

ФОРМАЛІЗАЦІЯ КЛІНІЧНОГО ДІАГНОЗУ І МОДЕЛЬ ЕЛЕКТРОННОЇ КЛАСИФІКАЦІЇ ДІАГНОЗІВ ДЛЯ МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ УКРАЇНИ

Аналіз останніх публікацій та постановка проблеми.

Проблема будівництва єдиної медичної термінології була і лишається однією з ключових у медичній інформатиці[1] [2]. Так, в [3][4], відзначається, що медичні дані, що зберігаються у комп'ютерних історіях хвороб, гостро вимагають більшої структурованості та можливості керування для задач вводу даних, їх обробки та відображення. З іншого боку, слід відзначити, зростання складності класифікаторів, що перетворюються в термінологічні бази знань (онтології) [2]. Так, у [5][6][2] практично доведена неможливість застосування нумеративної схеми або підходу у якості основи побудови класифікацій медичної термінології, такої, як процедури, діагнози на звісному рівні деталізації, що дозволяє відображати реальність лікувально-діагностичного процесу (clinically sensible and non-ambiguous за [7]).

Формалізація термінологічної системи тісно пов'язана з її мета-моделлю та є важливим кроком у розумінні системи і, як слід, потенціалу її використання[4]. В роботі [8] розглядається можливість використання апарату логіки предикатів та концептуальних графів[9] щодо формалізації термінологічних систем. Як результат формалізм концептуальних графів став фундаментом Snomed International.

Сучасні термінологічні системи базуються на формалізмі дескрипційної логіки, що є основою семантичної складової класифікації і впливає на дескриптивну(введення, завдання концептів) та функціональну(пошук, процес виводу) складові мови керування класифікацією, наприклад мова GRAIL класифікації GALEN [10].

В роботі [11] розглядається спроба формалізації п'яти найвідоміших термінологічних систем: МКХ[12][13]; NHS СТ(Код Ріда)[14][15]; SNOMED [16][6]; UMLS [17]; GALEN [18] [19]. Методика формалізації базується на використанні апарату Entity

Relationship Diagrams (ERD), як інструменту, що дозволяє наглядно відобразити концепти та їх стосунки, і логіки предикатів першого порядку, у ролі універсального, строгого математичного апарату, що доповнює діаграми з боку ясного відображення комплексних обмежень та уникнення багатозначності та інконсистентності.

Базуючись на критеріях щодо оцінки термінологічних систем, що надані в [20][21][22][2], можуть бути сформульовані загальні положення щодо ідеальної медичної термінологічної системи для МІС:

1. Термінологія має бути концептно-орієнтованою. Необхідне чітке відділення концептів від термінів
2. Формалізація означення концептів.
3. Термінологічна система має бути здібною відображати клінічні процеси (зміст та покриття термінології)
4. Кожен концепт має бути визначеним у системі тільки один раз і мати ясне, чітке означення
5. Відображення до існуючих адміністративних систем звітності (МКХ-10 та інші)
6. Композиційна модель
7. Підтримка синонімів
8. Існування атрибутів та надання механізмів підтримки модифікації або вдосконалення основного терміну.
9. Існування механізму ієрархій та спадкоємство, при цьому слід прийняти до уваги, що термін може мати декілька родителів.
10. Користування вільними від смислу ідентифікаторів
11. Унікальні ідентифікатори (код не може бути застосованим, якщо він є архаїчним).
12. Існування означень-пояснень концептів
13. Лінгвістична незалежність
14. Існування механізмів формування синтаксично коректних конструкцій на засадах заданих кодів.

Спроби формалізації напису діагнозу у вигляді формули були зроблені М.М. Амосовим та співробітниками у 1969 році[23], які взяли за основу родо-видові зв'язки, кодуючи діагноз, як сукупність літер та чисел, де класи мають код з літер, групи діагнозів та самі діагнози - числовий.

У роботі [24] надана модель уніфікованої клініко-статистичної класифікації діагнозів (далі УКСКХ), що була розроблена на базі проблемної лабораторії „нові інформаційні технології в медицині” ДДМА Дніпропетровська. В [25] розглядалася об’єктна модель фреймової класифікації діагнозів, створеної на базі УКСКХ.

