

**Харківський національний університет радіоелектроніки**

**Головянко Марія Валентинівна**



**УДК 004.89:004.415.5**

**МЕТОДИ І МОДЕЛЬ ВЕРИФІКАЦІЇ ЗНАНЬ ДЛЯ  
ІНТЕЛЕКТУАЛІЗАЦІЇ WEB-КОНТЕНТУ**

**05.13.23 – системи та засоби штучного інтелекту**

**Автореферат  
дисертації на здобуття наукового ступеня  
кандидата технічних наук**

**Харків – 2010**

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки, Міністерство освіти і науки України.

**Науковий керівник** доктор технічних наук, професор Терзіян Ваган Якович, Харківський національний університет радіоелектроніки, професор кафедри штучного інтелекту.

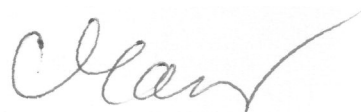
**Офіційні опоненти:** доктор технічних наук, професор Шабанов-Кушнарєнко Юрій Петрович, Харківський національний університет радіоелектроніки, професор кафедри програмного забезпечення ЕОМ, м. Харків; доктор технічних наук, професор Серков Олександр Анатолійович, Національний технічний університет «Харківський політехнічний інститут», професор, завідувач кафедри систем інформації, м. Харків.

Захист відбудеться «02» березня 2011 р. о 15<sup>00</sup> годині на засіданні Спеціалізованої вченої ради Д 64.052.01 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

З дисертацією можна ознайомитися у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

Автореферат розісланий «26» січня 2011 р.

Вчений секретар  
спеціалізованої вченої  
ради



С.Ф. Чалий

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

### **Актуальність теми.**

В умовах зростання популярності Web-орієнтованих систем, зокрема таких, що застосовуються для прийняття рішень, особливої важливості набувають програмні інтелектуальні засоби автоматичного вилучення релевантної інформації, достовірність якої можна формально перевірити або верифікувати.

Розроблені дотепер підходи до верифікації знань використовують переважно статистичні та синтаксичні методи перевірки цілісності та несуперечливості розміщених у окремії базі знань, що обумовлює нездатність існуючих методів верифікації впоратись із різномірністю, неструктурованістю та суперечливістю інформації, поданої у Web.

У зв'язку з цим актуальною є задача розробки моделі та методів семантичної верифікації знань на основі інформації, яка може вилучатись з розподілених різномірних джерел, що дозволить інтелектуалізувати Web-контент, доповнивши зміст Web-систем такими формальними описами, які нададуть можливість інтелектуальним програмам або агентам автоматично переконуватися у достовірності в певному контексті наданої системою інформації. Напрямок досліджень відповідає положенням задекларованої World Wide Web консорціумом ініціативи «Semantic Web», та покликаний розв'язати задачу підняття рівня довіри до Web-ресурсів.

В Україні стрімкі темпи науково-технічного прогресу є причиною особливої потреби в інтелектуалізації інформаційних систем таких галузей, як освіта і наука, які потребують новітніх технологій безперервного накопичення та обробки достовірної інформації для вирішення задач менеджменту різномірних ресурсів. Аналогічні проблеми є актуальними і для інших галузей, для яких характерними є великі обсяги неструктурованої інформації.

### **Зв'язок роботи з науковими програмами, планами, темами.**

Дисертаційна робота виконувалась у Харківському національному університеті радіоелектроніки на кафедрі штучного інтелекту відповідно до плану науково-дослідних робіт університету в межах держбюджетних тем №219 – «Розробка Web-орієнтованої системи для підтримки процедур акредитації та ліцензування вищих навчальних закладів України» (№ ДР0108U010139), №233 – «Розробка системи підтримки семантичних запитів до онтологічної бази акредитації і ліцензування» (№ ДР0109U001647), в яких здобувач Головянко М.В. взяла участь як виконавець. В межах зазначених тем здобувачкою як виконавцем були розроблені архітектура онтологічної розподіленої системи із верифікаційними компонентами, методи верифікації онтологічних знань для забезпечення цілісності, несуперечливості та коректності інформації, яка вилучається з розподілених гетерогенних джерел, нова модель подання онтології за рахунок додавання верифікаційної складової до онтологічних структур.

Як основи технічних механізмів, що розроблялись в межах дисертаційної роботи, було використано чисельні розробки та дослідження консорціуму «The World Wide Web Consortium», що проводяться в галузі створення інтелектуальних систем менеджменту різномірних ресурсів.

**Мета та задачі дослідження.** Метою дослідження є розробка моделі та методів верифікації знань, що вилучаються з розподілених різномірних джерел,

шляхом доведення їх достовірності згідно з концепцією Semantic Web для інтелектуалізації Web-контенту.

Досягнення поставленої мети здійснюється розв'язанням таких основних задач:

- аналіз методів верифікації Web-контенту та можливості застосування онтологічної моделі подання знань для їх верифікації;
- вдосконалення онтологічної моделі подання знань шляхом розробки верифікаційної складової моделі;
- розробка методу семантичної верифікації знань шляхом доведення достовірності автоматичного гібридного виведення знань на основі інформації, яку отримано з розподілених різнорідних джерел;
- розробка методу інтерпретації верифікаційних елементів;
- вдосконалення моделі інформаційних інтелектуальних систем верифікаційними компонентами;
- розробка структури та опис функцій програмних засобів семантичної верифікації на основі розроблених методів та моделей.

*Об'єктом дослідження* є процес верифікації знань в розподілених інформаційних інтелектуальних системах.

*Предметом дослідження* є методи та моделі верифікації знань для інтелектуалізації Web-контенту.

*Методи дослідження.* Дослідження, проведені в роботі, ґрунтуються на комплексному використанні теорії систем, теорії штучного інтелекту, теорії графів, формалізмів дескрипційних логік, що базуються на численні предикатів першого порядку, Web-технологій (зокрема, технологій Semantic Web) згідно з провідними дослідженнями консорціуму «The World Wide Web Consortium».

**Наукова новизна роботи.** В процесі розв'язання поставлених задач отримано такі результати:

1. Вперше розроблено метод семантичної верифікації знань, який полягає в обґрунтуванні достовірності знань, поданих у Web-просторі за допомогою онтологічної моделі, на основі аналізу інформації, яка автоматично вилучається з розподілених різнорідних джерел, що дозволило підвищити рівень довіри до Web-контенту.

2. Вперше розроблено метод інтерпретації верифікаційних елементів, який полягає у визначенні еквівалентності деревоподібних структур з урахуванням семантики вузлів. Це дозволило класифікувати онтологічні знання за способом їх виведення та віднаходити зміни в механізмах виведення знань.

