

## ІНТЕЛЕКТУАЛЬНІ АГЕНТИ ПОШУКУ РЕЛЕВАНТНИХ ПРЕЦЕДЕНТІВ НА ОСНОВІ АДАПТИВНИХ ОНТОЛОГІЙ

**Анотація.** У статті розглянуто функціонування інтелектуальних агентів на основі адаптивних онтологій, які використовують прецеденти. Розроблено математичне забезпечення діяльності таких агентів.

**Ключові слова:** інтелектуальний агент, онтологія, прецедент.

**Аннотация.** В статье рассмотрено функционирование интеллектуальных агентов на основе адаптивных онтологий, использующих прецеденты. Разработано математическое обеспечение деятельности таких агентов.

**Ключевые слова:** интеллектуальный агент, онтология, прецедент.

**Abstract.** The paper reviews the functioning of intelligent agents based on adaptive ontologies using precedents. A mathematical support of such agents is developed.

**Keywords:** intelligent agents, onology, precedent.

### 1. Вступ

Інтелектуальні агенти (ІА), засновані на прецедентах (Case-Based Reasoning, або CBR), широко використовуються для розв'язування слабоформалізованих задач. Виведення, засноване на прецедентах, є методом побудови ІА, які приймають рішення щодо поточної проблеми за наслідками пошуку аналогій, що зберігаються в базі прецедентів [1]. Такий аналог називають релевантним прецедентом. З математичної точки зору це означає, що серед елементів множини прецедентів  $Pr = \{Pr_1, Pr_2, \dots, Pr_N\}$  релевантним  $Pr_k$  є прецедент, для якого відстань  $d$  до поточної ситуації  $S$  є найменшою, тобто

$$Pr_k = \arg \min_i d(Pr_i, S).$$

Задачу пошуку релевантних прецедентів можна вважати задачею класифікації, де класами є прецеденти. Тоді задача полягає у віднесенні поточної ситуації до певного класу.

Тобто, вводиться метрика на просторі ознак. У цьому просторі визначається точка, відповідна поточному випадку, і в рамках цієї метрики знаходиться найближча до неї точка серед точок, що відображають прецеденти. Кожній ознаці призначають вагу, що враховує її відносну цінність. Повністю ступінь близькості прецеденту за всіма ознаками можна обчислити, використовуючи узагальнену формулу вигляду

$$\sum_k w_k \cdot \text{sim}(x_{ki}, x_{kj}), \sum_k w_k = 1,$$

де  $w_k$  – вага  $k$ -ої ознаки,  $\text{sim}$  – функція подібності (метрика),  $x_{ki}$  і  $x_{kj}$  – значення ознаки  $x_k$  для поточного випадку  $i$  та  $j$ -го прецеденту, відповідно. Після обчислення ступенів близькості всі прецеденти ранжуються. Поточна ситуація відноситься до прецеденту із найвищим рангом.

### 2. Аналіз літературних джерел

Вибір метрики (або міри близькості) вважається центральним моментом, від якого зале-

жить пошук релевантних прецедентів. У кожній конкретній задачі цей вибір відбувається по-своєму, з урахуванням головних цілей дослідження, фізичної і статистичної природи інформації, що використовується, тощо. Як методи розв'язання таких задач використовуються алгоритми типу Lazy-Learning [2], зокрема, відомі алгоритми найближчого сусіда і  $k$ -ближніх сусідів, нейронні мережі, генетичні алгоритми, байєсівські мережі [3], дерева рішень [4].

Основним недоліком нейромережевої парадигми є необхідність мати дуже великий об'єм навчальної вибірки. Інший істотний недолік полягає в тому, що ваги декількох со-тень міжнейронних зв'язків абсолютно не піддаються аналізу й інтерпретації людиною.

Популярність дерев рішень пов'язана з наочністю і зрозумілістю. Але для них дуже гостро стоїть проблема значущості. Річ у тому, що окремим вузлам на кожному новому побудованому рівні дерева відповідає все менша і менша кількість записів даних – дерево дробить дані на велику кількість окремих випадків, тому воно не даватиме статистично обґрунтованих відповідей. Як показує практика, в більшості систем, що використовують дерева рішень, ця проблема не знаходить задовільного рішення. Крім того, загальновідомо, і це легко показати, що дерева рішень дають корисні результати тільки у разі незалежних ознак. Інакше вони лише створюють ілюзію логічного виведення.

