

ЗАСТОСУВАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ ОЦІНЮВАННЯ КОМПЕТЕНТНОСТІ ФАХІВЦІВ З РОЗРОБЛЕННЯ НАЦІОНАЛЬНИХ СТАНДАРТІВ

А.Я. ГЛАДУН, Ю.В. РОГУШИНА

Анотація. Необхідність і актуальність упровадження національних стандартів та їх гармонізації з міжнародними стандартами проаналізовано на прикладі сфери інформаційних технологій і тих специфічних проблем, що виникають при цьому. Запропоновано об'єктивні методи оцінювання компетентності експертів та розробників національних стандартів України. Подані методи ґрунтуються на семантичному зіставленні опису стандарту та документів, що характеризують компетентність його розробників в обраній сфері. Запропоновано використовувати знання, що здобуваються з відкритого середовища Web: Вікі-довідники, наукометричні бази, офіційні сайти організацій і супровідні метадані, онтології предметних областей тощо. Для інтеграції цих інформаційних ресурсів розроблено спеціалізовану онтологію наукової діяльності, що дозволяє стандартизувати термінологічну базу для опису кваліфікації потенційних експертів.

Ключові слова: стандартизація, онтологія, наукометричні показники, компетентність.

ВСТУП

Натепер майже всі сфери діяльності людини важко уявити без використання стандартів різного призначення. Стандарти акумулюють передовий науково-технічний досвід багатьох країн, який націлений на забезпечення єдності вимог до продукції, що є предметом міжнародного товарообміну, і забезпечує взаємозамінність комплектувальних виробів, єдині методи випробування та оцінювання якості виробів. Стандартизація являє собою складну, комплексну, міжгалузеву, багатопланову та багатоаспектну проблему, яка потребує сучасних рішень для її успішного розвитку.

Основна мета міжнародної стандартизації — забезпечити міжнародний товарообмін та взаємодопомогу, а також розширити співпрацю у сферах інтелектуальної, наукової, технічної й економічної діяльності у світовому масштабі. Тому цьому питанню у зв'язку з процесами євроінтеграції приділяється велика увага в Україні. Світову систему стандартизації складають міжнародні, міждержавні, національні, галузеві та інші організації стандартизації: ISO; IEC; ITU; ETSI; IEEE; EIA; ANSI; SEI; ECMA; CEN; CENELEC тощо.

В ISO/IEC за розвиток і синхронізацію стандартів у сфері інформаційних технологій відповідає об'єднаний технічний комітет JTC 1 (Joint Technical Committee 1 – Information Technology), який має багато вузькопрофільних підкомітетів (SC — Subcommittee). Зокрема, один з них JTC 1/SC 6 — Telecommunications and information exchange between systems відповідає за розвиток стандартів з телекомунікацій та систем обміну інформа-

цією між системами – саме зі сфери його діяльності наведено подальші приклади в цій роботі. Діяльність зі стандартизації в Україні ґрунтується на правових нормах Закону України «Про стандартизацію» [1], Декрету Кабінету Міністрів України «Про стандартизацію і сертифікацію» та інших нормативних актах з урахуванням принципів і положень міжнародних організацій зі стандартизації [2, 3].

Етапи розроблення національного стандарту. Для того щоб оцінювати компетентність розробників стандартів, потрібно проаналізувати, якою саме діяльністю вони займаються на різних етапах розроблення стандартів та, які завдання вони мають вирішувати.

Організація-розробник попередньо проходить процедуру тендерних закупівель. Основною базою для реформування державних закупівель в Україні є новий Закон України «Про здійснення державних закупівель» № 1197-VII (1197-18) від 10.04.2014. Мета закону – створення конкурентного середовища у сфері державних закупівель, запобігання проявам корупції у цій сфері, розвиток добросовісної конкуренції.

