

УДК 378.147:004.9:303.032].001.573

ІНВАРІАНТНА МОДЕЛЬ ПОДАННЯ ЗНАНЬ У СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ НА ОСНОВІ ОБ'ЄКТНО ОРІЄНТОВАНОГО ПІДХОДУ

О.А. Рижов, А.Н. Попов

Запорізький державний медичний університет

Авторами розроблено інваріантну модель подання фармацевтичних і медичних знань у системах дистанційного навчання на основі інтеграції поняттєвої структури знань та технології об'єктно орієнтованого аналізу предметної галузі. Використання такої моделі дає змогу поставити завдання розробки алгоритмічної бази для аналізу структури знань предметної галузі.

ИНВАРИАНТНАЯ МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В СИСТЕМАХ ДИСТАНЦИОННОГО ОБУЧЕНИЯ НА ОСНОВЕ ОБЪЕКТНО ОРИЕНТИРОВАННОГО ПОДХОДА

А.А. Рыжов, А.Н. Попов

Запорожский государственный медицинский университет

Авторами разработана инвариантная модель представления фармацевтических и медицинских знаний в системах дистанционного обучения на основе интеграции понятийной структуры знаний и технологии объектно ориентированного анализа предметной области. Использование такой модели позволяет поставить задачу разработки алгоритмической базы для анализа структуры знаний предметной области.

INVARIANT MODEL OF KNOWLEDGE REPRESENTATION IN DISTANCE LEARNING SYSTEMS ON THE BASIS OF OBJECT-ORIENTED APPROACH

A.A. Ryzhov, A.N. Popov

Zaporozhye State Medical University

The authors have developed invariant model of pharmaceutical and medical knowledge representation in distance learning systems on the basis of integration of conceptual structure of knowledge with the results of technologies of object-oriented analysis of subject field. Such model gives the possibility to set the task of the development of algorithmic foundation for the analysis of structure of knowledge of subject field.

Сучасний етап розвитку систем дистанційного навчання характеризується активним використанням інтелектуальних технологій подання знань і розробкою систем адаптивного управління процесом навчання [1; 2; 3]. Розробка системи подання знань потребує аналізу категоріального апарату. Вибір категорії об'єкта як основна інваріантна структура дає змогу здійснити часткову формалізацію фармацевтичних, медичних, біологічних знань на основі технологій об'єктно орієнтованого підходу з використанням можливостей універсальної мови моделювання (UML) на етапі концептуального проектування [4]. UML сьогодні активно використовують у стандарті HL7, біоінформатиці, системній біології, описі моделей біохімічних процесів. Проте застосування UML в чистому вигляді не дає змоги ефективно відобразити

ієрархію структурних відносин у фармацевтичних і біологічних системах. Водночас це завдання вирішують при використанні математичної моделі подання знань у вигляді онтологій. Створення онтології засноване на аналізі поняттєвої структури знань предметної галузі (ПрГ), відображеної в системі професійної мови відповідного наукового напряму. У статті здійснено спробу створити узагальнену інваріантну модель подання знань для розробки інструментальних систем побудови баз знань фармацевтичного й медико-біологічного профілів в адаптивних системах дистанційного навчання, заснованих на знаннях.

Мета статті – розробка інваріантної моделі подання знань у системах дистанційного навчання на основі інтеграції технологій об'єктно орієнтованого

аналізу знань ПрГ і аналіз поняттєвої структури знань, зафіксованих професійною мовою ПрГ.

Основна частина

Архітектура сучасних інтелектуальних навчальних систем (ІНС) і адаптивних навчальних систем (АНС) включає, як важливий елемент своєї структури, модель знань предметної галузі (МЗПГ), яка реалізується у вигляді бази знань. Еволюційний розвиток таких систем привів до дивергенції МЗПГ і, відповідно, до формування декількох БЗ, що виконують різні завдання й забезпечують управління процесом навчання студента при взаємодії із системою. Автор Г.В. Рибіна у своєму огляді [2], присвяченому навчальним інтегрованим експертним системам, виділяє три основні моделі знань, які можуть входити до складу ІОС або АОД, це: модель того, кого навчають (М1), модель навчання (М2), модель пояснення (М3). Модель предметної галузі розглядається у складі М1 й обмежується завданнями навчальної дисципліни. Ми вважаємо, що модель предметної галузі необхідно виділити в окрему модель М0 і в її склад мають бути інтегровані знання всіх навчальних курсів з однієї предметної галузі, викладання яких здійснює кафедра. Кожна із цих моделей має забезпечуватися своєю БЗ. Виходячи з різноманіття моделей, які використовують в одній навчальній системі, набуває актуальності питання про використання інваріантної математичної моделі (структури) подання знань, що дає змогу використовувати стандартні механізми (процедури) логічного виводу для різних моделей.