Генетичні алгоритми теж мають ряд недоліків. Критерій відбору хромосом і використовувані процедури є евристичними і не гарантують знаходження «кращого» рішення. Крім того, ефективно сформулювати завдання, визначити критерій відбору хромосом під силу тільки фахівцеві. Через ці чинники сьогодні генетичні алгоритми треба розглядати швидше як інструмент наукового дослідження, ніж як засіб аналізу даних для практичного застосування.

На нашу думку, позбутися ряду вище наведених недоліків дозволяють онтології предметної області та онтології задач.

## **Постановка задачі**

Розробити підхід для оцінювання релевантності прецедентів на основі онтологій. Побудувати метрику для здійснення такої оцінки. Апробувати цю метрику під час функціонування ІА у прикладній предметній області (ПО).

## **3. Основний матеріал**

### **3.1. Розроблення метрики пошуку релевантних прецедентів на основі адаптивних онтологій**

Формально онтологія складається з термінів (понять, концептів), організованих в таксономію, їх визначень і атрибутів, а також пов'язаних з ними аксіом і правил виведення. Тому під моделлю онтології  $O$  розуміють трійку вигляду

$$O = \langle C, R, F \rangle,$$

де  $C$  – поняття,  $R$  – відношення між поняттями,  $F$  – інтерпретація понять та відношень (аксіоми). Аксіоми встановлюють семантичні обмеження для системи понять та відношень.

Для того, щоб можна було на онтологіях будувати метрику, нами запропоновано розширити цю модель за рахунок введення двох скалярних величин – ваги важливості понять та зв'язків [5, 6]. Коефіцієнт важливості поняття (зв'язку) – це чисельна міра, яка характеризує значущість певного поняття (зв'язку) у конкретній предметній області і динамічно змінюється за певними правилами у процесі експлуатації системи. Отже:

$$O = \langle C, R, F, W, L \rangle,$$

де  $W$  – важливість понять  $C$ ,  $L$  – важливість відношень  $R$ .

Визначену таким чином онтологію будемо називати адаптивною, тобто такою, що адаптується до ПО за рахунок модифікації понять та коефіцієнтів важливості цих понять і зв'язків між ними. Очевидно, що в залежності від прецеденту ваги понять різні. Тобто, насправді  $W$  – вектор вимірності кількості прецедентів  $W = (W_1, W_2, \dots, W_N)$ . Надалі будемо розглядати лише один прецедент, тобто, нижній індекс у вазі понять опускатимемо.

Побудуємо метрику для пошуку релевантних прецедентів на основі адаптивних онтологій. Нехай множина прецедентів  $Pr = \{Pr_1, Pr_2, \dots, Pr_N\}$  описується ознаками  $X = \{x_1, x_2, \dots, x_M\}$ .  $D_i$  – домен ознаки  $x_i$ ,  $w_i$  – коефіцієнт важливості ознаки  $x_i$  прецеденту  $Pr_i$ . Значення властивості  $x_i$  позначатимемо  $z_i = z(x_i)$ . Отже  $Pr_i \leftrightarrow X_i = \{x_{i_1} = z_{i_1}, x_{i_2} = z_{i_2}, \dots, x_{i_k} = z_{i_k}\}$ , де  $z_{i_j} \in D_{i_j}$ .

Позначимо  $I_i$  – множину індексів властивостей прецеденту  $Pr_i$ . Тоді відстань між прецедентом  $Pr_i$  та поточною ситуацією  $S$  визначається як

$$d_i = \sum_{i_j \in I_i} \varphi(z_{i_j}, z_{i_j}^S), \quad (1)$$

де  $z_{i_j}$  – значення властивості  $x_{i_j}$  прецеденту  $Pr_i$ ,  $z_{i_j}^S$  значення властивості  $x_{i_j}$  поточної ситуації  $S$ ,  $\bar{I}_i \subset I_i$  – підмножина індексів важливих властивостей прецеденту  $Pr_i$ ,  $\bar{I}_i = \bar{I}_{i1} \cup \bar{I}_{i2} \cup \dots \cup \bar{I}_{iN_i}$ ,  $N_i$  – кількість властивостей, які необхідно розглянути для прийняття рішення стосовно прецеденту  $Pr_i$ . Тобто,

$$\bar{I}_{i1} = \left\{ i_{s1} \mid i_{s1} = \arg \max_{i_j \in I_i} w_{i_j} \right\}, \bar{I}_{i2} = \left\{ i_{s2} \mid i_{s2} = \arg \max_{i_j \in I_i / i_{s1}} w_{i_j} \right\}, \bar{I}_{i3} = \left\{ i_{s3} \mid i_{s3} = \arg \max_{i_j \in I_i / i_{s1} / i_{s2}} w_{i_j} \right\}, \dots$$

