

МЕТОД ОЦІНЮВАННЯ ЯКОСТІ ОНТОЛОГІЙ БАЗ ЗНАНЬ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ НА ОСНОВІ ISO/IEC 25012

© Демчук А. Б., Гопяк М. Я., 2014

Розглянуто значення онтологій у межах понять бази знань та інтелектуальних систем. Для оцінювання якості онтологій баз знань інтелектуальних систем підтримки прийняття рішень запропоновано використовувати стандарт ISO/IEC 25012.

Ключові слова: онтологія, інтелектуальна система, оцінювання якості, дані, база знань.

The importance of ontology within concepts of the knowledge base and intelligent systems is considered in the article. Standard ISO/IEC 25012 has been offered for evaluation of ontologies qualities of knowledge bases of intelligent decision support systems.

Key words: ontology, intelligent system, evaluation of quality, data, knowledge base.

Вступ. Постановка проблеми у загальному вигляді

Обсяги даних та інформації, що обробляються комп'ютерними системами, зростають в усьому світі. Якість програмного забезпечення повинна бути достатньо високою, щоб забезпечити успіх електронного урядування, електронного бізнесу та електронної комерції. Аналогічно і дані, які обробляються за допомогою програмного забезпечення, повинні відповідати певним характеристикам якості даних (наприклад, послідовність, семантична та синтаксична точність, безпека доступу тощо). Інформаційна система повинна відображати зміни, що відбуваються в її предметній області. Очевидно, що факти, які накопичуються в системі (властивості або твердження про об'єкти) можуть виявитися неправильними, суперечливими або некоректними. Підтримка контенту в актуальному стані підвищує ефективність виконання системою своїх функцій, дає змогу ефективно використовувати комп'ютерні ресурси і знижує ймовірність виникнення помилок. Як стандарт інженерії знань використовують онтологічний інжиніринг, у результаті застосування якого отримують онтологію бази знань. Онтологія – це детальна формалізація деякої області знань, подана за допомогою концептуальної схеми. Така схема складається з ієрархічної структури понять, зв'язків між ними, теорем та обмежень, прийнятих у певній предметній області (ПО) [1].

Використання онтологій у складі баз знань допомагає вирішити проблеми методологічного та технологічного характеру, які виникають під час створення таких систем. Характерні проблеми полягають у відсутності концептуальної цілісності та узгодженості окремих прийомів і методів інженерії знань; нестачі кваліфікованих фахівців у цій галузі; жорсткості розроблених програмних засобів та їх низької адаптивної здатності; складності впровадження інтелектуальних систем, що зумовлено психологічними аспектами.

Аналіз останніх досліджень та публікацій

Результати наукових досліджень, які здійснювались два останні десятиліття, привели до активного використання онтологій як концептуальних схем реляційних баз даних, що покладено в основу інформаційних систем [1, 2]. У контексті цієї роботи поняття онтології та концептуальної схеми використовуються як рівнозначні. Онтологія предметної області задається у вигляді базових понять, організованих у таксономію, і сукупності зв'язків між ними. Дані подаються у вигляді множини різнотипних інформаційних об'єктів – екземплярів понять і відношень онтології. У сукупності об'єкти утворюють контент або інформаційне наповнення системи. Кожний об'єкт

визначається поняттям або відношенням онтології і, як екземпляр класу, має задану ним структуру. Нижче розглянемо ISO/IEC 25012 як основу для розроблення методу оцінювання якості онтологій баз знань інтелектуальних систем. Розглянемо основні показники стандарту та спробуємо адаптувати їх для оцінки якості онтологій.

Формування мети

Адаптувати характеристики ISO/IEC 25012 для оцінювання якості онтологій баз знань інтелектуальних систем підтримки прийняття рішень.

