

МЕТОД ОЦІНЮВАННЯ НОВИЗНИ ЗНАНЬ ПІД ЧАС НАВЧАННЯ ОНТОЛОГІЙ

A method of evaluating the novelty of knowledge while learning ontologies is considered. Formalization of the evaluation process innovation expertise to building effective procedures for solving problems using ontological approach is implemented.

Keywords: *ontology, artificial intelligence, information technology.*

Розглянуто метод оцінювання новизни знань під час навчання онтологій. Здійснено формалізацію процесу оцінювання новизни знань з метою побудови ефективних процедур розв'язування задач за допомогою онтологічного підходу.

Ключові слова: *онтологія, штучний інтелект, інформаційні технології.*

Наповнення онтології вручну є дуже дорогим та трудомістким завданням. Це вимагає досконалого знання певної предметної області, та в більшості випадків результат може бути неточним чи неповним. Дослідники намагаються подолати ці недоліки, використовуючи напівавтоматичні або автоматичні методи побудови та наповнення онтологій [3].

Найвідомішими методами, що використовують для побудови та наповнення онтологій за допомогою тексту, є такі.

Метод Енеко Агіра. Цей метод призначений для збагачення існуючих великих онтологій за допомогою тексту, що витягується з мережі Інтернет. Метою цього методу є подолання недоліків великої онтології: відсутність актуальних зв'язків між поняттями, різні смислові значення одного терміна тощо.

Спершу інформація, що міститься в онтології, використовується для побудови запитів. Для кожного поняття з інтернету отримуються документи, з яких витягуються тексти, що формуються в колекції. Кожна колекція порівнюється з іншими. Ці колекції документів використовують для побудови тематичних ознак та для побудови ієрархічних кластерів понять. У результаті формуємо ознаки тем, які мають збігатися із нашими запитами.

Цей метод пропонує 4 кроки:

1. Отримання відповідних документів для кожного поняття (запити будуються для кожного поняття).
2. Побудова ознак теми (документи в кожній колекції пов'язані з конкретним поняттям, повинні бути оброблені. Потім дані з однієї колекції порівнюються з даними іншої колекції. Слова, що є відмінними, групуються у список, який потім представляє ознаки тем для кожного поняття).
3. Кластеризація понять (різні ознаки тем порівнюються із загальними словами для визначення дублювання ознак).
4. Оцінка результатів [7].

Метод Наталі Асенак-Гілз. Метод дає змогу створити модель предметної області за допомогою аналізу корпусу тексту, використовуючи обробку природної мови та лінгвістичні методи. Центральну роль у цьому методі відіграють технічні тексти. Метод поєднує інструменти видобування знань, заснованих на методах моделювання, що дає можливість підтримувати зв'язок між моделями і текстами. Він використовує тексти і може використовувати інші існуючі онтології або термінологічні ресурси для створення онтології. Цей метод заснований на видобуван-

ні термінів на основі дистрибутивного аналізу. Текст для корпусу обирають дуже ретельно. Він має чітко відповідати вимогам користувача [8].

Дії, що пропонуються у цьому методі, описують такими кроками:

1. Корпус тексту (тексти вибираються з наявної технічної документації експертом із заданої предметної області).

2. Лінгвістичне навчання (лінгвістичний рівень складається з термінів та лексичних відносин, витягнутих із текстів за допомогою лінгвістичного аналізу. Відбувається вибір відповідних мовних засобів і методів та застосування їх до текстів).

3. Нормалізація (результатом цієї діяльності є концептуальна модель, виражена за допомогою семантичної мережі. Відбувається перевірка концептуальної моделі).

4. Формалізація (перевірка і реалізація онтології) [9].

Метод Удо Гана. Удо Ган і його колеги представляють метод для підтримки предметно-орієнтованих таксономій, побудованих на природній мові. Дані таксономії постійно оновлюються, як нові поняття, що взяті із реальних текстів.

Дії, що пропонуються у цьому методі, описують такими кроками:

1. Мова обробки (визначає структурну залежність інформації від граматичних конструкцій, у яких невідомі лексичні одиниці трапляються у термінах відповідного синтаксичного дерева).

2. Визначення знаків якості (по-перше, лінгвістичний рівень, що відображає структурні властивості патернів, і, по-друге, концептуальний, що впливає з порівняння структур гіпотез).

