

## ПОБУДОВА МОДЕЛІ ПРЕДМЕТНОЇ ОБЛАСТІ ДЛЯ ІНТЕЛЕКТУАЛЬНИХ МЕДИЧНИХ СИСТЕМ. ЧАСТИНА 1.

А.П. Алпатов, Ю.О. Прокопчук

*Український державний хіміко-технологічний університет, ІТМ НАНУ і НКАУ,*

*itk3@ukr.net*

У статті розглядаються питання побудови моделі природної предметної області для використання в інтелектуальних медичних системах. Основне завдання полягає в розробці достатньо простого і, разом з тим, ефективного формалізму для представлення інформаційних об'єктів предметної області. У першій частині розглядаються найбільш загальні питання побудови моделі предметної області, такі як загальна структура моделі, подання тестів і емпіричних даних.

**Ключові слова:** банк знань, конфігуратор тесту, метод граничних узагальнень, модель предметної області.

## ПОСТРОЕНИЕ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ ИНТЕЛЕКТУАЛЬНЫХ МЕДИЦИНСКИХ СИСТЕМ. ЧАСТЬ 1.

А.П. Алпатов, Ю.А. Прокопчук

*Украинский государственный химико-технологический университет, ИТМ НАНУ и НКАУ*

В статье рассматриваются вопросы построения модели естественной предметной области для использования в интеллектуальных медицинских приложениях. Основная задача состоит в разработке достаточно простого и, вместе с тем, эффективного формализма для представления информационных объектов предметной области. В первой части рассматриваются наиболее общие вопросы построения модели предметной области, такие как общая структура модели, представление тестов и эмпирических данных.

**Ключевые слова:** банк знаний, конфигурактор теста, метод предельных обобщений, модель предметной области.

## CONSTRUCTION OF SUBJECT DOMAIN MODEL FOR USE IN INTELLECTUAL MEDICAL SYSTEMS. PART 1.

A.P. Alpatov, Yu.O. Prokopchuk

*Ukrainian State Chemical-Technological University*

Questions of construction of a subject domain model for use in intellectual medical systems are considered. The main task is in the development of rather simple and effective formalism for representation of information object of subject domain. In the first part are considered the most common questions of creating the model of subject domain, in particular, general structure of the model, representation of tests and empiristic data.

**Key words:** model of subject domain, bank of knowledge, configurator of the test, method of limiting generalizations.

**Вступ.** Одним з важливих напрямків інформатики на сучасному етапі є створення багатоцільових банків знань (БкЗ). БкЗ є ресурсом, що об'єднує всю інформацію, яка використовується в наукових дослідженнях, освітній та практичній діяльності в рамках конкретної предметної області (ПрО). Отже концепція багатоцільового банку знань лежить у руслі створення Єдиного інформаційного науково-освітнього простору [1-10].

Призначення БкЗ у медицині: розв'язання широко-го класу інформаційних завдань у клінічній практиці, освіті та наукових дослідженнях; створення шпитальних систем нового покоління - інтелектуальних ШІС; створення телемедичних систем (ТМС) нового покоління - інтелектуальних ТМС; створення навчальних тренажерів нового покоління - інтелектуальних тренажерів; створення комп'ютерно-орієнтованих

описів клінічних теорій, які забезпечать вирішення попередніх завдань.

БкЗ є сховищем інформаційних об'єктів ПрО, з яким взаємодіють численні клієнтські додатки (шпитальні, телемедичні, навчальні, наукові), і повинен містити в собі, зокрема, такі розділи [3, 7]:

- формальні теорії і онтології клінічних ПрО й суміжних ПрО; формальні мови різного рівня для опису запитів і лікарських завдань; моделі оптимізуючих перетворень лікарських завдань; моделі стандартної лексики (професійної мови) [2-4];
- різноманітні моделі виводу і добування знань, включаючи алгоритми діагностики, лікування й прогнозування; різноманітні моделі консиліумів вирішальних правил і інтелектуальних агентів [4, 8]; моделі тестування знань і тренінгу;
- моделі органів і систем, моделі фізичних і хімічних процесів; моделі пацієнта й моделі лікаря [4];
- Банк моделей тестів (спостережень) [9]; Банк моделей захворювань; Банк моделей технологічних процесів, Банк математичних моделей [4, 7];
- довільні інформаційні ресурси, що стосуються суміжних питань.

