”; погане визначення – “номер посилання для ідентифікації статей”. У поганому визначенні використовується слово “статті”, що може бути формою множини і це можна зрозуміти так, ніби один номер може посилатися на кілька статей.

б) визначення повинне пояснювати, **чим є** наведене поняття, **а не тільки чим воно не є**. Наприклад, “розмір вартості фрахтування”: добре визначення – “розмір витрат, які несе вантажовідправник для переміщення товарів з одного місця до іншого”; погане визначення – “розмір витрат, що не належать до витрат на пакування, документальне оформлення, завантаження, розвантаження та страхування”. У поганому прикладі не вказано, що входить до поняття елемента даних.

с) визначення повинне **мати вигляд описової фрази або речення**. Речення необхідне для формування точного визначення, яке містить важливі характеристики поняття. Просте наведення одного або кількох синонімів не є достатнім. Наприклад, “ім'я агента”: добре визначення

– “назва сторони, яка уповноважена діяти від імені іншої сторони”; погане визначення – “представник”. “Представник” є синонімом імені елемента даних, який не може бути адекватним визначенням.

д) визначення повинне **містити лише широко відомі скорочення**. Розуміння значення скорочення, зокрема абревіатур та ініціалів, зазвичай обмежується певним середовищем. В іншому середовищі ті ж самі скорочення можуть викликати неправильне розуміння або непорозуміння. Таким чином, для запобігання неоднозначності, у визначеннях використовуються тільки повні слова без скорочень. Наприклад, “прилад для вимірювання щільності”: добре визначення – “прилад, який використовується для вимірювання концентрації рідини, в одиницях виміру маси до одиниці об'єму (м. д. о.) (тобто фунтів на кубічний фут; кілограмів на кубічний метр)” погане визначення “прилад, який використовується для вимірювання концентрації рідини в термінах м. д. о. (тобто фунтів на кубічний фут; кілограмів на кубічний метр)” Проте м. д. о. не є загальновідомим скороченням і його значення може бути незрозумілим для деяких користувачів. Скорочення має бути наведене повними словами.

е) визначення повинне **бути викладене без використання визначень інших даних або базових понять**. Визначення термінів має наводитись у відповідному глосарії. Якщо потрібне інше визначення, воно має додаватись як примітка після тексту первинного визначення або як окремий запис у словнику. Пов'язані визначення можна отримати за допомогою атрибутів посилання (перехресних посилань). Наприклад: “код типу зразка”: добре визначення – “код, який ідентифікує тип зразка”; погане визначення – “код, який ідентифікує тип обраного зразка. Зразок – це мала частка, вилучена для проведення експериментів. Він може бути як єдиним зразком для тестування, так і сурогатним зразком для контролю якості. Зразок для контролю якос-

ті – це сурогатний зразок, обраний для перевірки результатів тестування єдиних зразків”. Погане визначення містить два додаткових визначення – “зразка” та “зразка для контролю якості”.

Семантичні вимоги до визначень термінів тезауруса:

а) визначення має **відображати суттєвий зміст поняття**. Усі первинні характеристики поняття, мають бути відображені у визначенні з відповідним рівнем специфічності залежно від контексту. При цьому необхідно запобігати пояснення неважливих параметрів. Рівень деталізації залежить від потреб користувача системи та середовища.

Наприклад, “номер послідовності завантаження вантажу” (визначений контекст: будь-яка форма транспортування): добре визначення – “номер, що вказує на послідовність, в якій здійснюється завантаження до транспортного засобу або елемента транспортного середовища”; погане визначення – “номер, який відображає послідовність, в якій здійснюється завантаження до вантажівки” (у визначеному контексті вантажі можуть транспортуватись різними транспортними засобами: вантажівками, кораблями, вантажними потягами і не обмежене лише вантажівками).

Інший приклад: “сума за рахунком-фактурою”: добре визначення – “загальна сума, яку потрібно сплатити за рахунком-фактурою”; погане визначення – “загальна сума вартості всіх елементів, зазначених в рахунку-фактурі, включаючи усі відрахування, зокрема знижки та дисконти, та додаткові платежі, зокрема страхові, транспортні та накладні витрати тощо”. У поганому визначенні міститься зайва інформація.

б) визначення має бути **точним та однозначним**. Визначення має бути достатньо зрозумілим, щоб забезпечити його однозначну інтерпретацію. Наприклад, “дата отримання вантажу”: добре визначення – «дата, на яку вантаж передається отримувачу»; погане визначення – «дата, на яку здійснюється доставка вантажу». У

поганому визначенні не роз’яснюється, що таке “доставка”. Під “доставкою” можна зрозуміти як момент розвантаження товару в певному місті, так і факт передачі товару кінцевому отримувачу. Не включено, що кінцевий отримувач ніколи не отримає вантаж або його передача може здійснитися через кілька днів після розвантаження.