Постановка проблеми

На поточному стані, в Україні питання використання тієї чи іншої термінологічної системи лишається невирішеним. З одного боку, слід відзначити важкість впровадження іноземних термінологічних систем до вітчизняних установ, що потребує значних коштів та зусиль. З другого, існування ряду проблем пов’язаних з їх структурою та змістом[26]. Можливий шлях щодо вирішення цієї проблеми складається у розробці та впровадженні національного стандарту з приводу термінологічної системи, подібно до [27], враховуючи світовий опит і приймаючи до уваги подальший неминучий перехід до однієї з стандартних термінологічних систем. При цьому доцільно проведення цього процесу поетапно, і найбільш пріоритетною для впровадження, може вважатися термінологія щодо деталізованого опису діагнозів та процедур, як центральних елементів оцінки лікувально-діагностичної практики. Першим кроком стосовно розробки такої системи стає питання надання її формальної моделі.

Мета роботи

Метою даної роботи є спроба сформулювати вимоги та надати формалізовану модель електронної класифікації хвороб щодо використання в МІС України, враховуючи опит розробки та впровадження УКСКХ. Особлива увага зосереджена на шляхах взаємодії з передовими термінологічними системами.

Результатом модулювання слід вважати концептуальну модель класифікації з визначенням можливих стосунків(та їх характеру) з вже існуючими термінологічними системами.

Основна частина

Структура класифікації диктується здібністю забезпечити вимоги користувачів та її використання. З цього приводу можна визначити дві складових: семантичну і синтаксичну.

Сформулюємо основні правила щодо семантичної складової з наданням формального опису, використовуючи логіку предикатів першого порядку.

1. Існує множина концептів діагнозів, які є основою реальних формулювань, що використовуються у клінічній практиці. І кожен реальний діагноз (точніше, його репрезентація) може бути узагальненим до одного з концептів діагнозів.

Так, у номенклатурі SNOMED у якості основи для завдання клінічних діагнозів використовується вісь D(diagnosis), що має пряме відображення до класифікації ICD-9-CM. У класифікації NHS CT у якості термінологічної вісі діагнозів виступає класифікація МКХ-10, що є адміністративною термінологічною системою, тобто за якою проходить вся звітність ВОЗ (всесвітня організація здоров'я).

2. Існує множина-решітка додаткових характеристик-концептів що можуть поширювати, вдосконалити смисл концептів діагнозів. Композиції з концептів-діагнозів і концептів-характеристик можуть складати реальні клінічні описання хвороб. Подібна ситуація має місце у SNOMED RT , SNOMED CT, NHS CT, GALEN.

Так, можна записати, що існує термінологія, яка є об'єднанням множин C і D, а точніше є супремумом для решіток C і D.

$$T = C \cup D = \bigcup_{i=N} C_i \cup D \quad (1)$$

T – загальна множина концептів, C – множина концептів-характеристик розташованих по термінологічним категоріям , D – множина діагнозів.

3. Існує множина лінгвістичних термінів, що є пов'язаною з множиною концептів-характеристик та концептів-діагнозів, при цьому один з цих термінів є пріоритетним до використання. Пріоритетний термін використовується в NHS CT.

$$\forall_{c \in T} \exists!_{t \in L} \exists_{t_1 t_2 \dots t_s \in L} . hasTerm(c, t, t_1 t_2 \dots t_s) \quad (2)$$

інакше цю пропозицію можна описати

$$\forall_{c \in T} \exists!_{t \in L} \exists_{t_c \in L} . hasPrefferedTerm(c, t) \wedge t \notin t_c \wedge hasSynonim(c, t_c) \wedge |t_c| \geq 0 \quad (3)$$

де t – пріоритетний термін, t1t2...ts – терміни-синоніми що складають підмножину tc, L – множина лінгвістичних термінів, що залежить від мови.