Концепція БкЗ ґрунтується на понятті моделі предметної області.

**Метою дослідження є:** 1. Розробка формалізму для опису інформаційних об'єктів предметної області; 2. Розробка структури моделі предметної області; 3. Розробка ефективних методів обробки та узагальнення інформації; 4. Розробка методів системної реконструкції медико-біологічних процесів на організаційному і популяційному рівнях.

#### Основна частина.

**Опис емпіричних даних.** Нехай  $\{t\}$  - множина елементарних тестів, за допомогою яких описуються всі фактори, обставини і явища, що мають відношення до досліджуваної системи (у минулому, сьогодні й майбутньому). "Елементарність" тесту означає, що його результат може бути представлений у вигляді "тест = значення". Конкретний результат тесту  $t$  будемо позначати через  $t$ . Приклади: <Прізвище, Тарасов>, <Вік, 52>, <Скарги, Головний біль, блювання, озноб>, <Тиск, 120/78>, <Головний біль, Є>.

Результати тестів можуть вибиратися (формуватися) з різних доменів (базових множин). Для фіксації того, що в якості множини результатів тесту  $t$  використовується домен  $T$ , будемо використовувати нотацію:  $t/T$ . Використовуючи різні домени, можна управляти спільністю (масштабом) результату того самого тесту. Висновки (діагнози) також можуть вибирати-

ся з різних доменів, що можна зафіксувати за допомогою нотацій:  $d/D$ ,  $\{d\}/D$  (висновки є різновидом тестів). Приклади доменів для тесту  $\tau = \text{'Вік'}$ :  $B1 = [0 \dots 120]$ ;  $B2 = \{\text{юний, молодий, середній, літній, старечий}\}$ ;  $B3 = \{\text{до 20; 20-40; 41-60; старше 60}\}$ . Приклади результатів тестів: <Вік/  $B2$ , літній>; Вік/  $B1$ ? 87; Вік/  $B3 = \text{"старше 60"}$ . Правила перерахування ' $B1 \rightarrow B2$ ;  $B2 \rightarrow B3$ ' задають *орієнтований граф доменів* для тесту 'Вік'. Подібний граф може бути побудований для будь-якого тесту крім дихотомічного [4, 5].

Нехай  $t/\Lambda$  - виділений тест, що означає час. Приклади доменів  $\Lambda$ :  $\Lambda1 = \text{'Дата: час'}$ ;  $\Lambda2 = \text{'Дата: \{ранок, день, вечір, ніч\}}$ ;  $\Lambda3 = \text{'Дата'}$ ;  $\Lambda4 = \text{'Місяць'}$ ;  $\Lambda5 = \text{'Рік'}$ . Можливі і інші домени. Конкретне значення часу відповідно до прийнятої нотації будемо позначати:  $t/\Lambda$ . Приклади конкретних значень часу:  $t/\Lambda1 = \text{'10.01.09:08.30'}$ ,  $t/\Lambda2 = \text{'10.01.09: ранок'}$ ,  $t/\Lambda3 = \text{'10.01.09'}$ ,  $t/\Lambda4 = \text{'січень'}$ ,  $t/\Lambda5 = \text{'2009'}$ . Цілком очевидний спосіб завдання перерахування з одного домену в інший (з більшим номером), тобто визначений обчислювальний ланцюжок (орієнтований граф доменів):  $\Lambda1 \rightarrow \Lambda2 \rightarrow \Lambda3 \rightarrow \Lambda4 \rightarrow \Lambda5$ .