3. Удосконалено модель подання онтологій, яка відрізняється від існуючих моделей, побудованих на основі дескрипційних логік, наявністю верифікаційної складової, яка розширює множину базових концептів та відношень, та додаванням верифікаційного компонента до існуючої триплетної форми опису тверджень, що дало можливість обґрунтовувати та пояснювати гібридне виведення знань.

4. Вдосконалено модель Web-орієнтованих розподілених систем менеджменту різнорідних ресурсів, яка відрізняється від існуючих наявністю верифікаційних компонентів, що дозволило забезпечити виконання семантичної верифікації контенту в системах такого типу.

**Практичне значення отриманих результатів** полягає в тому, що вдосконалену онтологічну модель подання знань та запропоновані методи

семантичної верифікації було використано для розробки Web-орієнтованої розподіленої системи менеджменту різнорідних ресурсів з засобами перевірки достовірності інформації та програмного модуля верифікації Web-контенту.

На основі розробленого в роботі методу семантичної верифікації знань, який полягає в обґрунтуванні достовірності знань, поданих у Web-просторі за допомогою онтологічної моделі, було розроблено верифікатор Web-контенту у вигляді програмного модуля, що динамічно підключається до систем перегляду ресурсів Internet і дозволяє перевіряти достовірність інформації, а також наочно відображає підсумки виконаної перевірки, який впроваджено у виробничий процес на ТОВ «КОДА» (акт впровадження від 04.01.2010 р.).

Результати дисертаційної роботи також впроваджені в освітньому процесі на кафедрі штучного інтелекту ХНУРЕ у дисципліни Експертні системи, Корпоративні портали та Web 2.0 (Wiki), міжнародний проект Tempus SM\_SCM-T020B06-2006 (UA) «Towards Transparent Ontology-Based Accreditation» («Новітня інформаційна технологія забезпечення прозорості акредитації університетів») в межах програми TEMPUS, що передбачає співробітництво в межах Європи в галузі вищої освіти, та навчальну українсько-французьку програму «European Virtual Venturing» (акти впровадження від 12.05.2009 р.).

Запропоновані методи і модель можуть бути використані для підвищення ефективності роботи багатьох Web-сервісів, спрямованих на пошук та накопичення релевантної інформації, що вилучається з різнорідних джерел, здатних автоматично перевірити достовірність отриманої інформації. Також розроблені методи і модель найбільш актуально та ефективно застосовувати у таких галузях:

- розроблення експертних систем, побудованих на основі онтологічної моделі подання знань згідно з концепцією Semantic Web;
- проектування систем, заснованих на знаннях, для підтримки процесів прийняття рішень;
- побудова великих інформаційних порталів;
- створення мультиагентних середовищ із можливістю оцінки ступеня довіри між агентами.

**Особистий внесок здобувача.** Всі результати дисертації отримані автором самостійно. В роботі [1] розглянуто основні принципи побудови інформаційних систем на основі підходу Semantic Web; в [2] розглянуто методи аналізу онтології, описаної за допомогою OWL, з метою вилучення повної інформації про онтологічні об'єкти; наукова праця [3] подає аналіз методів семантичного порівняння онтологічних об'єктів та розробку гібридного методу порівняння на основі синтаксичного аналізу об'єктів, їх властивостей, місця в онтології та семантичного околу; в [4, 5, 6, 7] описано базові моделі та методи побудови інтелектуальних Web-сайтів із можливістю верифікації контенту; в [8] автору належить опис модуля забезпечення несуперечливості та достовірності знань в архітектурі інтелектуальної системи обробки інформації та документообігу з онтологічною базою знань; [9] – автору належить опис підходу до забезпечення верифікації знань Web-орієнтованої системи менеджменту національними освітніми та науковими ресурсами; в [10, 11] описано базові методи та модель верифікації знань в онтологічних системах. В [12] автором запропоновано організацію збереження та обробки знань в розподілених сховищах для можливості подальшої їх обробки (в тому числі верифікації).

**Апробація результатів роботи.** Проводилась на 7-му, 9-му та 10-му Міжнародному молодіжному форумі “Радіоелектроніка та молоду у ХХІ столітті” (м. Харків, 2003, 2005, 2006); на Міжнародній конференції «Современные информационные системы. Проблемы и тенденции развития» (м. Харків, 2007); на Другій Міжнародній науково-технічній конференції «Комп’ютерні науки та інформаційні технології» (м. Львів, 2007); на Міжнародному ІТ-Форумі (м. Ювяскюля, 2007); на Міжнародній конференції «Научная сессия МИФИ-2009» (м. Москва, 2009) та Конференції "ИНФОТЕХ-2009" (м. Севастополь, 2009).

**Публікації.** За результатами досліджень опубліковано 12 робіт, з них 4 статті в наукових спеціалізованих виданнях, згідно з переліком ВАК України, 1 стаття в наукових збірках та 7 публікацій у збірниках праць міжнародних наукових конференцій.

**Структура та обсяг дисертаційної роботи.** Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Повний обсяг дисертації становить 163 сторінки, з них 129 сторінок основного тексту. Містить 7 додатків; 29 рисунків; 1 таблицю; список використаних джерел, що містить 168 найменувань та займає 18 сторінок.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми, сформульовано мету та задачі дослідження, розкрито наукову та практичну цінність отриманих результатів.

**Перший розділ** охоплює аналіз існуючих підходів до проблеми верифікації знань. Детально розглянуто існуючі методи та метрики верифікації знань, в тому числі для знань, що подані за допомогою онтологічної моделі, виявлено їх недоліки. Зосереджено увагу на визначенні поняття «знання» в дескрипційних логіках, які використовуються для подання онтологічної моделі, об’єктах та предметах семантичної верифікації. Розглянуто верифікацію як формальну процедуру, що виконується на кожному етапі та між етапами циклу розробки програмних систем. Виокремлено проблему верифікації баз знань з урахуванням сучасних підходів до організації розподілених Web-орієнтованих програмних систем з компонентами інтелектуальної обробки інформації. Зроблено висновок щодо доцільності розробки методу верифікації знань, які автоматично вилучаються з розподілених різнорідних джерел інформації й накопичуються базою знань інформаційної системи. Розглянуто технології Semantic Web для реалізації поставленої мети та обґрунтовано їх застосування.

