

СЕМАНТИКО-СИНТАКСИЧНА МОДЕЛЬ КЛАСИФІКАЦІЇ КЛІНІЧНИХ ДІАГНОЗІВ

Постановка проблеми. Сучасна тенденція розвитку госпітальних інформаційних систем полягає в переході від вузько-функціональних систем до багатфункціональних, комплексних і відкритих інформаційних систем; при цьому актуальним стає питання інтеграції різноманітних систем у єдиному інформаційному просторі[1,2]. Ключовим моментом у цьому сенсі стають питання розробки та вдосконалення стандартів щодо надання та обміну інформацією[3].

Вирішення питань щодо інтеграції різноманітних госпітальних інформаційних систем, централізованої обробки, аналізу та дослідження даних, що поступають з різноманітних джерел, тісно пов'язано з вирішенням проблем стандартизації відносно надання, використання та керування біо-медичною інформацією.

При цьому у якості пріоритетних класифікацій, що використовуються в госпітальних інформаційних систем стають класифікації діагнозів, операцій, процедур, які виконують роль бази для аналітичної та статистичної оцінки лікувально-діагностичного процесу. Найважливішою серед означених слід вважати класифікацію клінічних діагнозів, яка лежить у основі і діагностичних систем, і оцінки результатів діяльності лікувальних закладів, а також є базою створення стандартів лікування.

На даний момент у більшості медичних інформаційних систем, розповсюджених на території України і в світі, як класифікація діагнозів використовується міжнародна класифікація хвороб 10-го перегляду. Однак є труднощі з її використанням в практичних і дослідницьких цілях. Так, при інформаційній обробці діагнозів не враховуються всі ознаки, властиві конкретному захворюванню. Процес адаптації інших термінологічних систем пов'язаний як зі значними витратами, так і з рядом недоліків, які роблять неефективною їх адаптацію до систем державних лікувальних закладів України.

Таким чином, дослідження всесвітнього досвіду, обґрунтування і побудова концептуальної моделі оптимальної класифікації діагнозів для лікувальних закладів України, а також розробка і побудова інструментарію керування цією класифікацією, враховуючи можливу інтеграцію з однією із сучасних термінологічних систем, і створення госпітальної системи, що враховує специфіку державних лікувальних установ, є важливим завданням, пов'язаним з інформатизацією державних медичних закладів України.

Аналіз публікацій та постановка задачі. На основі аналізу сучасних термінологічних систем[4-9] слід відзначити наступні особливості: широке коло медичної термінології, що передбачає можливість кодування любого медичного поняття з будь-якої області; концептно-орієнтована архітектура; композиційна схема кодування; існування обмежень на семантику заданої композиції.

Розглянемо основні принципи використання термінологічної системи.

Фактично процес завдання комплексного медичного концепту (клінічного діагнозу) складається з формування композиції на основі концептів-атомів, що відносяться до різноманітних осей системи; при цьому враховуються семантичні обмеження на зв'язок понять. В результаті отримується кодова комбінація, що задає консистентний термін за визначеною ТС. Подальший розвиток передбачає реалізацію механізмів трансляції натурально-мовних речень, що описують медичний термін, в кодову послідовність (композицію концептів) з обмеженням або взагалі без участі користувача.

З означеного вище, виходять три припущення: будь-яке натурально-мовне консистентне описання стосовно медичної термінології має бути відображене ТС на базі взаємодії „термін-синоніми-концепт”; процес кодування може бути здійсненим тільки на базі інформаційних технологій; користувачем системи має бути досвідчений спеціаліст.

З визначеного підходу виходять наступні негативні фактори щодо впровадження такої системи в ЛПЗ України навіть у разі їх перекладу на українську або російську мову:

1. Складність процесу кодування навіть у разі досвідченого лікаря. Так, у разі завдання комплексного концепту, лікар повинен знайти необхідні концепти-атоми, що належать різним осям.

Пошук може здійснюватися на основі дерева класифікації та за термінами-синонімами, але у обох випадках пов'язаний з тратою певного часу на ітерації „пошук-вибір”.

2. Складність побудови механізмів аналізу інформації. За означеним принципом одне складне поняття може бути закодованим декількома різноманітними засобами; таким чином, щоб вибрати хворих з відповідним класом захворювання, слід генерувати усі комбінації можливих кодів.
3. Неможливість користування ТС за межами інформаційної системи, тобто кодування лікарем клінічного діагнозу з подальшим вводом цього коду оператором є повністю неможливим.

Розглянемо механізм завдання семантичних обмежень. Слід відзначити, що в першу чергу основним об'єктом дослідження з цього питання було вибрано SNOMED CT як основний претендент на стандартизацію, при цьому доменом інтересу є виключно область клінічних діагнозів. Семантична складова системи, описується логікою *ELN* [10, 11] яка: підтримує спадкоємство концептів і відношень; надає можливість визначення концептів через сукупність інших, використовуючи оператор кон'юнкції; підтримує оператор можливості зв'язку концепту з іншим концептом. Наприклад, на рис. 1 показано визначення концепту „гостра виразка шлунка” (*acute gastric ulcer*) в системі SNOMED CT, яка визначена як гостра виразка, що знаходиться в шлунку (*finding site.stomach*) та характеризується швидким розвитком (*clinical course. sudden onset and short duration*). Важливою особливістю є існування асоціацій, наприклад, з поняттям важкості (*severity*), але на атрибути не можуть бути накладені ніякі обмеження. Так, не можливо виділити обов'язкові атрибути, без значень яких концепт не може вважатися семантично повним або навіть коректним, та можливі характеристики, які не визначають, але поширюють поняття. Другою особливістю є неможливість завдання обмежень на кількість значень певної характеристики. Описані проблеми тісно пов'язані з відсутністю конструкторів обмежень на кількість філерів ролі або атрибуту (асоціації), що надає можливість завдання багатозначних концептів. Також слід відзначити можливість існування залежності

характеристик одного концепту, що також не може бути заданим на базі існуючих механізмів семантичних обмежень.

Стосовно питання взаємодії з адміністративними системами слід відзначити, що у версії SNOMED CT (January 2007) існує прямий зв'язок концептів діагнозів з класифікацією МКХ-9, в той час як єдиним стандартом для звітної документації на території України з 1997 р. є остання версія класифікація МКХ-10.

