

УДК 004.5:004.6:004.89:007.51:528.933

О.Є. СТРИЖАК, В.В. ПРИХОДНЮК, С.І. ГАЙКО, В.Б. ШАПОВАЛОВ

ВІДОБРАЖЕННЯ МЕРЕЖЕВОЇ ІНФОРМАЦІЇ У ВИГЛЯДІ ІНТЕРАКТИВНИХ ДОКУМЕНТІВ. ТРАНСДИСЦИПЛІНАРНИЙ ПІДХІД

***Анотація.** Розглядається технологічна проблема інтерактивної взаємодії з великими обсягами мережових інформаційних ресурсів (проблема Big Data). Визначаються інструменти структуризації мережових документів, формування таксономічних відображень структури мережових документів та їх контекстної зв'язності між усіма концептами-поняттями, які формують термінополе користувача, на засадах рекурсивної редукції. Представлено онтологічну модель мережевого документу.*

***Ключові слова:** термінополе, трансдисциплінарність, онтологія, рекурсія, редукція, предикат, контекст, таксономія, інтерактивний документ.*

Вступ

Вирішення складних прикладних задач, що мають певну практичну значимість, багато в чому залежить від об'єктивності та достовірності інформації, яка використовується протягом всього процесу їх розв'язання. При цьому необхідно враховувати, що об'єкти (концепти), властивості яких визначають умови та етапи розв'язання більшості прикладних задач, можуть належати до різних за тематикою предметних областей (ПрО). Практично завжди у фахівця, який вирішує задачу, виникає необхідність в інтерактивному відображенні контекстів мережевої інтегрованої інформації, яка використовується, і даних, які її характеризують, на основі тематичних властивостей інформаційних одиниць, що визначають обрану стратегію її використання у процесі розв'язування прикладної задачі.

Розв'язання цих проблем лежить у напрямках, пов'язаних зі створенням та використанням різноманітних засобів обробки інформації, як пасивної системи мережових знань, які здатні обробляти розподілені політематичні великі масиви даних і тим самим надавати певну допомогу фахівцю у виборі та прийнятті конкретного рішення за заданою проблематикою.

Виділення набору дій на основі системи знань, описаної і представленої в інтегрованому мережевому інформаційному ресурсі, можливо на основі застосування до її природно-мовного тексту процедури структуризації. Для цього реалізуються процедури перетворення книжкового 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.

Трансдисциплінарність інформаційного середовища представима через прояв рекурсивних та рефлексивних властивостей множин таксономічних і операціональних особливостей онтологій предметних областей. Рекурсивність, як функціональна властивість таксономії, дозволяє визначити множинну часткову упорядкованість множин таксономічних і операціональних властивостей онтологічних моделей предметних областей, що відображають всі процеси і об'єкти документів, які представлено у мережевому середовищі [1–3].

Застосування таких трансдисциплінарних процедур орієнтоване на вирішення наступних задач:

- забезпечення можливості оперативної організації доступу до інформаційних джерел формування знань, що стосуються політематичних описів предметних областей або об'єднаних схожими інтересами сфер діяльності;

- підтримка взаємодії всіх учасників процесу в рамках упорядкованих множин предметних областей з можливістю їх розширення;

- забезпечення можливості розширення списку джерел і споживачів різнорідних політематичних інформаційних джерел формування знань в рамках певної предметної області або сфери інтересів;

- обмеження доступу до інформаційних ресурсів рамками конкретної предметної області або сфери інтересів у зв'язку з можливістю вирішення попередньої задачі;

- забезпечення можливості для кожного суб'єкта використання інформаційних ресурсів кількох предметних областей;

- забезпечення можливості оперативного пошуку джерела необхідних інформаційних ресурсів, що стосується контекстів предметних областей.

Таким чином, 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.

1. Основна частина

Для забезпечення трансдисциплінарності представлення усієї сукупності документів використовуються онтологічні моделі. При цьому усю цю сукупність документів слід розглядати на основі типізацій взаємодії мережевих інформаційних систем. І в цій взаємодії можна виділити певні онтологічні аспекти, якщо її розглядати у вигляді класу натуральних систем – SN , які первинно задаються упорядкованими множинами пар виду: $\{<дія \rightarrow \text{результати}>\}$ і які можуть бути представлені за умови існування непорожньої множини можливих наборів дій – F , де F розглядається в якості кінцевої

множини функцій інтерпретації, яка задана на певній предметній області (ПрО) [4]. Кожну ПрО безпосередньо складають певні концепти, які складають множину $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$, та множину їх властивостей R , яка утворюється множиною декартових добутків множини X самої на себе – $R = \prod_i^n X_i$. Множина дій F може бути утворена декартовим добутком множин X і R – $F = X \times R$.