Розглянемо функцію  $\varphi(\xi, \eta)$ . Очевидно, що  $\xi$  може бути діапазоном, тобто, нечіткою підмножиною  $\xi \subseteq D$ , де  $D$  – універсальна множина, числовим значенням або нечисловим. У залежності від цього  $\varphi(\xi, \eta)$  визначається по-своєму, а саме:

$$\varphi(\xi, \eta) = \begin{cases} 1 - \mu_\xi(\eta), & \xi - \text{нечітка множина,} \\ \lambda \cdot |\xi - \eta|, & \xi, \eta - \text{числові значення,} \\ 1 - \mu(\xi, \eta), & \xi, \eta - \text{нечислові значення,} \end{cases} \quad (2)$$

де  $\mu_\xi(\eta)$  – коефіцієнт впевненості того, що  $\eta$  належить нечіткій підмножині  $\xi$ ;  $\lambda$  – числова величина, яка залежить від ПО, щоб добуток  $\lambda \cdot |\xi - \eta| \in [0, 1]$  (розмірність величини  $\lambda$  обернено-пропорційна до розмірності величин  $\xi$  та  $\eta$ , тобто, якщо  $\xi$  та  $\eta$  є маса і вимірюється в кг, то  $\lambda$  вимірюється в  $\text{кг}^{-1}$ );  $\mu(\xi, \eta) \in [0, 1]$  – нечітка величина подібності значень  $\xi$  та  $\eta$ .

### 3.2. Розвинення ваг важливості понять на всю онтологію

Для визначення важливості ваг понять та відношень пропонується використовувати методи інтелектуального аналізу даних (ІАД), наприклад, дерева рішень. Використовуючи ІАД, ми визначимо ваги деякої підмножини понять, які називатимемо базовими. Потім, на основі онтології ПО, розвинемо отримані ваги на всю онтологію. Цю процедуру здійсимо для кожного прецеденту. Тоді для пошуку релевантного прецеденту використовуватимемо

значення тих  $N_i$  понять, які для відповідного прецеденту мають найбільші ваги. Що стосується ваг важливості відношень, то нами пропонується задати їх такими, як це наведено у табл. 1. Групи відношень взяті із [7]. Вважаємо, що вага вертикальних зв'язків (ієрархія, агрегація) рівна 1,2 (чим конкретніше, тим краще). Квантитативні відношення не розглядаються, оскільки синонімія та королювання ніяк не впливають на значення ознак. Відразу вважається, що це одна і та ж ознака.

Таблиця 1. Ваги важливості відношень

Група відношень	Відношення	Значення ваг важливості, $L$	Група відношень	Відношення	Значення ваг важливості, $L$
Ієрархія	Рід↔вид	1,2	Функціональні	Об'єкт дії↔дія↔ суб'єкт дії	1
	Ознака↔значення ознаки	1,2		Причина↔наслідок	0,9
	Інваріант↔варіант	1,2		Умова↔дія	0,9
Агрегація	Ціле↔частина	1,2		Явище↔дія	0,9
	Об'єкт↔простір реалізації (локалізації) об'єкта	1,2		Стан↔дія	0,9
	Об'єкт↔властивість/ознака	1,2		Явище↔стан	0,9
	рівень↔одиниця рівня	1,2		Інструмент↔дія	0,9
Семіотичні	Термін↔спосіб вираження	0,2		Дані↔дія	0,9
	Термін↔спосіб подання	0,2		Дані↔величини	0,9
	Термін↔метазнак терміну	0,2			

Загальний підхід, який пропонується для знаходження релевантних прецедентів, наведено на рис. 1. Він складається із трьох кроків.

Розглянемо перші два кроки детальніше. Нехай для розв'язування задачі класифікації побудовано дерево рішень (ДР). Вершини (ознаки) цієї гілки знаходяться на  $k$  рівнях. Очевидно, що, чим вищий рівень, тим значуща ознака, яка на цьому рівні знаходиться. Ця евристична думка має бути відображена в значеннях ваг цих ознак. Крім того, пропонується ці ваги нормувати, тобто, щоб їх сума для кожного прецеденту (гілки) була рівна 1. Для визначення ваг базових ознак, які задовольняють вище описані два припущення, пропонується один із двох способів.