Останнім важливим питанням є питання існування механізмів формування синтаксично-правильного опису концепту на базі заданого коду (наприклад, статистична карта хворого стаціонару). Реальним виходом у контексті цієї проблеми стосовно кодування клінічних діагнозів є лише дублювання закодованого терміну додатковим натурально-мовним описом без можливості автоматичного перекладу, що безумовно ускладнює роботу лікаря.

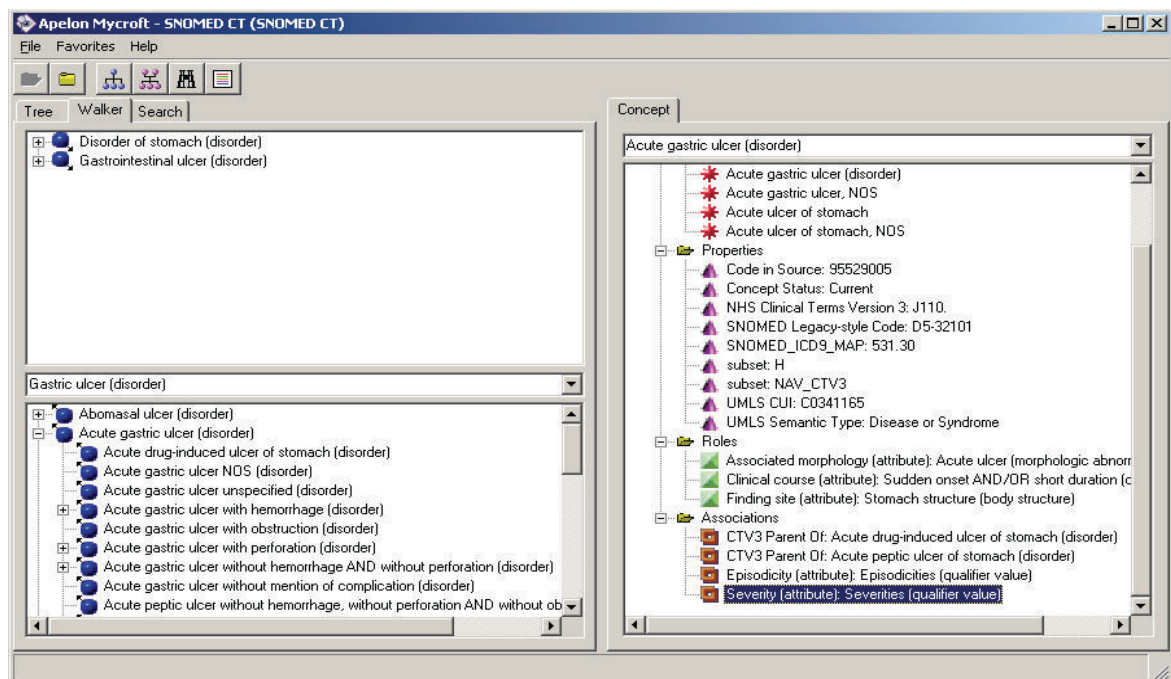


Рисунок 1 - Гостра виразка шлунка за SNOMED CT

Виходячи з даних проведеного аналізу, основними вимогами, що впливають на успішне використання клінічної класифікації в ЛПЗ України слід вважати: компактність надання множини клінічних діагнозів базуючись на композиційній архітектурі класифікації; врахування семантичних особливостей, пов'язаних із задачею формування лише семантично-коректних понять; наявність механізмів формування лише синтаксично-правильних описів діагнозів. Можливим рішенням може бути розробка або застосування

деякої надбудови над вказаною термінологічною системою, що враховувала би специфіку цих умов.

Задача:

В даній роботі ставиться задача надання моделі класифікації клінічних діагнозів, що відповідає вимогам ЛПЗ України.

Основна частина

На базі проблемної лабораторії „Нові інформаційні технології в медицині” Дніпропетровської Державної Медичної Академії було запропоновано форма прототипу і варіант класифікації клінічних діагнозів в домені гастроентерологічних хвороб [12]. Класифікація може бути представлена у вигляді мережевої структури, що складається з множини класів клінічних діагнозів, у ролі яких можуть виступати діагнози МКХ-10 та множини клінічних характеристик, що складають множини (осі) категорій клінічних характеристик. Елементами термінологічних осей є формули-макети або примітиви. Макети складаються зі слів (констант) та клінічних характеристик (змінних), які пов'язані з елементами інших термінологічних осей, виключаючи свою. Так макет діагнозу «виразка шлунку» може мати клінічну характеристику «ускладнення», який зв'язує даний діагноз з віссю «ускладнень». На зв'язок накладаються обмеження на значення і на число значень. Порядок клінічних характеристик у макеті та слова-константи, що виконують роль зв'язок клінічних характеристик, надають можливість створення натурально-мовного опису діагнозу. Концепти осі діагнозів, елементами якої являються діагнози МКХ-10, що мають макети, можуть бути пов'язаними зі всіма осями класифікації. Макети інших осей можуть включати клінічні характеристики всіх типів, окрім діагнозів. Макет діагнозу, пов'язаний з певним кодом МКХ-10, відображає всі можливі варіанти даного діагнозу і, як правило, поглинає варіанти даного діагнозу, представлені в МКХ-10.

У якості істотних особливостей клінічного діагнозу за наданими вимогами слід відзначити:

1. Композиція концептів, що складають комплексне поняття, може відбуватися тільки в рамках семантичних обмежень, накладених на клінічні характеристики, що складають змінну частину макету.

