

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ШАБАНОВА-КУШНАРЕНКО ЛЮБОВ ВОЛОДИМИРІВНА



УДК 519.715:004.89

**ПРЕДИКАТНІ МОДЕЛІ, МЕТОДИ ТА ТЕХНОЛОГІЯ ОБРОБКИ
ПРОЦЕСНИХ ЗНАНЬ В ІНФОРМАЦІЙНИХ СИСТЕМАХ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2016

Дисертацією є рукопис.

Роботу виконано на кафедрі інформаційних управляючих систем Харківського національного університету радіоелектроніки Міністерства освіти і науки України.

Науковий керівник доктор технічних наук, професор
Шостак Ігор Володимирович,
Національний аерокосмічний університет
ім. М.Є. Жуковського «ХАІ»,
професор кафедри інженерії програмного
забезпечення

Офіційні опоненти: доктор технічних наук, професор
Жолткевич Григорій Миколайович,
Харківський національний університет
ім. В. Н. Каразіна,
декан факультету математики і інформатики

кандидат технічних наук, доцент
Ситніков Дмитро Едуардович,
Харківська державна академія культури,
завідувач кафедри інформаційно-документних
систем

Захист відбудеться 30 червня 2016 р. о 13-00 годині на засіданні спеціалізованої вченої ради Д 64.050.07 в Національному технічному університеті «Харківський політехнічний інститут» за адресою:
61002, м. Харків, вул. Фрунзе, 21.

З дисертацією можна ознайомитись у бібліотеці Національного технічного університету «Харківський політехнічний інститут» за адресою:
61002, м. Харків, вул. Фрунзе, 21.

Автореферат розісланий « 27 » травня 2016 р.

Вчений секретар
спеціалізованої вченої ради

Дорофєєв Ю. І.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Проблема обробки процесних знань стала особливо важливою у зв'язку з появою і розвитком динамічних інформаційних систем, які здатні до зміни свого стану з плином часу, представниками яких є ІС підтримки прийняття рішень реального часу. Нині швидко розвиваються ІС, орієнтовані на відкриті динамічні предметні області, які в процесі свого функціонування оперують з великою кількістю інформації, що змінюється в часі.

Певний вклад в постановку і рішення задач обробки процесних знань внесли Wil van der Aalst, Richard T. Snodgrass, Arie Segev, J. F. Allen, Дж. Пітерсон, G. Bellinger, D. Nikovski, C. W. Holsapple, J. Liang, B. F. van Dongen, В. Н. Вагін, А. П. Єремеев і багато інших учених.

Існуючі підходи до обробки процесних знань і побудова на їх основі інформаційних систем для автоматизованої обробки інформації і управління не дозволяють отримати досить ефективні результати. Існуючі методи і моделі обробки процесних знань можуть бути вдосконалені на основі аналітичного аналізу лог-файлів інформаційних систем. Перспективною є обробка процесних знань за допомогою технологій інтелектуального аналізу процесів (англ.: Process Mining) і виведення знань на прецедентах (англ.: Case-Based Reasoning, CBR), вдосконалених на основі формального апарату алгебри скінченних предикатів і реляційних мереж. Вирішення цієї комплексної задачі забезпечує точність і повноту моделі процесних знань, відображення версій виконання процесу і адаптацію моделі до предметної області, що і визначає актуальність дисертаційного дослідження.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконана на кафедрі інформаційних управляючих систем Харківського національного університету радіоелектроніки у рамках держбюджетних НДР МОН України: «Розробка теорії і принципів побудови мозкоподібних структур» (№ ДР 0112 U000205) і «Теорія, методи і моделі управління життєвим циклом інтелектуальних інформаційних середовищ регіональних соціо-економічних об'єктів» (№ ДР 0115 U002430), в яких здобувач брала участь в якості виконавця.

Метою дисертаційної роботи є розробка предикатних моделей, методів та технології обробки процесних знань на основі аналізу лог-файлів з використанням апарату алгебри скінченних предикатів (АСП) і реляційних мереж, технології Process Mining і методів виведення знань на прецедентах.

Виходячи з мети, виділені основні задачі:

– провести аналіз засобів обробки процесних знань: технології Process Mining, методів виведення знань на прецедентах, формального апарату АСП і реляційних мереж;

– розробити модель оцінки близькості прецедентів з урахуванням функціональної залежності ваг атрибутів від значень атрибутів та метод лінеаризації шкал ваг;

– вдосконалити предикатну модель метрики на просторі прецедентів та метод його лінеаризації;

– розробити предикатну ієрархічну модель процесних знань та метод її конфігурування;

– експериментально перевірити отримані теоретичні результати дослідження на прикладі задач медичної діагностики;

– побудувати інформаційну технологію обробки процесних знань для комплексу автоматизованої медичної діагностики у Центральній клінічній лікарні Укрзалізниці м. Харкова.

Об'єкт дослідження – процеси обробки знань в інформаційних системах.

Предмет дослідження – методи і моделі обробки процесних знань в інформаційних системах.

Методи дослідження. Для вирішення поставлених задач використано: при розробці моделі метрики на просторі прецедентів та методу його лінеаризації – апарат АСП і теорію векторних просторів; при розробці моделі оцінки близькості прецедентів з урахуванням функціональної залежності ваг атрибутів від значень атрибутів та методу лінеаризації шкал ваг – методи виведення знань на прецедентах, апарат АСП та обчислювальну математику; при розробці методів побудови та конфігурування предикатної ієрархічної моделі процесних знань – методи Process Mining, апарат АСП, алгебру предикатних операцій, реляційні мережі; для експериментальної перевірки отриманих теоретичних результатів дослідження – обчислювальну математику, технології програмування і комп'ютерне моделювання для оцінки ефективності розроблених алгоритмів; при розробці інформаційної технології обробки процесних знань для комплексу автоматизованої медичної діагностики – методи Process Mining та виведення знань на прецедентах, апарат АСП та реляційних мереж, обчислювальну математику і технології програмування.

Наукова новизна отриманих в дисертації результатів:

– *вперше* розроблена модель оцінки близькості прецедентів з урахуванням функціональної залежності ваг атрибутів від значень атрибутів і метод лінеаризації шкал ваг, заснований на їх табулюванні, що дозволяє уточнити взаємодію прецедентів завдяки зміні ваг атрибутів в критичних діапазонах;

– *отримала подальший розвиток* предикатна модель метрики на прецедентах, заснована на лінеаризації простору атрибутів прецедентів за допомогою метризуючого предиката, а також операцій зовнішнього і внутрішнього рівноподілення, що дозволяє ввести евклідову метрику і підвищити точність класифікації прецедентів;

– *отримала подальший розвиток* ієрархічна модель процесних знань «дані-інформація-знання-метазнання», заснована на аналізі лог-файлів та формалізації структури процесних знань у вигляді реляційної метамережі, що спрощує конфігурування моделі;

– *вдосконалено* метод конфігурування ієрархічної моделі процесних знань у вигляді ієрархічної системи бінарних предикатів, що дозволяє адаптувати процес під предметну область та підвищити ефективність методів Process Mining.

Практичне значення одержаних результатів для інформаційних систем полягає у розробці методичних, алгоритмічних і інформаційних засобів аналізу процесних знань, які призначені для: ієрархічного аналізу лог-файлів, виконання і

контролю бізнес-процесів в електронному документообігу, аналізу зв'язків між даними (англ.: Link Analysis), що застосовуються в таких областях як біоінформатика, цифрові бібліотеки; пошук асоціативних правил та кластеризація даних в системах електронного бізнесу; побудова діагностичних медичних систем.