3. Оцінка якості (отримання доступного набору знаків якостей для кожної гіпотези. Результат є класифікованим списком понять гіпотез).

4. Аналіз роботи (емпірично оцінюється процес набуття знань з тексту за допомогою різних методів, які оцінюють точність і швидкість навчання) [10].

Метод Хунга Хванга. Цей підхід був розроблений для представлення та видобування інформації з великих текстових баз даних. Він заснований на використанні динамічної онтології, яка охоплює семантику інформації, наявної всередині документів. Онтологія організована у прості таксономії.

Метод складається з таких кроків:

1. Людина-фахівець вводить у систему невелику кількість слів.

2. Система обробляє вхідні документи і витягує тільки ті фрази, які містять так звані “засіяні слова”.

3. Система генерує відповідні поняття, поміщає їх у певне місце в онтології та попереджає про це експерта із заданої предметної області.

4. Витягаються кілька видів зв'язків. На кожному етапі фахівець перевіряє правильність понять. У разі необхідності експерт має право вносити виправлення в онтологію.

При побудові онтології цей метод також дає змогу індексувати документи для майбутнього пошуку. Проте є деякі проблеми для автоматичної генерації онтологій з цим підходом, наприклад, структурні синтаксичні неоднозначності, визначення різних фраз, які належать до однієї концепції, проблеми сенсу слова і т.д. [11].

Метод Джорджа Кітза. Цей метод використовують для виявлення специфічної онтології із заданих різномірних ресурсів за допомогою природомовного аналізу. Це напівавтоматичний процес – користувач бере у ньому безпосередню участь. У цьому методі автори використали збалансоване моделювання, де робота з побудови онтології розподіляється між кількома алгоритмами навчання і користувачем.

Цей метод складається з таких кроків:

1. Вибір джерел (користувач повинен зазначити, які документи повинні бути використані в наступних кроках, щоб удосконалити та розширити попередні онтології).

2. Вивчення понять (аналіз понять та вирішення, чи вони є специфічними для цієї онтології).

3. Зосередження на області застосування (скорочення ядра онтології шляхом усунення з неї загальних понять).

4. Вивчення зв'язків (часті поєднання понять у реченнях можна розглядати як зв'язки між поняттями в онтології).

5. Оцінка (оцінювання отриманої онтології та вирішення, чи варто повторювати весь процес знову) [12].

Метод Дерила Лонсдела. Цей метод має на меті побудову нової онтології з повторним використанням вже існуючої великої онтології. Він використовує як різні вхідні ресурси, так і документи з певної предметної області.

Були запропоновані 2 етапи для реалізації цього методу:

– перетворення великомасштабних термінологічних ресурсів у формат, який може бути використаний для видобування даних з онтології;

– система онтології об'єднує цю термінологічну інформацію з подібними ресурсами, а потім використовує базу знань, щоб проаналізувати вміст вхідних документів і виробити нові зв'язки.

Використовуючи цю різноманітність джерел інформації, область інтеграції даних словника вимагає принципового підходу до моделювання даних.

Метод можна зобразити такими кроками:

1. Попередня обробка джерел (потрібно обробити навчальні тексти, щоб видобувати з них потрібну інформацію).

2. Вибір понять (поняття вибирають за допомогою зіставлення між текстовим змістом і онтологічними даними).

3. Пошук зв'язків між поняттями (онтологія структурована як орієнтований граф, вузлами якого є поняття. Зв'язки між поняттями можуть бути зображені як шляхи між ними).

4. Виявлення обмежень (визначає обмеження на зв'язках, виявлених у попередньому кроці).

5. Уточнення результатів (остаточна онтологія повинна бути перевірена і уточнена експертом з певної предметної області) [13].

Метод Мішеля Місікофа. Цей метод застосовують для побудови дерева понять у певній предметній області і з'єднання його з існуючим ядром онтології.

У цьому методі використовують статистичний підхід, щоб визначити доречність терміна для певної предметної області, та семантичну інтерпретацію, засновану на методах машинного навчання, для визначення смислу термінів та семантичних зв'язків між ними.

Цей метод пропонує 3 основні кроки:

1. Видобування термінології (з проаналізованого тексту витягуються терміни та комбінації термінів. Метод пропонує використовувати заходи специфічності термінів через порівняльний аналіз різних корпусів).

