

УДК 61.001.8:57.089

ПРО ЗАГАЛЬНЕ ВПОРЯДКУВАННЯ МЕДИКО-БІОЛОГІЧНИХ ЗНАНЬ. СТВОРЕННЯ ОНТОЛОГІЇ

О. П. Мінцер, Л. Ю. Бабінцева

Національна медична академія післядипломної освіти імені П. Л. Шупика

Розглянуто питання узагальнення понять онтології для медико-біологічних досліджень. Підкреслено, що темпи впровадження онтологічних технологій залишаються повільними. Тому для забезпечення їх широкого впровадження пропонується будувати онтосистеми з використанням трьох основних принципів: підтримки модульності, колективної роботи, версій і системи погоджень; середовища відкритої мови роботи з онтологіями; формування внутрішнього представлення онтології.

Ключові слова: впорядкування знань, онтологія, медико-біологічні дослідження, онтографи, декларативні та процедурні знання.

ОБ ОБЩЕМ УПОРЯДОЧЕНИИ МЕДИКО-БИОЛОГИЧЕСКИХ ЗНАНИЙ. СОЗДАНИЕ ОНТОЛОГИИ

О. П. Минцер, Л. Ю. Бабинцева

*Национальная медицинская академия последипломного образования
имени П. Л. Шупика*

Рассмотрены вопросы обобщения понятий онтологии для медико-биологических исследований. Подчеркивается мысль о том, что темпы внедрения онтологических технологий остаются медленными. Основная причина – онтологии должны строиться высококвалифицированными специалистами в своей области, а языки представления онтологий являются сложными, техничными и далекими от этих областей знаний. Поэтому для обеспечения их широкого внедрения предлагается строить онтосистемы с использованием следующих трех основных принципов: поддержки модульности, коллективной работы, версий и системы согласований; среды открытого языка работы с онтологиями; формирования внутреннего представления онтологии.

Ключевые слова: упорядочение знаний, онтология, медико-биологические исследования, системы знаний, онтографы, декларативные и процедурные знания.

ON THE GENERAL RANKED MEDICAL AND BIOLOGICAL KNOWLEDGE. CREATION OF ONTOLOGY

O. P. Mintser, L. Yu. Babintseva

National Medical Academy of Postgraduate Education by P. L. Shupyk

The questions of generalization of the concepts of ontology for biomedical research are considered. It is emphasized that the pace of implementation ontotechnology remain slow. Therefore, to ensure their widespread implementation is proposed to build ontosystems using three basic principles: support for modularity, collaboration, versioning and permissions system; OSE language work on ontologies; the internal representation of the ontology.

Key words: ordering of knowledge, ontology, medical and biological research, ontographs, declarative and procedural knowledge.

Вступ. Термін онтологія прийшов із філософії (походить від спроби Аристотеля класифікувати предмети у світі), де він використовується для позначення системи знань, що відносяться до навколишнього світу (на противагу системі знань про внутрішній світ

людини). Іншими словами, онтологія – це наука про природу речей і взаємозв'язки між ними. В контексті інформаційних технологій представлення знань терміном онтологія можна визначити деякий механізм, спосіб, що використовується для опису певної області

знань (предметної області або домену), зокрема базових понять цієї області, їх властивостей і зв'язків між ними.

На жаль, на сьогодні не існує загальноприйнятого визначення онтології. Найчастіше користуються словами Грубера: «Онтологія – це явна специфікація концептуалізації» [1]. Тут концептуалізація означає абстрактне уявлення предметної області. Поширене також визначення: «Онтологія – спільне розуміння деякої області інтересу».

Загальноприйнято, що під онтологією можна розуміти такі смислові кластери понять, як семантичний базис у визначенні змісту; загальну логічну теорію, що складається зі словника та набору тверджень на мові логіки; основу для комунікації між людьми та комп'ютерними агентами.

Практично всі моделі онтологій у тій чи іншій мірі містять концепти (поняття, класи, сутності, категорії), властивості концептів (слоти, атрибути, ролі), відносини між концептами (зв'язки, залежності, функції) та додаткові обмеження (визначаються аксіомами, в деяких парадигмах фасетами).

Для представлення елементів у предметній області, тобто елемента даного концепту, використовується термін «екземпляр». Онтологія разом із безліччю окремих екземплярів складає базу знань. Зауважимо, що між онтологіями та базами знань грань нечітка. Не можна впевнено сказати, де «закінчується» онтологія та починається база знань.

Вважається, що онтологія являє собою базу знань, що описує факти, які передбачаються завжди істинними в рамках певної спільноти на основі загальноприйнятого сенсу використовуваного словника. База знань же може описувати факти та твердження, істинність яких залежить від стану змінних зовнішнього середовища. У даному питанні поки немає повної ясності.

Крім цього, однією із негативних властивостей деяких розроблених систем є те, що коло їх застосування досить вузьке. Останнє обумовлено або макетністю версії, що виконана для певних дослідницьких цілей, або обмеженістю локальною мережею певної установи.

Мета роботи: узагальнення ключових понять онтології для медико-біологічних досліджень.