2. Клінічні характеристики мають певний порядок у макеті, що впливає на генерацію синтаксично правильного опису.
3. З клінічною характеристикою може бути пов'язана деяка синтаксична складова, що містить ряд слів-констант, які забезпечують натуральність мови у описі діагнозу. У разі визначення клінічної характеристики слова константи, пов'язані з нею, з'являються у кінцевому описі варіанту діагнозу разом з її значенням. При цьому слід враховувати можливість існування механізмів, за яким може відбуватися редукція або зміна синтаксичної складової клінічної характеристики в макеті у разі вибору клінічної характеристики.
4. Існує залежність між клінічними характеристиками. Значення залежної характеристики задається тільки у разі визначення характеристики, від якої вона залежить.
5. Основним елементом макету є статистичний діагноз МКХ-10.
6. Кожен з визначених варіантів класу клінічного діагнозу, заданого макетом, має відповідність в МКХ-10.
7. Класифікація клінічних діагнозів має бути орієнтована на компактний опис множини варіантів заданого класу діагнозів і відповідно композицію необхідного поняття на базі даного опису, але не на пошук понять по всім термінологічним осям або трансформацію натурально-мовного речення в коди, що надає можливість виключити лікаря з процесу безпосередньої взаємодії з системою.

Приймаючи наведені вимоги та бажану форму і неформальний опис клінічної класифікації як вихідні дані для моделювання, доцільно провести формалізацію, побудову узагальненої семантико-синтаксичної моделі клінічного діагнозу.

В роботі [13] була приведена методика формування моделі клінічного діагнозу з застосуванням апарату логіки предикатів першого порядку. З точки зору семантики множина клінічних діагнозів може бути представлена у вигляді набору формул, оснований на аксіомах, що виступають в ролі метамови та вказують на засоби та шляхи побудови цих формул. Доменом інтерпретації цих формул є скінчена кількість класів клінічних діагнозів та їх клінічних характеристик.

Переваги даного підходу полягають в точному описі правил створення і використання концепту клінічного діагнозу, що надає можливість побудови алгоритмів перевірки як завдання концепту експертом, так і коду, введеного користувачем на базі певного макету діагнозу. Недоліками даного підходу є трудомісткість як з погляду побудови, так і з точки зору вдосконалення даної моделі.

Синтаксична складова форми клінічного діагнозу може бути віднесена до таких мов опису знань, що розглядає мову у якості множини правильно побудованих фраз, складених з елементів деякої фіксованої множини (алфавіту мови), не торкаючись при цьому семантики побудованої фрази.

Таким чином, можна сказати, що шаблон-макет клінічного діагнозу може бути представлений у вигляді формальної граматики, на базі якої генерується множина коректних речень діагнозів.

Так, у якості прикладу можна навести наступну граматику, яка відображає множину станів, пов'язаних с діагнозом «шлунково-стравохідний розривно-геморагічний синдром» (код К22.6 за МКХ-10), який має наступний вигляд у нотації УКСКХ „Синдром Маллорі-Вейса з кровотечею, {!I,{крововтрата !T ст. тяжкості}}”. Введемо до розширеної нотації знак „!”, який за нетермінальним символом буде означати можливість обов'язкової підстановки замість нього одного зі зв'язаних з ним значень. Для скорочення запису була виключена деяка кількість значень клінічних характеристик, та їх семантичні подробиці. Кожен шаблон клінічного діагнозу може бути наданим у нормальній формі, тобто у вигляді контекстно-незалежної граматики (зліва може стояти лише нетермінальний символ, теорема 5.1. за [14]). Наприклад:

К22.6 → Синдром Маллорі-Вейса з кровотечею S_1

$S_1 \rightarrow !S_2$

$S_2 \rightarrow$ крововтрата **T**

$I \rightarrow I_1|I_2$

$I_1 \rightarrow$ (активна струминна)

$I_2 \rightarrow$ (активна капілярна)

$T \rightarrow T_1|T_2|T_3$

$T_1 \rightarrow$ легкого ступеня важкості (*крововтрата 10-20%, пульс до 90 уд. за 1 хв., $AT > 120$ мм. рт. ст., шоковий індекс $P/AT - 0,54-0,78$; ...*)

$T_2 \rightarrow$ середнього ступеня важкості (*...додатковий опис*)

$T_3 \rightarrow$ важкого ступеня (...*додатковий опис*)

Загальну мета-модель (мета-граматику), що лежить в основі моделей (граматик), і здатна відображати як загальний процес трансформації деякої кодової комбінації в T -термінальний ланцюжок (натурально-мовне описання клінічного діагнозу), або порядок завдання деякого діагнозу за правилами граматики, так і можливу структуру шаблону діагнозу, що задає відповідну граматику.

$$K \rightarrow K_1 | K_2 | K_3 | \dots | K_n \quad (1)$$

Кожний існуючий діагноз (в ідеалі) може бути віднесеним до одного з шаблонів, пов'язаного з кодом МКХ-10. Це ствердження відображено продукцією

$$K_i \rightarrow \alpha_i S_i \quad \alpha_i \in VT^* \quad (2)$$

Правило (2) показує структуру шаблону, яка складається з початкового термінального рядка α_i , який може бути пустим, тобто віднесеним до множини терміналів з пустим ланцюжком, та набору слотів-атрибутів, пов'язаних з даним шаблоном.

$$S_i \rightarrow \Sigma_{i1} \Sigma_{i2} \dots \Sigma_{im} \quad (3)$$

Правило (3) показує структуру набору слотів-атрибутів розширеного опису діагнозу, пов'язаного з кодом K_i .

$$\Sigma_{ij} \rightarrow l_1 \alpha \Sigma_{ik} P_{ij} \beta \Sigma_{il} l_2 \gamma \quad (4)$$

Правило (4) показує синтаксичну структуру слоту, де l_1, l_2 – термінальні символи, що визначають границі слоту, P_{ij} – ядро слоту, α, β, γ – термінальні рядки, Σ_{ik}, Σ_{il} – залежні слоти з аналогічною структурою.

$$P_{ij} \rightarrow * \Theta_{ij} \varepsilon | ! \Theta_{ij} \varepsilon | * \Theta_{ij} \& | ! \Theta_{ij} \& \quad (5)$$

$$l_1 \alpha * \Theta_{ij} \varepsilon \beta l_2 \gamma \rightarrow \alpha N_{ij} \beta \gamma | \varepsilon \quad (6)$$

$$l_1 \alpha ! \Theta_{ij} \varepsilon \beta l_2 \gamma \rightarrow \alpha N_{ij} \beta \gamma \quad (7)$$

$$l_1 \alpha * \Theta_{ij} \& \beta l_2 \gamma \rightarrow \alpha (N_{ij}) + \beta \gamma | \varepsilon \quad (8)$$

$$l_1 \alpha ! \Theta_{ij} \& \beta l_2 \gamma \rightarrow \alpha (N_{ij}) + \beta \gamma \quad (9)$$