Оцінювання якості онтологій

Для оцінювання якості онтологій використаємо стандарт ISO/IEC 25012 [3]. Адапуємо характеристики цього стандарту до онтології бази знань. Якість даних можна відобразити за допомогою якісної моделі, яка класифікує атрибути якості даних на вісімнадцять характеристик: узгодженість, поширеність, повнота, чіткість, точність, конфіденційність, доступність, зворотність, зрозумілість, керованість, ефективність, мінливість, продуктивність, портативність, безпека, достовірність, зручність, дотримання встановлених норм [3]. До базових характеристик якості інтелектуальних систем за ISO/IEC 25012 є [4]: функціональність у використанні; достовірність; зрозумілість; зручність; портативність; відновлення; конфіденційність. З огляду на те, що розглядається клас інформаційних систем, які проєктуються на основі онтологічної моделі предметної області, ці характеристики напряду залежать від якості онтологій.

Функціональність у використанні. Функціональність – це здатність даних, яка дає змогу комп'ютерній системі задовольняти функціональні користувацькі вимоги та завдання. Функціональні вимоги до даних залежать від користувачів і, як правило, стосуються тих функцій (або можливостей), для котрих дані можуть призначатись. З погляду кінцевого користувача, повнота – це ступінь, відповідно до якого дані здатні задовольнити потреби користувача у кількісному розумінні. Визначення функціональності є однією з основних характеристик інформаційних систем, що залежить від повноти онтології, наскільки точно вона описує специфіку предметної області (ПО) та задач, які у ній виникають. Відповідно, повнота онтологій залежить від вміння давати правильні відповіді на запити до неї. Своєю чергою, це залежить від вміння системи оцінити новизну знань, які пропонується додавати до онтології. Мірою якості функціональності буде середній відсоток ненульових (нетривіальних) правильних відповідей на запити до онтології:

$$\chi_1 = \frac{M_q^p}{M_q} \cdot 100 \% , \quad (1)$$

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

Достовірність – це ступінь, відповідно до якого дані розглядають як істинні та надійні для користувачів [3]. Достовірність функціонування ІС – це відсоток достовірно розв'язаних задач системою. Це одна з базових характеристик якості ІС і вона залежить не тільки від якості БЗ, але й від моделі функціонування таких систем, отже, від побудованої метрики:

$$\chi_2 = \frac{M_z^p}{M_z} \cdot 100 \% , \quad (2)$$

де M_z – загальна кількість задач, які розв'язала ІС; M_z^p – кількість правильно розв'язаних задач.

Зрозумілість – це ступінь, відповідно до якого реальне значення даних є простим для користувачів; ступінь, відповідно до якого дані подано за допомогою відповідної мови, символів та одиниць і відповідно до якого ці значення є зрозумілими [5, 6].

Зручність – можливість отримувати доступ до даних, зокрема людям, яким необхідні допоміжні технології або спеціальна конфігурація через певні вади.

Портативність – можливість переміщати дані з однієї платформи на іншу; можливість встановлення та переміщення даних у призначеній платформі. Це особливо стосується однорідного

та узгодженого набору даних (тобто тих даних, які запитуються програмним додатком). Інформація про процеси, що відбуваються у зовнішньому середовищі, може характеризуватись великими обсягами і трудомісткістю первинного накопичення та актуалізації, що визначає необхідність її ретельного зберігання та регламентованої зміни. Критерієм портативності є швидкодія, яка виражається часом відгуку ІС на зовнішнє звертання (час реакції на зміну параметрів зовнішнього середовища, до яких чутлива система).

Відновлення – здатність даних для підтримання та збереження зазначеного рівня діяльності, фізичної та логічної цілісності, навіть у разі виходу системи з ладу.

Цей стандарт стосується якості даних, а не системи: характеристика зворотності пов'язана з даними, які повинні бути відновлюваними.

Конфіденційність – це доступ до даних та здатність їх інтерпретувати тільки авторизованим користувачам. Наприклад, дані, які належать до особистої або конфіденційної інформації (стан здоров'я або прибуток), можуть бути доступними лише авторизованим користувачам або повинні бути написані секретним кодом.

