

УДК 004.5
МПК 6 G11
№ держреєстрації 0114U002060
Інв. №

Національна академія наук України
Інститут проблем реєстрації інформації
(ІПРІ НАН України)

03113, м. Київ, вул. М. Шпака, 2; тел. (044) 454 21 51;
E-mail: vvp@ipri.kiev.ua

ЗАТВЕРДЖУЮ
Директор ІПРІ НАН України
академік НАН України

_____ В.В.Петров
31.12. 2018 р.

ЗВІТ

ПРО НАУКОВО-ДОСЛІДНУ РОБОТУ
Дослідити та розробити технології комп'ютерного моделювання
систем керування знаннями при проведенні аналітичної
діяльності («УЗОР-2014»)

(Заключний)

Науковий керівник НДР
завідуючий відділом
д-р техн. наук, професор,
Заслужений діяч науки і техніки України _____ О.Г.Додонов

Рукопис закінчено 26 грудня 2018 року.
Результати роботи розглянуті та схвалені Вченою Радою ІПРІ НАН України,
протокол № 14 від 04.12. 2018 р

Усі примірники звіту ідентичні за змістом

РЕФЕРАТ

Звіт про НДР (заключний): С. - 452, рис. 85, 290 джерел, табл. 31, 2 додатки – 2.

Тема: «Дослідити та розробити технології комп'ютерного моделювання систем керування знаннями при проведенні аналітичної діяльності («УЗОР-2014»).

Державний реєстраційний номер: 0114U002060

Мета роботи: – дослідження та розробка технологій комп'ютерного моделювання систем керування знаннями при проведенні аналітичної діяльності, у тому числі з обробкою стратегічно великих масивів даних (Big Data в розподіленому середовищі), розробка методології побудови системи керування знаннями при проведенні аналітичної діяльності, включаючи Web-орієнтовані системи.

Об'єкт дослідження: – системи керування знаннями в інформаційно-аналітичних системах для різних предметних областей, включаючи й Web-орієнтовані системи

Методи дослідження: Для досягнення поставленої мети в роботі були використані: системний аналіз, сценарний аналіз, методи семантичного моделювання, онтології, мови опису онтологій OWL, RDF, DL, SPARQL, уніфікованої моделі системи подання знань, теорія комп'ютерних інформаційних систем, методи теорії графів, методи математичного моделювання, методи та технології аналітичної обробки даних.

В результаті виконання НДР розроблено теоретичні та технологічні засади комп'ютерного моделювання систем керування знаннями при проведенні аналітичної діяльності, включаючи й Web-системи, зокрема:

- На підставі аналізу сучасних підходів та методів управління знаннями сформульовано цілі та задачі комп'ютерного моделювання систем керування знаннями. Проведено аналіз існуючих моделей опису знань для різних сфер аналітичної діяльності. Показано, що при створенні прикладних аналітичних систем необхідно об'єднання різних методів опису знань, оскільки жодний окремий метод не може забезпечити у повному обсязі опис моделі знань реальній аналітичній системи.
- Запропоновано та розроблено методологію побудови комп'ютерної інтегрованої моделі подання знань, яка враховує не тільки методи подання знань за різними моделями (онтологічної, мережевий, фреймової, продукційної та інших), але й засоби формування логічних правил, процесів та типових сценаріїв виведення нових знань та їх реалізації в інструментальному середовищі.

- Розроблено та реалізовано технологію комп'ютерного моделювання аналітичної діяльності на базі інтеграції сценарно-цільового підходу щодо організації аналітичної діяльності і системи керування знаннями, у тому числі при обробці Big Data. Технологія базується на застосуванні сучасних платформ: мови OWL 2, SPARQL, SWRL, редактора онтології Protégé 5, конверторів для перетворення різноманітних форматів даних – Ontop Mapping, машин логічного виводу – Reasoner, засобів виводу по прецедентах – Jcolibri, що дозволяє отримувати та досліджувати реальні моделі аналітичної системи, заснованої на знаннях.
- Запропоновано та створено дослідно-технологічний стенд для моделювання систем керування знаннями у вигляді корпоративного Web-порталу. Стенд включає різноманітні інструментальні засоби моделювання процесів аналітичної діяльності, а також реалізує методи та застосовує інструментальні засоби семантичної обробки даних Rapid Mainer, що дозволяє не тільки моделювати процеси обробки структурованих, неструктурованих, текстових, соціальних типів даних та знаходити оптимальні рішення для аналітичних додатків, а й отримувати практичні результати у вигляді реального програмного коду для аналітичних застосунків.