Правила (5-9) показують варіанти структури слоту. Додаткові знаки „*” та „!” вказують на обов'язковість підстановки одного з значень замість нетермінального символу та не входять до алфавіту V . Знак „&” вказує на можливість підстановки декількох значень нетермінального символу одночасно. Знак розширеної нотації формальних граматик „+” означає обов'язкову підстановку одного або декількох значень замість нетермінального символу.

$$\Sigma_{ik} \rightarrow l_1 \alpha \Sigma_{ik} * \Theta_{ij} \varepsilon \beta \Sigma_{il} l_2 \gamma \rightarrow$$

$$l_1 \alpha N_{ij} \beta l_2 \gamma | l_1 \alpha N_{ij} \beta \Sigma_{il} l_2 \gamma | l_1 \alpha \Sigma_{ik} N_{ij} \beta l_2 \gamma | l_1 \alpha \Sigma_{ik} N_{ij} \beta \Sigma_{il} l_2 \gamma | \gamma$$
(10)

$$\Sigma_{ik} \rightarrow l_1 \alpha \Sigma_{ik} ! \Theta_{ij} \varepsilon \beta \Sigma_{il} l_2 \gamma \rightarrow$$

$$l_1 \alpha N_{ij} \beta l_2 \gamma | l_1 \alpha N_{ij} \beta \Sigma_{il} l_2 \gamma | l_1 \alpha \Sigma_{ik} N_{ij} \beta l_2 \gamma | l_1 \alpha \Sigma_{ik} N_{ij} \beta \Sigma_{il} l_2 \gamma$$
(11)

Правила (10-11) показують варіант залежності слотів у шаблоні (слоти Σ_{ik} і Σ_{il} залежить від Σ_{ij}) у разі обов’язкової та необов’язкової клінічної характеристики. Відповідні правила для формул (3.19)-(3.20) будуються аналогічним чином.

$$N_{ij} \rightarrow \delta_{ij1} | \delta_{ij2} | \dots | \xi_{ij1} S_{ij1}^1 | \xi_{ij2} S_{ij2}^1 | \dots | \xi_{ij2} S_{ijn}^1$$
(12)

Правило (12) показує варіанти значень нетермінального символу, якими можуть бути як термінальні символи, так і певна кількість шаблонів-значення $\xi_{ij1} S_{ij1}^1 | \xi_{ij2} S_{ij2}^1 | \dots | \xi_{ij2} S_{ijn}^1$.

$$S_{ijk}^1 \rightarrow \Sigma_{ijk1}^1 \Sigma_{ijk2}^1 \dots \Sigma_{ijkn-1}^1$$
(13)

У випадку шаблону-значення він може бути організований з $m-1$ слотів-атрибутів, де m – кількість нетермінальних символів в алфавіті, що визначають клінічні характеристики діагнозу(правило 13).

Описана формальна модель дає наступні результати:

1. Визначена загальна синтаксична структура слоту клінічної характеристики, його компоненти, їх порядок у слоті та у фреймі.
2. Визначені чотири види слотів.
3. Визначені термінальні рядки, що можуть існувати перед клінічною характеристикою, після неї, та взагалі після слоту.
4. Визначені варіанти підстановки значень замість нетермінального символу-характеристики.
5. Описані правила підстановки у разі залежності між слотами.

З метою відображення семантичної складової класифікації моделі клінічного діагнозу та враховуючи досвід розробки сучасних термінологічних систем, доцільно розглянути модель семантичної складової класифікації на базі дескрипційної логіки. Це дозволить віднести мову опису класифікації до формалізованих мов, визначити її переваги і недоліки, можливі шляхи розвитку.

Термінологічна база знань у нотації дескрипційної логіки описується набором термінологічних аксіом, конструкторів, концептів і ролей. Слід особливо підкреслити, що від вибору конструкторів, аксіом та мови класифікації залежить не тільки

концептуальна модель і механізми керування класифікацією, але і її подальший розвиток. Вважаючи еволюційний підхід основою розробки, слід особливу увагу звернути на вибір необхідного мінімуму конструкторів та аксіом логіки, без протиріч сформованим вимогам.

Розглянемо аксіоми. В загальному випадку існують два види аксіом: аксіоми підпорядкування (inclusion axioms) та аксіоми визначення (equality axioms). Обидва типи аксіом стосуються ролей та концептів. З семантичної точки зору аксіоми підпорядкування говорять, що підпорядкування концептів або ролей у визначеному домені інтерпретації означає, що множина об'єктів, формуючих клас концептів або бінарних відношень формуючих клас ролей, є підмножиною більш загальної множини (класу) відповідно об'єктів та ролей $(C_1 \subseteq C_2)^{\mathfrak{A}} = C_1^{\mathfrak{A}} \subseteq C_2^{\mathfrak{A}}$.

В нашому випадку можна сказати, що доменом інтерпретації $\Delta^{\mathfrak{A}}$ є медична термінологія клінічних діагнозів, в якій чітко визначаються множини класів діагнозів, ієрархія яких задана на засадах МКХ-10, і класів клінічних характеристик $\Delta^{\mathfrak{A}} = \Delta_D^{\mathfrak{A}} \cup \Delta_C^{\mathfrak{A}}$. Класами ролей є: відношення „клас-суперклас” в суб-домені діагнозів $C_1^{\mathfrak{A}} \subseteq C_2^{\mathfrak{A}} \subseteq \Delta_D^{\mathfrak{A}}$; відношення асоціації „має зв'язок з клінічною характеристикою”, яке може бути заданим об'єднанням можливих бінарних зв'язків між доменами діагнозів і характеристик та можливих відношень в самому домені характеристик $R^{\mathfrak{A}} = (\Delta_D^{\mathfrak{A}} \times \Delta_C^{\mathfrak{A}}) \cup (\Delta_C^{\mathfrak{A}} \times \Delta_C^{\mathfrak{A}})$.