Завдання дослідження полягає в розробці моделі та методів інтелектуалізації Web-контенту шляхом семантичної верифікації знань, що містяться в базі знань інформаційної системи, поданих онтологією предметної галузі шляхом доведення достовірності результатів гібридного виведення знань на основі інформації, що отримують з розподілених семантично анотованих джерел, де онтологія – це структура вигляду:

$$O := (C, \leq_C, R, \sigma, \leq_R, I \leq_I, A), \quad (1)$$

де  $C$  та  $R$  – множини, елементами яких є ідентифікатори концептів та ідентифікатори відношень,  $\leq_C$  виконує часткове впорядкування множини  $C$ ,  $\sigma$  – сигнатура,  $\leq_R$  – частково впорядкована ієрархія відношень;  $I$  – множина екземплярів;  $A$  – множина аксіом.

У **другому розділі** дослідження запропоновано вдосконалену модель подання онтологій.

Семантична верифікація – процес перевірки достовірності знань, поданих за допомогою онтологічної моделі, заснованої на дескриптивних логіках, та сформованих на основі інформації, що видобувається з розподілених різнотипних джерел, як результат гібридного виведення.

Онтологічне гібридне виведення знань – це видобуток нових знань з наявної інформації за допомогою різних типів виведення:

- виведення на знаннях, закладених у структурі онтології;
- виведення, що базується на знаннях про екземпляри онтології;
- математичне обчислення;
- виведення за SWRL-правилами;
- виведення у відповідь на SPARQL- або RDQL-запит;
- отримання знань із зовнішніх джерел (в тому числі, від Web-сервісів).

Головна мета верифікації – надання безперечних свідочств достовірності дескрипційних тверджень, які формують знання.

Запропоновано базову структуру верифікації у вигляді:

1. Теза або знання (*VerifiedKnowledge*), що подано у вигляді множини дескрипційних тверджень у формі аксіом побудови екземплярів в термінах словника, заданого *TBox*-компонентом бази знань, достовірність якого слід обґрунтувати:

$$VerifiedKnowledge := a : C, (a, b) : R, \text{ або } a \neq b, \quad (2)$$

де  $a, b \in I$  – екземпляри онтології,  $R$  – семантичне відношення між екземплярами,  $C$  – концепт онтології.

2. Аргументи (*Arguments*) – множина тверджень (аксіом і фактів), що застосовувались для виведення тези.

3. Пояснення або демонстрація (*Explanation*) – спосіб організації обґрунтованого логічного зв'язку між тезою та аргументами, демонстрація шляху семантичного гібридного виведення певного знання.

Нехай знання *VerifiedKnowledge* потребує пояснення. Для пояснення застосовуються аргументи  $Argument_1, Argument_2, \dots, Argument_n$ , такі що з цих аргументів логічно виводиться *VerifiedKnowledge*. Вважаємо *VerifiedKnowledge* об'єктом пояснення, а  $Argument_1, Argument_2, \dots, Argument_n$  – базисом пояснення, тоді:

$$\{Argument_1, Argument_2, \dots, Argument_n\} \rightarrow VerifiedKnowledge. \quad (3)$$

Доведено необхідність вдосконалення існуючої онтологічної моделі подання знань, заснованої на дескрипційних логіках, розширенням існуючої триплетної форми тверджень та додаванням до неї верифікаційного компонента (рис. 1) – елемента

підтвердження, що дозволяє зберігати метадані, необхідні для формування підтвердження достовірності знань. Елемент підтвердження – одиниця пояснення природи знань, що зберігає опис впливу на систему, який призводить до змін у базі

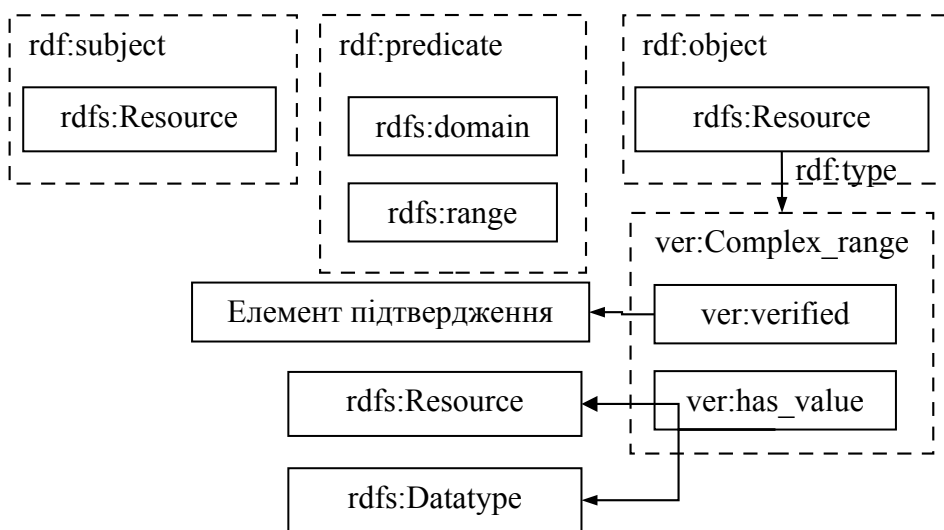


Рисунок 1 – Схема впровадження компонентів верифікації до існуючої моделі

знань. Верифікаційна модель – це онтологія, яка описується мовою подання онтологій OWL, завдяки чому є придатною для обміну між Web-сервісами та клієнтами, що застосовують RDF/XML синтаксис. Базовим для верифікаційної моделі є концепт «Дія» (*ver:Action*), що вказує на маніпуляцію, яка призвела до зміни середовища.

Згідно зі

способами формування знань дії поділяються на прості (отримання знання від явного джерела знань) та складні (виведення знання з наявних передумов шляхом застосування певного елемента виведення – правила, формули або запиту). Для кожного екземпляра класу «Дія», в залежності від його типу, визначається множина атрибутів, що описують даний екземпляр: дата (*date\_of\_action*), правило виведення (*has\_inference\_rule*), джерело дії (*has\_source*), формула, за якою відбулось обчислення (*calculated\_by\_formula*) та інші, які ставлять у відповідність даний дії множину об'єктів верифікаційної онтології. В роботі наведено результати проведеної оцінки розробленої моделі на повноту та точність.

В третьому розділі дослідження описано метод семантичної верифікації знань і метод інтерпретації верифікаційних елементів.