Твердження 1: завжди існує певний набір дій $F_k \subset F$ таких, що завжди існує хоча б одне непусте $f^i \in F_k$ таке, що існує також набір концептів X_j , для яких $f^i(x_1, x_2, \dots, x_n) \in F_k$. У зв'язку з цим, для елементів множини концептів X завжди знайдеться відповідний непустий набір дій з множини F . Таким чином, для кожної множини властивостей R у натуральній системі можливе визначення нового концепту x_{n+1} для множини X такого, що існує додаткова властивість r' , яка забезпечує виконання правила $f^i(x_1, \dots, x_n, x_{n+1}) \in F_k$. Такий набір дій визначається як узгоджений.

Конструктивність твердження 1 дозволяє визначити відношення натуральної системи SN з онтологічною системою O , яка відображає різні онтології O_n , починаючи від простого словника і таксономії до формальної структури концептуальної бази знань для високоінтелектуальних знання-орієнтованих систем, в основі яких знаходиться аксіоматизована теорія, і яка має вигляд впорядкованої четвірки виду

$$O = \langle X, R, F, A, (D, R_s) \rangle, \quad (1)$$

де X – множина концептів;

$X = \{X_1, X_2, \dots, X_i, \dots, X_n\}, i = \overline{1, n}, n = Card(X)$ – кінцева множина концептів;

$R = \{R_1, R_2, \dots, R_k, \dots, R_m\}, R \subseteq X_1 \times X_2 \times \dots \times X_n, k = \overline{1, m}, m = Card(R)$ – кінцева множина семантично значущих відношень між концептами ПрО, які визначають тип взаємодії між поняттями;

$F: X \times R$ – кінцева множина функцій інтерпретації, заданих на концептах та/або відношеннях. Кожна X_i в загальному випадку включає підмножину понять $\{x_{i-1}\}$, через які відношення, що зв'язує X_i з $\{x_{i-1}\}$, і множину атрибутів (ознак), властивих X_i ;

A – скінченна множина аксіом, які використовуються для запису завжди істинних висловлювань в термінах тематики ПрО;

D – множина додаткових визначень понять в термінах тематики ПрО;

R_s – множина обмежень, що визначають область дії понятійних структур визначеної тематики ПрО.

Онтологія як складова інформаційної системи (ІС) є формальним представленням концептуальних знань про предметну область. Процес

побудови такої складової ІС можна представити композицією певних висловлювань, суджень, тверджень, термінів-понять і відношеннями між ними, а його результат – основою для побудови складової частини аксіоматизованої теорії – онтологічної бази знань у заданій предметній області, описаній в декларативній формі.

Твердження 2. Онтології, які можуть бути утворені на основі концептів натуральної системи, можуть бути представлені наступними шістьма типами: неструктурований текст; глосарій; таксономія; тезаурус; проста онтологія; активна онтологія.

Особливу роль у формуванні та використанні онтологій грає категорія таксономії, як операціональна платформа онтологічної системи, яка може бути визначена для будь-якого складного концепту і утворюється класом класифікаційних відношень, на множині яких задається множинне бінарне відношення упорядкованості типу «група об'єктів – група об'єктів».

Категорію «таксономія» доповнено категорією «термінополе», під якою розуміється множина взаємопов'язаних дефініцій термінів, що визначають імена концептів ПрО.

Твердження 3. Кожне термінополе може бути представлено певною множиною таксономій.

Твердження 1, 2 і 3 характеризують конструктивізм визначень термінополя, натуральної системи і таксономії, що дає можливість встановити відповідність між категоріями натуральна система і онтологія.