Аксіоми підпорядкування та визначення концептів є базовими для будь-якої дискрипційної логіки. Виключення аксіоми підпорядкування приводить до виключення можливості побудови таксономії, що приводить до дублювання однієї і тієї інформації в системі, неможливості узагальнення висновків щодо заданих фактів. Таким чином, існування ієрархії в домені клінічних характеристик і відповідно аксіоми підпорядкування, слід вважати необхідною складовою класифікації клінічних діагнозів.

Розглянемо конструктори концептів класифікації. Концепт діагнозу визначається класом діагнозу (нозологія МКХ-10) та набором клінічних характеристик, серед яких слід виділити обов'язкові, тобто без завдання яких концепт не є семантично правильним,

необов’язкові та приймаючі декілька значень одночасно. Також була визначена залежність клінічних характеристик – значення певної клінічної характеристики має сенс лише при визначені другої.

Для визначення характеристик існують два механізми. Перший – це ролі звичайного апарату дескрипційної логіки. Другий – це характеристики (faetures) апарату логіки характеристик (feature logic). Логіка характеристик вийшла з досліджень в області обчислювальної лінгвістики і є логікою, що надає можливість задавати семантичні обмеження в уніфікаційній граматиці, яка може бути визначена як правила контекстно-незалежної граматики з семантичними обмеженнями, та має метою розбір та генерацію натурально-мовних речень [15, 16]. В основі логіки характеристик також знаходиться частина теорії предикатів з теорією моделей семантики Тарського, але основними недоліками логіки характеристик, стосовно термінології клінічних діагнозів, є відсутність механізмів обмежень на кількість значень атрибуту та спрямована орієнтація на роль додаткового рівня обмежень в синтаксико-семантичній системі уніфікаційної граматики. Таким чином, для опису характеристик доцільно обрати стандарт дескрипційної логіки.

Користуючись дескрипційною логікою, клінічний діагноз можна описати як $CN = D_i \sqcap R_j.C_i \sqcap \dots \sqcap R_k.C_k$, $R_j.R_k \sqsubseteq R, D_i \sqsubseteq T(D), C_i \dots C_k \sqsubseteq T(C)$, де D_i – клас діагнозу (наприклад за МКХ-10), $R_j \dots R_k$ – імена ролей-атрибутів, що відповідають категоріям клінічних характеристик, $C_i \dots C_n$ – класи характеристик, що відповідають даному діагнозу.

Тип клінічної характеристики (обов’язковість та ін.) може задаватися за допомогою конструкторів обмеження на кількість значень, пов’язаних з визначеним класом (\mathcal{Q} див. розділ 2).

Розглянемо причини виключення конструктору $CN = \exists R.C_1$, який є основою логіки SNOMED CT.

За визначенням цей конструктор означає: для множини об’єктів визначеного класу CN існує таке значення y атрибуту R , що належить класу C . В нотації теоретичної семантики це ствердження має відповідний запис $\{x \in \Delta^{\mathfrak{A}} \mid \exists y. \langle x, y \rangle \in R^{\mathfrak{A}} \wedge y \in C^{\mathfrak{A}}\}$. У якості прикладу [11] можна привести ствердження $\exists hasChild.Female$, яке означає

множину всіх об’єктів, що мають дитину жіночої статі, але це не означає, що об’єкти не мають дітей чоловічої статі.

На відміну від цього, конструктор $CN = \forall R.C_1$ означає, що всі об’єкти класу C пов’язані з класом CN відношенням R мають бути об’єктами цього класу. В нотації теоретичної семантики це ствердження має відповідний запис $\{x \in \Delta^{\exists} \mid \forall y. \langle x, y \rangle \in R^{\exists} \rightarrow y \in C^{\exists}\}$. Наприклад ствердження $\forall hasChild.Female$ означає клас об’єктів, які мають дітей тільки жіночої статі і ніяких інших.

У нашому випадку однією з основних вимог є строгість семантичних обмежень. Тобто, якщо ми говоримо, що діагноз „виразка шлунка” має атрибут „ускладнення”, ми визначаємо множину ускладнень виразки шлунку, але не всіх ускладнень взагалі, або ускладнень виразки шлунку та ще будь-яких інших ускладнень. За цим, виключення конструктору екзистенціальності вважаємо обґрунтованим.

Для визначення залежності клінічних характеристик в рамках дескрипції слід визначити конструктор матеріальної імплікації.

Таким чином, ми отримуємо наступний мінімальний набір конструкторів концептів клінічного діагнозу відповідно заданим вимогам (табл. 1). Приймаючи до уваги еквівалентність відношень $C_1 \rightarrow C_2 \equiv \neg C_1 \vee C_2$ або $C_1 \rightarrow C_2 \equiv \neg(C_1 \wedge \neg C_2)$, мову УКСКХ можна віднести до підмножини мов FL_0CQ або більш відомої логіки $ALCQ$ без оператора «обмеження на існування». В таблиці 2 наведено відповідності нотацій УКСКХ і DL $ALCQ$. Великими літерами з індексами внизу позначаються класи, які поєднують концепти певної осі, які можуть бути пов’язані з діагнозом, наданим фреймом, наприклад: фази захворювання діагнозу «виразка дванадцятипалої кишки». Розглянемо наприклад діагноз K40, який має такий шаблон „{!L} {!Q} пахова грижа, {ускладнена *O&.}”. Користуючись нотацією дескрипційної логіки даному концепту можна надати відповідну дескрипцію

$$F(K40) \doteq K40 \sqcap \forall^{-1} hasLocalization.L_i \sqcap \forall^{-1} hasClinicalCourse.Q_j \sqcap \forall^{20} hasComplexity.O_k \\ L_i \sqsubseteq T(L), Q_j \sqsubseteq T(Q), O_k \sqsubseteq T(O), OQL \sqsubseteq T(C)$$

За описаною моделлю, семантична складова діагнозу складається з двох частин: класу діагнозу та ролей-характеристик, пов’язаних з класами відповідних концептів характеристик, що відносяться до

певних категорій-осей. Значення ролі „локалізація” може бути лише з класу, що складається, наприклад, з об’єктів {лівостороння, правостороння, двостороння}, а ролі «клінічні прояви» з об’єктів {вправна, невправна, рецидивна}.