Матеріали та методи. Методологія дослідження базувалася на аналізі, систематизації та виявленні загальних закономірностей і чинників впливу на проблему впорядкування медико-біологічних знань.

Результати та їх обговорення. Потреба розроблення онтології виникла в зв'язку з необхідністю

спільного використання людьми або програмними агентами загального розуміння структури інформації; з можливістю повторного використання знань у предметній області; з можливістю робити допущення в предметній області явними; відокремлення знань у предметній області від оперативних знань; аналізу знань у предметній області тощо.

Іншими словами, онтології дозволяють представити нові поняття так, що вони стають придатними для комп'ютерного оброблення з метою знаходження зв'язку між ними й описами вже відомих класів, відносин, властивостей і об'єктів реального світу.

Досвід використання онтологій у науці. Системи, засновані на знаннях, представляють досить широкий напрям. Вони мають власні методи та принципи, істотно впливають на розвиток інформаційних технологій. Один із істотних принципів, що склалися в цій галузі, – поділ декларативних (непроцедурних) і імперативних (процедурних) знань і створення баз декларативних знань. Тенденція такого поділу в програмуванні привела до принципів об'єктно-орієнтованого та логічного програмування.

В базах даних декларативні знання виділяються у вигляді описів схем баз даних. Особливе місце бази декларативних знань отримали в зв'язку з розвитком Інтернет. Тому в 1991 році виникли умови для появи терміну «онтологія» [1]. Грубер Т. Р. ввів його для позначення зв'язкового фрагмента декларативного знання та використання в інформаційних технологіях.

За останні десятиліття побудова та впровадження онтологій переросли з суто філософської дисципліни в інформаційні технології, що інтенсивно розвиваються та вже знайшли своє застосування, в тому числі й у біології. Онтології є ефективним засобом навігації у величезних інформаційних масивах, таких як, наприклад, геномі, траскриптомні та протеомні бази даних. Незважаючи на величезний потік інформації в сучасній біології, багато її фундаментальних аспектів, зокрема біології розвитку рослин, досі не ясні.

Безперечно, що використання біоонтологій може бути ефективною ланкою між валовим отриманням експериментальних даних і структурованим науковим знанням. Сучасні біоонтології, такі як Gene Ontology та Plant Ontology, є величезними контрольованими словниками пов'язаних між собою термінів – концептів, що дозволяють досліднику оперативно знайти конкретні приклади – гени. Зручність таких «біологічних словників», зрозумілих для комп'ютера, безперечно, не тільки з точки зору формалізації біологічного знання та автоматизації процесу пошу-

ку, але й для вивчення практично будь-якої біологічної дисципліни або окремого біологічного процесу.

Як приклад, застосування онтологій дозволяє набагато швидше зорієнтуватися при вивченні дисципліни біологія розвитку рослин, особливістю якої є експоненціально зростаючий потік молекулярно-біологічної інформації. Онтологічний аналіз виступає високоефективною альтернативою стандартно вживаному пошуку в різних пошукових системах, що часто призводить до надмірності інформації.

Досвід використання онтологій у медицині. В багатьох дисциплінах сьогодні розробляються стандартні онтології, які можуть використовувати експерти з предметних областей для спільного застосування й анування інформації у своїй галузі. В медицині онтології гірше вивчені та про них мало що відомо. Лише останнім часом інтерес до онтологій в нашій країні зростає. Так, у галузі медицини створено великі стандартні структуровані словники – SNOMED і семантична мережа Системи Уніфікованого Медичної Мови (the Unified Medical Language System). Також з'являються великі загальноцільові онтології. Наприклад, програма ООН із розвитку (the United Nations Development Program) і компанія Dun & Bradstreet об'єднали зусилля для розроблення онтології UNSPSC, що надає термінологію товарів і послуг.

У роботі Ф. Москаленка, А. Клещова, М. Черняховської (2010) запропонована концепція комп'ютерного банку знань із медичної діагностики. Інформаційне наповнення спеціалізованої бази знань містить три онтології: онтологію спостережень у галузі медичної діагностики, онтологію бази знань (захворювань) з медичної діагностики й онтологію історій хвороби, а також три класи інформаційних ресурсів у різних галузях медицини – бази спостережень, бази знань, і бази даних (пацієнтів), що відповідають цим онтологіям.

Аналіз онтологічних систем для медичної діагностики, прогнозування результатів захворювань і лікування, які розробляються в останні роки, показав, що застосовувані в них онтології медичної діагностики є порівняно простими та не відображають такі повсюдно використовувані лікарями в своїй практиці знання предметної області, як знання про причини захворювань; знання про різні типи причинних зв'язків між ознаками та захворюваннями; знання про вплив подій на значення ознак при захворюваннях і в здорових пацієнтів; знання про різні варіанти змінення значень ознак, що залежать від анатомо-фізіологічних особливостей пацієнтів.