Твердження 4. Завжди можна виділити непорожній набір умов, заданих підмножиною відношень \tilde{R} концептів термінополя $\tilde{R} \subset R | R = X \times X$, за умови, що застосовність множини правил-інтерпретацій F_k , що формують операційне середовище натуральної системи SN , буде задано над концептами X термінополя, представленого множиною таксономій \tilde{T} . Тоді натуральна система може бути представлена онтологією виду

$$O = \langle X, \tilde{R}, F_k \rangle. \quad (2)$$

Таку властивість натуральних систем будемо називати пластичним перетворенням.

Твердження 5. Якщо непорожні декартові добутки множин X (концепти) і R (відношення) можуть утворювати певну підмножину функцій $F_t \subset F$, які є певними діями в термінах множини концептів, то на множині дій F можливо задати деяку непорожню множину істинних висловлювань типу {дія \Rightarrow результат}, яка також утворює натуральну систему SN . Тоді, справедливо наступне: будь-яка онтологія O_n , яку утворює натуральна система SN , може бути утворена на основі певної системи висловлювань, утворених множиною дій F над концептами X , і будь-яка натуральна система SN може бути утворена на основі певної системи висловлювань з концептів онтології O_n . Таку взаємодію між натуральними системами і онтологіями будемо називати двоїстою.

Твердження 6. Якщо певна онтологія O_n визначена у вигляді натуральної системи SN , то завжди є певна не порожня множина істинних висловлювань, яка утворюється концептами цієї онтології за умови, що ці концепти упорядковані між собою бінарними відношеннями виду:

$$r^m(x_i^j, x_l^k | x_i^j \in X_i; x_l^k \in X_l; r^m \in R \neq \emptyset), \quad (3)$$

які мають властивості: ациклічності – γ ; часткової впорядкованості – \tilde{p} ; лінійної упорядкованості – p :

$$x_i^j \gamma x_l^k \xrightarrow{\alpha} x_i^j p x_l^k \quad (4)$$

$$x_i^j \tilde{p} x_l^k \xrightarrow{\alpha} x_i^j p x_l^k . \quad (5)$$

Для множини таксономій \tilde{T} завжди можна знайти не порожню множину відображень \tilde{G} , що здійснює опис натуральної системи SN в онтологію O_n . Також завжди при заданих умовах формування множини таксономій \tilde{T}' можна знайти множину зворотних відображень \tilde{G}^{-1} , що переводять опис онтології O_n в опис натуральної системи SN

$$\tilde{G} : SN \Rightarrow O_n \quad (6)$$

$$\tilde{G}^{-1} : O_n \Rightarrow SN . \quad (7)$$

На основі універсальності певних властивостей онтологій можна зробити наступний висновок: всі семантичні утворення сформованого типу онтологічної моделі можна представити у вигляді множини істинних висловлювань та/або тверджень, що зв'язують концепти онтології. При цьому всі ці твердження можуть мати тривіальний вид, тобто бути представленими тільки двома пов'язаними концептами.

Концепти таксономії T можуть бути представлені у вигляді наборів послідовностей різної довжини, що становлять послідовність 2^X . Тоді в множині функціоналів F можна завжди визначити як мінімум одну впорядковану функцію F_p на послідовності 2^X . Функція F_p має властивість адитивності, порядкоадитивності і монотонності. Справедливим буде твердження, що серед концептів множини завжди можна знайти концепт $x \in X$, такий що:

$$F_p(X) = \max F_p(T). \quad (8)$$

Визначається не порожня, кінцева множина впливів Q , така, що її елементи $q_1, q_2, q_3, \dots, q_n$ визначають, при використанні концептів таксономії T , множину наслідків ψ , яка дозволяє визначити множину причинно-наслідкових відношень $-A$, між множинами X , R и F , і які представлені відображенням виду:

$$Q \times X \Rightarrow \psi \times T \Rightarrow A. \quad (9)$$

На основі множинних відношень між концептами таксономій визначається область застосування інваріантних дій при формуванні станів взаємодії онтологічної системи. Особливу роль тут грає дія вибору. Цей онтологічний інваріант визначається функцією вибору $-F_{sel}(X)$, що задана над концептами, між якими визначено множинне бінарне відношення часткової впорядкованості $R_p \subseteq R$:

$$Q \times R_p = F_{sel}(X) \Rightarrow \psi \times T. \quad (10)$$

Функція вибору виду (10) детермінована і забезпечує формування таксономій, на основі використання відношення множинної впорядкованості і причинно-наслідкових зв'язків $-A$. Сам вибір може бути представлений наступними методами: скалярно-оптимізаційний, векторно-оптимізаційний, графо-домінантний.