У роботі [5] проаналізовано можливості застосування в ІНС і АНС семи класів моделей подання знань: логічні, продукційні, фреймові, мережеві об'єктно орієнтовані, спеціальні, комплексні. Проведений аналіз дав змогу авторам [5] сформулювати основні вимоги до моделі знань:

1. Універсальність.
2. Наочність подання знань.
3. Однорідність.
4. Реалізація в моделі властивості активності знань.
5. Відкритість БЗ.
6. Відображення структурних відносин об'єктів ПрГ.
7. Використання системи семантичних шкал при реалізації проєкції знань.
8. Подання нечітких знань.
9. Багаторівневе подання знань на основі метамоделей.

Здійснюючи вибір математичної моделі подання фармацевтичних і медико-біологічних знань необхідно зважати на специфіку мислення фахівців даного

профілю. Розробкою бази знань навчальної системи в нашому випадку займатимуться викладачі кафедр відповідних дисциплін без залучення інженера-когнітолога. У зв'язку із цим в інтерфейсі інструментального середовища має бути відображена специфіка цієї галузі знань, а саме: системність, ієрархічність, наочність, вербальний опис (у межах професійної мови), обмеження на спільність застосування абстрактних понять (категорій), орієнтованих на реально існуючі об'єкти або процеси. Можливості математичної моделі подання знань онтології значною мірою дають змогу реалізувати вказані вимоги. Онтологія будується для реалізації яких-небудь конкретних цілей. Ми використовуємо системний принцип цільової функції системи, що розробляється. У цьому разі відразу накладаються обмеження на різні варіанти інтерпретації формалізованих знань, що є надзвичайно важливим для вирішення питання невизначеності. При побудові онтології можна використовувати методологію системного аналізу – стратифікацію [6], яка застосовує метод декомпозиції складних систем, а саме до такої категорії належать ці об'єкти. У результаті проведених операцій ми одержуємо ієрархічну структуру, описану в термінах предметної галузі. Проте онтологія, описуючи ієрархію понять або структуру складної системи (об'єкта), не відображає параметри й критерії, на основі яких ця ієрархія побудована. Онтологія спирається на інтерпретацію колективного суб'єкта – співтовариства професіоналів конкретної ПрГ. Водночас класифікації і систематика є найважливішими складовими фармацевтичних і медико-біологічних предметних галузей. Наявність у структурі БЗ ПрГ формалізованої інформації про ознаки і властивості даних систем або об'єктів розширює набір “інтелектуальних” операцій, які можна реалізувати на представленій БЗ, а отже, розширює і набір завдань, що вирішуються інтелектуальною системою.

Особливістю природничонаукових дисциплін, а саме до них належать ті, що розглядаються ПрГ, є експериментальний характер їх наукової діяльності. Знання, подані в БЗ і висновках, що одержують при вирішенні поставлених завдань, повинні мати можливість експериментальної перевірки. Технологія й методи цієї перевірки можуть бути відбиті в самій БЗ. Відомо, що точність даних, які одержує дослідник при вивченні об'єкта, значною мірою залежить від точності методу дослідження. Різні результати, отримані при використанні різних методів дослідження, можуть привести до різних висновків. Вирішення питання невизначеності й нечіткості представлення знань у

БЗ природничонаукових ПрГ може бути вирішено при вказівці методів (методик) дослідження параметрів або властивостей даних об'єктів, що має бути відбито в моделі подання знань.