Результати, отримані в процесі виконання дисертаційної роботи, впроваджені в Центральній клінічній лікарні Укрзалізниці (м. Харків) при розробці програмно-алгоритмічної частини і наповненні бази знань програмно-апаратного комплексу автоматизованої медичної діагностики (акт впровадження від 21.09.2015 р.).

Теоретичні результати дисертації впроваджено у навчальному процесі на кафедрі програмної інженерії ХНУРЕ при підготовці дисциплін «Теорія інтелекту», «Біоніка інтелекту» і «Системи штучного інтелекту» (акт впровадження від 10.11.2015 р.) та на кафедрі інтелектуальних комп'ютерних систем НТУ «ХПІ» при підготовці дисциплін «Математичні основи лінгвістики» і «Штучний інтелект: лінгвістичні проблеми» (акт впровадження від 22.02.2016 р.).

Особистий внесок здобувача. Дисертація є самостійною завершеною науковою роботою. Усі результати, представлені в дисертаційній роботі, отримані здобувачем особисто, серед них: предикатна ієрархічна модель процесних знань у вигляді реляційної метамережі, заснована на відомій ієрархічній моделі «дані-інформація-знання-метазнання»; метод побудови метрики на множині прецедентів з функціональною залежністю ваг від атрибутів прецедентів, який лінеаризує шкали виміру атрибутів; предикатна модель метрики на просторі прецедентів, яка дозволяє лінеаризувати нелінійний простір атрибутів і ввести на ньому евклідову метрику; метод конфігурування ієрархічної моделі процесних знань на основі аналізу лог-файлів, який дозволяє інтегрувати лог-файли процесу в одній ієрархічній моделі, що розширює сферу застосування методів інтелектуального аналізу процесів.

Апробація результатів. Основні положення і результати роботи докладалися і обговорювалися на: Міжнародній науково-технічній конференції «Інформаційні системи і технології» ICT-2012 (2012 р., Харків); VIII Міжнародній науково-практичній конференції «Військова освіта та наука: сьогодення та майбутнє» (2012 р., Київ); V Міжнародній науково-технічній конференції «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління» (2015 р., Полтава-Баку-Кіровоград-Харків); XX Міжнародному конгресі двигунобудівників (2015 р., Харків); V Всеукраїнській науково-практичній конференції «Інтелектуальні системи і прикладна лінгвістика» (2016 р., Харків).

Публікації. За результатами досліджень опубліковано 11 наукових праць, у тому числі 5 статей у наукових фахових виданнях України з технічних наук (1 – у виданні, включеному до міжнародних наукометричних баз), 1 – у закордонному періодичному фаховому виданні, 5 – у матеріалах науково-технічних конференцій.

Структура й обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел, додатків. Загальний обсяг дисертації становить 154 сторінки, з них 32 рисунків по тексту; 2 таблиці по тексту; списку використаних джерел з 164 найменувань на 17 сторінках, 2 додатки на 12 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** визначено основні проблеми інтелектуальної обробки процесних знань, обґрунтована актуальність теми, сформульовані мета і задачі дослідження, наведені відомості про зв'язок роботи з науковими програмами, планами і темами організації, у якій виконувалися дослідження. Надано характеристику об'єкту, предмету і методів дослідження, наукової новизни, практичного значення і використання отриманих результатів.

У **першому розділі** проаналізовано метод отримання рішення задач, заснованого на прецедентах, методи інтелектуального аналізу процесів. Досліджено основні особливості та структурні елементи моделі процесних знань; показано, що переваги моделі пов'язані з можливостями конфігурування, а також ієрархічного представлення процесу. Показано переваги апарату АСП як засобу для ідентифікації інформаційних процесів, що забезпечує можливість виділення структури моделі процесних знань (у тому числі ієрархічного аспекту) на основі аналізу лог-файлів.

Відзначено доцільність удосконалення моделі процесних знань для відображення її ієрархічного аспекту, а також розробки методу формування моделі процесних знань, який забезпечує побудову послідовності дій та ієрархічної складової.

У **другому розділі** наведено вдосконалену предикатну модель метрики на прецедентах, засновану на лінеаризації простору атрибутів.

Дано n -вимірний арифметичний простір N , на якому визначено предикат T

$$T(x_1, y_1, x_2, y_2) = D(r(\varphi(x_1), \varphi(y_1)), r(\varphi(x_2), \varphi(y_2))), \quad (1)$$

де r – евклідова метрика; φ – взаємно неперервне та взаємно однозначне відображення нелінійного простору N у лінійний простір N' , $x_1, y_1, x_2, y_2 \in N$.

Символ D означає предикат рівності, заданий на множині усіх відстаней між парами точок простору N' . Предикат T встановлює наявність або відсутність рівності відстаней між образами точок x_1, y_1, x_2, y_2 . Задача метризації простору N' зводиться до пошуку конкретного виду функції φ за значеннями предиката T .

Предикат R на просторі N вводиться через предикат $T \quad \forall x, y, z \in N$

$$\forall t \in N \quad R(x, y, z) = T(x, z, z, y) \wedge (T(x, t, t, y) \wedge T(z, x, x, t) \supset z = t).$$

Рівність $R(x, y, z) = 1$ означає, що точка $\varphi(z)$ лежить посередині відрізка прямої, що сполучає точки $\varphi(x)$ і $\varphi(y)$ в просторі N' . Відношення, що відповідає предикату $R(x, y, z)$, виражає внутрішнє рівноподілення точок x і y . У просторі N для будь-яких двох точок x і y завжди знайдеться єдина точка z , така, що точка y буде результатом внутрішнього рівноподілення інтервалу між точками x і z . Ця операція визначається як зовнішнє рівноподілення точок x і y . За допомогою метризуючого предиката T , операцій внутрішнього та зовнішнього рівноподілення побудована експериментальна сітка, яка лінеаризує простір атрибутів прецедентів.

Розроблено два варіанти метода експериментального знаходження точок, що визначають відображення φ . Перший метод полягає в побудові на площині N' сітки

з рівносторонніх трикутників із стороною e . Другий метод – сітки з квадратів із стороною e .

Процес лінеаризації простору N покажемо на прикладі першого методу, коли простори N і N' двовимірні. Задається число e , що визначає розмір комірки сітки. Нехай $a'_{11}, a'_{12}, a'_{13}$ – вершини одного з трикутників. На площині N вибирається довільна точка a'_{11} і визначається $\varphi^{-1}(a'_{11}) = a_{11}$. У подальшому вибирається довільна точка $a'_{12} \neq a'_{11}$ і визначається $\varphi^{-1}(a'_{12}) = a_{12}$. Наявність варіантів вибору точки a_{12} говорить про неоднозначність вибору масштабу і напряму по деякій координатній осі простору N' . Виберемо точку a_{13} так, щоб $T(a_{11}, a_{13}, a_{12}, a_{13}) = T(a_{11}, a_{12}, a_{12}, a_{13}) = 1$, що забезпечить рівність відстаней $r(a'_{11}, a'_{13})$, $r(a'_{12}, a'_{13})$ і $r(a'_{11}, a'_{12})$. Існують дві точки, які задовольняють цій умові. Вважаємо $\varphi^{-1}(a'_{13}) = a_{13}$. Подальша побудова точок сітки на площині N виконується однозначно.

Запропонована предикатна модель метрики на просторі прецедентів дозволяє лінеаризувати простір атрибутів прецедентів і ввести на них евклідову метрику. Завдяки цьому збільшується точність виміру відстаней між прецедентами.

На основі методів виведення знань на прецедентах та апарату АСП розроблено метод визначення метрики на множині прецедентів з урахуванням функціональної залежності ваг атрибутів. Такий підхід знайшов застосування при розробці комплексу автоматизованої медичної діагностики для уточнення оцінки близькості прецедентів завдяки зміні ваг атрибутів в їх критичних діапазонах, що є важливим фактором для багатьох інформаційних систем.

