

ТРАНСДИСЦИПЛІНАРНЕ ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ІНТЕРАКТИВНИХ ДОКУМЕНТІВ

О. П. Мінцер, В. В. Приходнюк¹, О. Є. Стрижак¹, О. М. Шевцова

Національна медична академія післядипломної освіти імені П. Л. Шупика

¹Інститут телекомунікацій та глобального інформаційного простору НАН України

Досліджено теоретичні узагальнення та можливості практичного використання онтології знань в охороні здоров'я, зокрема в діагностиці та виборі методу лікування при захворюваннях шлунково-кишкового тракту в дітей. Показано, що наслідком існування гомоморфізму між первинною структурою тексту та структурою інформації, що з нього виділяється, є можливість побудови спеціалізованої структури, призначеної для роботи не з об'єктами онтології, а лексемами первинної структури. Побудований на її основі інтерактивний документ $\langle T_{SM}, NS' \rangle$ може використовуватися для формування довільної мережевої інформаційної системи в предметній площині.

Ключові слова: міждисциплінарність, трансдисциплінарність, інтерфейси використання знань, політематичні інформаційні джерела, інтерактивні документи, предметні знання, онтологія знань.

TRANSDISCIPLINARY REPORTING OF INFORMATION WITH INTERACTIVE DOCUMENTS

O. P. Mintser, V. V. Prykhodnyuk¹, O. Ye. Stryzhak¹, O. M. Shevtsova

Shupyk National Medical Academy of Postgraduate Education

¹Institute of Telecommunications and Global Information Space of NAS of Ukraine

Theoretical generalizations and possibilities of practical use of ontology of knowledge in public health, in particular in diagnosis and choice of a method of treatment for diseases of the gastrointestinal tract in children, are investigated. It is shown that the consequence of the existence of a homomorphism between the primary structure of the text and the structure of information from it highlights the possibility of constructing a specialized structure designed to work not with ontology objects but with lexemes of the primary structure. The interactive document $\langle T_{SM}, NS' \rangle$ built on its basis can be used to form an arbitrary network information system in the object plane.

Key words: interdisciplinarity, transdisciplinarity, interfaces for the use of knowledge, polytematic information sources, interactive documents, subject knowledge, knowledge ontology.

ТРАНСДИСЦИПЛІНАРНОЕ ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ С ПОМОЩЬЮ ИНТЕРАКТИВНЫХ ДОКУМЕНТОВ

О. П. Минцер, В. В. Приходнюк¹, А. Е. Стрижак¹, Е. Н. Шевцова

Национальная медицинская академия последипломного образования имени П. Л. Шупика

*¹Институт телекоммуникаций и глобального информационного пространства
НАН Украины*

Исследованы теоретические обобщения и возможности практического использования онтологии знаний в здравоохранении, в частности в диагностике и выборе метода лечения при заболеваниях желудочно-кишечного тракта у детей. Показано, что следствием существования гомоморфизма между первичной структурой текста и структурой информации из него выделяется возможность построения специализированной структуры, предназначенной для работы не с объектами онтологии, а лексемами первичной структуры. Построенный на ее основе интерактивный документ $\langle T_{SM}, NS' \rangle$ может использоваться для формирования произвольной сетевой информационной системы в предметной плоскости.

Ключевые слова: междисциплинарность, трансдисциплинарность, интерфейсы использования знаний, политематические информационные источники, интерактивные документы, предметные знания, онтология знаний.

Вступ. Основною метою використання предметних знань, особливо медичних, є трансфер інформації, а також доказ її істинності. Важливим стає наявність малих термінів загально наукових слів, абстрактної лексики. При цьому судження мають вигляд конкретних висловлювань і тверджень, визначають набори дій, що можуть бути застосовані в процесі вирішення конкретних предметно-тематичних завдань. Трансдисциплінарності дослідження, захоплюючи зони прикордонних (демаркаційних) ареалів наукових дисциплін інтегрують сутнісні основи останніх, утворюючи так звані кластери конвергенції, в яких відбувається потужна синергетична взаємодія за рахунок взаємопроникнення парадигм і конкретних поточних результатів кожної з дисциплін, що входять у той чи інший кластер. Саме такі завдання існують, наприклад, у діагностиці та виборі методу лікування захворювань шлунково-кишкового тракту в дітей [4].

