

ТЕРМІНОЛОГІЯ ТА ЇЇ РОЛЬ У ПРЕДСТАВЛЕННІ ЗНАНЬ

© Ленков А. Р., 2009

Розглянуто роль термінології в представленні та поширенні знань. Коротко охарактеризовано сучасні комп'ютеризовані термінологічні збірки (термінологічні банки), а також сучасні комп'ютерні системи представлення знань, проаналізовано їх позитивні сторони та недоліки. Розглянуто роль і можливість використання термінологічних даних для побудови систем представлення знань.

The paper deals with the role of terms in knowledge representation. Short characteristic of modern computerized term banks is given; strong and weak sides from the knowledge representation point of view are analyzed. Modern knowledge presentation systems based on ontologies are revised to reveal their imperfection. Role of terms in knowledge representation systems is overviewed.

Вступ

Природну мову люди вживають для спілкування та збереження й передавання знань. Термінологія займає особливе місце в збереженні та передаванні знань, оскільки саме на неї припадає основне інформаційне навантаження.

Поява комп'ютерів спричинила швидкий розвиток науки про знання, створивши можливість легко зберігати, упорядковувати та розповсюджувати інформацію. Розроблялись моделі збереження та передавання знань, а також методології використання знань для різноманітних потреб, зокрема побудови систем штучного інтелекту та квазі-інтелектуальних систем спілкування та прийняття (ухвалення) рішень.

Програми та середовища, які використовують системи знань, вимагають великих ресурсів для реалізації, тому будують їх на різних машинах, часто з різними програмними платформами. Для успішної роботи такі системи мають задовольняти низку специфічних вимог. Вони повинні забезпечувати спілкування з користувачем природною мовою, вони мають розуміти інформаційні запити, використовувати доступні системи знань, забезпечувати обмін інформацією між різними платформами та передбачати можливість використання різних протоколів обміну даними.

Для продуктивного обміну знаннями необхідно розробляти інформаційні структури, у яких знання про світ представлені у вигляді спеціальних моделей, що мають бути зрозумілі для різних користувачів. Основне завдання таких моделей – полегшувати ефективне спілкування між користувачами шляхом забезпечення лаконічного, позбавленого неоднозначності, представлення знань для взаємного обміну. Таке моделювання дає змогу працювати зі знаннями як людям, так і автоматизованим системам.

Найчастіше користуються двома моделями представлення знань: онтологіями та термінологіями [1]. Онтології бувають дуже різноманітними – від простих класифікаторів, побудованих згідно із принципами наслідування, до складних багаторівневих концептуальних систем, тоді як термінології обмежено рамками галузевої лексики. У термінології, з погляду представлення знань бракує системності, у той час як в онтологіях термінологічна компонента є дуже бідно представлена. Ця робота має на меті переглянути роль термінології в представленні знань з метою вироблення рекомендацій для покращення існуючих систем.

I. Термінологія і представлення знань

Термінологію давно використовують для передавання та поширення знань, оскільки вона містить найточнішу та концентровану інформацію з певної галузі.

Корені термінології лежать у лінгвістиці та науці про пізнання. Використовуючи довколишній світ як об'єкт та мову як засіб, термінологія представляє фрагменти знання в усій їх складності та зумовленості зв'язків. З погляду представлення знань термінологія робить це засобами мови. Основними об'єктами є терміни. Якщо пильніше подивитись, можна зауважити різне інформаційне наповнення термінів. Деякі з термінів позначають прості поняття, а деякі – поняття, які мають набагато більше інформаційне наповнення. Останні часто називають концептами. Концептами вважають фрагменти знань, які дають змогу моделювати та пізнавати світ. З погляду значущості концепти є основними одиницями термінології, а системи концептів формують обсяги знань. З цього погляду терміни це лексичні одиниці, що допомагають означити концепти. Можна сказати, що терміни – це лексичні одиниці, що вживають в обмежених галузях і слугують для точного й лаконічного означення знань.