2. Семантична інтерпретація (визначення сенсу понять для кожного компонента комплексного терміна. Буде визначено область застосування понять).

3. Створення специфічної онтології (цей етап спрямований на інтеграцію таксономії з ядра онтології. Будується дерево понять і додається до існуючої онтології) [15].

Метод Дена Молдована і Роксани Гіру. Цей метод розроблений для виявлення специфічних понять і зв'язків для наповнення існуючої онтології новими

знаннями, отриманими з текстів. Джерелом нових понять є неспецифічний корпус тексту, який посилюється за допомогою словників. Користувач задає ряд специфічних для певної предметної області понять, які використовують для відкриття нових понять і зв'язків. Потім користувач виконує перевірку процесу і підтверджує правильність понять.

Для наповнення онтології пропонують такі кроки:

1. Вибір так званих “засіяних слів” (вони позначаються певним чином, цей список понять продовжується відповідними синонімами).

2. Дослідження нових понять (для кожного документа вилучаються речення, що містять “засіяні слова”, речення обробляються і витягуються нові поняття).

3. Відкриття лексико-синтаксичних шаблонів (метод пропонує створити новий корпус, що відрізняється від корпусу, використаного на попередньому кроці, щоб дослідити нові шаблони всередині нової групи речень).

4. Відкриття нових зв'язків між поняттями (для кожного нового поняття знаходяться всі синтаксичні зв'язки, отримані на 3 кроці. Перевірка виконується користувачем).

5. Класифікація та інтеграція (створюється нова таксономія понять, яка буде об'єднана з існуючими онтологіями, використовуючи нові зв'язки та поняття) [16].

Ефективність онтології безпосередньо залежить від новизни знань, які до неї додаються. Постає задача оцінки знань, які пропонують додавати в онтологію баз знань (БЗ) інтелектуальних систем (ІнтС), тобто побудови деякою мірою знань, подібної до відомої міри інформації, введеної Шенноном [6]. Однак на відміну від інформації для знань необхідно визначити який ефект від її використання. Отже, метою цієї роботи є розроблення методу визначення оцінки ефективності нових знань.

Постановка задачі. Розробити метод оцінювання новизни знань.

Викладення основного матеріалу. Стосовно видів кількісної оцінки інформації, то їх можна поділити на три групи: синтаксичні, семантичні та прагматичні.

1. *Синтаксичне вимірювання кількості інформації* базується на взаємозв'язку між інформацією і ентропією – мірою невизначеності. Якщо апріорно ситуація характеризується ентропією H_0 , а після одержання повідомлення S ентропія зменшилася до H_1 , то кількість інформації, що міститься у повідомленні S (позначення I_S), визначається різницею

$$I_S = H_0 - H_1. \quad (1)$$

Отже, кількість інформації вимірюється зменшенням невизначеності ситуації в результаті отримання повідомлення S . Чим більшою є кількість можливих повідомлень, тим більшою була апріорна невизначеність H_0 і тим більшу кількість інформації одержує адресат, коли ця невизначеність знижується.

Величина ентропії на один елемент повідомлення визначається за формулою К. Шеннона:

$$H = -\sum_{i=1}^m p_i \log p_i, \quad (2)$$

де m – кількість можливих станів елемента повідомлення; p_i – ймовірність того, що елемент перебуває в i -му стані.

2. *Семантичні міри інформації.* Ця категорія мір призначена для вимірювання змісту I_{sem} повідомлень, які одержує адресат інформації. Оскільки зміст повідомлень кожний конкретний їх адресат сприймає по-різному (залежно від об'єму інформації, якою він володіє), то саме “запас знань” адресата і кладуть в основу семантичних мір інформації. Отже, семантичні міри враховують не тільки зміст

повідомлень, але і те, який об'єм нової інформації несе повідомлення його адресату порівняно з тим, що він вже знав.

Методи кількісного вимірювання змісту інформації ще достатньо не розроблені. Найбільшого визнання здобула запропонована Ю. І. Шнейдером тезаурусна міра, у якій семантична ємність І_{сем} інформації пов'язується із здатністю її одержувача сприймати (асимілювати) відомості, що надходять, а це залежить від об'єму знань адресата інформації, або, як прийнято говорити, від його тезауруса.