Вводиться вектор ваг атрибутів a_i прецедента a – $v(a) = (v_1(a_1), v_2(a_2), \dots, v_m(a_m))$, m – кількість атрибутів прецедента a . Модель метрики на множині прецедентів з урахуванням функціональної залежності ваг атрибутів має вигляд

$$S = \sum_{i=1}^m |v_i(a_i) \cdot a_i - v_i(b_i) \cdot b_i|, i = \overline{1, m},$$

де a, b – пара прецедентів, для яких обчислюється оцінка близькості.

Для лінеаризації функцій $v_i(a_i), v_i(b_i)$ вводяться достатні умови існування шкал вимірювання аргументів, які лінеаризують модель об'єкту (рис. 1).

Задача полягає в знаходженні вектор-функції $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$, такої що

$$\begin{aligned} G(v(a) \cdot a, v(b) \cdot b) &= G'(\alpha(v(a)) \cdot a, \alpha(v(b)) \cdot b), \\ \alpha(v(a) \cdot a) &= (\alpha_1(v_1(a_1)) \cdot a_1, \alpha_2(v_2(a_2)) \cdot a_2, \dots, \alpha_m(v_m(a_m)) \cdot a_m), \\ \alpha(v(b) \cdot b) &= (\alpha_1(v_1(b_1)) \cdot b_1, \alpha_2(v_2(b_2)) \cdot b_2, \dots, \alpha_m(v_m(b_m)) \cdot b_m), \end{aligned} \quad (2)$$

де G – нелінійний, G' – лінійний предикати.

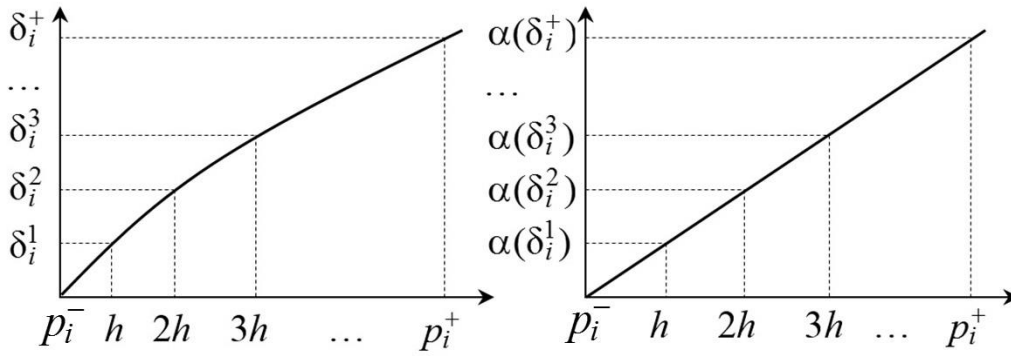


Рисунок 1 – Лінеаризація шкал вимірювання атрибутів a_i

Після лінеаризації здійснюється перехід від нелінійної моделі об'єкту G до лінійної G' . Вектор-функція α знаходиться по точках з експерименту, тому метод визначення метрики на множині прецедентів зводиться до її табулювання.

Якщо вектор-функція α є ізоморфізмом і хоча б два компоненти вектору $v(a)$ (наприклад, $v_1(a_1)$ і $v_2(a_2)$) не дорівнюють нулю, то процедура табулювання вектор-функції α наступна:

1. Визначення для вектор-функції $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ значень нижніх p_i^- і верхніх p_i^+ меж діапазонів α_i . Точність моделі відповідає числу експериментальних точок k , яке залежить від кроку табулювання h : $k = ((p_i^+ - p_i^-) / h) + 1$.

Визначимо величини $\delta_i^-, \delta_i^h, \delta_i^+$, для яких справедливі рівності:

$$\begin{aligned} \alpha(\delta_1^-, \delta_2^-, \dots, \delta_m^-) &= (p_1^-, p_2^-, \dots, p_m^-), \\ \alpha(\delta_1^-, \dots, \delta_i^h, \dots, \delta_m^-) &= (p_1^-, \dots, h, \dots, p_1^-), \\ \alpha(\delta_1^-, \dots, \delta_i^+, \dots, \delta_m^-) &= (p_1^-, \dots, p_i^+, \dots, p_1^-), i = \overline{1, m}. \end{aligned} \quad (3)$$

Тільки у цьому пункті потрібно використання явних значень вектор-функції α – величин h, p_i^+, p_i^- . Значеннями p_i^+ і p_i^- приймаються відповідно максимальні і мінімальні значення функцій α_i . В якості нулів функцій α_i беруться будь-які їх значення, оскільки вони визначають тільки положення системи координат цих функцій. Для спрощення алгоритму табулювання визначим їх рівними нижнім межах діапазонів p_i^- функцій α_i . Такий вибір нулів розташовує усі інші значення α_i в додатної області. Вибір значень δ_i^- і δ_i^+ досить довільний і визначає відповідність діапазонів областей визначення $[p_i^-, p_i^+]$ і областей значень $[\delta_i^-, \delta_i^+]$ функцій α_i .

2. Визначення величини δ_2^1 , рівновеликій величині δ_1^1 , тобто такої, що

$$\alpha_2(\delta_2^1) = \alpha_1(\delta_1^1). \quad (4)$$

З умови

$$G((\delta_1^1, \delta_2^0, \delta_3, \dots, \delta_m), (\delta_1^0, \delta_2^1, \delta_3, \dots, \delta_m)) = 1, \quad (5)$$

де $\delta_3, \dots, \delta_m$ – довільні фіксовані величини. Дійсно, перейдемо до предикату G' за формулою (2)

$$G'((\alpha_1(\delta_1^1), 0_2, \alpha_3(\delta_3), \dots, \alpha_m(\delta_m)), (0_1, \alpha_2(\delta_2^1), \alpha_3(\delta_3), \dots, \alpha_m(\delta_m))) = 1. \quad (6)$$

Спрощення предиката G' виконується наступним чином. Без обмеження загальності вважається, якщо $a_i > 0$, тоді $a_i = 1$, $i = \overline{1, m}$. Дійсно, якщо $a_i > 0$, то покладається $\alpha'_i(x_i) = \alpha_i(x_i) / a_i$. Очевидно, що функції α'_i утворюють ізоморфізм α' предикатів G і G_1 , такий що $\forall x, y \in R^m \ G(\alpha(x), \alpha(y)) = G_1(\alpha'(x), \alpha'(y))$, причому вектор-функція α' і предикат G_1 мають усі властивості вектор-функції α і предиката G відповідно. Значить, вважатимемо, що $a_i = 1$ або $a_i = 0$, $i = \overline{1, m}$. Користуючись останньою умовою, рівністю (3) і тим, що $a_1 \neq 0$ і $a_2 \neq 0$, здійснюється перехід від (6) до наступної рівності

$$\alpha_1(\delta_1^1) + 0 + \alpha_3(\delta_3) + \dots + \alpha_m(\delta_m) = 0 + \alpha_2(\delta_2^1) + \alpha_3(\delta_3) + \dots + \alpha_m(\delta_m).$$

Величини, залежні від $\delta_3, \dots, \delta_m$, взаємно знищуються і залишається незалежна від них рівність

$$\alpha_1(\delta_1^1) = \alpha_2(\delta_2^1), \quad (7)$$

що і треба було показати.