с) визначення має **бути коротким**. Слід запобігати використанню додаткових фраз описового характеру, подібних до “для забезпечення використання цього реєстру метаданих”, “терміни, що мають бути описані”. Наприклад, “ім’я набору символів”: – добре визначення “ім’я, що присвоюється набору фонетичних або ідеографічних символів, в які зашифровані дані”; погане визначення – “ім’я, що присвоюється набору фонетичних або ідеографічних символів, в яких зашифровані дані для забезпечення використання цього реєстру метаданих або, якщо говорити про загальний вжиток, спроможність системного обладнання і програмного забезпечення обробляти дані, зашифровані одним або декількома шифрами”. У поганому визначенні всі фрази після виразу “... в яких зашифровані дані” є зайвими.

д) визначення повинне **мати можливість використовуватися окремо**. Зміст поняття має бути наочним у визначенні. Для розуміння поняття не повинні бути потрібні додаткові роз’яснення. Наприклад, “назва міста розміщення школи”: добре визначення – “назва міста, де знаходиться школа”; погане визначення – “див. сайт школи”. Погане визначення не є самостійним, оскільки необхідно звернутися до додаткового джерела.

е) визначення повинне **бути поданим без використання пояснювальної інформації, функціонального використання або процедурної інформації**. Пояснення не слід включати до визначень, тому що вони містять зайву інформацію. У разі потреби такі пояснення можуть бути розміщені в інших атрибутах метаданих. Припустимо додати кілька прикладів після визначення. Наприклад, “мітка

поля даних”: добре визначення – “ідентифікація поля в індексі, тезаурусі, базі даних тощо”; погане визначення – “ідентифікація поля в індексі, тезаурусі, базі даних тощо, яка застосовується для таких елементів інформації, як примітки, колонки в таблицях”. У поганому визначенні містяться примітки, що стосуються функціонального використання. Якщо інформація, що починається зі слів “яка застосовується...” є необхідною, то вона має бути розміщена в іншому атрибуті.

f) визначення повинне запобігати циклічних посилань.

Два поняття не повинні бути розкриті одне через одне. Визначення одного поняття не може використовувати інше поняття як своє визначення, тому що це може призвести до ситуації, коли поняття визначається через інше поняття, яке, у свою чергу, визначається через перше поняття. Наприклад, два елементи даних з поганими визначеннями – “ідентифікаційний номер працівника – номер, що призначається працівнику; “працівник – людина, яка має відповідний ідентифікаційний номер працівника”. Визначення посилаються одне на одне, але в жодному з них не наведено зміст поняття.

g) визначення повинні використовувати однаковою термінологію та логічну структуру для пов'язаних визначень.

Для близьких або пов'язаних визначень має використовуватись одна й та ж сама термінологія та синтаксис. Наприклад, “дата відправлення товарів – дата, в яку товари були відправлені даній стороні”, “дата отримання товарів – дата, в яку товари були отримані даною стороною”. Використання єдиної термінології значно спрощує розуміння.

h) відповідати типу елемента метаданих, що визначається.

Різні типи елементів метаданих у реєстрі метаданих (приміром, поняття елемента даних, елемент даних, концептуальний домен, домен значення) відіграють різні ролі і це має відображатися у їх визначеннях.

Наприклад, поняття елемента даних: “Максимальний Розмір Заробітної Плати по Категоріях Робітників” (Визначення: Максимальний розмір заробітної плати, дозволений для відповідної категорії робітників). Поняття елемента даних не посилається на особливий домен значення.

Концептуальний Домен: “Кількість грошей” (Визначення: Кількість, яка може бути виражена у грошових одиницях). Визначення посилається на грошові одиниці, але не вказує на які.

Елемент даних 1: “ Максимальний Розмір Заробітної Плати по Категоріях Робітників в Європі”. Визначення: Максимальний розмір заробітної плати, дозволений для певної категорії робітників, виражений у євро”.

Елемент даних 2: “Максимальний Розмір Заробітної Плати по Категоріях Робітників у США” (Визначення: Максимальний розмір заробітної плати, дозволений для певної категорії робітників, виражений у доларах США”).

Визначення елементів даних можуть посилатись на відповідні домени значень.

Тезаурус як засіб опису термінологічних систем та знань

Тезаурус відноситься до спеціального типу словника нормативної лексики з точно визначеними зв'язками між термінами. Тезаурус можна розглядати як окремий випадок онтології.