Максимального значення семантичність інформації набуває при погодженості тезаурусів, коли інформація, що надходить, по-перше, є зрозумілою адресату, по-друге, несе йому невідомі раніше (відсутні у його тезаурусі) відомості.

Отже, кількість семантичної інформації у відомостях (тобто кількість нових знань, що їх дістає адресат) є величиною відносною: одне й те саме повідомлення може мати зміст для компетентного і бути беззмістовним (семантичним шумом) для некомпетентного адресата. Водночас інформація, яка хоча є зрозумілою, але відомою компетентному адресатові, є для нього також семантичним шумом. Іншими словами, відомості, які не є новими, несуть нульову кількість інформації.

Отже, зміст інформації можна вимірювати ступенем зміни тезауруса адресата інформації під впливом отриманого повідомлення.

У зв'язку з цим дуже важливими є задачі моделювання тезаурусів різних інформаційних систем, розроблення взаємоузгоджених тезаурусів у системі “людина–машина”, вирішення питань про те, яким повинен бути тезаурус у людей, що займаються тією чи іншою проблемною сферою (економікою, фінансами, математикою, програмуванням, інженерним проектуванням, медициною, біологією тощо).

3. *Прагматичні міри інформації.* Ці міри інформації відображають цінність інформації, її корисність і доцільність для досягнення поставленої мети.

Прагматичні міри, як і семантичні, є відносними, і вибір міри зумовлюється особливостями використання певної інформації у певній системі.

Здебільшого цінність інформації залежить від того, ким і з якою метою вона використовується. Наприклад, повідомлення, що завтра буде значна хмарність, має різну цінність для торгівельної організації і для авіації.

Відомі два основні методи визначення кількісної міри цінності інформації. Якщо мета напевне є досяжною і, до того ж, декількома можливими шляхами, то зручною є міра цінності, запропонована Р. Л. Стратоновичем. Вона полягає в оцінці умовних “штрафів” (або витрат часу, засобів, грошей та ін.) і вимірюється зменшенням витрат у результаті досягнення цілей. Так, наприклад, кількісною мірою цінності інформації, яка стосується поїздки у місто (розклад руху транспорту, вартість квитків тощо), може служити зекономлений час і (або) гроші.

Якщо, навпаки, досягнення мети малоімовірне, то зручніше користування критерієм, запропонованим Н. М. Бонгартом і А. А. Харкевичем. Мірою цінності при цьому є логарифм відношення ймовірностей досягнення мети до одержання інформації (P_{in}) і після цього (P_{fin}):

$$I_{\text{ö}} = \log \frac{P_{\text{fin}}}{P_{\text{in}}} . \quad (3)$$

Що стосується оцінки новизни знань, то у роботі [5] ми запропонували підхід, що ґрунтується на використанні лексичного словника WordNet [14]. Однак є проблема розроблення єдиного уніфікованого підходу до такої оцінки.

Проблема автоматизованого формування онтологій БЗ супроводжується виникненням таких проблем:

1) зменшення якості нарощуваного обсягу інформації (релевантності до заданої ПО);

2) втратою монотонності, тобто появою внутрішніх конфліктів, які порушують її цілісність;

3) появою неконтрольованої надлишковості.

Як відзначено у [2], онтологію формують вручну, напівавтоматизовано та автоматизовано. Якщо цей процес здійснюється автоматизовано, то необхідно здійснювати оптимізацію онтології [4].

Напрями оптимізації онтології напряму залежать від критерій якості ІнтС, оскільки якість таких систем напряму залежить від БЗ [1], а в розглянутому у цій роботі підході ядром БЗ є онтологія. Для оцінювання якості ІнтС використаємо стандарт ISO 9126 для інформаційних систем, оскільки спеціалізованого стандарту лише для інтелектуальних систем немає. Цей стандарт призначено або для стандартизації виробничого процесу (аналогія з життєвим циклом інформаційних систем), або для оцінки якості програмних засобів.

Характеристики ISO 9126 визначають напрями оптимізації онтологій, які полягають в оптимізації її структури, що динамічно формується під час експлуатації системи та оптимізації змісту онтології БЗ. Під змістом розуміємо інформаційне наповнення, яке має бути гнучким, тобто налаштовуваним під конкретну ПО і потреби користувача. Іншим підходом до оптимізації онтології є визначення оцінки новизни знань, яку ми плануємо додавати в онтологію порівняно з тими знаннями, які вже зберігаються в ній. Саме формальна постановка цієї задачі розглядається у цій роботі.