3. Табулювання функції α_1 складається із етапів:

3.1. Визначення величини δ_1^2 , такої що

$$\alpha_1(\delta_1^2) = 2\alpha_1(\delta_1^1), \quad (8)$$

з умови

$$G((\delta_1^1, \delta_2^1, \delta_3, \dots, \delta_m), (\delta_1^2, \delta_2^0, \delta_3, \dots, \delta_m)) = 1. \quad (9)$$

Дійсно, використовуючи рівності (2), (3) та умову $a_i = 1$ або $a_i = 0$, $i = \overline{1, m}$, отримаємо

$$\alpha_1(\delta_1^1) + \alpha_2(\delta_2^1) = \alpha_1(\delta_1^2). \quad (10)$$

Разом з рівнянням (4) це дає потрібне співвідношення (8).

3.2. Визначення величини δ_1^3 , такої що

$$\alpha_1(\delta_1^3) = 3\alpha_1(\delta_1^1), \quad (11)$$

з умови

$$G((\delta_1^2, \delta_2^1, \delta_3, \dots, \delta_m), (\delta_1^3, \delta_2^0, \delta_3, \dots, \delta_m)) = 1. \quad (12)$$

Доведення здійснюється аналогічно (10).

Табулювання триває, доки не буде пройдений весь діапазон значень, тобто до кроку 3.N, коли

$$\delta_1^N \geq \delta_1^{p^+}, \delta_1^{N-1} < \delta_1^{p^+}. \quad (13)$$

4. Табулювання функцій $\alpha_i, i = \overline{2, n}$, таких що $a_i \neq 0$. Для визначення величини δ_i^k , такої що

$$\alpha_i(\delta_i^k) = kh, \quad (14)$$

треба скористатися умовою

$$G((\delta_1^k, \delta_2, \dots, \delta_{i-1}, \delta_i^0, \delta_{i+1}, \dots, \delta_m), (\delta_1^0, \delta_2, \dots, \delta_{i-1}, \delta_i^k, \delta_{i+1}, \dots, \delta_m)) = 1. \quad i = \overline{2, m}, k = \overline{1, N}. \quad (15)$$

Доведення таке ж саме, що і в п. 3.1. Якщо $a_i = 0$, то значення δ_i і α_i не впливають на предикати G і G' . Загальна таблиця табулювання має вигляд:

$\alpha_i(x_i)$	0	h	$2h$...	Nh
x_i	δ_i^0	δ_i^1	δ_i^2	...	δ_i^N

Третій розділ присвячений удосконаленню методів та моделей обробки процесних знань, формалізованих за допомогою апарата АСП і реляційних мереж. Модель M процесу в максимально узагальненому виді – це система n бінарних предикатів R_j

$$M = \{R_j \mid j = \overline{1, n}\}.$$

Модель процесу, адаптована під предметну область, має вигляд

$$M^O = \{R_j \mid \forall R_j \in M \exists O_k \in O, k, j = \overline{1, n}\},$$

де O – предметна область, O_k – об'єкти предметної області.

Розроблено ієрархічну предикатну модель обробки процесних знань, що дозволяє представити «горизонтальний» процес у вигляді «вертикальної» структури в аналітичній предикатній формі. Перший рівень моделі – множина даних $\{x_1, x_2, \dots, x_n\}$ є результатом спостережень і не завжди має зручний для використання формат. На другому рівні реляційна мережа $\{R_1, R_2, \dots, R_m\}$ задає зв'язки між даними, які визначають структуру інформації. На третьому рівні накопичення знань на основі даних і інформації формалізується як додавання нових реляційних мереж $\{I_i(R_1, R_2, \dots, R_m)\}, i = \overline{1, k}$, заданих на множині початкових даних $\{x_1, x_2, \dots, x_n\}$. Четвертий рівень метазнань об'єднує усі попередні рівні, дозволяючи знаходити нове розуміння існуючого знання. Формально рівень метазнань має вигляд предикатної операції $W(I_1, I_2, \dots, I_k)$, яка задана на множині $\{I_1, I_2, \dots, I_k\}$, зв'язує всю формалізовану інформацію і в процесі її обробки отримує нову інформацію (рис. 2.).

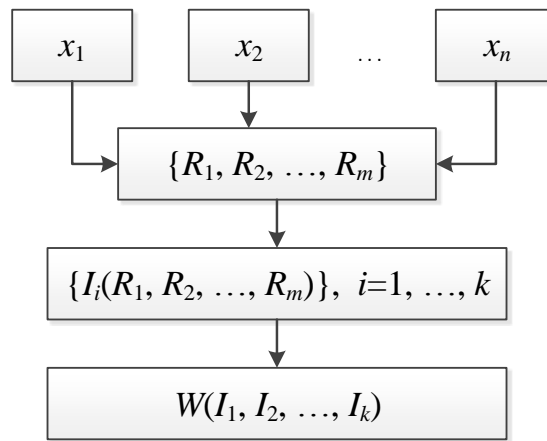


Рисунок 2 – Ієрархічна предикатна модель обробки процесних знань

Ієрархічна модель є реляційною метамережею. Вузлами метамережі є реляційні мережі, а зв'язками (дугами) – предикатні операції (рис. 3.). Предикатна модель виділяє фрагменти процесу і аналізує їх як підпроцеси основного процесу за заданими класифікаційними ознаками. Ієрархічне представлення містить набір допустимих варіантів реалізації (сценаріїв) з урахуванням умов предметної області, що зменшує складність моделі. Перевага запропонованої моделі полягає в усуненні проблеми «спагеті»-моделі, де граф процесу складний та заплутаний, внаслідок чого експерт не має можливості ефективно його аналізувати. Введена ієрархічна модель об'єднує лог-файли різних варіантів реалізації одного процесу, дозволяє аналізувати будь-який фрагмент моделі на різних рівнях ієрархії без втрати точності, програмно або за допомогою експерта.

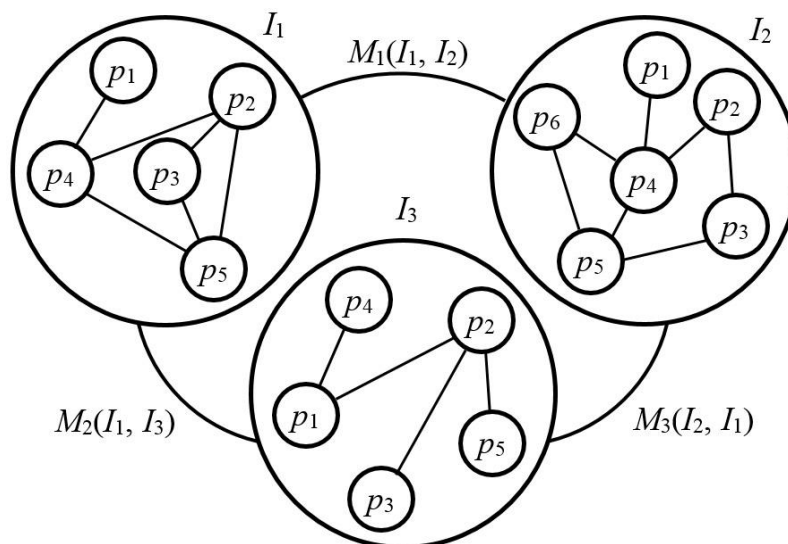


Рисунок 3 – Граф реляційної метамережі

Розроблено метод конфігурування ієрархічної моделі процесних знань на основі аналізу лог-файлів, який складається з наступних етапів:

Етап 1. Виключення надмірних підпроцесів. Виключаються підпроцеси, що розпочинаються з предиката R_p :

$$M^o = \langle S_k, R_i \rangle, i = \overline{1, m}, i \neq \{R_p, \dots, R_q\},$$

де S_k – предикатні операції; R_i – бінарні предикати; $\{R_p, \dots, R_q\}$ – предикати підпроцесу, що розпочинається з предиката R_p .