Одним зі складних питань розробки баз знань є питання про однозначну семантичну інтерпретацію формалізованих знань ПрГ. Семантична інтерпретація наукових знань, поданих у вербальному або у формалізованому вигляді штучною мовою даної ПрГ, насамперед це може бути мова математики, залежить від декількох складових – це наукова школа, у рамках якої здійснювалася робота, мета й завдання, в результаті яких були отримані знання, методології дослідження. Наприклад, мета й завдання при дослідженні гормональної системи людини при біохімічних, фізіологічних, клінічних, фармакологічних дослідженнях абсолютно різні, з чого виходить, що інтерпретація формалізованих моделей також може бути різною. Для звуження різноманіття інтерпретації формалізованих знань ми пропонуємо запроваджувати поняття контексту, яке має включати чинники, що впливають на семантику термінів, понять і моделей ПрГ.

Аналіз літературних даних показує, що для початкового етапу формалізації медико-біологічних знань сьогодні можна ефективно використовувати об'єктно-орієнтовані технології, що набули свого розвитку в CASE-технологіях проектування інформаційних систем, наприклад UML [4]. У нашому дослідженні здійснено розробку структури метаданих БЗ, яка дає змогу описати на етапі концептуального й логічного аналізу реальні фармацевтичні та медико-біологічні системи й об'єкти.

Інваріантна модель подання знань

Структура інваріантної моделі подання фармацевтичних і медико-біологічних знань була розроблена на основі інтеграції уявлень про онтологію і опис об'єкта засобами UML.

Стандартно онтологію можна формально подати впорядкованою трійкою [5; 7; 8]:

$$O = \langle C, R, F \rangle, \quad (1)$$

де C – скінченна множина понять предметної галузі;

R – скінченна множина зв'язків ПрГ, що накладаються на терміни;

F – скінченна множина правил або функцій інтерпретації C і R .

Для відображення “чистої” онтології в моделі універсального абстрактного класу “Universal Class” використовується клас лексем – “Lexeme” і клас, що описує структуру вузла онтології, – “Ontology Node”.

Основний масив знань ПрГ, що розглядаються, відображений множиною професійних мов L_p . Кожна професійна мова має свою термінологію T_p або терміносистему TS_p . У свою чергу, термінологія визначена на безлічі лексем Lex_p . Лексема складається з одного або декількох слів даної мови W_p . Тоді:

$$Lex_p \subseteq W_p; \quad (2)$$

$$T_p \subseteq Lex_p \mid T_p \subseteq TS_p; \quad (3)$$

Множина термінів ділиться на підмножину понять ПрГ C_p і множину інтерпретуючих термінів $T_{p,i}$:

$$C_p \subseteq T_p; \quad (4)$$

$$T_{p,i} \subseteq T_p. \quad (5)$$

“Поняття” в інтерпретації Н.Д. Ващенко [9] визначається як “фрагмент знань, що є узагальненою інформацією про безліч об'єктів, для яких визначені процеси розпізнання і генерації моделей конкретних об'єктів”. Кожне поняття номінується терміном відповідної ПрГ.

Коли $C_p \neq 0$ і $R \neq 0$, то онтологія O представляє граф:

$$O = \{C_p, R\}, \quad (6)$$

в якому вузли представлені поняттями, а зв'язки елементами множини R . В універсальному класі R зв'язки та відношення визначені на Lex_p :

$$r_i \in Lex_p. \quad (7)$$

Онтологія універсального класу будується на основі словника термінів, побудованого з використанням класу “Lexeme”, і самій онтології, яка створюється на основі класу “Ontology_Node”.

Клас “Lexeme” призначений для формування й розширення словника термінів при використанні операцій *Створити_лексеми()*, *Змінити_лексеми()* і *Видалити_лексеми()*. Для вирішення питання однозначної змістовної інтерпретації введена функція генерації унікального ідентифікатора *IdTerm*, яка враховує контекст, при якому використовується термін:

$$lex_{p,i} \in Lex_p; \quad (8)$$

$$IdTerm(lex_{p,i}, Context_n) \rightarrow Id_{lex} \quad (9)$$

де $Context_n$ – контекст, при якому застосовується термін;

Id_{lex} – унікальний ідентифікатор лексеми $lex_{p,i}$.