Відношення між таксономічними категоріями мають властивості гіпервідношення Gr виду $-YGrx$, де Y – множина всіх можливих множин концептів X таксономічної категорії \tilde{T} , а x – один з концептів цієї множини.

Відношення упорядкованості на множині всіх концептів і їх можливих множин в структурі таксономічних категорій \tilde{T} онтологічної системи O_n може бути представлено уже описаними типами бінарних відношень типу (4)–(5), які можуть бути задані над множиною всіх концептів X . Така бінарна гіпервпорядкованість може бути визначена у вигляді множин бінарних упорядкованостей P і представлена у вигляді наступного виразу – $P = \{p_n\} n \in N$:

$$xy \Leftrightarrow \forall n \in N : xp_n y. \quad (11)$$

Множинне гіпервідношення, Gr , має наступні властивості, що характеризують його як бінарне:

– агіперциклічність – якщо для Gr не існує гіперциклічної множини $X \subseteq U$ такої, коли:

$$\forall x \in X \exists Y \subseteq X : YGrx; \quad (12)$$

– іррефлексивність:

$$YGrx \Rightarrow (Y / \{x\})Grx; \quad (13)$$

– гіпертранзитивність:

$$YGrx, x \in X, XGrz \Rightarrow ((Y \cup X) / \{x\})Grz; \quad (14)$$

– регулярність:

$$YGrx, Y' \supseteq Y \Rightarrow Y'Grx. \quad (15)$$

Властивості гіпервідношення Gr та множина причинно-наслідкових відношень \mathcal{A} – (9)–(10) забезпечують формування процедури виділення з множини концептів X не порожньої підмножини концептів $\{x\}$, що має властивість бінарного відношення часткової впорядкованості \tilde{p} , яке і пов'язує їх між собою.

$$YGrx \Rightarrow \exists y \in Y : y\tilde{p}x. \quad (16)$$

При цьому часткова впорядкованість є елементом множини бінарних впорядкованостей – P , яку може бути розширено до гіпервідношення Gr і тим самим визначено множинну упорядкованість над підмножинами концептів $\{x\}$ у вигляді множинного відношення «група об'єктів – група об'єктів», що має властивості (11)–(16):

$$YGrx \Leftrightarrow n \in N \exists y \in Y : y p_n x : p_n \in P \subseteq Gr. \quad (17)$$

Предикативні вирази, що формулюються на основі концептів таксономічної категорії із заданим множинним відношенням упорядкованості виду (6)–(7) і (9)–(11), приймають лише значення істинності. Це дозволяє формувати, на основі термінів-концептів онтології, лінгвістичні вирази, що осмислено відображають стани кожної онтологічної системи O_n .

Гіпервідношення Gr , задане над множинними бінарними відношеннями упорядкованості, що визначають таксономію, дозволяє представити предикативні вирази, що формулюються на основі її концептів із заданим множинним відношенням упорядкованості виду (11)–(17) у вигляді рекурсивного предиката:

$$Pr(x_1, \dots, x_n) = \begin{cases} 1, \neg Pr(YGrx_i) \wedge Pr(x_1, \dots, x_n) \\ 0, Pr(YGrx_i) \end{cases}, \quad (18)$$

де $x_i \in X; 1 \leq i \leq n$, наявність семантичної властивості задається умовою:

$$\Pr(x_1, \dots, x_n) = 0 \Rightarrow \exists T \subseteq \tilde{T} : \forall x \in X \exists Y \subseteq X : T = YGrx, \quad (19)$$

а існування визначається умовами (8)–(11).