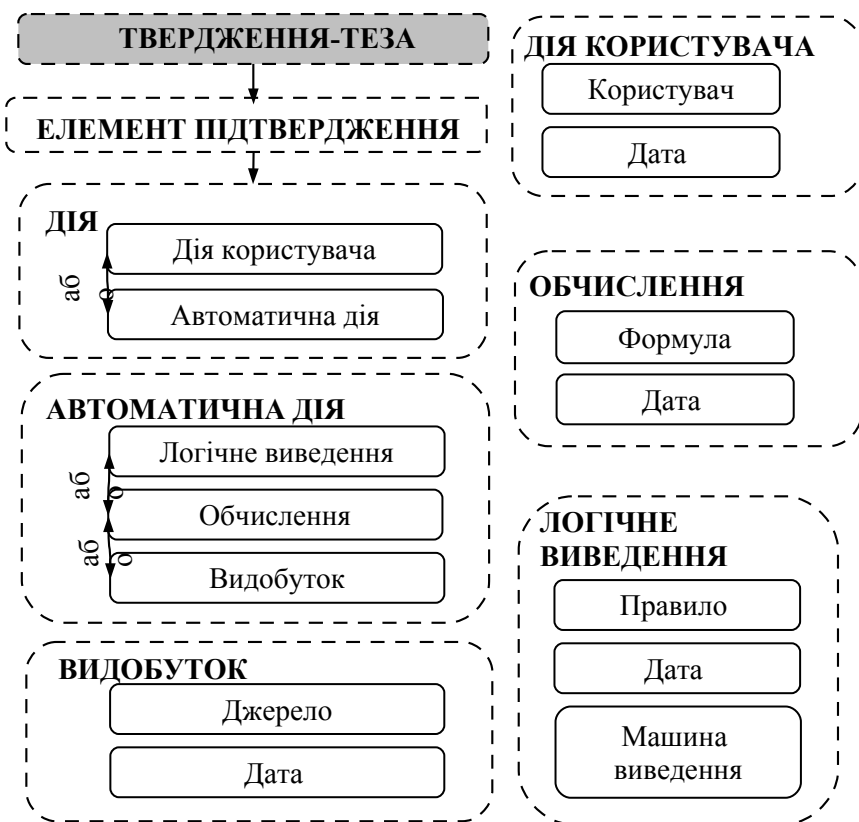


Рисунок 2 – Фрагмент структури верифікаційної моделі



Доведено, що верифікація знань може бути зведена до верифікації окремих дескрипційних тверджень.

Досліджено можливі типи змін в онтологічній базі знань, проведено їх класифікацію та запропоновано операції зміни.

Визначено онтологічну зміну як дію над онтологією, що має наслідок у вигляді онтології ( $Ont_{new}$ ), що відрізняється від оригінальної версії ( $Ont_{old}$ ).

Онтологічна операція зміни є функцією:

$$change\_operation : Ont_{old} \rightarrow Ont_{new} \quad (4)$$

де  $Ont$  – онтологія.

Дослідження можливих типів операції змін, що проводилось з метою класифікації операцій, відповідно до розробленої верифікаційної моделі, виявило три основні операції зміни: додавання певного визначення до онтології, модифікація вже існуючих онтологічних визначень та видалення визначень, які є зайвими для відповідної онтологічної бази знань. Під визначенням розуміються всі можливі компоненти онтологічної моделі: зокрема, концепти, властивості, екземпляри або онтології в цілому та їх складові, як наприклад, область визначення та область значень властивості.

Для кожного типу онтологічної зміни визначено операції – атомічні операції, базові для моделі, та складні операції зміни, які можна визначити за допомогою атомічних операцій із різною послідовністю застосування.

Дані операції є базовими для моделі:

– додавання аксіом (+). Визначимо семантику операції:

$$O + \alpha = O \cup \{ \alpha \}, \quad (5)$$

де  $\alpha$  є аксіомою, що додається;

– видалення аксіом (–):

$$O - \alpha = O \setminus \{ \alpha \}, \quad (6)$$

де  $\alpha$  є аксіомою, що видаляється.

Аксіома належить до скінченної множини онтологічних аксіом: аксіоми включення концепту  $C \subseteq D$ , аксіоми транзитивності  $Trans(R)$ , аксіоми включення ролі  $R \subseteq S$  та  $T \subseteq U$ , твердження концепту  $C(a)$ , твердження ролі  $R(a,b)$ , (не)еквівалентність екземплярів  $a = b$  та  $a \neq b$ , відповідно:

$$\alpha := C \subseteq D | Trans(R) | R \subseteq S | T \subseteq U | C(a) | R(a,b) | a = b | a \neq b, \quad (7)$$

де  $C$  та  $D$  – атомічні концепти онтології,  $R$  – абстрактна роль,  $S$  – абстрактна проста роль,  $T$  – конкретна роль,  $a$  та  $b$  – абстрактні та конкретні екземпляри онтології.

Множина базових операцій є достатньою для визначення всіх можливих змін. Для атомічних операцій зміни  $change\_operation_1, change\_operation_2, \dots, change\_operation_n$  визначаємо складні операції так:

$$\begin{aligned} change\_operation(x) &= change\_operation_n \circ \dots \circ change\_operation_1 = \\ &= change\_operation_n(\dots(change\_operation_1)(x)), \end{aligned} \quad (8)$$

де  $change\_operation_i$  – складна операція зміни.

Метод семантичної верифікації складається з етапів накопичення верифікаційних знань, ітеративного процесу, заснованого на прослуховуванні подій, та формування пояснення шляхом дедуктивного виведення у вигляді багатовимірної деревоподібної структури  $Exp = (N_{Exp}; E_{Exp})$ , де  $N_{Exp} \neq \emptyset$  – множина позначених вершин,  $E_{Exp} \subset N_{Exp} \times N_{Exp}$  – множина гілок.

У вершинах структури знаходяться множини тверджень у триплетній формі. Існує тільки один вузол  $N_{Root}$ , який є коренем структури, в якому знаходиться теза – твердження у триплетній формі, що піддається верифікації. Інші вузли, які містять аргументи з множини *Arguments* дерева містяться у  $k$  неперетинних парах множин (піддеревах)  $Exp_1, \dots, Exp_k$ , кожна з яких є орієнтованим деревом ( $k \geq 0$ ).

$$Exp = \{ \{N_{Root}\}, Exp_1, \dots, Exp_k \}. \quad (9)$$

Оскільки відносний порядок піддерев  $T_1, \dots, T_k$  є фіксованим, дерево є впорядкованим. На кожному рівні дерева  $Exp$  знаходяться множини вузлів, що мають однакову відстань до кореня дерева та описують стан предметної галузі в певний момент роботи системи. На основі тверджень, що знаходяться у вузлах структури, відбувається виведення тверджень, що знаходяться у батьківських вузлах. Гілки позначають переходи від одного стану інформаційної моделі до іншого. Відтак, у корені структури знаходиться твердження, що потребує пояснення шляху виведення, в листях – твердження, які визначають стан середовища до початкової події, яка призвела до активації виведення твердження, що потребує пояснення.