Отримані результати є теоретичним та експериментальним підґрунтям для побудови інформаційних технологій управління знаннями сучасних аналітичних систем різних предметних сфер в Україні.

За результатами НДР «УЗОР-2014» надруковано 42 печатні роботи, включаючи монографії, виступи на міжнародних конференціях та фахові видання.

КЛЮЧОВІ СЛОВА: СИСТЕМИ КЕРУВАННЯ ЗНАННЯМИ, МОДЕЛЬ ПОДАННЯ ЗНАНЬ, ІНФОРМАЦІЙНО-АНАЛІТИЧНІ СИСТЕМИ, КОМП'ЮТЕРНІ МОДЕЛІ, АНАЛІТИЧНА ДІЯЛЬНІСТЬ, БАЗИ ЗНАНЬ, ОНТОЛОГІЯ, BIG DATA.

ЗМІСТ

СПИСОК АВТОРІВ 2

РЕФЕРАТ 4

ЗМІСТ 6

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ ТА ПОНЯТЬ 10

ВСТУП 14

1.	АНАЛІТИЧНА ДІЯЛЬНІСТЬ ЯК ПРОЦЕС ОБРОБКИ ЗНАНЬ	17
1.1	Сутність аналітичної діяльності та її основні принципи	17
1.2	Інформаційна аналітика як засіб одержання нових знань	20
1.3	Системний аналіз як базова методологія аналітичної діяльності	26
1.4	Форми звітності як результат аналітичної діяльності	32
1.5	Управління знаннями в процесі здійснення аналітичної діяльності	39
1.6	Класифікація знань при здійсненні аналітичної діяльності	44
1.7	Властивості та інтелектуальна цінність знань	48
1.8	Особливості комп'ютерного моделювання систем керування знаннями при здійсненні аналітичної діяльності	51
1.8.1	Цілі та основні етапи моделювання систем керування знаннями	51
1.8.2	Методологічні принципи моделювання аналітичної діяльності, заснованої на управлінні знаннями	53
1.8.3	Огляд методології та засобів комп'ютерного моделювання і проектування програмних застосувань	55
1.8.4	Оцінка можливості CASE-технологій для моделювання знань	65
2.	АНАЛІЗ СУЧАСНИХ МЕТОДІВ ТА ЗАСОБІВ ФОРМАЛЬНОГО ОПИСУ ЗНАНЬ	68
2.1	Класифікація методів формального опису знань	68
2.2	Метод опису знань у вигляді логічної моделі	70
2.3	Метод опису знань на базі продукційної моделі	72
2.4	Фреймова модель подання знань	75
2.5	Метод опису знань з використанням семантичних мереж	79
2.6	Онтологічна модель знань	82
2.7	Формальна модель нечіткої онтології та бази знань на лінгвістичних змінних	84
2.8	Об'єктно-орієнтована модель знань	87
2.9	Критерії оцінювання вибору моделі опису знань	90
3.	АНАЛІЗ ТА ДОСЛІДЖЕННЯ МЕТОДІВ ОЦІНКИ ДОСТОВІРНОСТІ ТА ВЕРИФІКАЦІЇ ЗНАНЬ	95
3.1	Важіль достовірності та інформативності знань при здійсненні аналітичної діяльності	95
3.2	Класифікація джерел інформації в контексті аналітичної діяльності та	96
3.3	Методи оцінки достовірності та інформативності джерел	102
3.4	Методика відбору та оцінки неформалізованих знань за критерієм цінності	107
3.5	Верифікація знань методом встановлення коефіцієнтів важливості базових понять	109
3.6	Метод семантичної верифікації знань для Web-контенту	114
3.7	Метод автоматичної ідентифікації та корекції типових помилок у процесі здійснення аналітичної діяльності	120
3.8	Аудит знань	124
4.	ДОСЛІДЖЕННЯ ТА РОЗРОБКА МЕТОДІВ ПОБУДОВИ МОДЕЛЕЙ БАЗ ЗНАНЬ ДЛЯ ВИКОНАННЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ	127
4.1	Взаємозв'язок між базами знань та моделями представлення даних при побудові аналітичних систем	127
4.1.1	Модель подання структурованих даних	128
4.1.2	Інтерактивна аналітична обробка на основі моделі багатовимірних даних	129
4.1.3	Використання метаданих для опису неструктурованих даних	131
4.1.4	Модель представлення слабоструктурованих даних мовою XML-	133