Виділення набору дій на основі системи знань, описаної і представленої в клінічних настановах, можливо на основі застосування до її природномовного тексту процедури структуризації. Для цього реалізуються процедури перетворення книжкового та статейного тексту, представивши його не в звичному вигляді послідовного і за стилем узгодженого викладу інформації, а відобразивши його в сукупності конкретних висловлювань і тверджень. Конкретні предметні висловлювання / твердження, що мають тематичну спрямованість, можуть формувати пасивну базу знань, яка складається з контекстів тематики взаємодії.

Мета роботи: теоретично обґрунтувати інтерфейси використання знань при вивченні міждисциплінарних патологічних станів.

Результати та їх обговорення. Застосування міждисциплінарних та трансдисциплінарних процедур орієнтоване на вирішення таких завдань [8, 9]:

- забезпечення можливості оперативної організації доступу до інформаційних джерел формування знань, що стосуються політематичних описів предметних областей (ПрО) або об'єднаних схожими інтересами сфер діяльності;
- підтримка взаємодії всіх учасників процесу в рамках упорядкованих множин ПрО з можливістю їх розширення;
- забезпечення можливості розширення списку джерел і споживачів різномірних політематичних інформаційних джерел формування знань в рамках певної ПрО або сфери інтересів;

- обмеження доступу до інформаційних ресурсів рамками конкретної ПрО або сфери інтересів у зв'язку з можливістю вирішення попередньої задачі;
- забезпечення можливості для кожного суб'єкта використання інформаційних ресурсів кількох ПрО,
- забезпечення можливості оперативного пошуку джерела необхідних інформаційних ресурсів, що стосується контекстів ПрО.

Отже, головними принципами використання категорії трансдисциплінарності у процесі інтеграції інформаційних ресурсів є :

- множинна упорядкованість інформаційного середовища;
- орієнтація на використання властивостей об'єктів ПрО в процесі розв'язання складних політематичних прикладних задач;
- забезпечення можливості використання суперечливої інформації;
- орієнтація на використання контекстів інформаційних ресурсів;
- забезпечення виявлення новітніх властивостей у систем, які отримані на основі інтеграції їх складників;
- формування, за рахунок множинної упорядкованості, ієрархій контекстів тематики взаємодії.

Для забезпечення трансдисциплінарності представлення всієї сукупності документів використовують онтологічні моделі.

Трансдисциплінарне представлення множин онтологій базується на перетворенні між двома гіпермножинами:

$$f^{ct} : R \rightarrow F \quad (1)$$

де R, F — множини зв'язків і функцій інтерпретації певної онтології O .

Для множини онтологій можна побудувати гіпермножини:

$$\mathfrak{R} = \bigcup_i R_i, F = \bigcup_i F_i \quad (2)$$

де i — індекс, що визначає певну онтологію $O_i = \langle X_i, R_i, F_i \rangle$.

На цих гіпермножинах можна побудувати зворотне до f^{ct} перетворення:

$$f^{tt} : F \rightarrow \mathfrak{R} \quad (3)$$

Тобто певний зв'язок між об'єктами різних за тематикою онтологій може бути представлений

непустою множиною інтерпретуючих функцій з даних онтологій.

Дане перетворення можна розширити і на множини неструктурованих текстів, застосувавши до кожного з текстів оператор редукації (4), що є комбінацією чотирьох операторів, при цьому три з операторів виконують кроки перетворення, а один виконує допоміжну функцію.

$$F_{rd} = F_{l*} \circ F_x \circ F_{smr} \circ F_{ct} \quad (4)$$

де F_{l*} — оператор агрегації, що виконує допоміжну функцію, перетворюючи множини лексем L в множини конструктів L^* .

Конструкти є особливою формою лексем і об'єднують в собі послідовності слів або символів, зокрема словосполучення. Особливістю конструктів є те, що з погляду на подальшу обробку вони можуть розглядатись як лексеми. Отже, множина $L \cup L^*$ може використовуватись у тих самих випадках.

Розглянемо процедуру трансдисциплінарного представлення певної множини онтологій засобами інтерактивних документів. Дане представлення базується на функції (5).