Формування пояснення твердження полягає у покроковому виконанні аналізу верифікаційного компонента бази знань, автоматичному вилученні елемента підтвердження для кожного кроку виведення твердження та розміщенні відповідного елемента в загальній структурі у відповідь на зовнішній запит (від користувача, зовнішнього Web-сервісу або прикладання).

Для інтерпретації результатів верифікації розроблено метод інтерпретації верифікаційних елементів, який полягає у визначенні еквівалентності деревоподібних структур пояснення гібридного виведення з урахуванням семантики вузлів, що характеризується застосуванням декількох незалежних оцінок подібності.

Для структури пояснення визначено часткове впорядкування  $N_1 \not\approx N_2$  вузлів, якщо  $N_1 = N_2$  або  $N_1$  є нащадком для  $N_2$ .

Дано визначення поняттю повної, синтаксичної та семантичної еквівалентності деревоподібних структур.

Дві структури пояснення  $Exp_1$  та  $Exp_2$  є повністю еквівалентними ( $Exp_1 \equiv Exp_2$ ), якщо:

1.  $|Exp_1| = |Exp_2| = n$ ;
2. Впорядкування  $Exp_{11}; Exp_{12}; \dots; Exp_{1m}$  безпосередніх підструктур  $Exp_1$  та впорядкування  $Exp_{21}; Exp_{22}; \dots; Exp_{2m}$  безпосередніх підструктур  $Exp_2$  такі, що  $Exp_{1i} \equiv Exp_{2i}$  для всіх  $1 \leq i \leq m$ ;
3.  $N_{Exp_1 j} \equiv N_{Exp_2 j}$  для всіх  $1 \leq j \leq n$ , де  $n$  – кількість вершин дерева  $Exp_1$  або  $Exp_2$ ;
4.  $E_{Exp_1 k} \equiv E_{Exp_2 k}$  для всіх  $1 \leq k \leq l$ , де  $l$  – кількість гілок дерева  $Exp_1$  або  $Exp_2$ .

Дві структури пояснення  $Exp_1$  та  $Exp_2$  є семантично еквівалентними ( $Exp_1 =_{Sem} Exp_2$ ), якщо:

1.  $N_{Exp_1 j} =_{Sem} N_{Exp_2 j}$  для всіх  $1 \leq j \leq n$ , де  $n$  – кількість вершин дерева  $Exp_1$  або  $Exp_2$ ;
2.  $E_{Exp_1 k} =_{Sem} E_{Exp_2 k}$  для всіх  $1 \leq k \leq l$ , де  $l$  – кількість гілок дерева  $Exp_1$  або  $Exp_2$ ;
3. Впорядкування  $Exp_{11} \not\equiv Exp_{12} \not\equiv \dots \not\equiv Exp_{1m}$  безпосередніх підструктур  $Exp_1$  та впорядкування  $Exp_{21} \not\equiv Exp_{22} \not\equiv \dots \not\equiv Exp_{2m}$  безпосередніх підструктур  $Exp_2$  такі, що  $Exp_{1i} =_{Sem} Exp_{2i}$  для всіх  $1 \leq i \leq m$ ;

або:

1.  $Exp_1$  та  $Exp_2$  мають однакову кількість ( $m$ ) безпосередніх підструктур.
2. Впорядкування  $Exp_{11} \not\equiv Exp_{12} \not\equiv \dots \not\equiv Exp_{1m}$  безпосередніх підструктур  $Exp_1$  та впорядкування  $Exp_{21} \not\equiv Exp_{22} \not\equiv \dots \not\equiv Exp_{2m}$  безпосередніх підструктур  $Exp_2$  такі, що  $Exp_{1i} =_{Sem} Exp_{2i}$  для всіх  $1 \leq i \leq m$ .

Під підструктурою пояснення розуміємо деревоподібну структуру  $Exp'$ , вершини та гілки якої формують підмножину вершин та гілок структури  $Exp$  так, що  $Exp' \subseteq Exp$ . Лінійне впорядкування  $Exp_1 \not\equiv Exp_2$  виконується, якщо  $|Exp_1| < |Exp_2|$  або виконуються такі умови:

1.  $|Exp_1| = |Exp_2|$ ;
2. Для всіх підструктур  $Exp_{11}; Exp_{12}; \dots; Exp_{1m}$  існує  $Exp_{1i} \not\equiv Exp_{12} \not\equiv \dots \not\equiv Exp_{1m}$ , для  $Exp_{21}; Exp_{22}; \dots; Exp_{2n}$  існує  $Exp_{2i} \not\equiv Exp_{2i+1}$  для всіх  $i$ , тоді для  $i \leq \min(m, n)$ ,  $Exp_{1i} \not\equiv Exp_{2i}$  та для всіх  $1 \leq j < i$   $Exp_{1j} \equiv Exp_{2j}$  або  $m < n$  та  $Exp_{1i} \equiv Exp_{2i}$  для всіх  $1 \leq i \leq m$ .

Ступінь семантичної еквівалентності двох структур пояснення обчислюється за формулою 10.

$$\Xi (Exp_1, Exp_2) = w \left( \frac{num(Nodes_1 \cap Nodes_2)}{num(Nodes_1 \cup Nodes_2)} + \frac{num(Edges_1 \cap Edges_2)}{num(Edges_1 \cup Edges_2)} + \frac{num(CorNodes)}{num(Nodes_1 \cup Nodes_2)} \right). \quad (10)$$

$Nodes_1, Nodes_2$  – множини вершин структур пояснення;  $Edges_1, Edges_2$  – множини гілок структур пояснення.

$CorNodes$  – множина вершин в еквівалентних підструктурах.

Для встановлення ступеня еквівалентності об'єктів вдосконалено метод порівняння онтологічних об'єктів, розроблено гібридний метод, що застосовує 6 незалежних оцінок подібності.

Міра подібності – рефлексивна та симетрична функція:

$$Sim(a, b) = \sum_{k=1}^n w_k sig_k(sim_k(a, b) - 0,5), \quad (11)$$

де  $w_k \geq 0$  – вага коефіцієнтів подібності,  $sig(x) = \frac{1}{1 - e^{-ax}}$  – сигмоїдна функція, що виконує функціональне обчислення кожного з коефіцієнтів подібності,  $sim(a, b)$  – множина синтаксичних і семантичних оцінок подібності об'єктів, що обчислюється за формулою:

$$Sim_C(\lambda, \lambda') = \frac{\sum_{(C, C') \in (\lambda, \lambda')} Sim_C(C, C')}{\max(|\lambda|, |\lambda'|)}. \quad (12)$$

Обчислення синтаксичної подібності об'єктів заснована на формулі Левенштайна:

$$Sim_{synt}(L_i, L_j) := \max(0, \frac{\min(|L_i|, |L_j|) - ed(L_i, L_j)}{\min(|L_i|, |L_j|)}) \in [0, 1] \quad (13)$$

де  $L_i, L_j$  – лексичні одиниці,  $ed(L_i, L_j)$  – лексична відстань між двома об'єктами.