Визначимо дві *елементарних події*:  $e_1 = \langle \tau/T, t/\Lambda \rangle$  і  $e_2 = \langle -\tau/T, t/\Lambda \rangle$ . Подія  $e$  означає, що в момент часу  $t/\Lambda$  тест  $\tau/T$  мав значення  $\tau/T$ . Подія  $e_2$  означає, що в момент часу  $t/\Lambda$  тест  $\tau/T$  не мав значення  $\tau/T$ . Покладемо:  $e_1 = -e_2$ ,  $e_2 = -e_1$ . Приклади елементарних подій: <Кашель/{Є; Відсутній}? Є,  $t/\Lambda1$ ? 12.45>, <Le крові/{Знижено; Норма; Підвищена}? Норма,  $t/\Lambda3$ ? 23.02.08>.

На основі елементарних подій можна визначити *складені події* у такий спосіб:  $e = \langle \{\tau/T\}, t/\Lambda \rangle$  і  $e' = \langle -\{\tau/T\}, t/\Lambda \rangle$ . *Протяжною подією* назвемо подію  $e = \langle \{\tau/T\}, \delta/\Lambda \rangle$ , де  $\delta$  - довільний часовий інтервал (можливо не однозв'язний). Для інтервалів визначена операція об'єднання:  $e = \&_1 \langle \{\tau/T\}, \delta_1/\Lambda \rangle = \langle \{\tau/T\}, \delta/\Lambda \rangle$ , де  $\delta = \&_1 \delta_1$ . Приклад протяжної події: <Креатинін / {мг / 100 мол}? 1.3, [12.02.09; 14.02.09]/ $\Lambda3$ >, де  $\delta = [12.02.09; 14.02.09]$  (у термінах домену  $\Lambda3$ ).

Якщо для будь-якого моменту часу у розглянутому періоді спостереження значення деякого тесту  $\tau/T$  не змінюється, то такий результат назвемо *фактом*. Так, наприклад, перенесені в минулому операції або травми є фактами, захворювання батьків (спадковість) є фактами й т.д. *Життєвим циклом* довільного тесту, який не є фактом, назвемо множину  $\{\langle \tau/T, t/\Lambda \rangle \mid \tau/T \text{ фіксоване, } \Lambda \text{ фіксоване}\}$ .

Довільну ситуацію дійсності (клінічний випадок, дані про здоров'я популяції в регіоні) позначимо через  $\alpha = \alpha(\{\langle \tau/T, t/\Lambda \rangle\})$ , а множину ситуацій дійсності позначимо через  $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ .

**Модель предметної області.** Модель ПрО представимо у вигляді кортежу  $\langle O, \kappa \rangle$ , де  $O$  - модель онтології цієї предметної області, а  $\kappa$  - модель адекватної системи знань. Модель знань  $k$  містить у собі наступні компоненти:

$$k = k_c \cup k_M \cup k_A \cup k_O \quad (1)$$

де  $k_c$  - обчислювальні знання;  $k_M$  - опис моделей динамічних процесів (наприклад, захворювань та технологічних процесів);  $k_A$  - множина аксіом, фактів, висловлень із оцінкою їхньої істинності;  $k_O$  - інші знання. Онтологія  $O$  містить у собі, зокрема, Банк тестів разом з конфігураторами кожного тесту, Банк формалізованого опису захворювань, онтологічні угоди [4]. Модель знань  $k$  формується шляхом системної реконструкції множини ситуацій дійсності  $\Omega = \{\alpha\}$ .

У розгорнутому вигляді модель обчислювальних знань  $k_c$  представимо в такий спосіб [3]:

$$k_c = \{f/\mu: k^1 \rightarrow k^2 \mid \mu \in \{\mu\}_f\} \cup P_\kappa \quad (2)$$

де  $f/\mu$  - відображення, що реалізують ті або інші математичні моделі;  $\{\mu\}_f$  - різні механізми реалізації відображень (зі своєю енергетикою і ресурсами);  $k^1$  - вхідні дані завдання (опис інформаційного середовища і завдання);  $k^2$  - вихідні дані завдання;  $P_\kappa$  - правила композиції схем завдань, тобто правила, що описують способи об'єднання локальних завдань.