Тут необхідно підкреслити різницю між семантичної та синтаксичної складовими системи. Так, ім'ям концептів є

числа(ідентифікатори), при цьому вказується їх взаємовідношення. Терміни є характеристиками концептів.

4. Кожен концепт, в ідеалі, повинен мати одне ясне означення. Така можливість є в UMLS, GALEN, але відсутня у SNOMED.

$$\forall_{c \in C} \exists!_{descr \in L} .hasDescription(c, descr) \quad (4)$$

5. Концепти можуть бути визначеними за допомогою інших концептів і логічних операцій-зв'язок, або бути нащадками інших (generic relations). Такий механізм широко застосовується у UMLS, GALEN, NHS CT, SNOMED. У SNOMED застосовуються зв'язки І та АБО, символ “=” інтерпретується як відношення “is-a”. Так загальна формула може бути визначена як:

$$\exists_{c \in C_i} \exists_{c_1 c_2 \dots c_m \in T \setminus c} .c = r(c, c_1 c_2 \dots c_m) \quad C_i \subset T \quad (5)$$

Стосовно означень одного концепту через інші

$$\exists_{c \in C_i} \exists_{c_1 c_2 \dots c_m \in T \setminus c} .c =_{Def} (c_1 c_2 c_3) \vee \dots \vee (c_{m-1} c_m) \quad (7)$$

З боку відношень „підклас-суперклас”(generic relations) має місце наступне відношення

$$\exists_{c \in C_i} \exists_{c_1 c_2 \dots c_m \in C_i \setminus c} .c \sqsubseteq c_1 \sqsubseteq c_2 \sqsubseteq c_3 \sqsubseteq \dots \sqsubseteq c_{m-1} \sqsubseteq c_m \sqsubseteq \sqcup C_i \quad C_i \subset T \quad (8)$$

6. Концепти-діагнози та концепти-характеристики можуть мати атрибути, що визначають їх зв'язок з іншими концептами. Так, можна казати, що деякий атрибут v концепту c , має область значень $C_v = \{c_1, c_2, c_3, \dots, c_m\}$.

$$\exists_{c \in C_i} \exists_{v \in V} \exists_{c_1 c_2 \dots c_m \in C_v} .cv = c \oplus cc_1 \oplus cc_2 \oplus \dots \oplus cc_m, \quad C_i \subset T \setminus C_v, C_v \subset T \setminus C_i \quad (9)$$

v - атрибут, V - множина типів атрибутів. Оператор нерівнозначності (\oplus) використовується для відображення свойства атрибута приймати лише одне значення із відповідної області значень.

Далі, атрибути можуть бути обов'язковими або необов'язковими (виконується в GALEN, частково в NHS CT, але не в SNOMED). Так, у разі необов'язкового атрибута маємо

$$\exists_{c \in C_i} \exists_{v \in V} \exists_{c_1 c_2 \dots c_m \in C_v} .cv = c [1 \vee (c_1 \oplus c_2 \oplus \dots \oplus c_m)], \quad C_i \subset T \setminus C_v, C_v \subset T \setminus C_i \quad (9)$$

У разі обов'язкового v маємо

$$\exists_{c \in C_i} \exists_{v \in V} \exists_{c_1 c_2 \dots c_m \in C_v} .cv = c (c_1 \oplus c_2 \oplus \dots \oplus c_m), \quad C_i \subset T \setminus C_v, C_v \subset T \setminus C_i \quad (10)$$

7. Атрибути можуть прийняти декілька значень одночасно. Така ситуація вирішена в UKSKX. Так, у разі обов'язкового v маємо

$$\begin{aligned} \exists_{c \in C_i} \exists_{v \in V} \exists_{c_1 c_2 \dots c_m \in C_v} .c v = c (c_1 \vee c_1 c_2 \vee c_1 c_2 c_3 \dots \vee c_1 \dots c_{m-1} c_m), \\ C_i \subset T \setminus C_v, C_v \subset T \setminus C_i \end{aligned} \quad (11)$$