Як контекст може розглядатися наукова дисципліна, сфера практичної діяльності, навчальний курс тощо. У кожному випадку зміст і значення терміна можуть бути різними, хоча частина змісту в різних контекстах може бути спільною. Наприклад, поняття “система гормональної регуляції” в біохімії, фізіології, фармакології, клінічній дисципліні, курсі біології

середньої школи матиме різний зміст і глибину розкриття, оскільки в кожній предметній галузі реалізуються свої специфічні цілі й вирішуються певний клас завдань.

Абстрактний клас “*Ontology_Node*” призначений для побудови онтології на основі операцій *Додати_вузол()*, *Змінити_вузол()*, *Видалити_вузол()*. Можливо використовувати операції безліччю структурованих понять – *Додати_ніддереву()* і *Видалити_ніддереву()*. Логічний висновок на цій структурі можна організувати на основі множини операцій *Навігація()* і *Пошук()*. Для отримання інформації про пов’язані поняття (базові поняття, зміст поняття, сестринські вузли) досить відкрити заданий за унікальним ідентифікатором примірник і звернутися до відповідних властивостей, які подані у вигляді списків Id_{lex} : *Батьківські_вузли*, *Дочірні_вузли*. Інформація про його зв’язки знаходиться всередині заданого вузла і, в даному випадку, відсутня необхідність у додаткових операціях зі знаходження пов’язаних вузлів онтології. При цьому є певна надмірність даних, яка компенсується легкістю доступу до пов’язаних елементів графа.

Поняття, описане в онтології, залежно від рівня ієрархії даної онтології або від контексту може розглядатися як поняття різних типів: простий тип – поняття, визначене на графі онтології; складний тип – поняття-об’єкт, поняття-властивість, поняття-зв’язок, поняття-контекст, визначені на абстрактному класі “*Object*”.

Універсальність запропонованої моделі полягає також у тому, що поняття всіх типів можуть мати структуру, тобто бути вузлами дерева понять свого типу. При цьому поняття-об’єкт може розглядатися в локальних контекстах і також бути учасником різних автономних дерев, в яких воно розкрито по-різному, залежно від складності і повноти опису ПрГ. Особливість моделі полягає також у тому, що залежно від контексту властивість поняття-об’єкта може саме розглядатися як поняття-об’єкт і займати свою нішу в дереві цього контексту. При цьому значення властивостей також можна розглядати в контексті заданої деталізації. Таким чином, кожен вузол дерева в заданому контексті є примірником класу або об’єктом, у якому, з одного боку, поміщена інформація про положення об’єкта всередині семантичної мережі або його “координати”. Задавши список батьківських та дочірніх понять, ми отримуємо назву об’єкта в поточному контексті. З іншого боку, повне уявлення про об’єкт – його назва, статичні характеристики і його поведінка або методи дослідження. При використанні UML-нотації всі зв’язки поняття, його зміст, а також статичні характеристики об’єкта визначуваного по-

няття заносяться в слоти властивостей “Універсального класу”. За допомогою метаданих такої структури ми маємо можливість формалізації об’єктів слабоформалізованої природної ПрГ. При цьому знання ПрГ існують у вигляді онтології, у вузлах якої знаходяться повноцінні об’єкти, а також інформація про зв’язки, властивості й операції над цим об’єктом. Важливо також те, що ця модель дає змогу описувати ієрархічну структуру об’єктів.

Структура абстрактного класу “*Object*” відповідає представленням мови UML, тобто складається з імені *NameObj*, властивостей Pr_{Obj} і опису операцій Op_{Obj} або методів:

$$Object = \langle NameObj, Pr_{Obj}, Op_{Obj} \rangle. \quad (10)$$

Ім’я об’єкта відповідає поняттю, тобто імені вузла онтології:

$$NameObj \subseteq C_p. \quad (11)$$

Властивості об’єкта Pr_{Obj} представлені списком властивостей:

$$Pr_{Obj} = \langle pr_1, pr_2, \dots, pr_k, \dots, pr_m \rangle. \quad (12)$$

Кожна властивість представляє пару: ім’я властивості *NamePr* і значення *Value*:

$$pr_k = \langle NamePr, Value \rangle; \quad (13)$$

$$NamePr \in Lex_p. \quad (14)$$

Значення може бути просто терміном, об’єктом або універсальним класом:

$$Value = \exists Value(Value \in T_p \vee Value \in Object \vee Value \in UniversalClass). \quad (15)$$