1. Арифметичний спосіб. Визначаються як відношення різниці  $(k + 1)$  рівня дерева та рівня, на якому знаходиться ознака, до суми всіх рівнів гілки, тобто, базуються на сумі арифметичної прогресії:

$$w_i = \frac{k+1-i}{\sum_{j=1}^k j} = \frac{k+1-i}{(1+k)k} \cdot 2$$

2. Геометричний спосіб. Базуються на сумі геометричної прогресії. Ваги понять, які лежать на  $i$ -му рівні ДР, визначаються таким чином:

$$w_i = \frac{2^{k-i}}{2^k - 1} \cdot$$

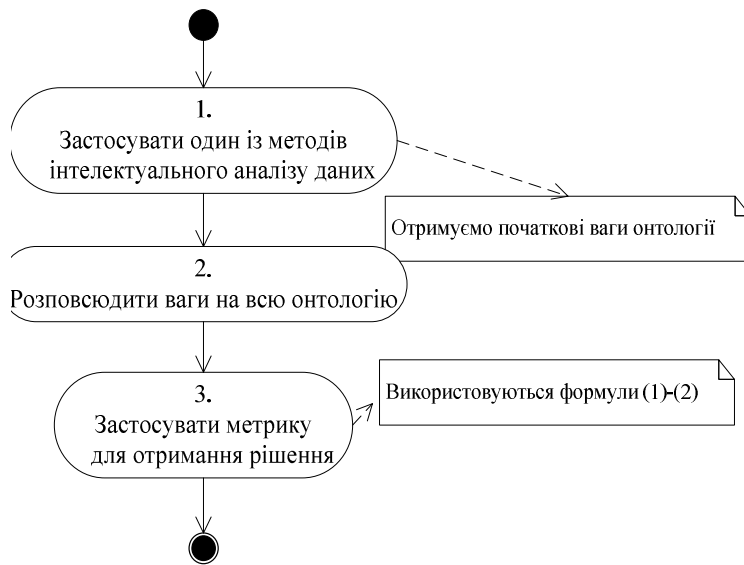


Рис. 1. Кроки функціонування ІА пошуку релевантних прецедентів на основі адаптивних онтологій

симетричних відношень рівні одиниці. Таким симетричним зв'язком є „Об'єкт дії↔дія↔суб'єкт дії”, який належить до групи функціональних зв'язків. Усі інші функціональні зв'язки є близькими до симетричних, тому їх ваги рівні 0,9. Отже, якщо відома вага  $W_i$  терміна  $C_i$ , і цей термін має симетричний зв'язок із терміном  $C_j$ , вага якого невідома, то  $W_j = L \cdot W_i$ . Для несиметричних зв'язків отримуємо аналогічне співвідношення:

$$W_j = L \cdot W_i, \quad (3)$$

якщо відомо вагу терміна  $C_i$ , і

$$W_i = \frac{W_j}{L}, \quad (4)$$

якщо відомо вагу терміна  $C_j$ .

### 3.5. Приклад функціонування інтелектуального агента діагностування ревматологічних захворювань

Розглянемо приклад функціонування ІА в області медицини, а саме ревматології. Як прецеденти було розглянуто сім захворювань:  $Pr = \{ \text{‘Ревматоїдний артрит (РА)’, ‘Деформуючий остеоартроз (ДО)’, ‘Системний червоний вовчак (СЧВ)’, ‘Анкілозуючий спондиліт (АС)’, ‘Гостра ревматична лихоманка (ГРЛ)’, ‘Подагричний артрит (ПА)’, ‘Системна склеродермія (СС)’} \}$ . Нами виділено 27 ознак, які впливають на ці захворювання. Очевидно, що не всі ознаки присутні в окремому взятому захворюванні. Тому, в залежності від прецеденту, важливість цих ознак різна. Саме таку вагу важливості нам необхідно визначити, щоб для діагностування використати формулу (1). При використанні різних ревматологічних довідників було побудовано онтологію ревматології. В онтології відображено взаємозв'язок між ознаками, їх вплив на захворювання в залежності від значення ознаки.