З урахуванням даних особливостей функція контекстної зв'язки матиме вигляд:

$$Q_c(x) = \bigcup_{l \in L_x} Q_S(Q_I(C), l) \quad (5)$$

де C — множина документів, що представляє інформаційне середовище, в рамках якого здійснюється зв'язка;

x — об'єкт, з яким здійснюється зв'язка;

L_x — текстове представлення контексту.

l — певна лексема.

Основним недоліком функції (5) є те, що в її результаті формується невпорядкована множина документів, яка може бути достатньо великою за розміром і тому незручною для оброблення експертом. Вирішити цю проблему можна двома способами:

Задати на множині результатів відношення порядку, яким може виступити відношення релевантності:

$$T_1 R_{rel} T_2 \Rightarrow card(L_{T_1} \cap L_x) > card(L_{T_2} \cap L_x) \quad (6)$$

де L_{T_1}, L_{T_2} — текстові представлення документів T_1, T_2 ;

L_x — контекст об'єкту x , з яким була здійснена контекстна зв'язка.

Виключити з результату документи, які мають недостатню релевантність, що можна зробити

застосуванням в функції (5) операції перетину замість операції об'єднання.

Для його здійснення необхідно виконати трансдисциплінарне перетворення за допомогою функції:

$$Q_{\Pi}(C) = \bigcup_{x \in X_C} \{Q_C(x)\} \quad (7)$$

де C — множина онтологій;

X_C — множина об'єктів, що належать об'єднанню онтологій $\bigcup_{O \in C} O$.

Застосування функції $Q_C(x)$ формує множини об'єктів з різних онтологій, що представляє певне гіпервідношення між відповідними об'єктами. За допомогою сформованих таким чином гіпервідношень можна побудувати трансдисциплінарне представлення O множини онтологій $C = \{ \langle X_i, R_i, F_i \rangle \}$:

$$C \xrightarrow{Q_{\Pi}} \langle \bigcup_i X_i, \bigcup_i R_i \cup Q_{\Pi}(C), \bigcup_i F_i \rangle \quad (8)$$

Перетворення (8) задає найбільш повне представлення наявної в C інформації, що не завжди зручно. Часто необхідно виконати представлення одної вибраної онтології O . В такому випадку необхідно змінити:

$$Q_{TO}(O, C) = \bigcup_{x \in X_O} \{Q_C(x)\} \quad (9)$$

де C — множина онтологій;

X_O — множина об'єктів, що належать онтології O .

Для побудови систем, що використовують трансдисциплінарне представлення інформації, важливими є такі твердження.

Твердження 1. На основі однієї онтології O , що належить множині документів C , можна сформулювати довільну кількість HC .

Твердження 2. Комбінація незалежних натуральних систем SN_i , що приймають на вхід один і той самий набір «дій», є натуральною системою.

Усі трансдисциплінарні перетворення — композиція операторів. Кожен з них виконує один крок перетворення. Повний цикл даного перетворення структурує частину інформації, що міститься у вхідному тексті, після чого перетворення рекурсивне викликається заново, доки не буде виділена вся інформація. Однак кожна зі складових оператора редукації також може бути розбита на складові.

В загальному випадку оператор виконання перетворення F задається базою правил G_R виконання

даного перетворення. Правило $g \in G_R$ має уніфіковану для всіх етапів структуру:

$$g = \langle f_{ap}^g, f_{ir}^g \rangle \quad (10)$$

де f_{ap}^g — функція застосовності, що визначає, чи може правило бути застосоване до певного набору вхідної інформації;

f_{ir}^g — функція перетворення, що задає перетворення вхідної інформації.

Задане правилом g перетворення $F_g: X \rightarrow Y$ має вигляд:

$$F_g(x) = \begin{cases} f_{ir}^g(x), f_{ap}^g(x) \\ x, \neg f_{ap}^g(x) \end{cases} \quad (11)$$

Кожна функція застосовності являє собою λ -терм виду [1, 2, 5]:

$$f_{ap} = (\lambda x_1, x_2 \dots x_{n_g}. t_{ap}(x)) a_1, a_2 \dots a_{n_g} = \\ = t_{ap}(a_1, a_2 \dots a_{n_g}) \quad (12)$$

де запис λx вказує, що дана конструкція являється λ -термом;

x_i — змінна, що приймає значення на множині $L \cup L^*$;

a_i — аргумент функції, що задає значення x_i ;

n_g — кількість аргументів, що повинні бути подані на вхід функції перетворення;

t_{ap} — умова застосовності, вираз, що містить n_g змінних.