Обчислення семантичної подібності об'єктів здійснюється:

- за синонімами, які визначають класи об'єктів, що вимагає наявності тезаурусів або словників, – бінарна функція  $sim_w$ ;
- за відмінними властивостями об'єктів:

$$sim_u(a, b) = w_p sim_p(a, b) + w_f sim_f(a, b) + w_a sim_a(a, b), \quad (14)$$

де  $sim_p, sim_f, sim_a$  – коефіцієнти подібності властивостей-частин, властивостей-атрибутів, властивостей-функцій;

– за семантичними відношеннями ( $sim_n$ ) між об'єктами, що будується на порівнянні семантичних оточень класів об'єктів:

$$N(a^o, r) = \{c_i^o\}, \text{ таке що } \forall d(a^o, c_i^o) \leq r, \quad (15)$$

де  $N$  – семантичне оточення класу  $a^o$ ,  $c_i^o$  – множина класів, від яких відстань  $d$  до класу  $a^o$  менше або дорівнює радіусу семантичного оточення класу  $r$ .

- за обмеженнями на об'єкти;
- за правилами, описаними для об'єктів.

Розроблений метод інтерпретації результатів верифікації дозволяє:

- створювати підтвердження достовірності знань;
- виявляти зміни в механізмах виведення знань;
- у випадку внесення змін до механізмів виведення знань у системі, перевіряти достовірність існуючих знань;
- перевіряти ідентичність формування знань.

У **четвертому розділі** запропоновано вдосконалену модель Web-орієнтованих розподілених систем менеджменту різнорідних ресурсів. Запропоновано новий підхід до інтелектуалізації Web-контенту шляхом верифікації. Введено поняття інтелектуального Web-сайта, який містить засоби автоматичного обґрунтування

достовірності інформації, що розміщено на ньому. Компоненти інформаційного контенту сайта підлягають семантичній верифікації, що реалізується шляхом аналізу верифікаційних знань, що зберігаються, як всередині самого Web-сайта, так і вилучаються із зовнішніх джерел інформації. Відтак інтелектуальний контент – це контент Web-системи, модель подання якого розширено верифікаційними компонентами, що дозволить перевірити достовірність контенту внутрішніми засобами самої системи, яка містить явне (у

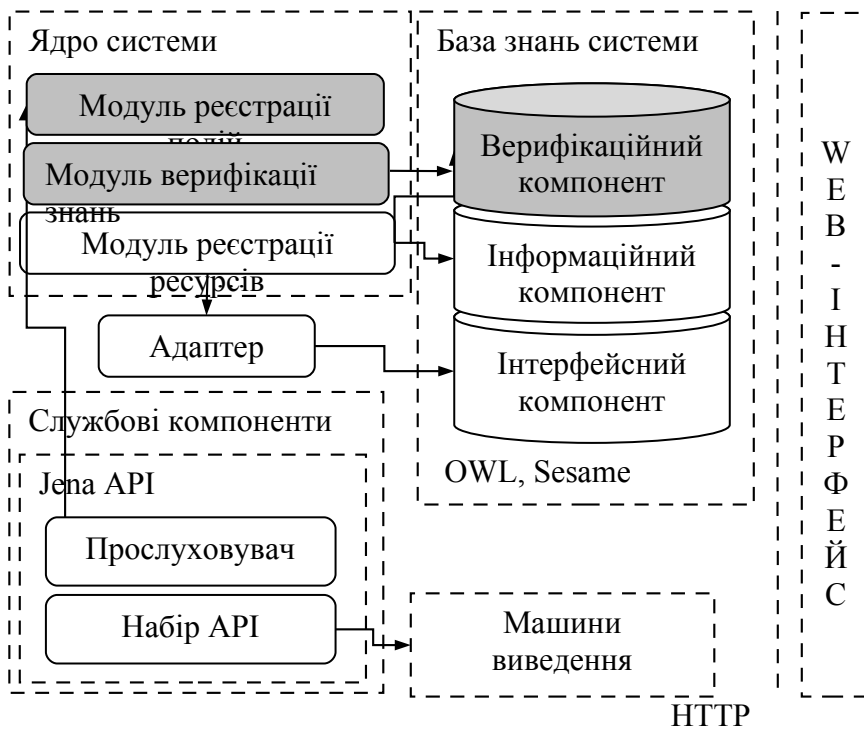


Рисунок 3 – Розширена модель Web-орієнтованих розподілених систем менеджменту різнорідних ресурсів

вигляді онтології) подання власної структури й моделі предметної галузі та реалізує збір і публікацію інформації про галузь у форматах Semantic Web.

Представлено модель інтелектуальних Web-систем з компонентами верифікації знань, зокрема, семантичних порталів, які застосовуються для менеджменту різнорідних розподілених ресурсів.

Описано структуру, взаємодію та призначення компонентів подібних систем згідно з технологіями, розробленими консорціумом World Wide Web. Розглянуто засоби автоматичної верифікації контенту інформаційних систем, які дозволять користувачам, інтелектуальним програмам або агентам пересвідчитись у достовірності наданої системою інформації. Увага в розділі також зосереджена на менеджменті верифікаційних знань в системах із описаною архітектурою.

Елементи обґрунтування зберігаються в межах самої системи, в її базі знань, розширюючи її основну структуру новим верифікаційним компонентом:

$$KB := (R, T, A, V), \quad (16)$$

де  $R$  – множина відношень,  $TBox$   $T$  – скінчена множина аксіом-включення,  $ABox$   $A$  – скінчена множина тверджень,  $V$  – множина верифікаційних тверджень, причому:

$$V := \{Argument_1, \dots, Argument_n\}. \quad (17)$$

База знань, розширена верифікаційним компонентом, набуває функції автоматичного самообґрунтування.

В роботі описано реалізацію та тестування прототипу інтелектуальної підсистеми верифікації знань, впровадженого в межах розробки інтелектуальної системи збереження та обробки інформації про організаційні, людські, інформаційні, матеріально-технічні ресурси України під управлінням Міністерства освіти – «Онтологічний портал менеджменту та оцінки освітніх і наукових ресурсів України» та у вигляді верифікатора Web-контенту, що динамічно підключається до систем перегляду ресурсів Internet та дозволяє перевіряти достовірність інформації, сформованої заснованими на знаннях Web-додатками, та наочно відображати підсумки виконаної перевірки.

У **висновках** сформульовані теоретичні та практичні результати роботи.