Позначимо елементарні тести через  $a, b, c, d, \{a\}$  і т.д. Нехай  $W\{c/C\}$  - деяке різноманіття на множині результатів тестів  $\{c/C\}$ . З урахуванням прийнятої моделі емпіричних даних будь-яке відображення моделі знань  $k_c$  можна представити в такий спосіб:

$$f/\mu_f: \{b/B\} \rightarrow \{a/A\},$$

для  $\{c/C\} \in W(\{c/C\}), \mu_f \in \{\mu\}_f$ . (3)

Модель обчислювальних знань (2) - (3) повністю відповідає синергетичним принципам: являє собою взаємодію (у процесі рішення завдання) великої кількості квантів знань, що представляються відображеннями; будь-який обчислювальний процес істотно нелінійний, тому що є композицією різних відображень (детермінованих, нечітких, нейромережевих, агентних і т.д.); система відкрита, тому що активно здійснюється обмін із зовнішнім середовищем (участь у консиліумах, безперервний пошук інформації в агентному середовищі і в Інтернет); кванти знань у загальному випадку мають фрактальну структуру; метод граничних узагальнень (МПУ) описує процеси горизонтальної й вертикальної самоорганізації, формуючи мінімальні ненадлишкові моделі знань на будь-якому рівні спільності [5, 6]. "Параметри порядку" у знаннях - перелік тестів мінімаль-

ної ненадлишкової моделі знань із доменами максимального рівня спільності. Таким чином, МПУ створює ступені свободи в описі ситуацій дійсності, що, відповідно до принципів синергетики, створює передумови для виникнення процесів самоорганізації.

Наведемо специфікації завдань деяких класів моделей обчислювальних знань ( $\underline{t}/T$  - результати тестів;  $\underline{t}/\Lambda$  - моменти часу;  $\langle \underline{t}/T, \underline{t}/\Lambda \rangle$  - події;  $d/D$  - висновки, діагнози;  $h/H$  - прогностичні гіпотези;  $r/R$  - програми керування;  $T, L, D, H, R$  - сорти або домени;  $J$  - оператор оцінки істинності довільної формули  $\phi$ ):

$F_1 = \{f/\mu: \{\underline{t}/T\}_1 \rightarrow \{\underline{t}/T\}_2\}$  - клас моделей обчислювальних знань;

$F_2 = \{f/\mu: \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\} \rightarrow d/D\}$  - клас моделей діагностичних знань;

$F_3 = \{f/\mu: \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\} \rightarrow -d/D\}$  - клас моделей знань, що описують область заборон;

$F_4 = \{f/\mu: \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\}, \{d/D\} \rightarrow \{h/H\}\}$  - клас моделей прогностичних знань;

$F_5 = \{f/\mu: \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\}, \{d/D\}, \{h/H\} \rightarrow \{r/R\}\}$  - клас моделей знань із оптимізації керування;

$F_6 = \{f/\mu: \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\} \rightarrow \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\}'\}$  - клас моделей знань, що представляє собою сукупність причинно-наслідкових зв'язків (як структурних, так і часових);

$F_7 = \{f/\mu: \{J_\beta \phi\}, \{\langle \underline{t}/T, \underline{t}/\Lambda \rangle\} \rightarrow \{J_\gamma \phi\}'\}$  - клас моделей знань, що дозволяє змінювати оцінку істинності формул - відображень залежно від тих або інших подій дійсності (відображення можуть належати кожному із класів  $F_1 - F_6$ ). Оцінка істинності в загальному випадку має фрактальну структуру.

Загальна модель обчислювальних знань  $k_c$  містить у собі всі згадані вище класи моделей, а саме:  $F_1 \cup F_2 \cup F_3 \cup F_4 \cup F_5 \cup F_6 \cup F_7 \subseteq k_c$ . Опис практичної реалізації представленої моделі обчислювальних знань наведено в роботах [3, 5, 7].