Критерій γ видалення підпроцесу:

$$\gamma(R_k) = \begin{cases} 0, \exists O_k \notin O, \\ 1, \forall O_k \in O, \end{cases}$$

$$R_k(x_i, x_j) \in M, \quad i, j = \overline{1, n}, \quad k = \overline{1, m}.$$

Етап 2. Розпізнавання розвилки процесу. Ознака розвилки процесу – наявність в моделі M структури з двох предикатів R_p і R_r , перша змінна яких однакова, а також фрагментів процесу, що містять виконання тільки одного підпроцесу, що починається з предиката R_p або R_r (рис. 4):

$$R_p(x_i, x_j), R_r(x_i, x_k) \in M, \quad j, k = \overline{1, n}, \quad p, r = \overline{1, m}.$$

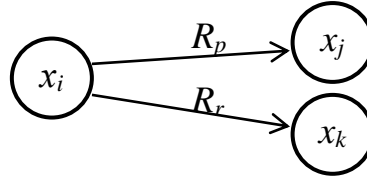


Рисунок 4 – Розвилка реляційної мережі

Етап 3. Вибір найбільш ефективного варіанту підпроцесу (рис. 5).

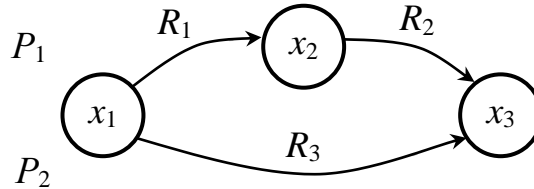


Рисунок 5 – Аналогічні підпроцеси P_1 і P_2

Етап 4. Пошук фрагментів процесу, що містять неоднозначні послідовності виконання операцій (рис. 6). Дія R_1 починається в момент часу t_1^1 та закінчується в t_2^1 . Дія R_2 починається в момент часу t_1^2 , закінчується в t_2^2 . Варіанти виконання дій R_1 і R_2 :

- дія R_2 починається після початку R_1 : $t_1^2 \in [t_1^1, t_2^1]$;
- дія R_2 починається після закінчення R_1 : $t_1^2 > t_2^1$.

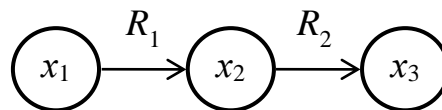


Рисунок 6 – Нечітка послідовність виконання дій

Етап 5. Усунення неоднозначності при виконанні послідовності дій. Вводиться предикатна операція $S_k(R_1, R_2)$, яка визначає послідовність дій.

Етап 6. Пошук фрагментів процесу, що містять паралельне виконання дій процесу, тобто паралельне виконання підпроцесів, що починаються з предикатів R_p і R_r (рис. 3).

Етап 7. Пошук і формалізація циклічного виконання дій процесу (рис. 7).

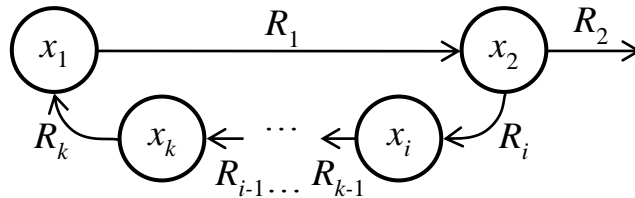


Рисунок 7 – Циклічне виконання дій процесу

Предикатна операція $S_i(R_i, n, z)$ визначає кількість циклів, де $\{R_1, \dots, R_k\}$ – бінарні предикати, що становлять тіло циклу; n – кількість повторів циклу; z – умова зупинки циклу.

Етап 8. Пошук і спрощення конструкцій логічних операцій XOR та OR (рис. 8.). Ознакою наявності в моделі процесу конструкції вибору XOR є виконання тільки однієї з дій R_1 і R_2 в різних фрагментах моделі, обмежених подіями x_1 і x_2 . Якщо ж в моделі зустрічаються конструкції, що містять між подіями x_1 і x_2 обидві дії R_1 і R_2 , то це означає наявність в конструкції вибору OR. Спрощення обох конструкцій полягає у виборі однієї з двох дій, що має більшу ефективність.

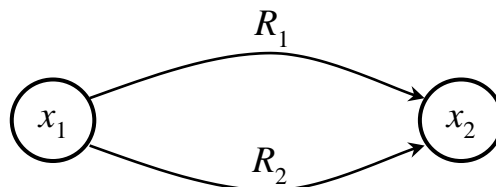


Рисунок 8 – Вибір логічних операцій XOR ($R_1 \oplus R_2$) та OR ($R_1 \vee R_2$)

Таким чином, наданий метод конфігурування ієрархічної моделі процесних знань дозволяє виконати розбиття моделі на як можна дрібніші смислові фрагменти, з подальшим складанням з них нової моделі, але при цьому використовуються тільки ті фрагменти, які відносяться до заданої предметної області, що дає значну перевагу над аналогами у точності і повноті.

Основна ідея побудови моделі процесних знань полягає у формалізації загального рішення, яке надалі може бути багаторазово використане для близьких предметних областей шляхом адаптації на основі правил налаштування (конфігурування). При конфігуруванні процесної моделі, яка має вигляд графу, де вершини графу відповідають подіям моделі, а дуги – діям, використовуються два основні оператори: маскування і виключення. Якщо деяка дуга маскується, то дія, яка їй відповідає, завжди виконується. Зазвичай маскована дія прив'язана до суміжної дії, вони завжди виконуються разом, тому для спрощення графа моделі вони не відображаються. Виключення дуги означає, що дія, яка їй відповідає, і дії

підпроцесу після неї не повинні виконуватися.

Наявність варіантів дій процесу в моделі представлена розвилками, в яких декілька дуг виходять з однієї вершини (рис. 9).

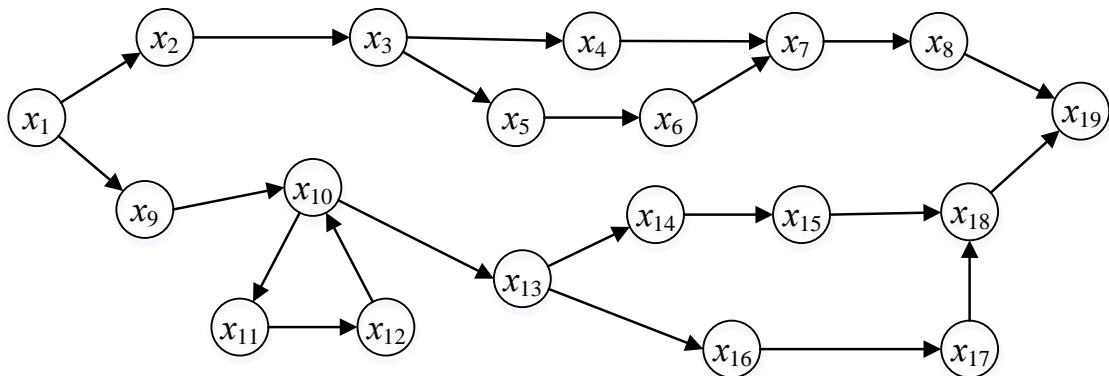


Рисунок 9 – Приклад моделі процесного представлення знань

У прикладі на рис. 9 початок виконання процесу – вершина x_1 , а закінчення – вершина x_{19} . Граф моделює фрагмент процесу виконання замовлення на розробку програмного забезпечення (ПЗ).

x_1 – вступ замовлення на розробку ПЗ;

x_2, x_9, x_{12} – аналіз технічне завдання (ТЗ);

x_3 – формування питань і зауважень до ТЗ;

x_4 – узгодження із замовником зауважень до ТЗ і програмної документації;

x_5 – узгодження зауважень до ТЗ;

x_6, x_{13} – узгодження програмної документації;

x_7, x_{16} – узгодження вартості і термінів виконання замовлення;

x_8, x_{18} – узгодження контролю і приймання замовлення;

x_{10} – узгодження ТЗ із замовником;

x_{11} – отримання відредагованого ТЗ;

x_{14} – узгодження вартості замовлення;

x_{15} – узгодження термінів виконання замовлення;

x_{17} – узгодження етапів виконання замовлення;

x_{19} – передача ТЗ в розробку.