Теоретично класи описаних об’єктів можуть бути віднесеними і до іншого концепту, і можуть бути розширеними. У цьому випадку можна говорити про успадкування одного класу клінічних характеристик від іншого, у якості прикладу можна навести ступінь важкості в контексті кровотечі.

Таблиця 1

Конструктори концептів УКСКХ

Оператор	Нотація дескрипційної логіки
Кон’юнкція	$CN = C_1 \sqcap C_2$
Імплікація	$CN = C_1 \rightarrow C_2$
Обмеження на значення	$CN = \forall R.C_1$
Обмеження на мінімальне число значень	$CN = \geq nR.C_1$
Обмеження на максимальне число значень	$CN = \leq nR.C_1$

Таблиця 2

Таблиця відповідностей конструкторів концептів УКСКХ та мови ALCQ

Нотація УКСКХ	Приклад	Нотація логіки	Опис
*Axis	{*F}	$\forall^{\leq 1} hasPhase.F_i$ $F_i \sqsubseteq F$	Необов’язкова клінічна характеристика діагнозу
{*Axis1 {*Axis2}}	{*T{*F}}	$\neg(\forall^{\leq 1} hasPhase.F_i \sqcap$ $\neg \forall^{\leq 1} hasSeverity.T_j)$ $F_i \sqsubseteq F, T_j \sqsubseteq T$	Залежність характеристики F від T
Axis&	{*O&}	$\forall^{20} hasComplexity.O_i$ $O_i \sqsubseteq O$	Можливість завдання декількох значень
!Axis	{!T}	$\forall^{-1} hasSeverity.T_k$ $T_k \sqsubseteq T$	Обов’язкова характеристика

Семантико-синтаксична модель класифікації. Серед формалізмів, здатних об’єднати синтаксичну і семантичну складові, слід відзначити фреймову парадигму опису знань[17,18,19], що має тісний зв’язок з об’єктно-орієнтованим підходом програмування. У такому сенсі фреймова парадигма може служити мета-мовою, здатною відобразити семантичні та синтаксичні особливості, визначені на певних рівнях формалізації. Результатом моделювання може вважатися концептуальна модель структурної та функціональної

складових класифікації, методики її інтеграції з існуючими термінологічними системами.

Іншим методом об'єднання семантичної та синтаксичної складових може вважатися застосування апарату уніфікаційної граматики, що базується на застосуванні правил контекстно-вільної граматики поширених логікою характеристик[15]. Але її спрямованість на синтаксичну складову (застосування графів характеристик) та відсутність засобів щодо завдання необхідних семантичних обмежень є вагомими недоліками при розробці клінічної класифікації діагнозів.

На засадах методики описання фреймових систем слід відзначити, що моделлю класифікації можна вважати композицію наступних складових: аксіоми та перелік конструкторів, здатних описати фрейм-концепт; структура слоту з визначенням відповідних граней; визначення функціональної складової класифікації, що може бути визначена як набір команд мови керування класифікацією та базується на перших двох складових. Аксіоми та конструктори концептів були визначені в попередньому параграфі. Структуру слоту, тобто множину його граней, що об'єднує семантичну та синтаксичну складові, наведено в табл. 3. Необхідно відзначити існування термінального рядка, який не належить жодному слоту.

Основу функціонального рівня складають команди управління фреймовою мережею класифікації на засадах описаної структури. Серед особливостей відзначимо команди завдання абстрактних фреймів, які складаються з певної множини слотів характеристик, успадкування від абстрактних фреймів і т.п.

Таблиця 3

Загальна структура слоту фрейму УКСКХ

Ім'я грані слоту	Опис грані слоту
SuperSlot	Слот, від якого наслідуються даний
MemberSlot	Ідентифікатор слоту
SlotIsDependentOf	Слот, від якого даний залежить
ValueClass	Ім'я концепту (фрейму), з яким задається відношення
Cardinality.Min	Обмеження на мінімальну кількість значень
Cardinality.Max	Обмеження на максимальну кількість значень
OrderInFrame	Порядок слоту у фреймі
OrderInSlot	Порядок підпорядкованого слоту в слоті
DirectionForOrderInSlot	Показчик напрямку місцезнаходження підпорядкованого слоту в слоті
TerminalString1	Термінальний рядок перед КХ у фреймі
TerminalString2	Термінальний рядок за КХ у фреймі
TerminalString3	Термінальний рядок, який стоїть за слотом у фреймі

Розглянемо більш детально особливості механізму успадкування синтаксичної складової. Проблема полягає в тому, що при звичайному методі успадкування такі синтаксичні особливості наданого опису, як відмінок та множина, втрачаються. Можливими рішеннями можуть бути: успадкування тільки семантичної складової, з повторним завданням синтаксичних термінів, враховуючи описані особливості; надання додаткового механізму трансформації синтаксичних термінів філерів слотів абстрактного фрейму у відповідну форму на засадах додаткових ознак. Перший метод приводить до збільшення часу, що витрачається експертом на редагування класифікації, другий, у чистому вигляді, до складності логіки програми керування.

Таким чином, оптимальним може вважатися інтеграція описаних підходів, за яким за заданими правилами здійснюється генерація та подальше копіювання заданих термінів до множини значень відповідного слоту; при цьому експерту надається можливість змінювати ці значення. Важливо підкреслити, що з семантичної точки зору концепти лишаються незмінними.

Формально, процес успадкування може бути описаним наступним чином: $\forall_{C \leq F} \forall_{s \in F.S} .s \in C.S \equiv Copy(GT(s, F.s.GTRules), C.S)$, де F абстрактний фрейм, C фрейм, $F.S$ – множина слотів супер-фрейму, $Copy$ – предикат, що описує процес копіювання відповідної частини мережі, що пов’язана з успадкованим слотом, GT – предикат, що описує синтаксичну трансформацію термінів при успадкуванні на основі описаних правил $GTRules$ для даного слоту.