Базові характеристики якості ІнтС за ISO 9126:

- функціональна придатність до використання;
- коректність або достовірність;
- ресурсна економічність;
- практичність;
- супроводжувальність;
- мобільність.

Оскільки ми розглядаємо клас інтелектуальних систем, призначених для підтримки прийняття рішень, ядром БЗ яких є онтології, то ці характеристики напряму залежать від якості онтологій, оскільки відповідно якість ІнтС залежить від її БЗ.

Функціональна придатність ІнтС залежить від повноти онтологій, наскільки вона точно описує специфіку ПО, та задач, які у ній виникають. Відповідно повнота онтологій залежить від вміння давати правильні відповіді на запити до неї. Це залежить від вміння системою оцінювати новизну знань, яку пропонується додавати до онтології. Тому надалі розглянемо метод оцінки новизни знань, який ґрунтується на використанні лексичного словника WordNet. Мірою якості функціональної придатності буде середній відсоток нетривіальних (ненульових), правильних відповідей на запити до онтології. Тобто

$$\chi_1 = \frac{N_q^p}{N_q} \cdot 100\% ,$$

де N_q – кількість всіх запитів до онтології БЗ; N_q^p – кількість правильних відповідей на запити.

Визначення функціональної придатності є однією з базових характеристик ІнтС.

Коректність або достовірність функціонування ІнтС – це відсоток правильно розв'язаних задач інтелектуальною системою. Це основна характеристика якості ІнтС, і вона залежить не тільки від якості БЗ, але й від моделі функціонування таких систем, тобто від побудованої метрики. Отже,

$$\chi_2 = \frac{N_z^P}{N_z} \cdot 100\% ,$$

де N_z – кількість задач, які розв'язала ІнтС; N_z^P – кількість правильно розв'язаних нею задач.

Використовуваність ресурсів (або *ресурсна економічність*) у стандартах відображається зайнятістю ресурсів центрального процесора, оперативної, зовнішньої та віртуальної пам'яті, каналів введення-виведення, терміналів і каналів зв'язку. Для покращення цієї характеристики розглянемо оптимізаційну задачу, критерієм якої буде фізичний об'єм пам'яті, яку займає онтологія. Очевидно, що онтологія займає найменший об'єм пам'яті, якщо в ній немає жодного поняття. Тому цей критерій необхідно скомбінувати з іншим критерієм, а саме з функціональною придатністю ІнтС. Отримуємо задачу оптимальної кількості понять онтології, яку розв'язуємо нижче.

Практичність – важко формалізоване поняття, яке визначає функціональну придатність і корисність застосування ІнтС для певних користувачів. У цю групу показників входять субхарактеристики, які з різних сторін відображають функціональну зрозумілість, зручність освоєння, системну ефективність і простоту використання ІнтС. Така придатність повинна базуватись на цілісності онтології, тобто відсутності в її тілі взаємозаперечувальних тверджень і дублювання, а також на збалансованості ПО, яка виражається у рівномірному поданні її окремих підрозділів в онтології.

Супроводжуваність ІнтС відображається зручністю і ефективністю виправлення, удосконалення або адаптації структури та змісту онтології БЗ залежно від змін у зовнішньому середовищі застосування, а також у вимогах і функціональних специфікаціях замовника. Узагальнено якість супроводжуваності ІнтС можна оцінювати як потребу ресурсів для її забезпечення функціональності та для її реалізації. Сукупність субхарактеристик супроводжуваності програмної системи, подана в стандарті ISO 9126, цілком застосовна для описання цієї якості інтелектуальних систем, в основному, тими ж організаційно-технологічними субхарактеристиками.

Мобільність характеризується тривалістю і трудомісткістю інсталяції інформаційних продуктів, адаптації та заміщення при перенесенні на інші апаратні та операційні платформи. Інформація про процеси, що відбуваються у зовнішньому середовищі, може мати великі обсяги і трудомісткість первинного нагромадження та актуалізації, що визначає необхідність її ретельного зберігання та регламентованої зміни. Критерієм мобільності є швидкодія, яка виражається часом відгуку ІнтС на зовнішнє звертання (час реакції на зміну параметрів зовнішнього середовища, до яких чутлива система).