Для того щоб оцінити корисність факту для інформаційної системи, необхідно визначити його трастову метрику або достовірність. Фактом, у нашому випадку, називається мінімальне знання про об'єкт, тобто це або значення атрибута об'єкта, або його зв'язки з іншими об'єктами. Достовірність (trustworthiness) визначає ступінь довіри до факту рядового користувача інформаційної системи. Для оцінки використовують характеристики джерел факту і враховують час його існування в інформаційній системі. Ці характеристики описані нижче.

Нехай F – деякий факт, T_i – i -й документ, в якому згадується факт F . Позначимо експертну оцінку документа T_i як $D_i \in [0,1]$. Експертна оцінка характеризує рівень довіри експерта до інформації з документа на підставі знань про джерело цього документа і, можливо, якоїсь додаткової інформації, якою володіє експерт. Межі інтервалу, в яких міститься значення експертної оцінки, відповідає граничним випадкам: повній довірі, якщо $D_i = 1$, й відповідно, повній недовірі за $D_i = 0$. Значення $D_i = \frac{1}{2}$ відповідає відсутності інформації про джерело в експерта. Значення за замовчуванням у випадку відсутності експертної оцінки обчислюють за формулою $D_i = \frac{(N-1)}{N}$, де N – кількість різних джерел, що містять документ.

Достовірність джерела визначається як ймовірність отримання від нього істинного твердження $D_n = P(s = 1)$. Апостеріорна достовірність в процесі поступової перевірки істинності s наданих n -м джерелом тверджень визначається за формулою:

$$D_{n,i+1} = \frac{D_{n,i}}{(2-s)} + \frac{1-D_{n,i}}{2} \cdot s, \quad (3)$$

де s – істинність твердження, що набуває значення 1 у випадку, якщо твердження істинне, або 0 – у протилежному випадку, i – номер кроку підтвердження/заперечення істинності одного з тверджень n -го джерела.

Зміна достовірності факту F в інформаційній системі описується ланцюжком пар $\langle j, \pi_j^F \rangle$, де j – момент часу, π_j^F – достовірність у момент часу j , тобто дискретною множиною або точкою у двовимірному просторі. Ухвалення рішення щодо факту на основі тільки поточного значення достовірності неефективне внаслідок того, що достовірність може опуститися нижче від мінімально допустимого значення у випадку похибки в оцінці, низького авторитету вибраного джерела або інших факторів. Для зменшення ступеня впливу подібних збурювань необхідно аналізувати окіл поточної точки. Аналіз дискретних околів також виявився неефективним, тому що не дає змоги прийняти рішення у випадку коливань достовірності навколо середнього значення. У цьому випадку ми можемо усереднити й оцінити значення в проміжних точках, апроксимуючи або інтерполюючи наявну множину точок або її підмножину гладкою кривою. Окремі сегменти кривої дають змогу оцінити рівень довіри в заданому околі поточного моменту часу без врахування впливу попередньої історії. Для розв'язування цієї задачі проаналізовано різні методи апроксимації й

інтерполяції кривими і за підсумками вибрано метод апроксимації В-сплайном [7]. Щоб прийняти рішення про подальшу долю факту, довіра до якого на поточний момент часу опустилася нижче від мінімально допустимого, пропонується виділяти загальну тенденцію поведінки, ґрунтуючись при цьому на аналізі кривої апроксимації хвоста з декількох значень.

Онтологічна модель нозологій

Для розв'язання задачі проблеми доступу до інформаційних ресурсів особам з різними фізичними вадами побудовано онтологію нозологій (рис. 1). Джерелом для побудови онтології нозологій стала їх класифікація та підходи до подолання наслідків різних вад (рис. 2)