Слід відзначити також правило відносно змін у абстрактному фреймі. Без перевірки фактів, заданих на основі фрейму C відносно змінених слотів, зміна в слотах не можлива, що може бути виражено як $\forall_{C \leq F} \forall_{s \in F.S} \forall_{c \in C \wedge c.s.v \neq \varepsilon} .Update(F.s) \rightarrow Update(C.s) \rightarrow Check(c.s.v)$, де c – екземпляр-факт C з заданим значенням відповідного слоту, s – слот, v – філер слоту в екземплярі-об’єкті.

Розглянемо перехід від фреймової до об’єктно-орієнтованої моделі класифікації. Як було описано в розділі 2, перехід може здійснюватися наступним чином: структурна частина отриманої фреймової репрезентації може бути відображена на даталогічну модель, а функціональна – на об’єктно-орієнтовану модель, що описує контур керування класифікацією.

Концептуальна модель класифікації представлена на рис. 2. Основою класифікації є композиція класів *Frame* та *Slot* (табл. 4, 5), що відповідає за репрезентацію фреймової мережі клінічних діагнозів. Особливістю моделі є використання циклічного зв'язку на базі шаблону *Composite*, що дозволяють репрезентувати мережу з необмеженою кількістю рівнів ієрархії „частина-ціле”. Залежність слотів у рамках фрейму (конструктор імплікації слотів) визначається циклічним зв'язком класу *Slot*, що дозволяє задати нескінчену кількість вкладених слотів. Класи *AFrameFrame* та *AFrameSlot* відповідають за реалізацію механізму успадкування фреймів-концептів від абстрактних фреймів, при цьому передбачається можливість реалізації часткового успадкування (лише деяких слотів абстрактного фрейму). Також важливою особливістю при реалізації успадкування є існування механізму граматичної трансформації філерів визначеного слоту, при якому зі значенням слоту абстрактного фрейму можуть бути пов'язані схеми граматичної трансформації, які визначаються композицією класів *GT*, *GTItem*, *GTTypeID*: клас *GT* визначає схему трансформації; клас *GTItem* визначає елементарну трансформацію в рамках фрейму-значення; *GTType* визначає тип трансформації. Елементарною трансформацією можна вважати заміну вказаних маркерів (*Marker*) в значенні атрибуту *TS0* відповідного фрейму на фразу (*Description*) за визначеною схемою. Клас *Axis* відповідає за репрезентацію осей класифікації, тісно пов'язаних з визначенням слотів фреймів. Класи *ICD*, *ICDFrame* відповідають за зв'язок клінічної класифікації з МКХ-10. Приклад завдання фрейму діагнозу «шлунково-стравохідний розривно-геморагічний синдром» (К 22.6 по МКХ-10) з макетом {, (!I) {, крововтрата !T ст. тяжкості}} показано в таблицях 6, 7.

Таблиця 4

Структура таблиці «фрейми»

Атрибут	Призначення
FrameID	Ідентифікатор фрейму
SlotID	Ідентифікатор слоту
Element	Номер характеристики даного класу для даного слоту
TS0	Значення у випадку примітивного концепту або термінальний рядок, який передує першому слоту у випадку фрейму
Note	Додатковий атрибут для більш детального опису

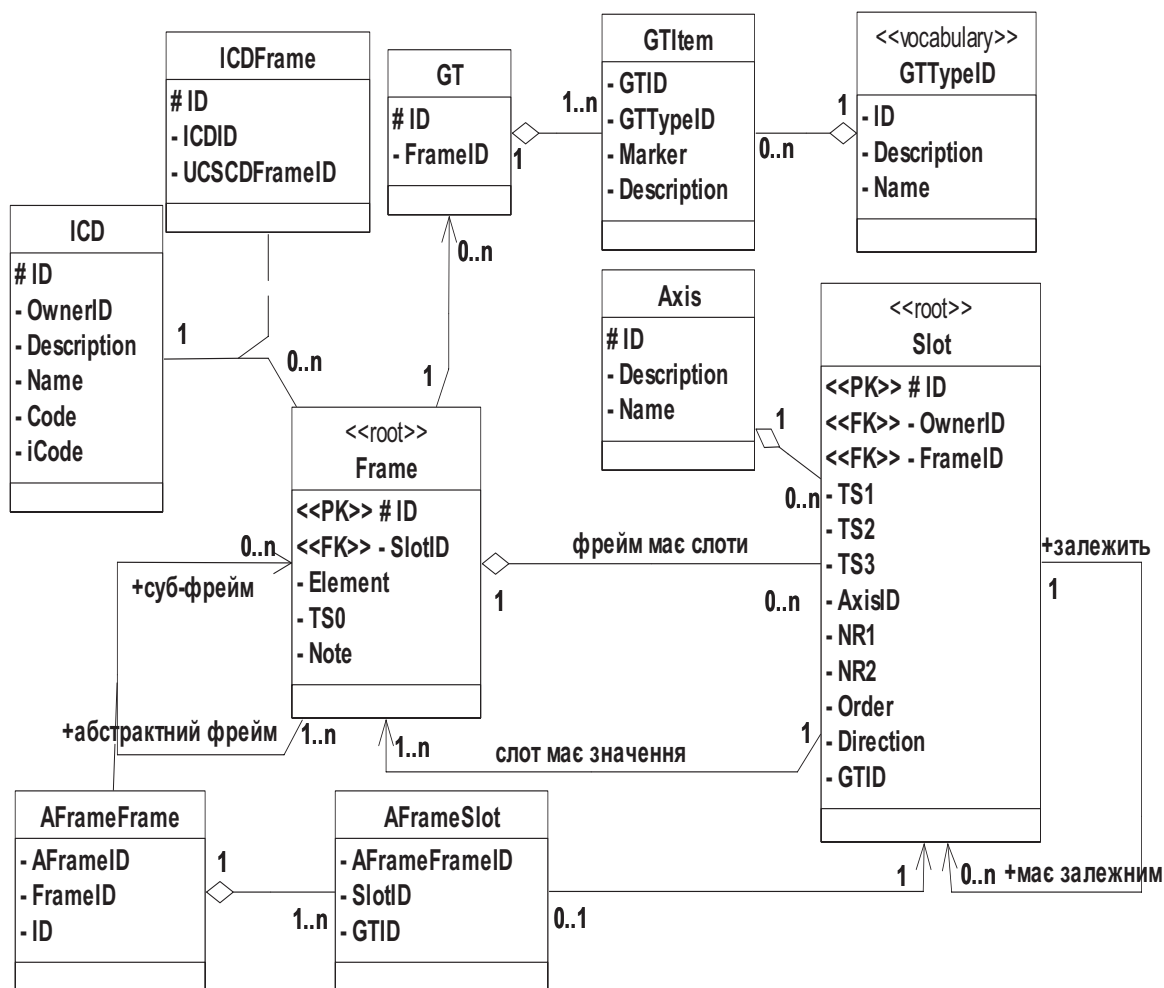


Рисунок 2 - Концептуальна модель УКСКХ

Таблиця 5

Структура таблиці «слоти»