Отже, основною характеристикою якості будь-якої ІнтС є достовірність отриманого розв'язку цією системою χ_2 . Якщо ІнтС побудована на основі онтології, то вагомою є й перша характеристика χ_1 .

Для підвищення ефективності вищенаведених шести характеристик необхідно розв'язати задачу оцінки новизни онтологічних знань.

Нехай Z – задача, яка розв'язується за допомогою знань, поданих у вигляді онтології.

$Z_j \subset Z$ – деякий підклас задачі Z .

Для розв'язування задачі Z використовують множину методів $M^Z = \{M_1^Z, M_2^Z, \dots, M_n^Z\}$.

Якість методів, які використовують для розв'язування задачі Z , оцінюється множиною параметрів $P^Z = \{p_1^Z, p_2^Z, \dots, p_m^Z\}$.

Новий метод розв'язування задачі Z : M_{new}^Z .

Виграш цього методу порівняно з деяким іншим i -м методом, який використовують для розв'язування підкласу Z_j задачі Z : $M_i^{Z_j}$, задають як відображення в підмножину елементів, які є парою (параметр, перевага):

$u(M_{new}^{Z_j}, M_i^{Z_j}) \rightarrow U_{i,new}^{Z_j} = \{(p_{i_s}^{Z_j}, \alpha_i)\}_{s=1}^{k_i}$, α – кількісний або якісний опис переваги.

Аналогічно програш $v(M_{new}^{Z_j}, M_i^{Z_j}) \rightarrow V_{i,new}^{Z_j} = \{(p_{i_s}^{Z_j}, \beta_{i_s})\}_{s=1}^{k_i}$.

Функція виграшу нового методу порівняно з методом $M_i^{Z_j}$: $\varphi(U_{i,new}^{Z_j})$, функція програшу нового методу порівняно з методом $M_i^{Z_j}$: $\psi(V_{i,new}^{Z_j})$.

Читаємо як: розроблено новий метод M_{new}^Z для розв'язування задачі Z , який на відміну від іншого методу M_i^Z дає змогу отримати виграш $\varphi(U_{i,new}^{Z_j})$ та програш $\psi(V_{i,new}^{Z_j})$.

Приклад. Нехай задача Z полягає у розв'язанні квадратного рівняння $ax^2 + bx + c = 0$.

Z^1 – підклас задачі, для якої $D = b^2 - 4ac > 0$, $a, b, c \in Z$, $a = 1$, $b \neq 0$, $c \neq 0$.

Z^2 – підклас задачі, для якої $D = b^2 - 4ac > 0$, $a, b, c \in Z$, $a \neq \{0, 1\}$, $b \neq 0$, $c \neq 0$.

$M^Z = \{M_1^Z = 'a\hat{e}\hat{n}\hat{e}\hat{d}\hat{e}\hat{i} \hat{z} \hat{a}\hat{i} \hat{o} ', \dots\}$, $P^Z = \{p_1^Z = 'z\hat{a}\hat{n} \hat{d}\hat{i} \hat{c}\hat{a}' \hat{y}\hat{c}\hat{o}\hat{a}\hat{i} \hat{t} \hat{y}', \dots\}$.

$M_{new}^Z = 'd \hat{a}\hat{i} \hat{d}\hat{a} \hat{a} \hat{A}^{\hat{z}o} \hat{o} \hat{a}'$. Тоді

$u(M_{new}^{Z_1}, M_1^{Z_1}) \rightarrow U_{1,new}^{Z_1} = \{(p_1^{Z_1}, 30c)\}$, $v(M_{new}^{Z_2}, M_1^{Z_2}) \rightarrow V_{1,new}^{Z_2} = \{(p_1^{Z_2}, 40c)\}$.

Тобто $\alpha_1 = 30c$, $\beta_1 = 40c$.

Функція виграшу $\varphi = \alpha_1 = t_1 - t_{new}$ – різниця між часом розв'язування задач старим та новим методом.