Відповідно до затвердженої Програми робіт з національної стандартизації організація-розробник призначає відповідальних виконавців і визначає термін виконання етапів роботи. Потім створюється *технічне завдання* на розроблення стандарту. Для цього потрібно зібрати, вивчити та проаналізувати матеріали про об'єкт стандартизації з урахуванням досягнень вітчизняного та закордонного досвіду, зокрема виконувати патентні дослідження [4]. *Технічний комітет* зі стандартизації для виконання робіт визначає відповідний *підкомітет*, який формує робочу групу або залучає фахівців, найкомпетентніших у цій предметній області (ПрО). Підкомітет готує проект першої редакції стандарту згідно з договором і технічним завданням на розроблення стандарту. Усі етапи розроблення національного стандарту технічним комітетом стандартизації подано в ДСТУ 1.2:2015 [3].

Особливу увагу під час розроблення стандартів фахівці технічного комітету приділяють термінології. *Термінологія* — сукупність термінів, тобто слів та/або словосполучень, що виражають специфічні поняття з певної ПрО (галузі), а також сукупність усіх термінів, наявних у тій чи іншій мові. Від звичайних слів терміни відрізняються точністю семантичних меж [5].

Галузеві термінології (тобто сукупності термінів конкретних галузей) називають термінологічними системами (терміносистемами). Характеризуючи терміносистему в сфері інформаційних технологій (ІТ), необхідно відзначити динаміку її виникнення та інтенсивність збагачення новими лексичними одиницями. Фахівець з розроблення проекту стандарту повинен досконально знати ПрО розроблюваного стандарту, а також мати значний запас термінів з цієї спеціалізованої ПрО. Особливо це стосується процесу розроблення стандарту, гармонізованого зі стандартами, викладеними іншими мовами світу.

ПОСТАНОВКА ЗАВДАННЯ

Створення об'єктивних методів оцінювання компетентності претендентів на основі використання й оброблення знань про ПрО для прогнозування успішного розроблення стандартів є сьогодні актуальною і важливою науковою проблемою.

Оцінки можна отримати на основі аналізу формальних характеристик розробників стандарту (дипломів, кількості публікацій у зазначеній Про, стажу роботи тощо), однак вони є надто приблизними і не враховують семантику виконуваного завдання. Тому необхідно розробити методи, які здатні обробляти знання про ці характеристики, враховуючи динамічність, розподіленість та нечіткість таких знань. Пропонується використовувати для цього засоби онтологічного аналізу.

Використання онтологій для формалізованого подання знань про компетентність учасників виконання проекту (розробників, наукових редакторів, експертів тощо) потребує розроблення засобів зіставлення семантичного опису стандарту та відомостей про учасників, що характеризує їх знання і досвід у Про стандарту та враховує їх знання і досвід.

Таке зіставлення зводиться до аналізу двох множин природномовних документів і пов'язаних з ними метаданих та знань. Запропонований підхід дозволяє інтегрувати персоніфіковані онтологічні знання про експертів із Про і логічні методи їх оброблення, використовуючи онтологічний аналіз та методи семантичного пошуку.

Мета роботи — розроблення об'єктивних алгоритмів та методів оцінювання компетентності учасників розроблення стандартів, що ґрунтуються на здобутті знань з відкритого середовища Web та документів, які описують стандарти. Запропонований підхід дасть змогу інтегрувати відомості про розробників у певному домені IT та стандартизувати термінологічну базу.

ПРОФЕСІЙНА КОМПЕТЕНЦІЯ ТА КОМПЕТЕНТНІСТЬ ФАХІВЦІВ ЗІ СТАНДАРТИЗАЦІЇ

Термінологічні проблеми в сфері стандартизації. Основна складність перекладу текстів у сфері IT, зокрема переклад термінологічних одиниць, полягає у розкритті та передачі засобами української мови іншомовних понять. Інтенсивний розвиток науки і техніки разом із процесами глобалізації в сучасному суспільстві сприяє безперервному збагаченню термінологічної системи словами та словосполученнями, пов'язаними зі сферою IT. Комп'ютерна термінологія є найдинамічнішою із термінологічних систем, і в ній навряд чи колись можна буде поставити крапку та вважати її дослідженою, оскільки інновації в комп'ютерній техніці розвиваються досить швидко [6].