Оскільки термінологія загалом є неоднозначною, побудова однозначних відповідних систем термінів має величезне значення, яке особливо зросло в останні десятиліття. Це пов'язане з тим, що різко зріс обсяг знань у всіх галузях, а також обсяг розповсюдження знань не тільки серед спеціалістів але й між нефаківцями. Розповсюдження знань відбувається із широким залученням перекладу, для якого роль термінів є особливою. Відомо, що двомовні глосарії складають основу будь-якої системи перекладу із залученням чи без залучення людини. Крім того, широке розповсюдження комп'ютерів привело до збільшення інтенсивності обміну, опрацювання та поширення знань, а також до виникнення систем управління інформацією та ухвалення рішень. У всіх цих випадках критичним є використання термінологічних сукупностей.

Набір термінів складає основу кожної термінологічної системи, тому збір та опис термінів це основа термінологічної роботи в кожній галузі. Вибір термінів з письмових та усних мовних джерел – дуже складна, тривала, багата на помилки робота. Успіхи останніх років у автоматизації обробки текстової інформації спонукали посилення діяльності в напрямку розроблення автоматизованих систем збирання термінів. Ці системи вимагають підготування великих обсягів текстів для опрацювання комп'ютерами, а також розроблення складного програмного забезпечення для розпізнавання термінів.

Хоча успіхи в цьому напрямі не такі вже й великі, проте сам він є перспективним, оскільки відкриває нові можливості для досліджень, знімаючи часові обмеження, зменшуючи кількість помилок та необхідної людської праці.

Збирання та впорядкування термінів – це тільки початок термінологічної роботи. Не менш важливим є відслідковування термінів від часу їх запозичення із загальної мови, протягом існування, змін, виходу з ужитку [2]. Для проведення цієї роботи створюють системи оперування термінологією, відомі як термінологічні банки даних.

Використання таких систем роботи з термінами дає змогу формувати бази термінологічних даних підвищеної складності, які містять інформацію не тільки про значення та впорядкування, але й адміністративні дані (час та особа що ввели дані), лексичні та граматичні дані, концептуальні дані та описові дані (де і наскільки часто вживають). Наявні термінологічні банки даних набагато простіші й не підтримують всіх можливостей комп'ютерних засобів. Пояснюється це тим, що вони виникали тоді, коли потреба в такій організації матеріалу ще не стала критичною, а комп'ютерні засоби були набагато дорожчими й менш продуктивними.

Проте самої термінології недостатньо для передавання чи розповсюдження знань, оскільки знання в людському суспільстві є культурно зумовлені. Тому для представлення взаємовідносин одиниць інформації (які подано термінами) потрібні загальнозживані слова, що передбачає використання природної мови й веде до появи неоднозначностей та нечіткостей під час використання чи передавання. Найвиразніше це помітно коли знання передають у багатомовному середовищі.

II. Системи представлення знань на основі онтологій

Зрозуміло, що використання термінології для передавання знань вимагає введення додаткової інформації і, можливо, зміни структури представлення її. Для розв'язання нових задач, які передбачають використання систем представлення знань, необхідно розробляти нові підходи організації та подання інформації. Одним з таких підходів є укладання та використання онтологій [3].

Термін онтологія сьогодні часто вживають у зв'язку з представленням знань, проте його використання може спричинити непорозуміння. Загальновідомим є значення, яке вкладали в це поняття стародавні греки (вивчення природи чи суті наявного), але в царині штучного інтелекту цей термін широко використовують зі зовсім іншим змістом (формальне представлення набору концепцій у певній галузі) із другого боку.

У системі представлення знань онтологію розуміють, як докладний понятійний (концептуальний) опис певної галузі. Основним в онтології є концептуальність. Концептуальність передбачає набір концептів (об'єкти, події, стани справ, вірування) та відповідних відношень між ними. Онтологія передбачає повний опис концептів та відношень за допомогою формальної мови. Онтологічний аналіз певної галузі передбачає концептуалізацію (встановлення концептів та відношень між ними) разом з укладанням відповідного словника для опису її. Якщо для побудови словника використовують побутову мову, то в результаті концептуалізації ми отримуємо словник конкретних значень. Такий повний та формалізований опис значень уможливує робити судження та поповнювати обсяги знань, використовуючи для цього штучні системи. Це сприяє створенню та підтримуванию онтологій та баз знань.

Залежно від призначення розрізняють застосовні онтології, онтології певних галузей, генеративні онтології та онтології представлення. Найяскравішим прикладом є онтології певних галузей, які забезпечують їхній опис. Генеративні онтології містять опис взаємозв'язку загальних концептів відносно часу та простору. Онтології представлення описують способи представлення знань.