Кожна властивість об’єкта обслуговується трьома типами операцій:

$$Op_{Obj} = \{ Op_{Obj, Serv, k}, Op_{Obj, Model, k}, Op_{Obj, Pract, k} \}, \quad (16)$$

де перший індекс *Obj* позначає унікальний ідентифікатор даного об’єкта;

другий індекс позначає тип операції;

третій індекс *k* – позначає властивість *k* даного об’єкта.

Тип операцій з індексом *Serv* – представляє скінченна множина стандартних службових операцій, що обслуговують властивість об’єкта.

Враховуючи особливості подання декларативних знань, для опису яких призначена ця інваріантна модель, блок операцій UML об’єкта розширюється додатково двома типами операцій: перший тип з індексом *Model*, визначений як скінченна множина операцій і на рис. 1 названий *Теоретичний_метод()*. Він дає змогу описати теоретичну модель операцій

над властивістю k . Другий тип з індексом $Pract$ визначений як скінченна множина операцій і на рис. 1 названий *Практичний_метод()*. Він посилається

на конкретний практичний метод даної ПрГ і є інтерфейсом для виходу на діалог з користувачем для опису значення властивості k .

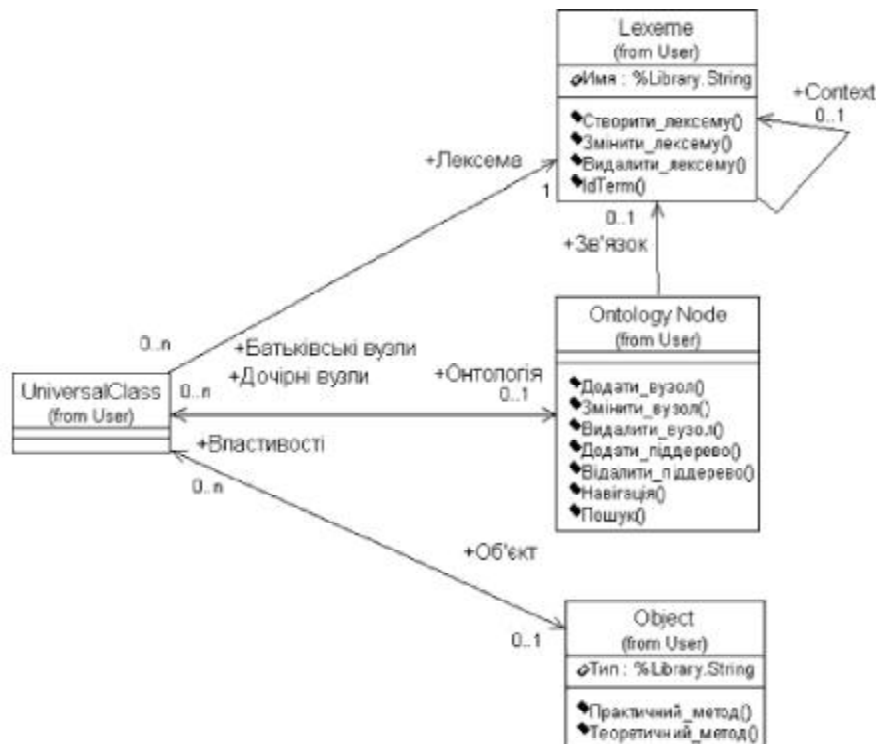


Рис. 1. UML-діаграма класів медико-біологічної моделі знань.