Граф на рис. 9 містить два аналогічні підпроцеси, записаних в одному лог-файлі (журналі подій), але виконуваних в двох різних відділах однієї фірми. Для наочності модель виконання процесу на рис. 10 розділена на два сценарії, що відповідають цим двом підпроцесам.

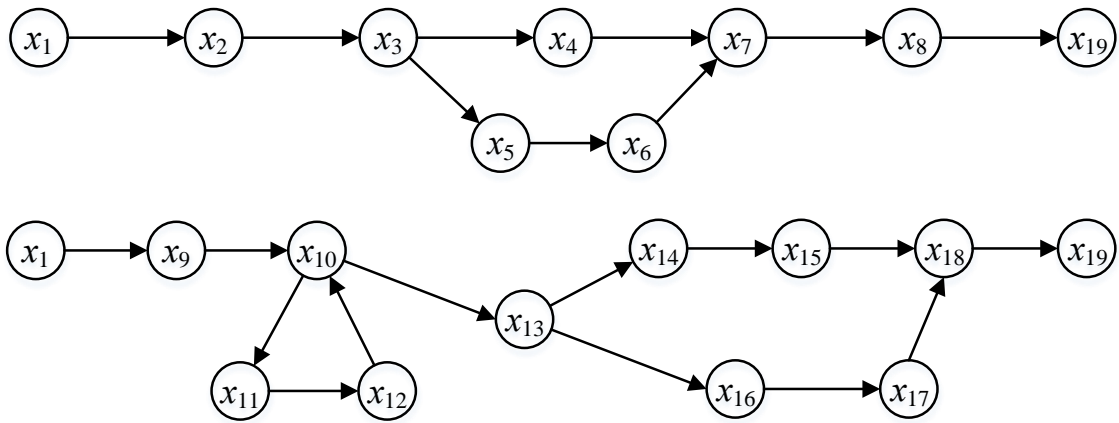


Рисунок 10 – Варіанти виконання сценарію процесу

Операції процесу (дуги графа), що виключаються, на рис. 11 відображаються символом D, а масковані операції - символом M.

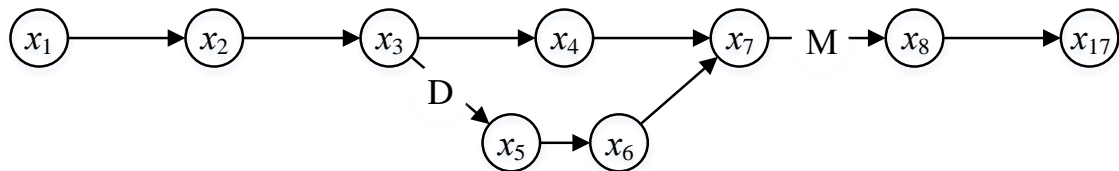


Рисунок 11 – Конфігурування процесної моделі

Конфігурована процесна модель представлена на рис. 12.



Рисунок 12 – Конфігурована процесна модель

У четвертому розділі наведені результати експериментальної перевірки теоретичних досліджень та застосування інформаційної технології обробки процесних знань.

Технологія Process Mining заснована на побудові моделі фактичної послідовності подій за лог-файлом виконання процесу. Лог-файл програмно аналізується за допомогою ProM, або аналогічними програмами, і представляється у вигляді графа процесу, який у подальшому використовується експертом для контролю або вдосконалення процесу. Однією з важливих проблем, що виникає при аналізі лог-файлів ІС, є забезпечення точності моделі. Проблема полягає в тому, що лог-файли досить складні, тому їх моделювання в повному об'ємі дає ефект «спагеті»-моделі. Для вирішення проблеми методами Process Mining скорочують початкові дані і спрощують отримувану модель.

Фрагмент реляційної мережі ієрархічної процесної моделі представлено на рис. 13.

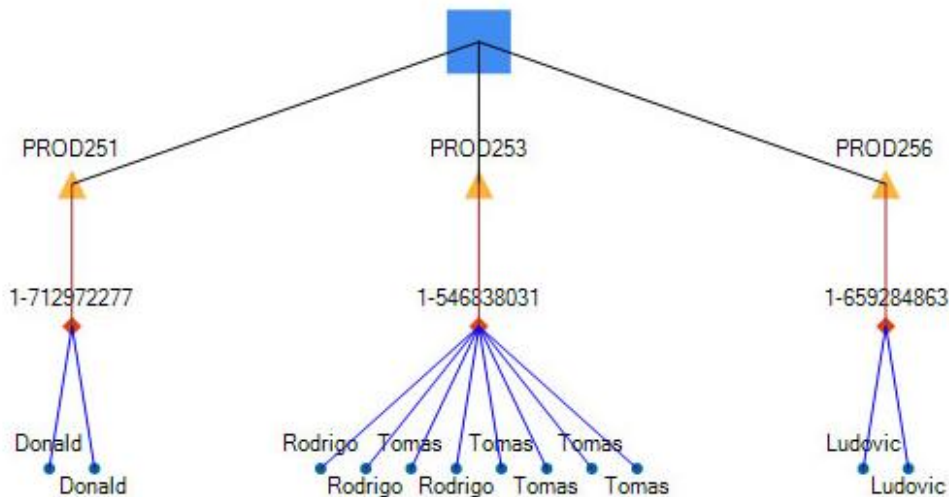


Рисунок 13 – Фрагмент реляційної мережі ієрархічної процесної моделі

Сформульовано функціональні вимоги до програмного забезпечення, що призначене для підтримки роботи ІС, та включає реалізацію наступних функцій: завантаження і аналіз лог-файлу; виділення базових елементів процесу на основі порівняльного аналізу варіантів процесу; побудова ієрархії базових елементів з урахуванням додаткової інформації з лог-файлу; візуальне подання ієрархічної предикатної моделі процесу; підтримка прийняття рішень на основі пошуку та адаптації прецедентів.

Розроблені програмні модулі: Diagnosis для підтримки прийняття рішень, заснованих на прецедентах; ProjectAnalyzer, що реалізує модель процесних знань на основі аналізу лог-файлів. Модулі застосовані при розробці програмно-апаратного комплексу автоматизованої медичної діагностики, який створюється в Центральній клінічній лікарні Укрзалізниці м. Харкова.

На прикладі роботи системи при діагностиці різних форм захворювань жовчних шляхів проведена верифікація теоретичних розробок. Так, атрибути завдання прецеденту (вектор ознак):

[1Біль, 2Апетит, 3Початок захворювання, 4Тривалість захворювання, 5Жовтяниця, 6Вага, 7Біль під ребром справа, 8Випорожнення, 9Уробилин, 10Цитологія, 11Холестерин крові, 12Білірубін, 13Альбуміни, 14Проба на екскреторну функцію, 15Рентгенологічне дослідження, 16Артеріальний тиск, 17Пультс].