Онтології відрізняються за глибиною, від простого списку до розгалужених мереж. Існують термінологічні онтології (лексикони, класифікатори), інформаційні онтології (записи в базах даних) та онтології знань [4].

Про складність онтології говорить набір відношень між концептами. Основним видом відношень є відношення часткове/ загальне. Такий вид відношень дає змогу будувати ієрархію структур та застосовувати механізм наслідування, що спрощує опис. Іншим параметром, що говорить про складність онтології є рівень окремих концептів. Найпростішими є онтології, які представлені концептами у вигляді слів, складніші можуть бути подані концептами у вигляді ситуацій чи подій.

Якщо вважати онтологію сукупністю словника та набору значень, то можна розрізнити їх за ступенем формальності. Неформальні онтології використовують звичайну мову у вигляді словника, напівформальні використовують визначення разом із загальноживаними словами. Існують формальні онтології, у яких всі складники словника мають чітке формалізоване визначення.

Зрозуміло, що будь-яка програма опрацювання природної мови, яка повинна працювати зі значеннями, потребує онтології, що слугують семантичним лексиконом для них. Такі онтології повинні уможливлувати подання значень текстів у незалежній від мови формі, зокрема допомагати розв'язувати неоднозначності. Найвідоміші практичні застосування – машинний переклад та інформаційний пошук. Найвідоміші на сьогодні онтології такого виду є WordNet [5], Mikrokosmos [6], SENSUS [7].

Використання онтологій в інформаційному пошукові дає змогу розширити його можливості за рахунок розширення/доповнення інформаційних запитів за допомогою синонімів, омонімів та семантично пов'язаних концептів. Використання онтологій дає змогу шукати відповідні матеріали за змістом, а не за відповідністю до ключових слів.

Створення онтологій уможливує використовувати знання в різноманітних системах та галузях, зменшуючи витрати на розроблення систем опрацювання інформації. Системи знань, побудовані на основі онтологій, легше вбудовувати в різноманітні середовища опрацювання та пошуку інформації. Як приклад можна назвати використання онтології SENSUS для побудови бази знань в іншій галузі [8]

Усі онтології використовують терміни для опису концептів та відношень між ними. Проблема полягає в тому, що для опису використовують термінологію, яка може бути спільною для різних концептів, тому, описуючи концепти, не вводять паралельно термінології. Якщо ми хочемо використовувати термінологію більш ґрунтовно, доцільно спочатку визначити спільне та відмінне у використанні її та онтології для представлення знань.

Як термінологія, так і онтології виконують однакову функцію – подають концептуальні поняття з певної галузі у вигляді, який дає змогу обмінювати чи передавати знання в комплексній формі. Обидві системи подають знання підвищеної складності у формі, яка знімає (або зменшує) залежність від мови. Обидві системи використовують комп'ютерні засоби для збереження, підтримання та представлення інформації.

Якщо в термінології для визначення окремих термінів (понять) використовують звичайну мову, то для побудови онтологій використовують спеціальну, формалізовану мову, яка дозволяє уникнути неоднозначності. Комп'ютеризовані термінологічні бази даних не можуть безпосередньо представляти знання, тоді як онтології мають розроблені оболонки для доповнення, представлення та обміну знань. Якщо термінологічні бази даних орієнтовано на користування людиною, то онтології чудово забезпечують користування як людьми, так і автоматизованими системами. Інформація в термінологічних базах даних представлена у формі, яка є зрозумілою та зручною для людини, в онтологіях інформація представлена в спеціальній формалізованій мові, яку важко розуміти людині, особливо без спеціальної підготовки.

Було здійснено декілька спроб наблизити термінологічні бази даних до онтологій. Відомий італійський проект *Blueterm*, у якому кожний елемент словника представлено як концепт, що займає одну комірку в базі даних. У цій же комірці міститься вся відповідна інформація. Це тлумачення терміна і його синоніми на інших мовах. У тлумаченні терміна наведено інформацію про зв'язок цього терміна як поняття (концепту) з іншими. Також наведено галузь у якій термін вживають, а також міститься контекстуальна (приклади вживання) та граматична інформація. Наведено також інформацію про джерело походження терміна, означення, контекстні приклади. Не зважаючи на таку повноту, є і недоліки. До них належить неможливість прямого доступу до концептуальної інформації, єдиний спосіб доступу є алфавітний, через термінологію. Неможливо вийти на концептуально споріднені терміни. Також неможливо перевірити зв'язність системи після внесення нової інформації.