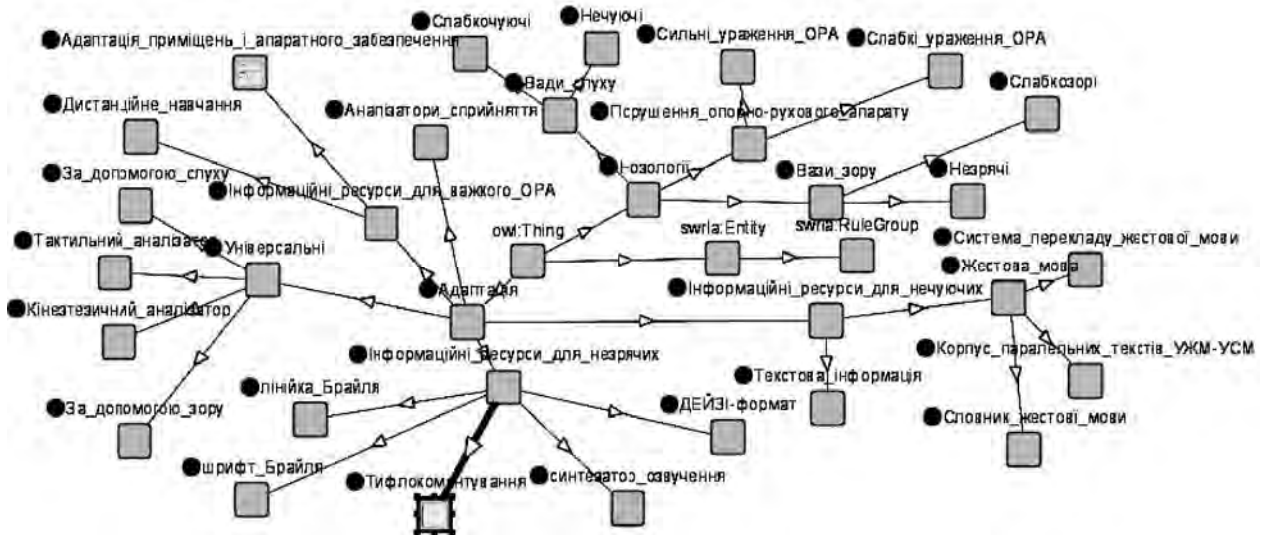


Рис. 1. Онтологія нозологій для уможливлення доступу до інформаційних ресурсів особам з різними типами вад

Відповідна класифікація та підходи переведені у програмний засіб Protégé.

Protégé – це вільний, відкритий редактор онтологій і фреймворк для побудови баз знань для різних предметних областей. Платформа Protégé підтримує два основних способи моделювання онтологій за допомогою редакторів Protégé-Frames і Protégé-OWL. Онтології, побудовані в Protégé, можуть бути експортовані в безліч форматів, зокрема RDF (RDF Schema), OWL і XML Schema.

Для реалізації цієї онтології у редакторі Protégé розроблено ієрархію класів (рис. 3) та відповідні властивості (див. рис. 4). На рис. 4 наведено властивість:

доступ_зір_сильні

де доменом (областю визначення) є термін (клас):

Незрячі

а ренгом (множиною значень) є терміни (класи):

Універсальні та Інформаційні ресурси для незрячих.

Онтологія містить понад 40 понять, 40 % понять є визначеними. Запити до онтології показали, що показник її функціональної придатності – близько $\chi_1 = 90$ %.

Як зазначено вище, для отримання показників періоду експлуатації та вартості робіт використаємо мову запитів SPARQL до онтології. Приклад таких запитів наведено нижче.

Приклад 1. Запит:

```
PREFIX table: <http://www.owl-ontologies.com/Ontology1253189272>
SELECT *
FROM <http://www.owl-ontologies.com/Ontology1253189272.owl>
where { {
$Cleaning rdfs:comment $value.
$Cleaning rdfs:subClassOf <#Нозології>}}
ORDER BY ASC(?value)
```

повертає типи нозологій: Вад слуху, Вад зору, Порушення опорно-рухового апарату.