4.1.5	Графова модель подання інформації та знань	136
4.2	Використання семантичного Web для формування моделі знань	138
4.3	Аналіз сучасних інструментальних засобів для побудови моделі знань на основі онтології	145
4.4	Інструментальна платформа для дослідження графових моделей – Gephi	150
4.5	Інтегрована модель опису знань для здійснення аналітичної діяльності	153
4.6	Портали знань як засоби об'єднання різноманітної інформації в аналітичних системах	160
4.6.1	Модель функціонування порталу знань на основі сценарного підходу	162
4.6.2	Метод динамічного формування сценаріїв на порталі знань	166
5.	АНАЛІЗ ТА ДОСЛІДЖЕННЯ РОЗШИРЕНОГО ПОШУКУ ДЛЯ ЗДІЙСНЕННЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ	169
5.1	Процес здобуття нових знань з сирих даних	169
5.2	Розширений пошук інформації та знань на основі семантичній моделі	172
5.3	Пошук неявних знань, заснованих на нейронних мережах	176
5.4	Мультиагентна модель пошуку інформації та знань у Web просторі	181
5.5	Проблеми витягу нечитких знань	186
5.6	Інформаційні технології підтримки процесів пошуку та доступу до джерел знань	188
5.6.1	Технологія пошуку інформації по ключових словах	189
5.6.2	Застосування технологій Web 2.0 для роботи зі знаннями	192
5.6.3	Технології семантичного Web в управлінні знаннями	194
5.7	Формування сценаріїв роботи аналітика в Web-середовищі	197
6.	ДОСЛІДЖЕННЯ І РОЗРОБКА МЕТОДІВ ПОБУДОВИ МОДЕЛЕЙ УПРАВЛІННЯ ЗНАННЯМИ ДЛЯ ЗДІЙСНЕННЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ	202
6.1	Проблеми управління знаннями та підходи до їх вирішення	202
6.2	Підходи до управління знаннями при здійсненні аналітичної діяльності	210
6.3	Аналіз ризиків та стратегії впровадження систем управління знаннями	214
6.4	Підвищення ефективності систем управління знаннями при здійсненні аналітичної діяльності	219
6.5	Висновки	222
7.	РОЗРОБКА МЕТОДОЛОГІЇ ПОБУДОВИ СИСТЕМ УПРАВЛІННЯ ЗНАННЯМИ НА ОСНОВІ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ	224
7.1	Особливості семантичного моделювання сучасних систем управління знаннями	224
7.2	Методологія побудови систем управління знаннями на основі семантичного моделювання	226
7.3	Визначення глибини і масштабу моделі предметної області	230
7.4	Методологія побудови концептуальної моделі онтології предметної області	234
7.4.1	Побудова концептуальної моделі	234
7.4.2	Формування ієрархії основних класів – таксономії предметної області	236
7.4.3	Формування підмножини властивостей для семантичного опису понять предметної області	241
7.4.4	Формування примірників бази знань із встановлення числових або логічних обмежень	242
7.5	Встановлення відносин між концептами предметної області та їх властивостями	245
7.5.1	Теоретичні положення OWL 2 та механізм зв'язування понять моделі предметної області	245
7.5.2	Визначення словника дієслів для формування предикатів	246
7.5.3	Формування відношень між екземплярами класів	248
7.5.4	Зв'язування понять онтології та формування аксіом	249
7.6	Формування графічного відображення концептуальної моделі онтології предметної області	251
7.7	Дослідження ефективності мов логічного висновку при здобутті нових знань	255
7.7.1	Дескриптивна логіка як формальна семантика моделі онтології	255
7.7.2	Розширення виразності мови OWL 2	258