Інший відомий проект *KBS Hyperbook* розроблено в Німеччині. Він є колекцією гіпертекстових документів, об'єднаних за принципом онтологій. Онтологічна частина написана на спеціальній об'єктно-орієнтованій мові *O-Telos* і використовується як метайнформація, яка допомагає структурувати та пов'язувати дані між собою. Цей підхід має багато подібного до семантичного моделювання, що використовують у гіпертексті, але відрізняється можливістю розділення метаданих (онтологій) і даних (документів). Кожний концепт має візуальну форму екранної сторінки, розділеної на дві частини так, що в одній частині міститься інформація пов'язана з концептом, а у другій частині показано зв'язок цього концепту з іншими в системі. Недоліками системи є: простота онтологічної частини – вона служить лише для зв'язку концептів між собою і вказує тип зв'язку. Крім того кожний концепт має невелику кількість атрибутів.

III. Використання термінології для представлення знань

Використання комп'ютерної техніки дає змогу перетворювати збірки термінів у бази знань, надаючи для цього необхідні технічні можливості. Проте такою інформацією важко користуватись, оскільки в основі її організації лежать принципи, які не передбачали системного представлення. Термінологічна інформація історично була орієнтована на використання людиною й представлення у формі друкованих збірок. Оптимальною формою групування була вибрана алфавітна, оскільки вона забезпечувала найшвидший доступ до інформації й уможлилювала реалізацію в друкованому вигляді.

Використання комп'ютерних засобів створює можливості представлення та маніпулювання термінологічною інформацією, яка може бути значно доповнена й розширена аспектами необхідними з погляду теорії представлення знань за допомогою онтологій. Існують відомі приклади реалізації за таким принципом, проте можливості представлення знань у таких системах є дуже обмеженими.

Можна доповнювати термінологічну чи онтологічну системи, що не змінює їх суті. З іншого боку, можна використати переваги обох систем, забезпечуючи одночасно доповнення їх додатковими можливостями чи інформацією. Це може бути варіант паралельної побудови двох систем, термінологічної бази даних й онтології в певній галузі знань із забезпеченням їх взаємного «зшиття», тобто побудови оболонки, яка дасть змогу здійснювати перехід з рівня термінології на рівень онтології з будь-якого елемента кожної з систем (термін чи поняття-концепт). Така робота передбачає доопрацювання термінологічної бази на понятійному рівні й визначення структури онтологічного представлення окремої галузі.

Опрацювання термінологічної бази передбачає виявлення та фіксацію онтологічних відношень між окремими термінами. Наукова мова використовує іменники для позначення предметів, об'єктів та явищ. Кожний із предметів чи явищ має певні властивості, здатний вступати у відношення взаємних стосунків з іншими предметами та явищами, що описується за допомогою прикметників, які означають:

- а) властивості предмета, що висловлює його причетність до певного явища, процесу або іншого предмета, зокрема до неперехідної дії;
- б) здатності чи призначеності предмета активно діяти;
- в) пасивні здатності чи призначеності предмета піддаватися дії;
- г) стани предмета внаслідок дії на нього, яка ще триває;
- г) стани предмета, спричинені дією, що закінчилася [9].

Предмети та явища можуть мати й часовий зв'язок, зумовлений причинно-наслідковими залежностями. Крім того, явища відбуваються в часі й можуть мати характеристики, що характеризують їх перебіг. Тому використовують дієслова та їх словосполучення з іншими частинами мови, які визначають причинно-наслідкові зв'язки та часові характеристики явищ та об'єктів.

Зрозуміло, якщо будувати систему представлення знань на основі термінології, необхідно значно розширити описову частину термінологічної інформації. Кожний термін у такому випадку повинен бути описаний не тільки семантично, синтаксично, але й онтологічно, тобто описано його відношення до концептів, у які він входить, а також до концептів, у які він не входить, але з якими концепт, у який він входить, має семантичні, причинно-наслідкові та інші зв'язки.