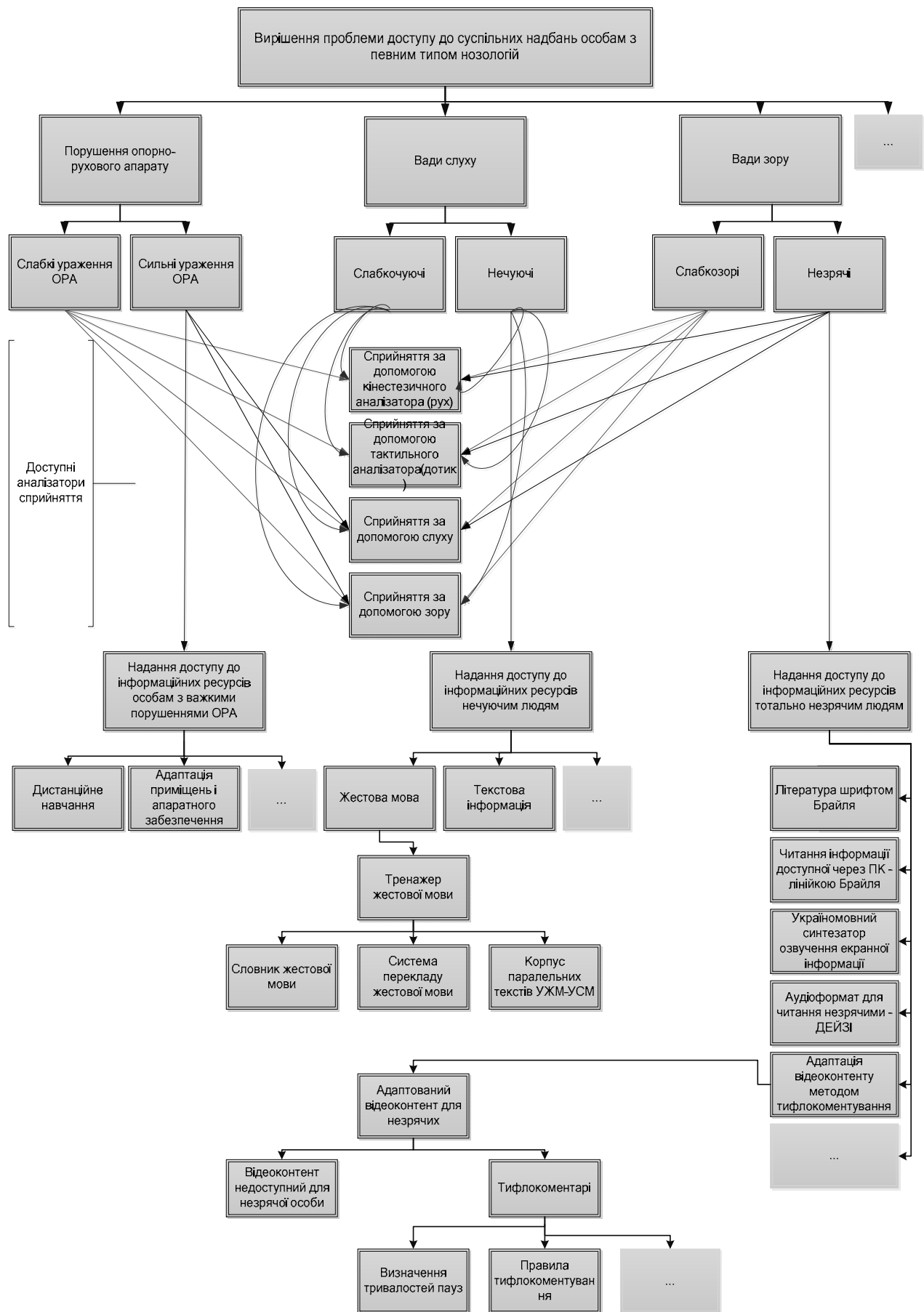


Рис. 2. Класифікація різних типів нозологій і можливості доступу до інформаційних ресурсів за допомогою аналізаторів, що працюють

Приклад 2. Запит:

```
PREFIX table: <http://www.owl-ontologies.com/Ontology1253189272>  
SELECT *  
FROM <http://www.owl-ontologies.com/Ontology1253189272.owl>  
where {{  
$Method rdfs:comment $value.  
$ Method_sight rdfs:subClassOf <# Method_sight>.  
$Period > 20.}}  
ORDER BY ASC(?value)
```

повертає методи доступу до відеоконтенту для осіб з вадами зору, яким вже більше ніж 20 років, тобто тифлокоментування.