Врахування властивостей множини причинно-наслідкових відношень – A (9)–(10), спільно з властивістю індуктивності рекурсивного предикату виду (18) і умов існування таксономії (3)–(5) і (9)–(10) дозволяє визначити умови стійкості онтологічної системи O_n . Ці умови визначаються типом відношень бінарної упорядкованості (11)–(17) і забезпечують незалежність значень характеристичної функції рекурсивного предиката (18) від послідовності контекстів індуктивного вибору. В складних системах, якими є онтології, виділяють наступні умови стійкості:

Умова спадкування:

$$\exists T' \subseteq T \Rightarrow \exists f(T') \supseteq f(T) \cap T', \quad (20)$$

Умова незалежності:

$$\exists f(T') \subseteq T' \subseteq T \Rightarrow f(T') = f(T)', \quad (21)$$

Умова згоди:

$$\bigcap_n f(T_n) \subseteq f(\bigcup_n T_n). \quad (22)$$

Таким чином, при використанні певних концептів онтології можливо виконувати підстановку множин висловлювань, які задаються на множині концептів онтології на підставі їх тематичних властивостей і бінарних відношень упорядкованості виду (16)–(17) і (19)–(22). При цьому всі зазначені концепти мають бути представлені у вигляді предикативного виразу (18).

Твердження 7. Якщо певна онтологія O_n формується на основі таксономії \tilde{T} , то завжди є певна не порожня множина істинних висловлювань $\{Pr(x_1, \dots, x_n)\}$, яка утворюється концептами x цієї онтології, за умови, що ці концепти пов'язані між собою бінарними відношеннями виду (3)–(5).

Побудова висловлювань з концептів реалізується на підставі кінцевих наборів правил з множин Rul , які визначають порядок застосування множинного відношення часткової впорядкованості \tilde{p} як над концептами з множини X , так і над семантичними відношеннями з множини R_{sem} . Також відношення \tilde{p} дозволяє нам формувати з концептів термінополя певні таксономії T , над концептами якої задається множинне бінарне відношення часткової впорядкованості, яке, по суті, повністю еквівалентне відношенню \tilde{p} .

Трансдисциплінарне представлення множин онтологій базується на перетворенні між двома гіпермножинами:

$$f^{ct} : R \rightarrow F, \quad (23)$$

де R, F – множини зв'язків і функцій інтерпретації певної онтології O .
Для множини онтологій можна побудувати гіпермножини:

$$\mathfrak{R} = \bigcup_i R_i, F = \bigcup_i F_i, \quad (24)$$

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:

$$f^{tt} : F \rightarrow \mathfrak{R}. \quad (25)$$

Тобто певний зв'язок між об'єктами різних за тематикою онтологій може бути представлений не пустою множиною інтерпретуючих функцій з даних онтологій.

Дане перетворення можна розширити і на множину неструктурованих текстів, застосувавши до кожного з текстів оператор редуції (26), який являє собою комбінацію чотирьох операторів, при цьому три оператори виконують кроки перетворення, а один виконує допоміжну функцію.

$$F_{rd} = F_{l*} \circ F_x \circ F_{smr} \circ F_{ct}, \quad (26)$$

де F_{l*} – оператор агрегації, що виконує допоміжну функцію, перетворюючи множину лексем L в множину конструктів L^* . Конструкти є особливою формою лексем і об'єднують в собі послідовності слів або символів, зокрема, словосполучення. Особливістю конструктів є те, що з точки зору подальшої обробки вони можуть розглядатись як лексеми (тобто як одне слово чи символ). Таким чином, множина $L \cup L^*$ може використовуватись у тих же випадках.

Розглянемо процедуру трансдисциплінарного представлення певної множини онтологій засобами інтерактивних документів. Дане представлення базується на функції (27).

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

$$Q_c(x) = \bigcup_{l \in L_x} Q_S(Q_l(C), l), \quad (27)$$

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 x ;

l – певна лексема.

Основним недоліком функції (27) є те, що в її результаті формується неупорядкована множина документів, яка може бути достатньо великою за розміром і тому незручною для обробки експертом. Вирішити цю проблему можна двома способами:

1) Задати на множині результатів відношення порядку, в якості якого може виступити відношення релевантності R_{rel} :

$$T_1 R_{rel} T_2 \Rightarrow card(L_{T_1} \cap L_x) > card(L_{T_2} \cap L_x), \quad (28)$$

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

L_x – контекст об'єкта x , з яким була здійснена контекстна зв'язка.

2) Виключити з результату документи, що мають недостатню релевантність, за допомогою застосування у функції (27) операції перетину замість операції об'єднання.