В загальному випадку умова застосовності t_{ap} в загальному випадку означає існування гомеоморфізму між орієнтованим графом, утвореним вхідною послідовністю лексем (а також синтаксичними зв'язками між ними), і певним еталонним орієнтованим графом G_{ap} , що являє собою вибраний користувачем підграф первинного представлення T_{sn}^e певного тексту. Як T_{sn}^e може виступати первинне представлення як поточного тексту T_{sn} , так і будь-якого іншого тексту (наприклад, тезауруса **ПДО**). Умова має структуру (13) і складається з *предикатів ідентифікації* [3, 6, 7]. Такі предикати дозволяють ідентифікувати контексти певної лексеми і на основі цього робити висновок про необхідність або відсутність необхідності виконання перетворення. Кожен з предикатів задає певну умову, й умовою застосовності правила є виконання всіх умов, заданих кожним із предикатів. Кількість предикатів у виразі задає число n_g .

$$t_{ap} = c_{p_1}(x_1) \& \dots \& c_{p_n}(x_{n_g}) \& r_{k_{11}}(x_1, x_1) \& \dots \& \\ \dots \& r_{k_{n_g n_g}}(x_{n_g}, x_{n_g}) \quad (13)$$

Одномісні предикати, присутні в виразі, — це предикати ідентифікації лексем. Такий предикат задає умову, якій повинна відповідати певна лексема (або конструкт) з вхідної множини. Предикат має структуру (14):

$$c_p(l) = \begin{cases} 1, p = 0 \vee p = l^T \vee p \in P_l \\ 0, p \neq 0 \wedge p \neq l^T \wedge p \notin P_l \end{cases} \quad (14)$$

Робота такого предикату залежить від *шаблонного параметра* p . Залежно від типу даного параметра предикат може бути:

1. Стандартним предикатом ідентифікації. В такого предиката p — це морфологічна характеристика лексеми. Такий предикат визначає, чи має вхідна лексема задану характеристику ($p \in P_l$).
2. Предикатом ідентифікації ключових слів. В таких предикатів p — це текстове представлення необхідної лексеми, і це значення порівнюється зі значенням вхідної ($p = l^T$). Даний предикат завжди має значення 0 для конструктів.
3. Нульовим предикатом ($p=0$). Такий предикат завжди має значення 1, незалежно від поданої на вхід лексеми.

Двомісний предикат — це предикат ідентифікації зв'язків. Такий предикат визначає, чи міститься між двома заданими лексемами зв'язок заданого типу. Предикат має вигляд (15):

$$r_k(l_1, l_2) = \begin{cases} 1, k = 0 \vee \langle l_1, l_2, k \rangle \in R_{sn} \\ 0, k \neq 0 \wedge \langle l_1, l_2, k \rangle \notin R_{sn} \end{cases} \quad (15)$$

Як і предикат ідентифікації лексем, даний предикат має нульову модифікацію, яка має значення незалежно від вхідних даних. При цьому для коректного використання умови (13) повинна виконуватись задана структурою лексичного аналізатора умова:

$$k_{ij} = 0, i = j \quad (16)$$

База правил задає перетворення F , що має вигляд:

$$F_G(L) = \bigcup_{\tilde{L} \in P(L)} \bigcup_{g \in G} F_g^*(\tilde{L}) \quad (17)$$

де $P(L)$ — множина всіх підмножин L ;

F_g^* — модифікована функція (11), доповнена додатковими умовами.

Додаткових умов накладається дві.