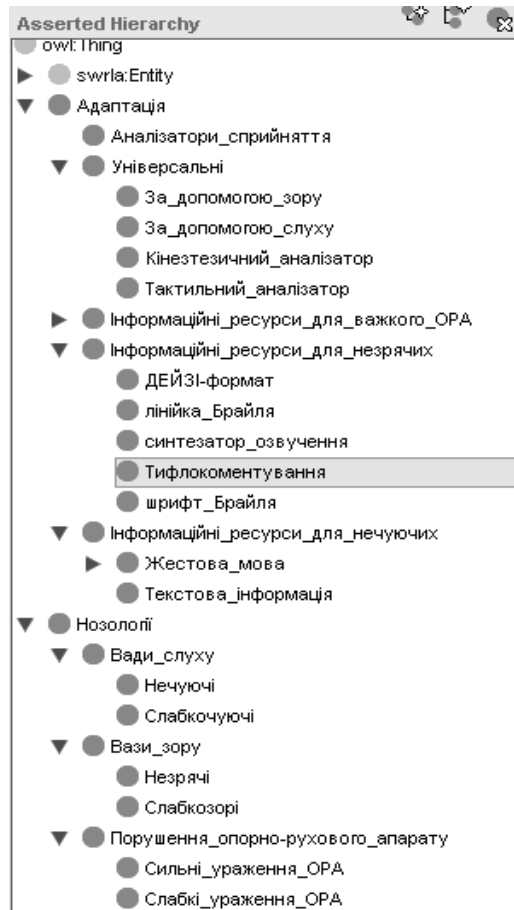


Рис. 3. Ієрархія класів

Рис. 4. Відповідні властивості

Висновки і перспективи подальших наукових розвідок

Отже, проаналізовано стан досліджень та розробок у галузі онтологічного інжинірингу. Запропоновано використати ISO/IEC 25012 для оцінювання якості онтологій баз знань інтелектуальних систем. Розглянуто основні показники стандарту та адаптовано їх для оцінки якості онтологій. Побудовано онтологію нозологій для уможливлення доступу до інформаційних ресурсів особам з різними типами фізичних вад. Подальше дослідження спрямоване на оцінювання новизни знань, які пропонують додавати в онтологію.

1. Gruber T. *A translation approach to portable ontologies* / T.Gruber // *Knowledge Acquisition*. – 1993. – № 5 (2). – P. 199–220. 2. Литвин В. В. *Бази знань інтелектуальних систем підтримки прийняття рішень: монографія* / В. В. Литвин. – Львів: Видавництво Львівської політехніки, 2011. – 240 с. 3. *Інтелектуальні системи, базовані на онтологіях: монографія* / Д. Г. Досин, В. В. Литвин, Ю. В. Нікольський, В. В. Пасічник. – Львів : Цивілізація, 2009. – 414 с. 4. ISO/IEC 25012:2008 <https://www.iso.org/obp/ui/#iso:std:iso-iec:25012:ed-1:v1:en>. 5. AS/NZS ISO/IEC 25012:2013 *Software engineering—Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model*. 6. Литвин В. В. *Підхід до побудови інтелектуальних систем підтримки прийняття рішень на основі онтологій* // *Проблеми програмування: наук. журн. / Національна академія наук України; Інститут програмних систем*. – Київ, 2013. – № 4. – С. 43–52. 7. Литвин В. В. *Метод автоматизованої розбудови та оцінювання якості онтологій баз знань* / В. В. Литвин, М. Я. Голяк, А. Б. Демчук // *Автоматизовані системи управління та прилади автоматики : Всеукр. міжвід. наук.-техн. зб.* – Харків, 2012. – № 161. – С. 62–69. 8. Де Бор К. *Практическое руководство по сплайнам*. – Москва: Радио и связь, 1985. – 304 с.