Сучасна українська науково-технічна мова не завжди містить релевантні терміни для перекладу нових понять або понять, що набули нового значення в іншій Про. Обов'язкова умова повноцінного перекладу будь-якого спеціального тексту — це повне розуміння його перекладачем. Дослівний переклад термінів без проникнення у їх сутність без знання самих явищ, процесів та механізмів, про які йдеться в оригіналі, може призвести до грубих помилок у перекладі та спотворення його семантики. Основними способами відтворення українською мовою комп'ютерної термінології є: калькування, транскодування, описовий переклад, еквівалентний переклад.

Під *калькуванням* розуміють переклад частин іншомовного слова (складного, похідного або словосполучення) з подальшим складанням елементів в одне ціле. Приклади калькування — переклад таких термінологічних словосполучень, як «access code» — «код доступу»; «abstract semantic network» — «абстрактна семантична мережа». Такий вид перекладу досить часто призводить до створення незрозумілих слів, які ускладнюють розуміння тексту.

Транскодування (транскрипція) — це передача звукової або графічної форми слова вихідної мови засобами абетки мови перекладу. Наприклад, «semantic web» перекладається як «семантичний веб» (а не «змістове павутиння»), «installation» — як «інсталяція», «notebook» — як «ноутбук», «startup» — як «стартап». Такий прийом найчастіше використовують для перекладу слів та словосполучень, які на практиці широко вживаються саме похідною мовою і не мають безпосередніх аналогів в українській мові. Нерідко в процесі перекладу транскрипцію та калькування застосовують одночасно.

Описовий переклад, або експлікація (від лат. explicatio — роз'яснення) — це лексико-граматична трансформація, за якої лексична одиниця мови-оригіналу замінюється словосполученням, яке дає пояснення або визначення цієї одиниці. Наприклад, «native mode» — «режим роботи у власній системі команд»; «processor-specific code» — «програма, прив'язана до певного процесора».

Якщо значення англійського терміна повністю та однозначно відповідає значенню українського, то це — *еквівалентний переклад*. Цей вид перекладу досить поширений, хоча в українській мові немає достатнього обсягу термінології, наприклад, у галузі ІТ.

Вибір способу перекладу досить часто залежить від компетентності перекладачів у Про тексті, які його перекладають: найкваліфікованіші з них здатні вибрати найвідповідніші терміни української мови і лише за їх відсутності звертатися до калькування та транскрипції.

Професійна компетенція розробників. Проблема розроблення стандартів присвячено багато методичних розробок [7], а аналіз їх актуальності та прогнозування ефективності є завданням експертів відповідних Про [8].

Варто розрізнати поняття «компетенція» і «компетентність». Компетенція — це знання, які повинна мати особа; коло питань, з якими особа повинна бути обізнана; досвід, який треба мати для успішного виконання певних завдань відповідно до встановлених прав, законів тощо, тобто поняття, що в загальному випадку не пов'язане з конкретною особою. Компетентність — це відносини між особою і компетенцією, що означає, що певна людина має цю компетенцію. У розгляданому випадку компетенція — це характеристики та завдання, які потребують вирішення під час розроблення проекту зі стандартизації, а компетентність — властивість розробника, його характеристики, досвід та можливості.

Елементи компетентності — знання і навички, життєвий досвід, здатності, риси характеру, інтелект тощо, об'єднані в різних конфігураціях для вирішення особою конкретних завдань. *Компетентність* — рівень досягнень (досвіду, знань, звичок) особи в певній Про. Компетентність можна визначати на підставі аналізу діяльності фахівця, його поінформованості

щодо досягнень науки і техніки, його розуміння досліджуваних проблем і можливих способів їх вирішення.

Варто враховувати, що в загальному випадку експерт — це компетентний фахівець. У сучасних умовах, коли відбувається експонентне збільшення обсягів наукової інформації, виконувати експертне оцінювання стає дедалі важче, і тому важливо якомога чіткіше визначати межі компетентності кожного потенційного експерта [9]. У сфері розроблення стандартів існують як формальні, так і неформальні вимоги до експертів.