**Лексичні дерева.** БкЗ зберігає значну частину інформації в текстових файлах певної структури. Основним засобом для подання інформації в БкЗ є лексичні дерева (ЛД), які можуть бути декількох типів. Залежно від типу ЛД є можливість аналізу додаткових характеристик ЛД, властивих конкретному типу [2 - 4].

Спочатку лексичні дерева були розроблені для формалізації професійної лексики в рамках шпитальних систем [2]. Їхнє використання дозволило кардинально змінити підхід до автоматизації ведення медичних записів (вдалося відмовитися від фіксованих шаблонів медичних протоколів). Пізніше вони були використані для формального запису алгоритмів діагностики, лікування і прогнозування [3]. У роботах [3, 4] сфера зас-

тосування ЛД розширена до опису клінічних теорій (множин тестів, клінічних картин захворювань, клінічних ситуацій і т.д.) і об'єктів Банку математичних моделей. Таким чином, розвиток концепції банку знань йшов від потреб практики й практика залишається критерієм корисності пропонованих рішень.

Простота базових принципів побудови ЛД - їхня головна перевага. Приклад фрагменту лексичного дерева:

#### СКАРГИ НА

```
|постійний; нападпоподібний| !головний біль {голов_біль
|помірний; інтенсивний|
головний біль
що |виникає; посилюється| при
підвищенні тиску
фізичному навантаженні
зміні положення |голови; тіла|
< ? >
протягом _____років; року; міс; дн|
|постійне; нападпоподібне| ! запаморочення {запаморочення} .
```

Наведемо запис довільного відображення моделі знань (3) з використанням ЛД і Банку тестів (ФВ - функціональне відображення):

```
<Повне ім'я ФВ> -f {ФВ_<Абревіатура для ФВ>
[Умова {cond_
<тести - умова> - {c/C}}]
Вхід {inp_
<вхідні тести> - {b/B}}
Вихід {out_
<вихідні тести> - {a/A}}
Методи {met_
<методи обробки> - {μ}_}}
Відображення поєднуються в семантичні групи
різного рівня вкладеності [3]:
<Ім'я предметної області> {ПРО_< Ім'я >
[Визначення
<Скорочене ім'я тесту> - <Повне ім'я тесту>
[ {Посилання} ]
[<ПРО_Ім'я>]
<ФВ_1>
<ФВ_N>
} [ПрО]
```

**Конфігуратори тестів.** Базовим доменом назвемо домен, за допомогою якого описуються максимально точні значення тесту (область можливих значень - ОМЗ). Для таких тестів як, наприклад, "Температура", "Ріст", "Вага", "Швидкість",

"Прискорення", "Час" у якості базового домену можна прийняти відрізок числової прямої. Для тесту "Вік" базовим доменом буде інтервал від 0 до 120 і т.д. Дискретний домен максимальної спільності назвемо *БІ-арним конструктом тесту*, де  $N$  - число елементів домену. Приклади бінарних конструктів: {+; -}, {Норма: Відхилення}, {Ефективно; Неефективно}, {Скарги є; Скарг немає}, {Сприятливий; Несприятливий} і т.д. Для того самого тесту може бути потенційно задано нескінченно багато БІ-арних конструктів. Між базовим доменом тесту і БІ-арним конструктом також може бути задано як завгодно багато проміжних доменів. Усередині ієрархії доменів задаються правила перерахування елементів одного домену в елементи домену більш високого рівня спільності. Правила перерахування можуть бути самими різними: детермінованими, нечіткими, нейромережевими і т.д.