Для його здійснення необхідно виконати трансдисциплінарне перетворення за допомогою функції:

$$Q_{\Pi}(C) = \bigcup_{x \in X_C} \{Q_C(x)\}, \quad (29)$$

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 множину 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 $C = \{< X_i, R_i, F_i >\}$:

$$C \xrightarrow{Q_{\Pi}} < \bigcup_i X_i, \bigcup_i R_i \cup Q_{\Pi}(C), \bigcup_i F_i >. \quad (30)$$

Перетворення (30) задає найбільш повне представлення наявної в C інформації, що не завжди зручно. Часто необхідно виконати представлення однієї вибраної онтології O . В такому випадку (29) необхідно змінити:

$$Q_{TO}(O, C) = \bigcup_{x \in X_O} \{Q_C(x)\}, \quad (31)$$

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 8. На основі однієї онтології O , що належить множині документів C , можна сформуувати довільну кількість НС.

Твердження 9. Комбінація незалежних натуральних систем SN_i , що приймають на вхід один і той же набір «дій», є натуральною системою.

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

У 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:

$$g = \langle f_{ap}^g, f_{tr}^g \rangle, \quad (32)$$

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

f_{tr}^g – функція перетворення, що задає перетворення вхідної інформації.

Задане правилом g перетворення $F_g : X \rightarrow Y$ має вигляд (33).

$$F_g(x) = \begin{cases} f_{tr}^g(x), f_{ap}^g(x) \\ x, \neg f_{ap}^g(x). \end{cases} \quad (33)$$

Кожна функція застосовності являє собою лямбда-терм виду [5–7]:

$$f_{ap} = (\lambda x_1, x_2 \dots x_{n_g}. t_{ap}(x)) a_1, a_2 \dots a_{n_g} = t_{ap}(a_1, a_2 \dots a_{n_g}), \quad (34)$$

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} 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. В якості T_{sn}^e може виступати первинне представлення як поточного тексту T_{sn} , так і будь-якого іншого тексту (наприклад, тезауруса ПдО). Умова має структуру (35) і складається з *предикатів ідентифікації* [8–10]. 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, і умовою застосовності правила є виконання всіх умов, заданих кожним з предикатів. Кількість предикатів у виразі задає число n_g .

$$t_{ap} = c_{p_1}(x_1) \& \dots c_{p_n}(x_{n_g}) \& r_{k_{11}}(x_1, x_1) \& \dots r_{k_{n_g n_g}}(x_{n_g}, x_{n_g}). \quad (35)$$

Одномісні предикати, присутні у виразі – це предикати ідентифікації лексем. Такий предикат задає умову, якій повинна відповідати певна лексема (або конструкт) з вхідної множини. Предикат має структуру (36).

$$c_p(l) = \begin{cases} 1, p = 0 \vee p = l^T \vee p \in P_l \\ 0, p \neq 0 \wedge p \neq l^T \wedge p \notin P_l \end{cases} \quad (36)$$

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

1) Стандартним предикатом ідентифікації. У 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. У 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) Нульовий предикат ($p = 0$). 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 (37).

$$r_k(l_1, l_2) = \begin{cases} 1, k = 0 \vee \langle l_1, l_2, k \rangle \in R_{sn} \\ 0, k \neq 0 \wedge \langle l_1, l_2, k \rangle \notin R_{sn} \end{cases} \quad (37)$$

Як і предикат ідентифікації лексем, даний предикат має нульову модифікацію, яка має значення незалежно від вхідних даних. При цьому для коректного використання умови (35) повинна виконуватись задана структурою лексичного аналізатора умова:

$$k_{ij} = 0, i = j \quad (38)$$

База правил задає перетворення F , що має вигляд:

$$F_G(L) = \bigcup_{\tilde{L} \in P(L)} \bigcup_{g \in G} F_g^*(\tilde{L}), \quad (39)$$

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

F_g^* – модифікована функція (33), доповнена додатковими умовами.

Додаткових умов накладається дві:

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

$$f_{ord}(x) = \begin{cases} 1, \forall x_1, x_2 \in x, x_1 \prec x_2 \vee x_2 \prec x_1 \\ 0, \exists x_1, x_2 \in x, x_1 \notin x_2 \wedge x_2 \notin x_1. \end{cases} \quad (40)$$

Відношення порядку, що можуть використовуватись в якості G :

- 1) Відношення слідування \prec ;
- 2) Транзитивне замикання відношення R ;
- 3) Транзитивне замикання відношення, заданого зв'язками R_{sem} .