Крім того, ми пропонуємо розглядати тезаурус як модель термінологічної системи (ТС). Термінологічна система є більш високим рівнем організації термінів, тобто це «складна динамічна стабільна система елементами якої є відображені по певних правилах лексичні одиниці будь-якої природної мови, структура якої ізоморфна структурі логічних зв'язків між поняттями спеціальної області знань і діяльності, а функція полягає в тому, щоб слугувати знаковою (мовною) моделлю в цій області знань і діяльності» [10]. У цьому визначенні вже підтримується ідея моделювання ТС за допомогою тезауруса. Саме властивості термінів і ТС - системність, стійкість та регулярність зв'язків,

відсутність експресії, направленість на об'єктивність опису - уможливають моделювання ТС за допомогою тезаурусів.

Специфіка нашого підходу до тезаурусів визначається наступними моментами: 1). Тезаурус перш за все - модель ТС, а вже потім - частина текстового редактора, інформаційно-пошукової системи, бази даних тощо; 2). Зміст тезаурусу - терміни. У порівнянні із загальноживаною лексикою терміни мають переваги як основні елементи тезаурусу. Термінам властиво: регулярність та стійкість семантичних відношень, емоційна нейтральність, можливість виділити відносно замкнуті утворення – ТС; 3). Терміни у тезаурусі зв'язані різними типами семантичних відносин з досить великого набору. Сильно диференційовані семантичні зв'язки дозволяють здійснити досить тонке налаштування моделі; 4). Операції й процедури роботи є необхідною складовою частиною тезаурусу. Цей підхід можна порівняти з об'єктно-орієнтованим стилем проектування, що забезпечує адекватність моделі (тезауруса) динамічного об'єкта (ТС). Для подальшого застосування онтологій та тезаурусів нам потрібно вивчити, як співвідносяться тезаурус, терміно-система, система знань.

Висновки

Опису певної предметної області та створення інтелектуальних систем та мереж велике значення має розробка нових алгоритмів та методик формування тезаурусів та онтологій. Для адекватного відображення інтероперабельних моделей подання знань в роботі запропоновано методику, яка використовує мереологічний підхід для створення визначень термінів тезаурусу, що дозволило сформувати більш якісні знання.

Список літератури

1. *Gruber T.R.* A translation approach to portable ontology specifications // *Knowledge Acquisition*, 1993, V. 5. – P. 199–220.

2. *Musen M.* Domain Ontologies in Software Engineering: Use of Protege with

the EON Architecture // *Methods of Inform. in Medicine*, 1998. – P. 540–550.

3. *Гладун А.Я., Рогушина Ю.В.* Онтологический подход к поиску веб-сервисов в распределенной среде Интернета // *Информатика*, Минск, 2006. – № 4 – С. 116–127.

4. *Gómez-Pérez A., Moreno A., Pazos J., Sierra-Alonso A.* Knowledge Maps: An essential technique for conceptualisation // *Data & Knowledge Engineering*, 2000, V.33(2). – P. 169–190.

5. *Нариньяни А.С.* Кентавр по имени ТЕОН: Тезаурус + Онтология. – <http://www.artint.ru/articles/narin/teon.htm>.

6. *Лесьневский С.* Об основаниях математики. – <http://www.philosophy.ru/library/logic/lesnewxi.html>.

7. *Рогушина Ю.В., Гладун А.Я.* Мереологические аспекты онтологического анализа интеллектуальных Web-сервисов // Зб.праць VII Міжнар.конф. „Інтелектуальний аналіз інформації” IAI-2007 К.: – 2007. – С. 312–321.

8. *Гладун А.Я., Рогушина Ю.В.* Онтологии и мультилингвистические тезаурусы как основа семантического поиска информационных ресурсов Интернет // *The Proc. of XII-th Intern. Conf. KDS'2006*, Varna, Bulgaria. – P. 115–121.

9. *Андреев А.М., Березкин Д.В., Симаков К.В.* Особенности проектирования модели и онтологии предметной области для поиска противоречий в правовых электронных библиотеках. – <http://www.inteltec.ru/publish/articles/textan/RCDL2004.shtml>.

10. *Браславский П.И., Гольдштейн С.Л., Ткаченко Т.Я.* Тезаурус как средство описания систем знаний // *Информационные процессы и системы*. – 1997. – № 11 (Серия 2). – С. 16–22.

11. *Ненейвода Н.Н.* Мереология. – <http://www.logic.ru/Russian/events/ifras/nnn.pdf>.

12. IDEF5 – Ontology Description Capture Method – (www.idef.com/IDEF5.html).