Умова порядку означає, що всі елементи вхідної підмножини повинні бути лінійно впорядковані певним відношенням строгого порядку \mathbf{G} , і має вигляд:

$$f_{ord}(x) = \begin{cases} 1, \forall x_1, x_2 \in x, x_1 \prec x_2 \vee x_2 \prec x_1 \\ 0, \exists x_1, x_2 \in x, x_1 \epsilon x_2 \wedge x_2 \epsilon x_1 \end{cases} \quad (18)$$

Відношення порядку, що можуть використовуватись як \mathbf{G} :

1. відношення слідування \prec
2. транзитивне замикання відношення \mathbf{R}
3. транзитивне замикання відношення, заданого зв'язками \mathbf{R}_{sem}

Умова узгодженості f_{ap+}^g визначає, чи є даний елемент $\tilde{L} \in P(L)$ придатним для обробки правилом \mathbf{g} :

$$f_{ap+}^g(x) = \begin{cases} 1, card(x) = n_g \\ 0, card(x) \neq n_g \end{cases} \quad (19)$$

З урахуванням даних умов перетворення (11) перетворюється на :

$$F_g^*(x) = \begin{cases} f_{tr}^g(x) \cdot f_{ord}(x) \wedge f_{ap+}^g(x) \wedge f_{ap}^g(x) \\ x \cdot \neg f_{ord}(x) \vee \neg f_{ap+}^g(x) \vee \neg f_{ap}^g(x) \end{cases} \quad (20)$$

Вказані представлення перетворення текстів є зручнішими для користувача, оскільки легше представляти в текстовій формі. Вирази (1)–(18) фактично накладають певні умови на топологію графової структури, сформованої вхідною множиною і зв'язками між її елементами кожного текстового документу. Оскільки як конструкти, так і концепти в рамках виконуваних над ними операцій можуть розглядатись як аналог лексем, то і сформована ними графова структура може розглядатись як аналог структури, сформованої лексемами, а саме — первинної структури тексту. Важливим наслідком даного факту є те, що користувачу легко розробити процедуру автоматизованого створення, оскільки така процедура зводиться до простої в реалізації функції вибору довільного контексту. Важливо зазначити, що може використовуватись і первинна структура тексту, що обробляється, — тобто формування інтерактивного документу може виступати в ролі додаткового кроку після синтаксичного аналізу і перед агрегацією довільної текстової структури.

Висновок. Наслідком існування гомеоморфізму між первинною структурою тексту та структурою інформації, що з нього виділяється, є можливість

побудови спеціалізованої структури, призначеної для роботи не з об'єктами онтології, а лексемами первинної структури. Побудований на її основі інтерактивний документ $\langle T_{sm}, NS \rangle$ може здійснюватися для формування довільної мережної інформаційної системи в предметній площині.

Література.

1. Величко В. Ю. Деякі способи виділення відношень між термінами в природномовному тексті / В. Ю. Величко, В. В. Приходнюк // Системний аналіз та інформаційні технології : зб. наук. праць за матеріалами XV конференції (Київ, 27–31 травня 2013 р.). — К. : НТУУ «КПІ», 2013. — С. 406.
2. Величко В. Ю. Інструментарій автоматизованого виділення відношень з текстів технічної тематики / В. Ю. Величко, С. В. Сирота, В. В. Приходнюк // Системний аналіз та інформаційні технології : зб. наук. праць за матеріалами XVI конференції (Київ, 26–30 травня 2014 р.). — К. : НТУУ «КПІ», 2014. — С. 346–347.
3. Величко В. Ю. Спосіб автоматизованого виділення відношень між термінами з природномовних текстів технічної тематики / В. Ю. Величко, В. В. Приходнюк // Knowledge — Dialogue — Solution : зб. праць XX міжнародної конференції. — К. : ІТНЕА, 2014. — С. 27–28.
4. Мінцер О. П. Прийняття рішень в трансдисциплінарних задачах / О. П. Мінцер, Л. Ю. Бабінцева, О. М. Шевцова // Системи підтримки прийняття рішень. Теорія і практика : одинадцята дистанційна наук.-практ. конф. з міжнар. уч. : зб. пр., 5 черв. 2017 р., Київ. — К. : ІПММС НАНУ, 2017. — С. 99–100.
5. Приходнюк В. В. Автоматизоване формування електронних шарів геоінформаційних систем на основі структурованої і неструктурованої інформації / В. В. Приходнюк // Геоінформаційні технології в територіальному управлінні : зб. наук. праць за матеріалами XIII міжнар. наук.-практ. конф. (Одеса, 17–18 вересня 2015 р.). — Одеса : Одеський регіональний інститут державного управління [та ін.], 2015. — С. 73–76.
6. Приходнюк В. В. Онтологічна ГІС, як засіб впорядкування геопросторової інформації / В. В. Приходнюк, О. Є. Стрижак // Наука і техніка Повітряних Сил Збройних Сил України. — 2017. — Т. 27, № 2. — С. 167–174.
7. Приходнюк В. В. Таксономізація естественно-языковых текстов / В. В. Приходнюк // Information models and analyses. — 2016. — Т. 5, № 3. — С. 270–285.
8. Стрижак А. Е. Инвариантные задачи онтологических систем / А. Е. Стрижак // Information technologies & knowledge. — 2014. — № 8. — С. 356–360.