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

$$f_{ap+}^g(x) = \begin{cases} 1, card(x) = n_g \\ 0, card(x) \neq n_g. \end{cases} \quad (41)$$

З урахуванням даних умов перетворення (33) перетворюється на F_g^* :

$$F_g^*(x) = \begin{cases} f_{tr}^g(x) & , f_{ord}(x) \wedge f_{ap+}^g(x) \wedge f_{ap}^g(x) \\ x & , \neg f_{ord}(x) \vee \neg f_{ap+}^g(x) \vee \neg f_{ap}^g(x). \end{cases} \quad (42)$$

Вказані представлення перетворення текстів є зручнішими для користувача, оскільки їх легше представляти в текстовій формі. Вирази (23)–(42) 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.

Іншим важливим наслідком існування гомоморфізму між первинною структурою тексту і структурою інформації, що з нього виділяється, є можливість побудови спеціалізованої НС NS' , призначеної для роботи не з об'єктами онтології, а лексемами первинної структури [10]. Побудований на її основі інтерактивний документ $\langle T_{sn}, NS' \rangle$ може використовуватися для формування довільної мережевої інформаційної системи.

Висновок

Таким чином, трансдисциплінарна структуризація мережевої інформації створює технологічні умови щодо реалізації контекстної зв'язності між усіма концептами-поняттями, які формують термінополе користувача та визначають кожен мережевий інформаційний документ. І вже вузли онтографу, що представляють цю зв'язність, відображають усі стани інтерактивної взаємодії з множиною мережевих документів, які використовуються у процесі розв'язання конкретної задачі.

СПИСОК ЛІТЕРАТУРИ

1. Стрижак О. Є. Трансдисциплінарна інтеграція інформаційних ресурсів: дис. д-ра техн. наук. К., 2014. 470 с.
2. Стрижак А. Е. Инвариантные задачи онтологических систем. *INFORMATION TECHNOLOGIES & KNOWLEDGE*, 2014. №. 8. С. 356–360.
3. Трансдисциплінарне представлення інформації за допомогою інтерактивних документів / О.П. Мінцер, В.В. Приходнюк, О.Є. Стрижак О.М. Шевцова // Медична інформатика та інженерія, Випуск 1 (41), 2018 – С. 47–52.
4. Малишевский А. В. Качественные модели в теории сложных систем. – М.: Наука. Физматлит. 1998. – 528 с.
5. Величко В. Ю., Приходнюк В. В. Деякі способи виділення відношень між термінами в природномовному тексті. *Системний аналіз та інформаційні технології*: зб. наук. праць за матеріалами XV конференції (м. Київ, 27–31 травня 2013 р.). К. : НТУУ «КПІ», 2013. С. 406.
6. Величко В. Ю., Сирота С. В., Приходнюк В. В. Інструментарій автоматизованого виділення відношень з текстів технічної тематики. *Системний аналіз та інформаційні технології*: зб. наук. праць за матеріалами XVI конференції (м. Київ, 26–30 травня 2014 р.). К. : НТУУ «КПІ», 2014.
7. Приходнюк В. В. Автоматизоване формування електронних шарів геоінформаційних систем на основі структурованої і неструктурованої інформації. *Геоінформаційні технології в територіальному управлінні*: зб. наук. праць за матеріалами XIII міжнар. наук.-практ. конф. (м. Одеса, 17–18 вересня 2015 р.). Одеса : Одеський регіональний інститут державного управління [та ін.], 2015. С. 73–76.
8. Величко В. Ю., Приходнюк В. В. Спосіб автоматизованого виділення відношень між термінами з природномовних текстів технічної тематики. *Knowledge – Dialogue – Solution* : Збірник праць XX міжнародної конференції. К. : ІТНЕА, 2014. С. 27–28.
9. Приходнюк В. Таксономизация естественно-языковых текстов. *Information Models and Analyses*, 2016. №. 5. С. 270–284.
10. Приходнюк В. В., Стрижак О. Є. Онтологічна ГІС, як засіб впорядкування геопросторової інформації. *Наука і техніка Повітряних Сил Збройних Сил України*, 2017. №. 2(27). С. 167–174.

Стаття надійшла до редакції 22.08.18.