ВИСНОВКИ

Здійснено формальну математичну постановку задачі оцінювання новизни онтологічних знань з погляду підвищення ефективності функціонування інтелектуальних систем, ядром баз знань яких є онтологія. Визначено основні характеристики інтелектуальних систем згідно з ISO 9126, що дало змогу обґрунтувати доцільність розв'язування такої задачі. Оцінка новизни знань на відміну від оцінки новизни інформації базується на функції виграшу, яку може отримати користувач системи, використовуючи ці знання. Наведено приклад, який ілюструє розроблену математичну модель задачі.

1. Гаврилова Т. А., Хорошевський В. Ф. Базы знаний интеллектуальных систем. – СПб.: Питер., 2001. – 384 с.
2. Інтелектуальні системи, базовані на онтологіях / Д. Г. Досин, В. В. Литвин, Ю. В. Нікольський, В. В. Пасічник. – Львів: "Цивілізація", 2009. – 414 с.
3. Литвин В. В. Базы знаний интеллектуальных систем підтримки прийняття рішень. – Львів: Вид-во "Львівська політехніка", 2011. – 240 с.

4. *Литвин В. В.* Задачі оптимізації структури та змісту онтології та методи їх розв'язування // Вісн. Нац. ун-ту "Львівська політехніка". Сер.: Інформаційні системи та мережі. – 2011. – № 715. – С. 189–200.
5. *Литвин В. В., Мельник А. С., Крайовський В. Я.* Оцінка новизни знань під час автоматичної розбудови онтологій // Вісн. Нац. ун-ту "Львівська політехніка". Сер.: Інформаційні системи та мережі. – 2011. – № 699. – С. 343–353.
6. *Шеннон К.* Работы по теории информации и кибернетике. – М.: 1963. – 830 с.
7. *Agirre E., Ansa O., Hovy E.* Enriching very large ontologies using the WWW // Proc. of the Workshop on Ontology Construction of the European Conf. of AI (ECAI-00). – Spain: Donostia. – 2000. – P. 7.
8. *Aussenac-Gilles N., Biebow B., Szulman S.* Corpus Analysis For Conceptual Modelling // Workshop on Ontologies and Text, Knowledge Engineering and Knowledge Management: Methods, Models and Tools, 12th Int. Conf. EKAW'2000, Juan-les-pins, France. – Paris: Springer-Verlag, 2000. – P. 8.
9. *Aussenac-Gilles N., Biebow B., Szulman S.* Revisiting Ontology Design: A Methodology Based on Corpus Analysis // 12th Int. Conf. in Knowledge Engineering and Knowledge Management (EKAW'00). – Paris, 2000. – P. 9.
10. *Hahn U., Schnattinger K.* Towards text knowledge engineering // IAAI '98 Proc. of the 15th National Conf. on Artificial Intelligence & 10th Conf. on Innovative Applications of Artificial Intelligence. – Madison, Wisconsin, July 26–30, 1998. – P.8.
11. *Hwang, C. H.* Incompletely and imprecisely speaking: Using dynamic ontologies for representing and retrieving information // Proc. of the 6th Int. Workshop on Knowledge Representation meets Databases. – Linköping, Sweden, July 29–30, 1999. – P.9.
12. *Kietz J. U., Maedche A., Volz R.* A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet / Eds.: Aussenac-Gilles N, Biebow B, Szulman S // EKAW'00 Workshop on Ontologies and Texts. – France: Juan-Les-Pins, 2000. – P. 14.
13. *Pepperling* Knowledge Sources with SALT; Boosting Conceptual Content for Ontology Generation / D. Lonsdale, Y. Ding, D. W. Embley, and A. Melby // Proc. of the AAAI Workshop on Semantic Web Meets Language Resources. – Edmonton, Alberta, Canada, July 2002. – P. 7.
14. *Miller G. A.* WORDNET: A lexical database for English // Communications of ACM (11). – 1995. – P. 39–41.
15. *Missikoff M., Navigli R., and Velardi P.* The Usable Ontology: An Environment for Building and Assessing a Domain Ontology // Research paper at International Semantic Web Conf. (ISWC) 2002, June 9–12th, Sardinia, Italia. – P. 39–42.
16. *Moldovan D. I., Girju R. C., Rus V.* Domain-Specific Knowledge Acquisition from Text // Proc. of the Applied Natural Language Processing Conf. – Seattle, WA., April–May, 2000. – P. 8.