Також дотепер не відпрацьовано механізм обліку залежності значення медичної інформації від часу її використання. Водночас темпоральні (часові) аспекти відіграють істотну роль у медичному знанні. Хвороби характеризуються складними паттернами симптомів, що з'являються й зникають у певній часовій послідовності. В багатьох випадках цей темпоральний паттерн симптомів є істотним для прийняття рішень про характер патологічного процесу.

На жаль, заповнення бази знань здійснюється без необхідних для лікарів компонентів, а саме: оцінювання валідності отриманої інформації, релевантності та пертинентності відомостей. Очевидно, будь-які онтології повинні супроводжуватися підсистемами довіри, що показують користувачам (лікарям, біологам), якими знаннями володіє система діагностики, а також підсистемами пояснень, що роз'яснюють користувачеві, на основі яких міркувань і знань системою пропонуються ті чи інші рішення.

Проблеми формування та використання бібліотек онтологій. Темпи впровадження онтологічних технологій залишаються повільними, оскільки онтології повинні будуватися висококваліфікованими фахівцями у своїй галузі, а мови представлення онтологій є складними, технічними та далекими від цих галузей знань.

Для формування найпростіших онтологій у вигляді класифікацій були побудовані графічні редактори, що спрощують роботу з такими онтологіями та роблять їх наочнішими. Це визначило активність побудови класифікаційних онтологій у багатьох галузях знань. У свою чергу, після обґрунтування логіки класифікаційних онтологій стали з'являтися потреби і в складніших онтологіях, у бібліотеках онтологій, у нових методологіях їх побудови.

Інша причина відсутності масового використання онтологічних технологій у Веб у даний час полягає в тому, що масовий користувач не бачить безпосереднього ефекту від використання онтологій, а від нього ці технології вимагають великих зусиль по семантичній розмітці тієї інформації, яку він виставляє в Веб. Практично безмежним є подання онтологій у вигляді онтографу. Очевидно, необхідні подальші дослідження щодо можливої кластеризації існуючих графічних уявлень онтологій. Це ж стосується проблем семантичних реалізацій онтологій.

Як мінімум, для подолання перелічених труднощів потрібно розробити Веб-середища та інструменти, в яких користувачі зможуть створювати власні семантично розмічені сторінки та мови запитів до них із новою функціональністю у вигляді інформаційних систем.

Нові перспективи розвитку онтологій можуть з'явитися в зв'язку з використанням Wiki-технології для формування та використання бібліотек онтологій. Як відомо, в цій технології вироблено об'єднання основних принципів Інтернет і Веб: принципу розподілу ресурсів, відсутності єдиного центру управління, наповнення змісту мережі самими користувачами мережі Інтернет тощо.

Перевага технологій Wikipedia для створення та використання бібліотек онтологій, насамперед, полягає у їхньому соціальному характері. Друга, і дуже важлива, – це можливість семантично розмітити лише сторінки-шаблони, звільняючи інших користувачів від неприємної роботи семантичної розмітки своїх сторінок. Зрозуміло, ця розмітка безглузда, якщо немає мови запитів, для відповідей на які вона використовується.

Література

1. Gruber T. R. A Translation Approach to Portable Ontology Specification / T. R. Gruber // Knowledge Acquisition. – 1993. – Vol. 5. – P. 199–220.
2. Палагин А. В. Онтологические методы и средства обработки предметных знаний : монография / А. В. Палагин, С. Л. Кривый, Н. Г. Петренко. – Луганск : ВНУ им. В. Даля, 2012. – 324 с.
3. Мінцер О. П. Використання принципів медичної онтології для побудови сценарних моделей післядипломної освіти лікарів і провізорів / Мінцер О. П., Денисенко С. В., Стрижак О. Є. // Медична інформатика та інженерія. – 2013. – № 2. – С. 18–23.

Висновки. 1. Онтосистеми й онтопроекти створюються та розвиваються вже більше 10 років. Успіх і значимість цього напрямку очевидні. Проте темпи впровадження онтотехнологій, особливо в медичному напрямі, все ще невеликі.

2. Для широкого впровадження онтотехнологій пропонується будувати онтосистеми з використанням таких основних трьох принципів: з підтримкою модульності, колективної роботи, версій і системи погоджень; середовища відкритої мови роботи з онтологіями, що формується самими користувачами, в міру поповнення бази онтологій; формування внутрішнього представлення онтології, для семантичного аналізу виразів мови, при формуванні відповідей на запити до онтології та її налагодженні, для міжкомп'ютерного обміну онтологіями та для застосування онтологій у додатках.

4. Клещёв А. С. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия онтология / А. С. Клещёв, И. Л. Артемьева // НТИ. – Сер. 2. – 2001. – № 2. – С. 20–27.
5. Клещёв А. С. Математические модели онтологий предметных областей. Часть 2. Компоненты модели / А. С. Клещёв, И. Л. Артемьева // НТИ. – Сер. 2. – 2001. – № 3. – С. 19–28.
6. Клещёв А. С. Математические модели онтологий предметных областей. Часть 3. Сравнение разных классов моделей онтологий / А. С. Клещёв, И. Л. Артемьева // НТИ. – Сер. 2. – 2001. – № 4. – С. 10–15.