Значення атрибутів (ознак):

- 1 [Напади:1, Найчастіше постійні:2, Постійні:3];
- 2 [Мінливий:1, Найчастіше поганий:2, Поганий:3];
- 3 [Найчастіше гостре:1, Поступове:2];
- 4 [<рік:1, >рік:2];
- 5 [Немає:1, Іноді:2, Після нападу:3, Постійно:4];
- 6 [Норма:1, Підвищена:2, Знижена:3];
- 7 [Іноді:1, Постійно:2, Напруження:3, Тиск:4];
- 8 [Нестійке:1, Запор:2, Пронос:3];
- 9 [Немає:1, Є:2];
- 10 [Норма:1, Запалення:2, Ракові клітини:3];
- 11 [Норма:1, Іноді:2, Підвищений:3];
- 12 [Норма:1, Іноді:2, Підвищений:3];

13 [Норма:1, Знижені:2];
 14 [Норма:1, Затримка:2];
 15 [Норма:1, Порушення функції жовчного міхура:2, Камені:2, Пухлина:3];
 16 [АТС:?, АТД:?, Дата:DATA, Час:TIME];
 17 [Частота:?, Ритмічність[17.2.ритм.; 17.2.аритм.], Напруження[17.3.тв., 17.3.ср., 17.3.мяг.], Наповнення[17.4.ср., 17.4.повн., 17.4.пуст.], Форма[17.5.шв., 17.5.скач., 17.5.пов.], Дефіцит[17.6.1, 17.6.2], Дата:DATA, Час:TIME].

Варіанти значень атрибутів рішення прецеденту (діагнозу):

- 1 дисфункція жовчних шляхів,
- 2 хронічний холецистит,
- 3 рак жовчного міхура.

Вибірка з трьох прецедентів з бази знань:

- 1 (1, 1, 1, 1, 2, 2, 1, 4, 2, 1, 1, 2, 1, 1, 1, [135, 85, 15.01.16, 09-35], [76, 1, 2, 3, 2, 1])
- 2 (2, 1, 2, 1, 3, 2, 2, 3, 1, 2, 2, 2, 2, 1, 2, [150, 80, 23.12.15, 10-20], [84, 1, 1, 2, 3, 1])
- 3 (2, 3, 2, 2, 4, 3, 4, 3, 1, 3, 2, 3, 2, 2, 3, [125, 80, 28.12.15, 15-45], [72, 2, 2, 2, 2, 2])

У комплексі автоматизованої медичної діагностики реалізовано дві моделі оцінки прецедентів – модель H_1 з постійними вагами атрибутів

$$H_1 = \sum_{i=1}^m v_i \cdot a_i^k, i = \overline{1, m}, \quad (16)$$

та модель H_2 з вагами атрибутів, функціонально залежними від значень атрибутів

$$H_2 = \sum_{i=1}^m v_i(a_i^k) \cdot a_i^k, i = \overline{1, m}, \quad (17)$$

де H , a_i^k та v_i – оцінки, атрибути та ваги атрибутів заданих прецедентів.

Значення ваг атрибутів та результати діагнозу визначали експерти-медики. Розроблена у дисертації модель H_2 (17) показала більшу точність діагностики, коли значення атрибутів знаходились у критичних діапазонах.

У додатках наведено акти впровадження теоретичних і практичних результатів дисертаційної роботи.

ВИСНОВКИ

У дисертаційній роботі вирішено науково-практичну задачу створення технології обробки процесних знань в інформаційних системах на основі алгебрологічної формалізації моделей і методів інтелектуального аналізу процесів і виведення знань на основі прецедентів. Розроблена технологія пройшла апробацію у системі підтримки прийняття рішень автоматизованого комплексу медичної діагностики.

Отримані наступні основні наукові результати:

1. Проведено аналіз результатів застосування технології Process Mining та методів виведення знань на прецедентах для обробки процесних знань, який виявив їх недоліки. На основі проведеного аналізу поставлено задачу розробки методів, моделей та технології обробки процесних знань, які базуються на застосуванні апарату АСП і реляційних мереж.

2. Розроблена модель оцінки близькості прецедентів із урахуванням функціональної залежності ваг атрибутів від значень атрибутів і метод лінеаризації шкал ваг, заснований на їх табулюванні, що дозволяє уточнити оцінку близькості прецедентів завдяки зміні ваг атрибутів в їх критичних діапазонах і підвищити ефективність пошуку прецедентів в ІС.

3. Розроблена узагальнена предикатна модель метрики на прецедентах, заснована на лінеаризації простору атрибутів прецедентів за допомогою метризуючого предиката, а також операцій зовнішнього і внутрішнього рівноподілення, що дозволяє ввести евклідову метрику і підвищити точність класифікації прецедентів.

4. Вдосконалена ієрархічна модель процесних знань «дані-інформація-знання-метазнання», заснована на аналізі лог-файлів та формалізації структури процесних знань у вигляді реляційної метамережі, що спрощує конфігурування моделі. Розроблений метод конфігурування ієрархічної моделі процесних знань у вигляді ієрархічної системи бінарних предикатів, що дозволяє адаптувати процес під предметну область та підвищити ефективність методів Process Mining.

5. Проведено експериментальну перевірку отриманих теоретичних результатів дослідження на прикладі задач медичної діагностики, яка показала перевагу отриманих засобів обробки процесних знань над існуючими методами.

6. Розроблено технологію обробки процесних знань для комплексу автоматизованої медичної діагностики в Центральній клінічній лікарні Укрзалізниці м. Харкова.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Шабанова-Кушнарєнко Л. В. Реляционная метасеть выделения неявных знаний / Л. В. Шабанова-Кушнарєнко, С. Ф. Чалый // АСУ та прилади автоматики. – Харків : ХНУРЕ. – 2013. – № 164. – С. 80-84.

Здобувачем запропонована математична модель ієрархічної структури «дані-інформація-знання-метазнання».

2. Шабанова-Кушнарєнко Л. В. Метод табулювання метрики прецедентів с функціональною залежністю ваг атрибутів прецедентів / Л. В. Шабанова-Кушнарєнко // Системи обробки інформації. – Харків : ХУПС. – 2015. – №10(135). – С. 152-155.

3. Шабанова-Кушнарєнко Л. В. Построение предикатной аксиоматической модели метрики на пространстве прецедентов / Л. В. Шабанова-Кушнарєнко // Системи обробки інформації. – Харків : ХУПС. – 2015. – №12(137). – С. 93-96.

4. Шабанова-Кушнарєнко Л. В. Конфигурирование иерархической модели представления процессных знаний на основе анализа лог-файлов / Л. В. Шабанова-

Кушнарєнко, С. Ф. Чалый // Уральский Научный Вестник. – Уральск : «Фирма Сервер + ». – 2015. – № 23(154). – С. 62-66.

Здобувачем запропоновано метод конфігурування ієрархічної моделі процесних знань на основі аналізу лог-файлів.

5. Шабанова-Кушнарєнко Л. В. Построение структуры линейного пространства для предикатной модели метрики / Л. В. Шабанова-Кушнарєнко // Системи обробки інформації. – Харків : ХУПС. – 2016. – №1(138). – С. 118-121.

6. Шабанова-Кушнарєнко Л. В. Разработка гибридной модели формирования координирующих решений при реализации производственного процесса на приборостроительных предприятиях / Л. В. Шабанова-Кушнарєнко, А. П. Собчак, И. В. Шостак // Технологічний аудит та резерви виробництва. – Харків : НВП ПП «Технологічний Центр». – 2016. – № 2/6(28). – С. 18-24.

Здобувач розробила модель типового фрагменту СППР при розробці складальних технологічних процесів.

7. Шабанова-Кушнарєнко Л. В. Построение процессно-ориентированных моделей с использованием реляционных сетей / Л. В. Шабанова-Кушнарєнко, С. Ф. Чалый // Материалы Международной научно-технической конференции «Информационные системы и технологии» ИСТ-2012. – Харьков : ХНУРЭ. – 2012. – С. 89-105.

Здобувачем запропоновано математичну модель процесних знань у вигляді реляційних мереж.