7.7.3	Базові положення мови запитів SPARQL	261
7.7.4	Механізми формування SPARQL- запитів до бази знань	264
7.7.5	Механізми формування запитів до бази знань на основі синтаксиса Manchester OWL	266
7.7.6	Порівняльний аналіз мов запитів до бази знань за критерієм зручності –usability	271
7.8	Оцінка якості моделі онтології та бази знань	275
7.9	Дослідження методів та технологій інтеграції онтологічної моделі з реляційними даними	280
7.10	Моделювання аналітичної діяльності на основі отології та прецедентів	291
8.	МЕТОДИ ТА ТЕХНОЛОГІЇ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ ДЛЯ ВИТЯГУ НОВИХ ЗНАНЬ	297
8.1	Визначення великих даних – Big Data	297
8.2	Особливості аналізу Big Data в аналітичних системах при отриманні нових знань	298
8.3	Теоретичні основи аналізу великих обсягів даних методами Data mining	301
8.3.1	Основні задачі Data Mining	302
8.3.2	Ітераційний процес здійснення Data Mining	305
8.3.3	Навчання як невід’ємний процес аналізу даних методами Data Mining	308
8.4	Аналіз факторів, що затримують запровадження засобів обробки Big Data в аналітичних системах	312
8.5	Моделювання процесів витягу знань з великих даних	314
8.5.1	Візуальний метод побудови моделей аналізу даних з метою витягу знань	314
8.5.2	Підходи до вирішення завдань отримання знань методами кластерного аналізу	319
8.5.3	Моделювання процесу групування даних методами кластеризації в середовищі RapidMiner	322
8.5.4	Побудова моделі прогнозування в середовищі RapidMiner	330
8.5.5	Дослідження технології аналізу великих обсягів даних засобами XLMiner	333
8.5.6	Мова розмітки для прогнозного моделювання та інтелектуального аналізу неструктурованих даних	336
8.6	Застосування методів нечіткої кластеризації при вирішенні задач класифікації	338
8.7	Імітаційна оптимізація в задачах аналізу ризиків мережевих моделей	342
8.7.1	Методологія ризик-аналізу в середовищі Analytic Solver Platform	344
8.7.2	Оптимізаційне моделювання мережевих структур, що потребують реконфігурації	349
8.8	Метод та сценарна модель аналізу багатовимірних даних на прикладі моніторингу виконання бюджетного процесу	355
9.	ДОСЛІДЖЕННЯ МЕТОДІВ І РОЗРОБКА ТЕХНОЛОГІЇ ФОРМУВАННЯ КОМП’ЮТЕРНОГО СЕРЕДОВИЩА ДЛЯ МОДЕЛЮВАННЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ ІЗ СИСТЕМОЮ УПРАВЛІННЯ ЗНАННЯМИ	367
9.1	Огляд сучасних технологій формування комп’ютерного середовища для моделювання аналітичної діяльності	367
9.1.1	Формування комп’ютерного середовища для моделювання аналітичної діяльності на базі Cloud-технології	367
9.1.2	Технологія розподілених обчислень MapReduce	368
9.1.3	Використання шаблонів проектування Model-view-controller для прискорення розробки програмних застосунків	372
9.1.4	Технологія django python для швидкої розробки Web-додатків	374
9.1.5	Технологічні засоби конфігурування та адміністрування розподілених компонентів як інструменти формування елементів сценаріїв аналітичної діяльності	376
9.2	Прототип інструментального засобу для моделювання систем керування знаннями при проведенні аналітичної діяльності	379
9.2.1	Особливості організації інструментальних засобів моделювання на базі корпоративних Web-порталів	380
9.2.2	Архітектура інтегрованої платформи моделювання систем керування знаннями при проведенні аналітичної діяльності	382

9.2.3	Модель користувача інтегрованої платформи моделювання систем керування знаннями	389
9.2.4	Інтелектуальний інтерфейс взаємодії системи моделювання з користувачем	393
9.3	Дослідження методів і розробка технології адміністрування та управління системою керування знаннями в сучасних умовах	395
9.3.1	Дослідження моделей управління контентом корпоративного Web-порталу	396
9.3.2	Система управління знаннями інтегрованої платформи інтелектуального аналізу даних на основі корпоративного Web-порталу	401
9.3.3	Управління знаннями корпоративного Web-порталу за технологією Joomla	403
9.3.4	Дослідження інформаційних загроз в роботі корпоративного Web-порталу	406
9.3.5	Адміністрування корпоративного Web-порталу знань за технологією SharePoint	409
	ВИСНОВКИ	412
	ПЕРЕЛІК ПОСИЛАНЬ	415
	ДОДАТОК 1.	429
	Порівняльна таблиця сучасних інструментальних засобів семантичного моделювання	429
	ДОДАТОК 2.	437
	Огляд інструментальних засобів Data Mining з відкритим кодом	437