Конкретну ієрархію доменів, основою якої служить базовий домен, вершиною - карний конструкт, з фіксованими правилами перерахування, назвемо *конфігуратором тесту*. Конфігуратор тесту  $t$  будемо позначати в такий спосіб:

$$fr_t: OMZ(t) \succ \delta_A(t), \quad (4)$$

де  $OMZ(t)$  - базовий домен,  $\delta_A(t)$  -  $N$ -арний конструкт. Множину конфігураторів для тесту  $t$  позначимо  $\{fr\}_t$ . Для числових тестів конфігуратор може задавати основу у вигляді фізичного фракталу типу "Канторівський пил". Конфігуратори зберігаються в Банку тестів. Приклади конфігураторів (Б1 - Б4 - домени):

```
Anemum {
D1 {Апетит нормальний ^N; Апетит змінений ^b}
D2 {Відраза до їжі ^b; Апетит знижений ^b; Апетит нормальний ^N; Апетит підвищений ^b}
Рівень гемоглобіну ^Hb {
D1 {Нормальний рівень Hb ^N; Ненормальний рівень Hb ^a b}
D2 {Анемія ^a. Нормальний рівень Hb ^N; Підвищений рівень Hb ^b}
D3 {Різко виражена анемія ^a [20; 50]; Виражена анемія ^a [51; 90]; Помірна анемія ^a [90; 119]; Нормальний рівень Hb ^N [120; 160]; Рівень Hb помірно підвищений ^b [161; 170]; Підвищений рівень Hb ^b [171; 180]; Різко підвищений рівень Hb ^b [181; 200] }
D4 {[20; 200]}
D4 ->> D3 ({параметри скейлінгу})}
```

**Висновок.** Головною перевагою запропонованого формалізму є його відносна простота й прозорість.

Більшість інформаційних об'єктів моделі ПрО на процедурному рівні описується за допомогою лексичних дерев, що забезпечує загальну методичну основу. Розроблений формалізм дозволяє описувати ме-

дичні тести будь-якої складності, професійну лексику, клінічні картини захворювань, клінічні критерії й клінічні ситуації.

#### **Література**

1. Мінцер О.П., Вороненко Ю.В. На шляху до суспільства медичних знань // Актуальні питання фармацевтичної та медичної науки та практики. Зб.наук. праць Вип. XVI. - Запоріжжя: У ЗДМУ, 2006. - С. 3-15.
2. Алпатов А. П., Прокопчук Ю. О., Костра В. В. Шпитальні інформаційні системи: архітектура, моделі, рішення. - Дніпропетровськ: УДХТУ 2005. - 257 с.
3. Алпатов А. П., Прокопчук Ю. О., Юденко О. В., Хорошилов С. В. Інформаційні технології в освіті та охороні здоров'я. - Дніпропетровськ: ІТМ НАНУ і НКАУ 2008. - 287 с.
4. Прокопчук Ю.О. Інтелектуальні медичні системи: формально-логічний рівень. - Дніпропетровськ : ІТМ НАНУ й НКАУ, 2007. - 259 с.
5. Прокопчук Ю. О. Метод граничних узагальнень - ефективний принцип роботи обчислювального інтелекту // Штучний інтелект. - 2008. - № 3. - С. 727 - 736.
6. Prokopchuk Yu.O. Method of Limiting Generalizations for Solving Logical and Computing Tasks // Proceedings of 2nd International Conference on Inductive Modelling (September 15 - 19, 2008, Kyiv, Ukraine). - Kyiv: IRTC ITS, 2008. - P. 62-65.
7. Прокопчук Ю. О. Багатоцільовий банк знань в області клінічної медицини // Вісник ХНТУ - 2008. - № 1(30). - С. 14-20.
8. Прокопчук Ю. О. Проведення лікарських консилиумів за участю інтелектуальних систем // Вісник ХНТУ - 2007. - № 4 (27). - С. 198-202.
9. Черняховская М.Ю., Клещев А.С., Кулаков Ю. В., Онищенко С. Б. База знаний для системы интеллектуальной поддержки обследования больных для врача терапевта. - Владивосток: ИАПУ ДВО РАН, 1998. - 46 с. (Препринт/ИАПУ ДВО РАН; № 16).
- Орлов В. А., Клещев А. С. Компьютерные банки знаний. Многоцелевой банк знаний // Информационные технологии. - 2006. - № 2. - С. 2-8.