Здається, що онтологію можна було б будувати на основі термінології, адже кожен термін має визначення (тлумачення), але це неможливо, оскільки між концептами-термінами немає чітких логічних зв'язків, вони можуть мати декілька значень і використовуватись для опису різних концептів.

Крім того, слід урахувати, що терміни в системі знань виконують подвійну роль – як номінатори концептів і як номінатори властивостей, рис концептів. Концепти, номіновані термінами, можуть мати ієрархію, деякі з них можуть бути складовою частиною інших концептів.

Якщо мовна інформація щодо термінів є доступною (міститься в термінологічних словниках), то онтологічної інформації немає. Понадто, онтологію тої чи іншої галузі треба будувати, приймаючи рішення про концептуальну структуру, ієрархію й концептуальну сукупність.

Таким чином, для представлення системи знань спочатку треба будувати онтологію, на яку потім можна «накласти» термінологічну верству. Побудована таким чином система знань чимось нагадує друковану книжку, де спочатку йде онтологічна інформація (розділи, параграфи, абзаци, кожен з яких несе певну кількість знань) а в кінці – словник термінів з посиланнями на місце в онтологічній структурі.

Структуру онтологічного представлення визначають галузю, для якої будують, та термінологією, якою представляється знання в цій галузі. Вона може бути закритою (обмежуватись лише даною галуззю) чи відкритою. Вона може містити окремі об'єкти різного рівня загальності й терміни, що описують їх: класи, зв'язки, функції, об'єктні константи. Це дає змогу описати кожний термін додатковими параметрами, що визначають місце його в ієрархії онтології, наслідування властивостей та їх обмеження [10].

IV. Висновок

Термінологія сама по собі не може забезпечити представлення знань з тієї чи іншої галузі, оскільки вона не покриває всього обсягу знань, з одного боку, і не уможливорює систематизувати інформацію без використання метатекстових засобів. Проте як засіб інтерфейсу й забезпечення доступу в системах представлення знань вона може бути дуже корисною, оскільки забезпечує легкий семантичний доступ до окремих фрагментів.

1. *Guarino, N. Ontologies and Knowledge Bases. Towards a Terminological Clarification [Text] / N. Guarino, P. Giarretta // Towards Very Large Knowledge Bases. – IOS Press, Amsterdam, 1995. – P.14–21.*
2. *Ahmad, K. Terminology workbenches and the engineering of special languages [Text] / K. Ahmad // Proceedings of the 3rd TermNet Symposium on Terminology in Advanced Microrcomputer Applications. – TermNet, Vienna, Austria. 1994. – P. 5–52.*
3. *Gruber Thomas R. A Translation Approach to Portable Ontology Specifications [Text] / T. Gruber // Knowledge Acquisition. – 1993. – 5(2). – P.199–220.*
4. *van Heijst, G. Using explicit ontologies in KBS development [Text] / G.van Heijst, A. Schreiber, B. Wielinga // International Journal of Human and Computer Studies. – 1997. – 46(2/3). – P.183–292.*
5. *Fellbaum, C., WordNet – An Electronic Lexical Database [Text] / C. Fellbaum editor. – MIT Press, 1998. – 423 p.*
6. *Mahesh, K. A situated ontology for practical NLP / K. Mahesh and S. Nirenburg // Proceedings of the IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada. – 1995.*
7. *Knight, K. Building a large-scale knowledge base for machine translation [Text] / K. Knight, S. K. Luk // Proceedings of the 12th National Conference on Artificial Intelligence (AAAI'94), Seattle, Washington. – 1994. – V. 1. – P. 773–778.*
8. *Swartout, B., Toward distributed use of large-scale ontologies [Text] / B. Swartout, R. Patil, K. Knight, T. Russ // Proceedings of the 10th Workshop on Knowledge Acquisition for Knowledge-Based Systems, Banff, Canada. – 1996.*
9. *Ярема С. Структура термінного гнізда та термінні прикметники [Текст] / С. Ярема // Проблеми української термінології: зб. наук. праць – Львів: Ліга-Прес, 2004. – 228 с.*
10. *Gruber Thomas R. The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases [Text] / T. Gruber // Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA: Morgan Kaufmann, 1991. – P. 601–602.*