8. Шабанова-Кушнарєнко Л. В. Применение продукционной модели знаний в системе ограничений для процессов преобразования ресурсов сетей / Л. В. Шабанова-Кушнарєнко, И. Б. Буцукина, С. Ф. Чалый // Матеріали VIII Міжнародної науково-практичної конференції «Військова освіта та наука: сьогодення та майбутнє». – Київ : ВІКНУ. – 2012. – С. 31.

Здобувач розробила продукційну модель процесів перетворення інформаційних ресурсів.

9. Шабанова-Кушнарєнко Л. В. Многоуровневый подход к выявлению неявных знаний / Л. В. Шабанова-Кушнарєнко, С. Ф. Чалый // Матеріали V Міжнародної науково-технічної конференції «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління». – Харків : ХУПС. – 2015. – С. 8.

Здобувачем запропонована предикатна модель ієрархічної структури процесних знань для виявлення неявних знань.

10. Шабанова-Кушнарєнко Л. В. Рекурсивный метод построения иерархической предикатной модели процесса на основе предикатного дерева / Л. В. Шабанова-Кушнарєнко // Материалы XX Международного конгресса двигателестроителей. – Харьков : НАУ «ХАИ». – 2015. – С. 83.

11. Шабанова-Кушнарєнко Л. В. Алгебро-логічне моделювання процесних знань в інформаційних системах / Л. В. Шабанова-Кушнарєнко, І. В. Шостак // Матеріали V Всеукраїнської науково-практичної конференції «Інтелектуальні системи і прикладна лінгвістика». – Харків : НТУ «ХПІ». – 2016. – С. 15-16.

Здобувачем запропоновано метод адаптації ієрархічної моделі процесних знань до заданої предметної області.

АНОТАЦІЇ

Шабанова-Кушнарєнко Л. В. Предикатні моделі, методи та технологія обробки процесних знань в інформаційних системах. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Харківський національний університет радіоелектроніки Міністерства освіти і науки України, Харків, 2016.

Мета дисертаційного дослідження – розробка моделей процесних знань на основі аналізу лог-файлів з використанням апарату АСП і реляційних мереж, а також технологій Process Mining і виведення знань на прецедентах. Основні результати: вдосконалена предикатна модель метрики на просторі прецедентів, яка допускає практичну перевірку її існування за допомогою системи умов та введення евклідової метрики на лінеаризованому просторі атрибутів. Вперше розроблено метод оцінки близькості прецедентів, заснований на табулюванні шкал атрибутів та привласненні атрибутам ваг з урахуванням функціональної залежності ваг від атрибутів прецедентів, що дозволяє значно підвищити точність оцінки близькості прецедентів завдяки врахуванню зростання реальної ваги атрибутів в їх критичних діапазонах. Отримав подальший розвиток метод побудови предикатної ієрархічної моделі процесних знань у вигляді реляційної метамережі, яка заснована на моделі «дані-інформація-знання-метазнання». Ієрархічне представлення спрощує конфігурацію моделі процесу і доповнює її неявними зв'язками. Удосконалено метод конфігурування ієрархічної моделі процесних знань на основі аналізу лог-файлів, яка має вигляд ієрархічної системи бінарних предикатів, що задають логіку поведінки процесу. Метод включає налаштування під предметну область, що підвищує ефективність методів Process Mining.

Ключові слова: інформаційна технологія, логічне виведення знань, алгебра скінченних предикатів, предикатна модель, процесні знання, метрика на прецедентах, ієрархічна модель процесних знань.

Шабанова-Кушнарєнко Л. В. Предикатные модели, методы и технология обработки процессных знаний в информационных системах. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Харьковский национальный университет радиоэлектроники Министерства образования и науки Украины, Харьков, 2016.

Цель диссертационного исследования – разработка моделей процессных знаний на основе анализа лог-файлов с использованием аппарата АКП и реляционных сетей, а также технологий Process Mining и вывода знаний на прецедентах, способных повысить эффективность решения задач анализа

темпоральных знаний в интеллектуальных системах.

Проведен анализ существующих методов интеллектуального анализа процессов и метода получения решений, основанного на прецедентах. Исследованы основные особенности и структурные элементы процессной модели знаний. Показано, что преимущества данной модели связаны с возможностями конфигурирования, а также иерархического представления процесса. Показаны преимущества аппарата АКП как средства выделения структуры процессной модели знаний. Показана актуальность разработки средств моделирования процессных знаний в ИС.

Усовершенствована предикатная модель метрики на пространстве прецедентов, которая допускает практическую проверку ее существования с помощью системы условий и введение евклидовой метрики на линеаризованном пространстве атрибутов. Впервые разработан метод оценки близости прецедентов, основанный на табулировании шкал атрибутов и присвоении атрибутам весов с учетом функциональной зависимости весов от атрибутов прецедентов, что позволяет значительно повысить точность оценки близости прецедентов за счет учета роста реального веса атрибутов в их критических диапазонах. Получил дальнейшее развитие метод построения предикатной иерархической модели процессных знаний в виде реляционной метасети, которая основана на модели «данные-информация-знания-метазнания». Иерархическая структура упрощает конфигурацию модели процесса и дополняет ее неявными связями. Усовершенствован метод конфигурации иерархической модели процессных знаний на основе анализа лог-файлов, которая имеет вид иерархической системы бинарных предикатов, задающих логику поведения процесса. Метод включает настройку под предметную область, что повышает эффективность методов интеллектуального анализа процессов.

Результаты диссертационной работы практически применены в Центральной клинической больнице Укрзалізничці г. Харькова при разработке программно-алгоритмической части и наполнении базы знаний программно-аппаратного комплекса автоматизированной медицинской диагностики заболеваний. Теоретические результаты диссертации были использованы в учебном процессе на кафедре интеллектуальных компьютерных систем НТУ «ХПИ» и на кафедре программной инженерии ХНУРЭ.

Ключевые слова: информационная технология, логический вывод знаний, алгебра конечных предикатов, предикатная модель, процессные знания, метрика на прецедентах, иерархическая модель процессных знаний.

Shabanova-Kushnarenko L. V. Predicate models, methods and technology of process knowledge treatment in informative systems. – As Manuscript.

Thesis for a candidate degree in technical sciences, specialty 05.13.06 – Information Technologies – Kharkiv National University of Radio Electronics of the Ministry of Education and Science of Ukraine, Kharkiv, 2016.

Improved predicate model metric in the space of precedents, which, unlike the existing ones, allows practical test of its existence with the help of conditions and the introduction of the Euclidean metric in the space of linearized attributes. For the first time

developed a method for estimating the proximity of precedents based on the tabulation attributes scales and the attribution of weights based on the functional dependence of the weights of the attributes of precedents that can significantly improve the accuracy of estimates of the proximity of precedents by taking into account the weight of the real growth of the attributes in their critical bands. It has continued to develop a technique of constructing a predicate hierarchical model of process knowledge in the form of a relational metanet, which is based on the model of "data-information-knowledge-meta-knowledge." The hierarchical structure simplifies the configuration process model and complements its implicit connections. Improved methods of hierarchical model configuration process knowledge based on the analysis of log files, which, unlike the existing ones, has the form of a hierarchical system of binary predicates that define the behavior of the logic of the process. The method includes tuning a subject area, which increases the efficiency of Process Mining techniques.

Keywords: information technology, knowledge inference, algebra of finite predicate, the predicate model, process knowledge, metric on precedents, hierarchical model of process knowledge.



Підп. до друку 20.05.16 р. Формат 60×90 1/16
Папір офсетний. Друк – різнографія. Гарнітура Times New Roman.
Умовн. друк. арк. 0,9. Наклад. 100 прим. Зам. № 066832

Надруковано у ТОВ «Планета-Принт».
ЄДРПОУ № 31235131 від 19.12.02 р.
61002, м. Харків, вул. Фрунзе, 16