Насамперед обчислимо початкові вагові коефіцієнти. Для цього побудуємо дерево рішень (ДР) на основі архіву даних про захворювання колишніх пацієнтів. Отримане ДР наведено на рис. 2. Для задання ваг базових ознак прецедентів скористаємось арифметичним способом їх визначення. Так, для РА отримуємо такі значення:  $W_{12}^0 = \frac{1}{2}$ ,  $W_1^0 = \frac{1}{3}$ ,

Тепер необхідно ці ваги розвинути на всю онтологію ПО. Для ознак, які приймають участь у ДР, до ваг важливості понять відповідного прецеденту додаємо вагу, отриману на основі ДР. Потім розглядаємо функціональні та семіотичні зв'язки, якщо вони визначені в онтології. Функціональні відношення діляться на симетричні  $R_S$  (деяка підмножина горизонтальних зв'язків) та несиметричні  $R_N$  (вертикальні зв'язки, інша підмножина горизонтальних зв'язків). Очевидно, що елементи, які приймають участь у симетричних зв'язках, є рівносильними. Тому ваги  $L$

$W_7^0 = \frac{1}{6}$ . Верхній індекс (0) вказує, що ваги є початковими, нижній індекс – ключ ознаки в базі даних (БД). Аналогічно визначаємо ваги ознак для інших захворювань.

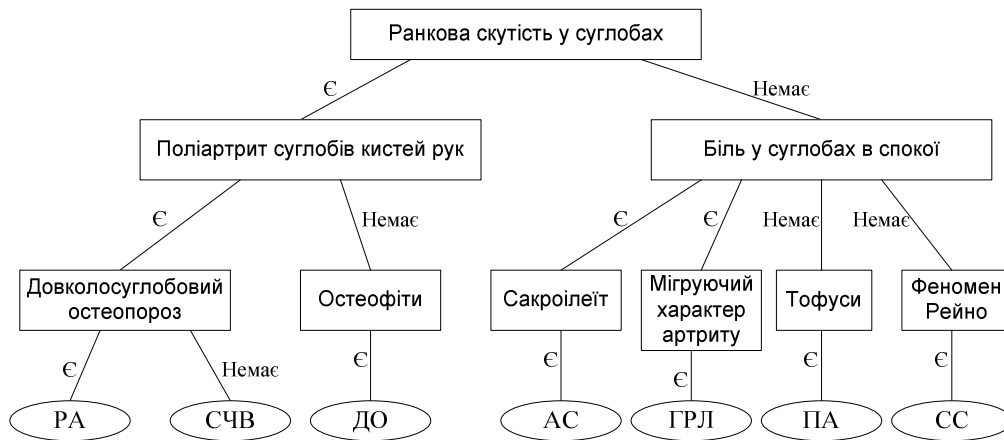


Рис. 2. Дерево рішень, отримане на основі аналізу даних ревматологічних захворювань

Основними для перерахунку ваг важливості понять онтології є функціональні та кореляційні відношення. Окремі з них наведено на рис. 3а-в. У двох відношеннях доменом є ознака „Ерозивний артрит” (ключ 6 у БД). Множиною значень у першому випадку є „Ранкова скутість у суглобах” (ключ 12 у БД) та „Довколосуглобовий остеопороз” (ключ 7 у БД). Обидва відношення є функціональні та симетричні.