У **додатках** наведено акти впровадження отриманих теоретичних і прикладних результатів, приклади використання розроблених методів та моделей.

## ВИСНОВКИ

В дисертаційній роботі наведено результати, які відповідно до поставленої мети є вирішенням науково-технічної задачі розробки методів і моделей семантичної верифікації онтологічних знань для інтелектуалізації Web-контенту, що ґрунтуються на застосуванні принципів Semantic Web.

Протягом наукових досліджень отримано такі результати:

1. Внаслідок аналізу існуючих методів верифікації знань, у тому числі для знань, поданих за допомогою онтологічної моделі, виявлено недоліки, які знижують ефективність їх застосування. Так більшість методів використовують переважно

статистичну та синтаксичну перевірку розміщених в окремій базі знань, що обумовлює нездатність існуючих методів верифікації впоратись із різноманітністю, неструктурованістю та суперечливістю інформації, поданої у Web. У зв'язку із цим доцільною є розробка моделі та методів семантичної верифікації знань на основі інформації, яка може вилучатись з розподілених різноманітних джерел.

2. Вперше розроблено метод семантичної верифікації знань, який полягає в обґрунтуванні достовірності знань, поданих у Web-просторі за допомогою онтологічної моделі. Обґрунтування достовірності формується у вигляді деревоподібних структур на основі аналізу інформації, яка автоматично вилучається з розподілених різноманітних джерел. Це дозволило підвищити рівень довіри до Web-контенту.

3. Вперше розроблено метод інтерпретації елементів верифікації, що використовує визначення семантичної, синтаксичної та повної еквівалентності деревоподібних структур обґрунтування достовірності на основі декількох незалежних характеристик подібності онтологічних об'єктів, таким чином дозволяючи врахувати семантику вузлів.

4. Удосконалено модель подання онтологій, яка відрізняється від існуючих моделей, побудованих на основі дескрипційних логік, наявністю верифікаційної складової, елементи якої дозволяють описати процес накопичення та виведення нових знань. Це надало можливість обґрунтовувати достовірність знань.

5. Вдосконалено верифікаційними компонентами модель Web-орієнтованих розподілених систем менеджменту різноманітних ресурсів. Це дозволило забезпечити виконання семантичної верифікації контенту в системах такого типу.

6. Проведено апробацію запропонованих методів: методу верифікації онтологічних знань, впровадженого у програмний модуль верифікації Web-контенту, розроблений на ТОВ «Кода». Використання модуля верифікації показало практичну застосовність та ефективність моделі верифікації знань і методів побудови пояснення гібридного виведення порівняно з існуючими методиками забезпечення якості знань в інтелектуальних системах менеджменту ресурсів. Результати дисертаційної роботи впроваджені в освітній процес на кафедрі штучного інтелекту ХНУРЕ в дисциплінах: Експертні системи та Semantic Web та Корпоративні портали та Web 2.0 (Wiki), Комерціалізація інтелектуальних систем, що викладається в межах міжнародної українсько-французької студентської навчальної програми «European Virtual Venturing», а також в міжнародний проект Tempus SM\_SCM-T020B06-2006 (UA) «Towards Transparent Ontology-Based Accreditation» ("Новітня інформаційна технологія забезпечення прозорості акредитації університетів») в межах програми TEMPUS, що передбачає співробітництво в межах Європи в галузі вищої освіти (акти впровадження від 12.05.2010).

7. Розроблені моделі та методи можуть бути ефективно використані для розроблення принципів, методів та архітектурних розв'язань побудови експертних систем на основі онтологічної моделі подання знань згідно з концепцією Semantic Web, проектування систем, заснованих на знаннях, для підтримки процесів прийняття рішень, побудови великих інформаційних порталів, створення мультиагентних середовищ із можливістю оцінки ступеня довіри між агентами.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Климова М.В. Метод забезпечення прозорості інтелектуального виведення знань в онтологічних системах / М.В. Климова // Східно-Європейський журнал передових технологій. – 2008. – Вип. 6/2(36). – С. 8–12.
2. Шевченко О.Ю. Метод побудови інтелектуальних систем обробки інформації та документообігу за допомогою онтологічної бази знань / О.Ю. Шевченко, М.В. Климова // Штучний інтелект. – 2009. – Вип. 2. – Донецьк. – С. 91–97
3. Климова М.В. Розробка методу та моделі верифікації знань в онтологічних системах / М.В. Климова // Східно-Європейський журнал передових технологій. – 2009. – Вип. 4/8(40). – С. 32–36.
4. Шевченко О.Ю. Онтологічна система менеджменту національними освітніми ресурсами / О.Ю. Шевченко, М.В. Климова // "Системи обробки інформації". – 2009. – Вип. 4(78). – С. 146–151.
5. Рябова Н.В. Розробка загальної апаратно-програмної архітектури розподіленої версії онтологічного порталу / Н.В. Рябова, О.Ю. Шевченко, М.В. Білоіваненко, М.В. Головянко, Н.О. Волошина, О.В. Шубкіна // Інформаційні технології в освіті. – Вип. 4. - Херсон: ХДУ. – 2009. С. 164–178.
6. Шевченко А.Ю. Применение новых подходов к построению информационных систем / А.Ю. Шевченко, М.В. Климова, М.Г. Стукаленко // Радиоэлектроника и молодежь в XXI веке: 7-й Международный молодежный форум: материалы форума. – Харьков: ХНУРЭ. – 2003. – С. 76.
7. Климова М.В. Интерпретация объектов с использованием онтологии, выраженной на языке OWL / М.В. Климова // Радиоэлектроника и молодежь в XXI веке: 9-й Международный молодежный форум: материалы форума. – Харьков: ХНУРЭ. – 2005. – С. 343.
8. Климова М.В. Разработка методов и механизмов семантического сравнения объектов / М.В. Климова // Радиоэлектроника и молодежь в XXI веке: 10-й Международный молодежный форум: материалы форума. – Харьков: ХНУРЭ. – 2006. – С. 268.
9. Климова М.В. Разработка системы контроля на основе онтологий в интеллектуальных сайтах / М.В. Климова // Вторая международная конференция «Современные информационные системы. Проблемы и тенденции развития»: тезисы докл. – Х., 2007. – С. 349.
10. Klymova M.V. Intelligent sites with the services providing transparent reasoning based on ontologies / M.V. Klymova // Друга міжнародна науково-технічна конференція «Комп'ютерні науки та інформаційні технології»: тези доповіді. – Л., 2007. – С. 229–232
11. Климова М.В. Метод формирования SWRL метаправил для обеспечения целостности и непротиворечивости знаний / М.В. Климова // Международная конференция «Научная сессия МИФИ-2009»: тезисы докл. – Москва: МИФИ. – С. 81.
12. Климова М.В. Метод верифікації знань в онтологічних системах / М.В. Климова // Конференция "ИНФОТЕХ-2009": тези доповіді. – С., 2009. – С. 21–24.