у разі необов'язкового

$$\begin{aligned} \exists_{c \in C_i} \exists_{v \in V} \exists_{c_1 c_2 \dots c_m \in C_v} .c v = c (1 \vee c_1 \vee c_1 c_2 \vee c_1 c_2 c_3 \dots \vee c_1 \dots c_{m-1} c_m), \\ C_i \subset T \setminus C_v, C_v \subset T \setminus C_i \end{aligned} \quad (12)$$

8. Атрибути можуть бути залежними один від одного. Так, може існувати атрибут u , значення якого може бути заданим тільки у разі надання значення атрибуту v . Така ситуація вирішена в УКСКХ.

$$\exists_{c \in T} \exists_{v \in V \setminus u} \exists_{u \in V \setminus v} .c (u \supset v) \quad (13)$$

9. Існує обмеження на тимчасову присутність концептів-характеристик у рамках одного формулювання діагнозу.

$$\begin{aligned} \exists_{c \in C_i} \exists_{v \in V \setminus u} \exists_{u \in V \setminus v} \exists_{c_1^u c_2^u c_3^u \dots c_n^u \in C_u} \exists_{c_1^v c_2^v c_3^v \dots c_m^v \in C_v} .c u v = \\ c \wedge (c_1^u \oplus c_2^u \oplus c_3^u \dots \oplus c_k^u) \wedge (c_1^v \oplus c_2^v \oplus c_3^v \dots \oplus c_m^v) \wedge \\ \wedge \neg (c_i^u c_k^v \oplus c_k^u c_l^v), \\ C_i \subset T \setminus (C_u \cup C_v), C_u \subset T \setminus (C_i \cup C_v), C_v \subset T \setminus (C_i \cup C_u) \end{aligned} \quad (14)$$

Так для двох атрибутів з областями значень C_u і C_v можна вказати дозволені і/або заперечені набори значень.

Дана модель відповідає вимогам наданим [20],[21], [2], а також вдосконалює ряд питань відповідно до семантичних відношень, пов'язаних з обмеженнями при завданні та використанні термінологічної системи, виявленими під час користування електронної версії УКСКХ.

Розглянемо синтаксичну складову термінологічної системи. З лінгвістичної точки зору, описання клінічного діагнозу є граматично правильним висловленням, яке може бути описано як упорядкована множина(ланцюг) з термінів та зв'язок. Проблема граматично-коректних висловлювань має місце у всіх існуючих термінологічних системах[20][21][26][2], особливо це стосується впровадження систем подібних SNOMED в Україні. У якості засобів які призначені вирішувати цю проблему можна навести модуль SPECIALIST lexicon UMLS[17] [28], що матиме дозволяти конвертувати натуральномовну фразу у код термінологічної системи, що підтримується. Але при декодуванні як правило лікар має справу не з натурально-мовним висловлюваннями. Побудова універсальних інтерпретаторів, які здібні перекладати семантичне описання в натурально-мовне речення будь-якої мови є складною задачею комп'ютерної лінгвістики. В УКСКХ

для цієї мети пропонується включити до слотів фреймів додаткові синтаксичні грані, відображаючи таким чином фрейм у своєрідний граматичний шаблон, фрейм об'єднує семантичні складові з синтаксичними використовуючи механізм гранів(facets). Так, наприклад, діагноз К40, який має, за нотацією УКСКХ, такий шаблон „{!L} {!Q} пахова грижа, {ускладнена *O&.}” може бути описаним наступною граматиною:

$K40 \rightarrow S_1 S_2$ пахова грижа S_3 ?

$S_1 \rightarrow L$

$L \rightarrow L_1 | L_2 | \dots$

$L_1 \rightarrow$ Лівостороння

$L_2 \rightarrow$ Правостороння

$S_2 \rightarrow Q$

$Q \rightarrow Q_1 | Q_2 | \dots$

$Q_1 \rightarrow$ вправна

$Q_2 \rightarrow$ невправна

$S_3 \rightarrow$ ускладнена O

$O \rightarrow O_1 | O_2 | \dots$

$O_1 \rightarrow$ защемлення R?