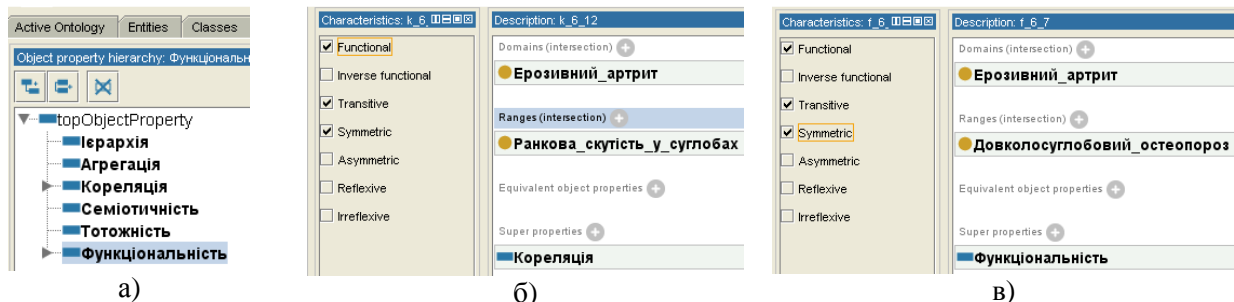


Рис. 3. Види відношень онтології ревматології

Перерахувавши ваги важливості понять згідно з формулами (3)–(4), отримаємо такі ознаки для різних захворювань, які необхідно використати під час обчислення відстаней за формулою (1) (табл. 2). Ми взяли не більше, як три ознаки для кожного прецеденту. У випадку, якщо вага 2-го або 3-го поняття була менша за 0,5, то, відповідно до розгляду, бралась одна або дві ознаки (так як у випадку ПА та СС). Окрім того, використання такої властивості, як Disjoint в редакторі онтологій Protégé-OWL (тобто, заперечення), привело до того, що відсутність ознаки також має значну вагу. Такий результат ми отримали у випадку ГРЛ, коли відсутній „Ерозивний артрит”.

Аналіз табл. 2 в порівнянні з ДР, яке наведено на рис. 2, показує, що для ПА та СС ознаки практично не змінились. Для інших захворювань помінялись в одній або двох позиціях. Перевірка реальних захворювань показала, що використання ДР правильно класифікує захворювання в 64% випадків, а класифікація, отримана згідно з онтологією ревматології, в 79% випадків.

Таблиця 2. Залежність захворювань від ознак

Прецедент	Ознака 1	Ознака 2	Ознака 3
Ревматоїдний артрит (РА)	Поліартрит суглобів кистей рук	Ерозивний артрит	Довколосуглобовий остеопороз
Деформуючий остеоартроз (ДО)	Остеофіти	Субкортикальні кісти без ерозій	Біль у суглобах при фізичному навантаженні
Системний червоний вовчак (СЧВ)	Біль у суглобах у спокої	Антитіла до двохспіральної ДНК у сироватці крові	Поліартрит суглобів кистей рук
Анкілозуючий спондиліт (АС)	Сакроілеїт	Біль у суглобах у спокої	Експресія HLA-D27 антигену
Гостра ревматична лихоманка (ГРЛ)	Моноолігоартрит великих суглобів	Мігруючий характер артрити	Ерозивний артрит (відсутність)
Подагричний артрит (ПА)	Тофуси	Сечова кислота	
Системна склеродермія (СС)	Феномен Рейно		

#### 4. Висновки

Розроблено математичну модель функціонування інтелектуальних агентів на основі адаптивних онтологій для пошуку релевантних прецедентів. Ця модель базується на метриці. Для побудови такої метрики використано адаптивну онтологію. З цією метою у загальноприйнятій трьохелементній кортеж, який задає онтологію (множина понять, відношень та їх інтерпретація), нами додано дві скалярні величини (важливість понять та відношень), які використовуються для обчислення відстаней. Розглянуто способи задання початкових коефіцієнтів важливості понять та відношень на основі інтелектуального аналізу даних та їх розповсюдження на всю онтологію. Розроблено загальний підхід для функціонування інтелектуальних агентів, які використовують розроблену модель, що складається із трьох кроків. Розглянуто приклад функціонування такої системи для діагностики ревматологічних захворювань. Отримані результати показують вищу ефективність розробленого методу в порівнянні з деревами рішень.

#### СПИСОК ЛІТЕРАТУРИ

1. Funk P. Advances in Case-Based Reasoning / P. Funk, P.A. González-Calero // 7th European Conference, ECCBR 2004. – Madrid, Spain, 2004. – P. 375 – 380.
2. Wettschereck D. A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms / D. Wettschereck, D. Aha, T. Mohri // Artificial Intelligence Review. – 2008. – N 11. – P. 273 – 314.
3. Рассел С. Искусственный интеллект / С. Рассел, П. Норвиг. – М., С.-П., К.: Вильямс, 2006. – 1408 с.
4. Цветков А.М. Разработка алгоритмов индуктивного вывода с использованием деревьев решений / А.М. Цветков // Кибернетика и системный анализ. – 1993. – № 1. – С. 174 – 178.
5. Інтелектуальні системи, базовані на онтологіях / Д.Г. Досин, В.В. Литвин, Ю.В. Нікольський, В.В. Пасічник. – Львів: Цивілізація, 2009. – 414 с.
6. Литвин В.В. Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології / В.В. Литвин // Радіоелектроніка. Інформатика. Управління. – 2009. – № 2 (21). – С. 120 – 126.
7. Найханова Л.В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования / Найханова Л.В. – Улан-Удэ: Издательство БНЦ СО РАН, 2008. – 244 с.

*Стаття надійшла до редакції 29.03.2011*