9. Стрижак О. Є. Трансдисциплінарна інтеграція інформаційних ресурсів : дис. ... д-ра техн. наук : 05.13.06 / Стрижак Олександр Євгенійович. — К., 2014. — 470 с.

References.

1. Velichko, V. Yu., & Prikhodnyuk, V. V. (2013). Deyaki sposobi vidilennya vidnoshen' mizh terminami v prirodnomovnomu teksti [Some ways to highlight the relationship between terms in the natural language text]. In *Sistemniy analiz ta informatsiini tekhnologii (System analysis and information technology): a collection of scientific papers on the materials of the XV conference (Kyiv, 27–31 May 2013)* (p. 406). Kyiv: NTUU KPI.
2. Velichko, V. Yu., Sirota, S. V., & Prikhodnyuk, V. V. (2014). Instrumentarii avtomatizovanogo vidilennya vidnoshen' z tekstiv tekhnichnoi tematiki [Instrumentation of automated separation of relations from technical texts]. In *Sistemniy analiz ta informatsiini tekhnologii (System analysis and information technology): a collection of scientific papers on the materials of the XVI conference (Kyiv, 26–30 May 2014)* (pp. 346–347). Kyiv: NTUU KPI.
3. Velichko, V. Yu., & Prikhodnyuk, V. V. (2014). Sposib avtomatizovanogo vidilennya vidnoshen' mizh terminami z prirodnomovnikh tekstiv tekhnichnoi tematiki [The method of automated allocation of relations between terms from natural language texts of technical subjects]. In *Knowledge — Dialogue — Solution: a collection of scientific papers on the materials of the XX international conference* (pp. 27–28). Kyiv: ITHEA.
4. Mintser, O. P., Babintseva, L. Yu., & Shevtsova, O. M. (2017). Priinyattya rishen' v transdistsiplinarnikh zadachakh [Decision making in transdisciplinary tasks]. *Sistemi pidtrimki priinyattya rishen'. Teoriya i praktika (Decision support systems. Theory and practice)*, 99–100.
5. Prikhodnyuk, V. V. (2015). Avtomatizovane formuvannya elektronnikh shariv geoinformatsiinih sistem na osnovi strukturovanoi i nestrukurovanoi informatsii [Automated formation of electronic layers of geoinformation systems on the basis of structured and unstructured information]. In *Geoinformatsiini tekhnologii v teritorial'nomu upravlinni (Geoinformation technologies in territorial management): a collection of scientific papers on the materials of the XIII international scientific and practical conference (Odesa, 17–18 September 2015)* (pp. 73–76). Odesa: Odessa Regional Institute of Public Administration [et al.].
6. Prikhodnyuk, V. V., & Strizhak, O. Ye. (2017). Ontologichna GIS, yak zasib vporiadkuvannya geoprostorovoi informatsii [Ontological GIS as a means of ordering of geospatial information]. *Nauka i tekhnika Povitryanikh Sil Zbroinikh Sil Ukraini (Science and technology of the Air Force of the Armed Forces of Ukraine)*, 27(2), 167–174.
7. Prikhodnyuk, V. V. (2016). Taksonomizatsiya estestvenno-yazykovykh tekstov [Taxonomy of natural language texts]. *Information models and analyses*, 5, 270–285.
8. Strizhak, A. E. (2014). Invariantnye zadachi ontologicheskikh sistem [Invariant tasks of ontological systems]. *Information technologies & knowledge*, 8, 356–360.
9. Strizhak, O. Ye. (2014). *Transdistsiplinarna integratsiya informatsiinih resursiv [Transdisciplinary integration of information resources]*. (Doctoral dissertation).