## АНОТАЦІЯ

Головянко М.В. «Методи та модель верифікації знань для інтелектуалізації Web-контенту». – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки, Харків, 2011.

Дисертаційна робота Головянко М.В. присвячена розробці методів інтелектуалізації Web-контенту на основі верифікації знань. Удосконалено модель подання онтологій, яка відрізняється від існуючої моделі наявністю верифікаційної складової, що дало можливість обґрунтувати та пояснювати гібридне виведення знань. Вперше розроблено інтелектуальний метод верифікації онтологічних знань, який полягає у доведенні коректності та поясненні автоматичного гібридного виведення знань на основі інформації, яка вилучається з розподілених різнорідних джерел із застосуванням підходів Semantic Web, що дозволило підвищити ефективність та надійність інтелектуальної обробки інформації в розподілених інформаційних системах та інтелектуалізувати Web-контент. Вперше розроблено метод інтерпретації елементів верифікації, що використовує визначення семантичної, синтаксичної та повної еквівалентності деревоподібних структур обґрунтування достовірності на основі декількох незалежних характеристик подібності онтологічних об'єктів, таким чином дозволяючи врахувати семантику вузлів.

Удосконалено архітектуру Web-орієнтованих розподілених систем менеджменту різнорідних ресурсів, яка відрізняється від існуючих наявністю верифікаційних компонентів, що дозволило забезпечити виконання семантичної верифікації контенту в системах такого типу. Запропоновані методи та модель реалізовано у вигляді програмного модуля семантичної верифікації онтологічних знань

**Ключові слова:** інтелектуалізація, Web-контент, верифікація, онтологія, база знань, семантичне порівняння, гібридне виведення знань, дескрипційна логіка, пояснення

## АННОТАЦИЯ

Головянко М.В. «Методы и модель верификации знаний для интеллектуализации Web-контента». – Рукопись.

Диссертация на соискание научной степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. Харьковский национальный университет радиоэлектроники, Харьков 2011.

Диссертационная работа Головянко М.В. посвящена разработке методов интеллектуализации Web-контента. В работе проведен анализ существующих подходов к верификации, детально рассмотрены методы и метрики верификации знаний, представленных с помощью онтологической модели, выявлены их недостатки. Обоснована необходимость разработки методов и моделей семантической верификации знаний, накапливаемых в базе знаний системы в результате гибридного вывода на основе информации, полученной из распределенных разнородных источников, с использованием технологий, предложенных в рамках инициативы Semantic Web.

Представлена онтологическая модель представления знаний, основанная на

дескрипционных логиках, расширенная верификационной составляющей, элементы которой позволяют описать процесс накопления знаний для подтверждения достоверности полученных системой решений. Модель описана на языке представления онтологий OWL, что гарантирует возможность обмена верификационными знаниями между Web-системами, которые используют RDF/XML синтаксис. Это позволило дополнить Web-контент такими формальными описаниями, которые позволяют повысить эффективность и надежность обработки информации в распределенных информационных системах за счет возможности автоматической проверки достоверности предоставленной системой.

Исследованы возможные типы изменений в онтологической базе знаний, проведена их классификация по типам, которые основаны на различиях между разными уровнями интерпретации онтологии. Предложен базовый набор операций изменений, на основе которых можно описать любые изменения, происходящие в онтологической базе знаний.

Впервые разработан метод верификации онтологических знаний, который заключается в генерации древовидных структур пояснения для автоматического гибридного вывода знаний на основе информации, которая извлекается из распределенных разнородных источников. Доказано, что верификация знаний сводится к верификации множества утверждений, представленных в виде триплетов. Для интерпретации результатов верификации предложен метод определения эквивалентности древовидных структур, который отличается от существующих возможностью учета семантики узлов дерева и характеризуется использованием нескольких независимых оценок подобия онтологических объектов.

Предложен новый подход к интеллектуализации Web-контента на основе программных средств семантической верификации. Представлена модель интеллектуальных Web-систем с компонентами верификации знаний на примере семантических порталов, которые используются для менеджмента разнородных распределенных ресурсов. Описана структура, взаимодействие и назначение компонентов систем такого типа. Рассмотрены средства автоматической верификации контента, которые позволяют пользователям, интеллектуальным программам или агентам удостовериться в достоверности предоставленной информации. Внимание в разделе также уделяется менеджменту верификационных знаний.

В работе описана реализация подсистемы верификации знаний в рамках разработки прототипа интеллектуальной системы «Онтологічний портал менеджменту та оцінки освітніх і наукових ресурсів України». Описан также верификатор Web-контента, который в виде программного модуля динамически подключается к Internet-обозревателям. Внедрение верификатора выполнено на производстве в рамках разработки Web-ориентированной системы документооборота коммерческой организации.

**Ключевые слова:** интеллектуализация, Web-контент, верификация, онтология, база знаний, семантическое сравнение, гибридный вывод знаний, дескрипционная логика, пояснение, достоверность.

**ABSTRACT**

Golovianko M.V. "Methods and a model of knowledge verification for Web-content intellectualization" - Manuscript.

Thesis for the candidate's degree in technical sciences, specialty 05.13.23 – systems and tools of artificial intelligence. Kharkiv National University of Radioelectronics, Kharkiv 2011.

The thesis is devoted to the development of methods for Web-content intellectualization. For this purpose the existing ontology-based knowledge representation model was improved by a verification component, which allowed to justify and explain the hybrid knowledge inference. For the first time a verification method for ontological knowledge is developed, which performs explanation of hybrid automatic inference of knowledge based on data extracted from distributed heterogeneous sources based on Semantic Web approaches, thereby improving the efficiency and reliability of intelligent information processing in distributed information systems. New methods for interpretation of verification elements (results) were proposed based on the method comparing the ontological objects, which is characterized by the use of several semantic and syntactic independent evaluations of similarity, which allowed to classify ontology-based knowledge by the type of its inference and to find changes in the inference. Improved architecture of Web-oriented distributed systems for management of heterogeneous resources is represented, which differs from the existing by verification components that would provide the verification of semantic content in the systems of this type. The proposed methods and the model is implemented in a module for verification and in a prototype of intelligent system «Ontology-based portal for management and evaluation of educational and scientific resources of Ukraine».

**Keywords:** proof, trust, Semantic Web, Web-content, verification, hybrid knowledge inference, ontology, knowledge base, semantic comparison, description logic.