$R \rightarrow R_1 | R_2 | \dots$

$R_1 \rightarrow$ пасма великого сальнику

$R_2 \rightarrow$ тонка кишка

$O_2 \rightarrow$ гангрена кишки

Таким чином, моделлю класифікації буде множина граматик, кожна з яких являє собою компактний запис усіх варіантів визначеного діагнозу. Концепти лишаються незалежними від лінгвістичних складових, але поряд з семантичним описанням концепту, будується додатковий лінгвістичний шаблон, що здатен враховувати синтаксичні особливості репрезентації. Ці шаблони можуть перекладатися на різні мови, змінюючи при цьому порядок слотів, включаючи чи виключаючи додаткові термінальні ланцюжки, які виконують роль граматичних зв'язок, змінюючи граматичні особливості значень. При цьому семантична складова фрейму лишається незмінною.

Концептуальна модель. Безсумнівним фактом є те, що сучасні передові термінологічні системи мають важливу підтримку, поширюються та розвиваються, і люба вітчизняна розробка з приводу

атрибутам. І відповідає за пріоритетну семантичну складову загальної системи. У разі відсутності деякої термінологічної системи УКСКХ приймає ці обов’язки на себе. УКСКХ відповідає за додатковий рівень обмежень щодо семантиці описаного діагнозу (правила 7,8,9), та за граматичну інтерпретацію. Надбудова УКСКХ пов’язана з термінологічною системою за правилом: кожний фрейм(значення) УКСКХ має відображення до концепту термінологічної системи та кожний слот має бути відображеним до атрибуту ТС. Слід особливо відзначити блок класів FrameConstraint, SlotConstraint, ValueConstraint, TypeOfConstraint, які вирішують задачу консистентності на тимчасову присутність концептів-характеристик у рамках одного формулювання діагнозу (правило 9) . Так, FrameConstraint вказує на існування обмежень та на їх характер у вигляді: всі комбінації значень усіх атрибутів є консистентними; всі консистентні/інконсистентні окрім; існують консистентні/інконсистентні. SlotConstraint – надає обмеження на комбінацію слотів: з’явлення любых значень двох слотів є інконсистентним/консистентним; всі консистентні/інконсистентні окрім; існують консистентні/інконсистентні. ValueConstraint – обмеження на комбінацію значень таким чином: одне значення одного слоту і кожне значення другого слоту є консистентними/інконсистентними; одне значення одного слоту і одне значення другого слоту є консистентним/ інконсистентними.

Висновки. Сформульовані основні правила щодо семантичної складової з наданням формального опису, використовуючи логіку предикатів першого порядку. Надана концептуальна модель класифікації з визначенням можливих стосунків(та їх характеру) з вже існуючими термінологічними системами. Так, у випадку застосування деякої термінологічної системи, основним претендентом якої є SNOMED CT, модифікація існуючої УКСКХ можна розцінювати як надбудову з приводу описання діагнозів. При цьому весь процес кодування надається у вигляді основної термінологічної системи.

ЛІТЕРАТУРА

1. Sittig DF. Grand challenges in medical informatics. J Am Med Inform Assoc. 1: 412-413, 1994.
2. Alan L. Rector. Clinical terminology: Why is it so hard? Methods Inf. Med., 38:239-252, 1999.