Атрибут	Призначення
FrameID	Ідентифікатор фрейму
SlotID	Ідентифікатор слоту
HSlotID	Ідентифікатор слоту, від якого залежить даний.
TS1	Термінальний рядок символів перед гранню Axis
Axis	Вісь
TS2	Термінальний рядок символів після грані Axis
TS3	Термінальний рядок символів за слотом
Order	Порядок слоту в фреймі/слоті(У разі залежного слоту)
Direction	Показчик місця (має сенс у разі залежного слоту).
NR1	Обмеження на мінімальне число значень
NR2	Обмеження на максимальне число значень

Таблиця 6

Фрейми і значення

ID	SlotID	Axis	Element	T ₀ /Value
100	0			Синдром Малорі-Вейса з кровотечею
120	210	I	1	активна струминна
121	210	I	2	активна капілярна
...				

144	211	T	1	легкого
145	211	T	2	среднього

Таблица 7

Слоты

FrameID	SlotID	HSslotID	TS ₁	Axis	TS ₂	TS ₃	Order	oif	Direction	NR1	NR2
100	210	0	,	I)	.	0	1	0	1	1
100	211	210	,крово- втрата	T	ступінь тяжкості		1	0	0	1	1

Висновки. В даній роботі надано семантико-синтаксична модель класифікації клінічних діагнозів, що відповідає основним вимогам щодо успішного використання в ЛПЗ України, якими є: простота і компактність надання множини клінічних діагнозів базуючись на композиційній архітектурі класифікації; врахування семантичних особливостей, пов'язаних із задачею формування лише семантично-коректних понять; наявність механізмів формування лише синтаксично-правильних описів діагнозів. Формальна модель клінічного діагнозу, як елементу класифікації, включає семантичну і синтаксичну складові, які описані на базі формальних граматик і дескрипційної логіки. Концептуальна модель класифікації діагнозів основана на формальній семантико-синтаксичній моделі і надається з використанням фреймового і об'єктно-орієнтованого підходів до опису знань.

ЛИТЕРАТУРА

1. Гриценко В.И. Медицинское информационное пространство. Госпитальные информационные системы / В.И. Гриценко, А.М. Сердюк, М.И. Вовк, А.Б. Котова. – К.: Международный науч.-учеб. центр ЮНЕСКО\МПИ информ. технологий и систем, 1998. – 16 с.
2. Гриценко В.И., Вовк М.И., Котова А.Б. Введение в архитектуру информационного пространства. Модели. Проблемы развития. Киев, Наукова думка, 2003. – 168 с.
3. Емелин И.В. Интеграция стандартов медицинской информатики. – М.: Кремлевская медицина, 2000. – С. 26–36.
4. Keizer N.F., Abu-Hanna A., Zwetsloot J.H.M. Understanding Terminological Systems (I): terminology and typology // Methods of information in Medicine. – 2000. – № 39. – P. 16–21.

5. Spackman K.A., Campbell K.E. Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies // Proceedings AMIA Annual Fall Symp. – 1998. – P. 740–744.
6. Rothwell D. J. SNOMED-based knowledge representation // Meth Info Med. – 1995. – № 34. – P. 209–213.
7. Rothwell D., Coty R. Managing Information with SNOMED: Understanding the model // SCAMC, 1996. – P. 80–83.
8. Read JD, Benson TJR. Comprehensive coding. British Journal of Healthcare Computing. – 1986 – №3 – p. 622-625.
9. Ceusters W., Smith B., Flanagan J. Ontology and Medical Terminology: Why Description Logics Are Not Enough // Towards an Electronic Patient Record (TEPR 2003), San Antonio, Boston, MA: Medical Records Institute (CD-ROM publication), 10–14 May, 2003.
10. Baader F., Brandt S., Lutz C. Pushing the EL envelope // Proc. of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers. – p.364-369.
11. The description logic handbook: Theory, implementation, and applications / Edited by Franz Baader, Deborah L. McGuinness etc. – Schneider. – 2003. – 574 p.
12. Библиотека практического врача. Унифицированная клинико-статистическая классификация болезней. Болезни пищеварения / Под ред. чл.-корр. АМН Украины Филиппова Ю.А. – Днепропетровск, 1998. – 40 с.
13. Ханецкий В. С., Литвинов А. А. Методика формирования модели диагноза на базе унифицированной клинико-статистической классификации болезней // Информационные технологии. – 2002. – №4. – С. 49–54.
14. Глушков В.М. Алгебра. Языки. Программирование. – 1978. – 320 с.
15. Smolka G. Feature constraint logics for unification grammars. Journal of logic programming. – vol.12. – 1992 – P.51-87.
16. Nebel B., Smolka G. Attributive description... and the rest of the World. In Textunderstanding in LILOG: Integrating computational linguistic and artificial intelligence. Springer-Verlag, Berlin, Heidelberg, New York. – 1991. – P.439-452.
17. Marvin Minsky. A Framework for Representing Knowledge. MIT-AI Laboratory Memo 306, June, 1974
18. Kifer M., Lausen G., Wu J. Logical Foundations of Object-Oriented and Frame-Based Languages. // Journal of the Association for Computing Machinery. – vol. 42 -- №4 -- 1995. – P. 741-843.
19. Karp, PD,. The Design Space of Frame Knowledge Representation Systems. Technical Report 520, SRI International Artificial Intelligence Center, 1992. 49 p.

Получено 22.11.2007 г.