Тип операцій с індексом $Model$ представляє математичну або логічну модель у вигляді безлічі формул.

$$Op_{Obj,Model,k} = \exists op_{Obj,Model,k} (op_{Obj,Model,k} \in Lex_p \vee op_{Obj,Model,k} \in Object). \quad (17)$$

Тип операцій с індексом $Pract$ – представляє назву методу, методики або алгоритму, який має бути реалізований виконавцем для набуття значення властивості k у реальному житті:

$$Op_{Obj,Pract,k} = \exists op_{Obj,Pract,k} (op_{Obj,Pract,k} \in Lex_p \vee op_{Obj,Pract,k} \in Object). \quad (18)$$

Імена всіх даних структурних елементів мають бути визначені словником термінів ПрГ. Розширене ім'я об'єкта представляє унікальний ідентифікатор, що дає змогу отримати доступ до всієї інформації про об'єкт. Внутрішнє подання об'єкта в базі знань відображається віртуальним денотатом [10]. Такий підхід дає змогу використовувати різні вхідні словники різними мовами – природними і формальними при побудові унікальної структури об'єкта, не залежної від мовної форми його подання.

Висновки. Представлена інваріантна модель, отримана в результаті структурної інтеграції онтології й об'єктно орієнтованого опису знань, розширює можливість інтерпретації понять, а також дає змогу роз-

робити інструментальні засоби для створення баз знань обмеженою природною мовою. Використання цієї моделі можливе для вирішення ряду питань і завдань, що виникають при розробці й супроводі інтелектуальних і адаптивних навчальних систем.

1. Інваріантна модель дає можливість проводити зіставлення структури знань про об'єкт у різних контекстах і різних ПрГ.

2. Об'єктно орієнтовані методи проектування можуть використовуватися для формалізації знань фармацевтичних, медико-біологічних ПрГ. Наявність стандартів, що досить повно описують об'єкти, які представляють структуру, можна розглядати як специфікацію для розробки інструментальних систем, що дозволяють візуалізувати процес аналізу знань.

3. Використання такої моделі даних розширює коло фахівців з фармації, медицини і біології, які можуть брати участь у створенні баз знань на основі розглянутої моделі.

4. Трудомісткість розробки таких баз знань виправдовується можливістю неодноразового використання БЗ і баз навчальних елементів у різних курсах і системах, а також реінженерингу навчальних елементів і БЗ, розробки алгоритмів предметної інтеграції знань при плануванні програм навчальних курсів.

5. На основі цієї моделі можлива розробка змістовної моделі знань студента фармацевтичного й меди-

ко-біологічного профілю, яка надалі може бути використана для якісної оцінки знань вивченого.

Література

1. Brusilovsky P. Adaptive and Intelligent Web-based Educational Systems / P. Brusilovsky // *International Journal of Artificial Intelligence in Education*. – 2003. – № 13. – P. 156-169.
2. Рыбина Г.В. Обучающие интегрированные экспертные системы: некоторые итоги и перспективы / Г.В. Рыбина // *Искусственный интеллект и принятие решений*. – 2008. – № 1. – С. 22-46.
3. Федорук П.І. Адаптивна система дистанційного навчання та контролю знань на базі інтелектуальних Інтернет-технологій / П.І. Федорук. - Івано-Франківськ : Видавничо-дизайнерський відділ ЦІТ Прикарпатського національного університету ім. Василя Стефаника, 2008. – 315 с.
4. Рамбо Дж. UML : специальный справочник / Дж. Рамбо, А. Якобсон, Г. Буч. – СПб. : Питер, 2002. – 656 с.
5. Башмаков А.И. Интеллектуальные информационные технологии : учеб. пособие / А.И. Башмаков, И.А. Башмаков. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.
6. Теория систем и системный анализ в управлении организациями: справочник : учеб. пособие / под ред. В.Н. Волковой и А.А. Емельянова. – М. : Финансы и статистика, 2006. – 848 с.
7. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, Хорошевский. – СПб. : Питер, 2000. – 384 с.
8. Палагин А.В. Архитектура онтологоуправляемых компьютерных систем / А.В. Палагин // *Кибернетика и системный анализ*. – 2006. – № 2. – С. 111-124.
9. Ващенко Н.Д. Автоматизация процессов формирования понятий / Н.Д. Ващенко // *Автоматизация информационного обеспечения научных исследований* / под ред. А.А. Стогния. – Киев : Наукова думка, 1990. – С. 96-111.
- Рижов О.А. Модель представлення знань на основі понять для комп'ютерних систем навчання / О.А. Рижов // *Медична інформатика та інженерія*. – 2008. – № 2. – С. 83-88.