3. Moorman P, van Ginneken A, van der Lei J, van Bommel JH. A model for structured data entry based on explicit descriptive knowledge. *Yearbook of Medical Informatics* 1995: 195-204.
4. de Keizer NF, Abu-Hanna A, Zwetsloot JHM. Understanding Terminological Systems (I): terminology and typology, *Methods of information in Medicine* 39:16-21, 2000.
5. Spackman KA, Campbell KE, Cфтй RA. SNOMED-RT: a reference terminology for health care. *Proc AMIA Annu Fall Symp.*:640-4,1997.
6. Spackman KA, Campbell KE. Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. *Proceedings/AMIA Annual Fall Symposium.* :740-4, 1998.
7. De Keizer N, Abu-Hanna A, Cornet R, Zwetsloot-Schonk J, Stoutenbeek C. Analysis and Design of an Ontology for intensive care diagnoses. *Method Inform Med* 1999; 38: 102-12.
8. Campbell K, Das A, Musen M. A logical foundation for representation of clinical data. *J Am Med Informatics Assoc* 1994; 1:218-32.
9. Sowa JF. *Conceptual Structures*. Reading, MA: Addison-Wesley, 1984.
10. THE DESCRIPTION LOGIC HANDBOOK: Theory, implementation, and applications. *Edited by* Franz Baader, Deborah L. McGuinness, Daniele Nardi, Peter F. Patel-Schneider, 2003
11. de Keizer NF, Abu-Hanna A. Framework for understanding Terminological Systems (II): Experience with Conceptual and Formal Representation of Structure. *Methods of Information in Medicine*, 39:22-29, 2000.
12. ICD-9, *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, World Health Organisation, Geneva , (1977).
13. ICD-10, *International Statistical Classification of Diseases and Related Health Problems - 10th Revision*, World Health Organisation, Geneva, (1993).
14. Robinson D, Comp D, Schulz E, Brown P, Price C. Updating the Read Codes: Userinteractive Maintenance of a Dynamic Clinical Vocabulary. *J Am Med Inform Assoc* 1997; 4 (6): 465-72.
15. Schulz E, Price C, Brown P. Symbolic Anatomic Knowledge Representation in the Read Codes Version 3: Structure and Application. *J Am Med Inform Assoc* 1997; 4: 38-48.
16. D. J. Rothwell, SNOMED-based knowledge representation, *Meth Info Med*, 34, 209-13, 1995.
17. D. A. Lindberg, B. L. Humphreys, A. T. McCray. The Unified Medical Language System. *Methods Inf Med* 1993 Aug;32(4):281-91.
18. Rector A, Solomon W, Nowlan W, Rush T, Zanstra P, Claassen W. A Terminology Server for medical language and medical information systems. *Method Inform Med* 1995; 34: 147-57.

19. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The Grail concept modelling language for medical terminology. *Artif Intell* 1997; 9: 139-71.
20. Cimino J, Clayton P, Hripcsak G, Johnson S. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J Am Med Informatics Assoc* 1994; 1: 35-50.
21. Campbell J, Carpenter P, Sneiderman C, et al. Phase II evaluation of clinical coding schemes: completeness, definitions and clarity. *J Am Med Inform Assoc* 1997; 4: 238-51.
22. Cimino J. Desiderata for controlled medical vocabularies in the twenty-first century. *Meth Inform Med* 1998; 37 (4-5): 394-403.
23. Основные задачи медицинской кибернетики/ Н.М. Амосов, А.А. Попов, В.Г. Мельников, Н.А. Гватуа, Р.М. Птуха, А.М. Кочетов, А.М. Мигай// Труды семинара «Медицинская кибернетика», вып. 1. – Киев. – 1969. – с.3-23.
24. В. С. Хандецкий, А. А. Литвинов, Методика формирования модели диагноза на базе унифицированной клинико - статистической классификации болезней. Информационные технологии. №4 2002. Москва. Машиностроение.
25. Литвинов А.А. Использование фреймовой классификации диагнозов в медицинской информационной системе. Сборник тезисов докладов по материалам Международной научно-практической конференции «Математическое и программное обеспечение интеллектуальных систем». Днепропетровск, 2004., с.83-84.
26. W. Ceusters, B. Smith, J. Flanagan, Ontology and medical terminology: Why Description Logics are not enough, Proceedings of TEPR 2003 - Towards an Electronic Patient Record. San Antonio, Texas, May 10-14, 2003 (2003) (CD-ROM publication).
27. Anna Vikström, Ylva Skenjör, Lars-Erik Strender, Gunnar Nilsson Intermapper Consistency between Coders mapping SNOMED CT Concepts to the Swedish Primary Health Care Classification KSH97-P. SEMANTIC MINING CONFERENCE ON SNOMED CT. October 1 - 3 2006. Copenhagen. Denmark
28. MCCRAY AT. The nature of lexical knowledge. *Methods of information in medicine* 37:4-54-5, 353-360, Schattauer, 1998.

Получено 09.01.07 г