Відкриті джерела відомостей про ПрО та фахівців. Основним джерелом відомостей про проект розроблення стандарту є його *природомовний* опис (заявка, технічне завдання, публікація міжнародного стандарту, програма робіт з національної стандартизації тощо), а також додаткові зовнішні інформаційні ресурси (ІР) — онтології, Вікі-сторінки тощо, які містять структуровані та семантично значущі відомості про ПрО.

Складнішим завданням є аналіз інформації про учасників розроблення стандартів. Частина відомостей про них є чітко формалізована і може бути однозначно оцінена без урахування семантики ПрО та додаткових знань про стандарт (рівень освіти, їх стаж роботи у відповідній сфері, наявність раніше розроблених стандартів). Але часто цих відомостей недостатньо для визначення компетентності щодо розроблення стандарту в нових і динамічно змінюваних ПрО. Одним з найбільш об'єктивних критеріїв оцінювання сфери компетентності наукових співробітників є аналіз їх публікацій, поданих у Web: наукових статей, доповідей, звітів і презентацій.

Таким чином, як джерела відомостей про сферу компетентності авторів проекту пропонується розглядати такі ІР: офіційні документи, що підтверджують освіту та досвід роботи (наприклад, диплом про вищу освіту, науковий ступінь, сертифікати та нагороди); ІР, що розкривають семантику офіційних документів розробників (паспорти спеціальностей та дисциплін, вимоги до здобування наукових звань та вчених ступенів, посадові інструкції, таксономія національних учених ступенів тощо); тексти опублікованих статей, тез доповідей, монографій, підручників, технічних звітів, патентів та інших об'єктів інтелектуальної власності, поданих у вигляді природомовних документів, опублікованих у Web, з урахуванням рейтингів наукометричних баз даних (НМБД); Вікі-сторінок авторів розроблення стандартів, що містяться у різних Вікіпедіях, і Вікі-довідників, які забезпечують структуроване подання інформації; Вікі-сторінок організацій та установ, з якими пов'язані автори розроблення (наприклад, членство в технічному комітеті зі стандартизації, співпраця з Національною академією наук України, навчальними закладами) [10].

НАУКОМЕТРИЧНІ ХАРАКТЕРИСТИКИ НАУКОВО-ДОСЛІДНОЇ ДІЯЛЬНОСТІ

Ефективність наукової діяльності окремих осіб, колективів та організацій може оцінюватися з використанням як якісних, так і кількісних показників. Основу якісних оцінок становлять висновки експертів. Однак суб'єктивність подібних оцінок значно зменшує достовірність отримуваних результатів, а відсутність кількісного вираження ускладнює їх використання.

Наукометричні показники є зручними для оцінювання результатів фундаментальних досліджень, затребуваність яких оцінюють через відгуки наукового співтовариства на публікації авторів. Статистичним дослідженням структури та динаміки потоків наукової інформації займається наукометрія — галузь наукознавства, яка вивчає еволюцію науки через числові вимірювання наукової інформації, такі як кількість наукових статей за певний проміжок часу, цитування тощо. Термін «наукометрія» введено в 1969 р. В.В. Налімовим [11]. Підвищений інтерес до наукометричних показників спричинено насамперед можливістю автоматизації процесу оцінювання результатів наукової діяльності [12].

Натепер під час формування рейтингу дослідника застосовують різні параметри — кількість публікацій — сумарно або окремо за типами (монографії, статті, тези, публікації у виданнях, проіндексовані у Web of Science, SCOPUS або Google Scholar тощо) і посилань на них. Іноді враховують обсяг публікацій. На їх основі формують інтегральні критерії, у яких бали за публікацію визначають її типом та імпаکت-фактором видання, у якому її опубліковано.

Ефективність наукової діяльності можна оцінювати, застосовуючи як якісні, так і кількісні показники. У праці [13] проаналізовано найефективніші та найпоширеніші характеристики продуктивності наукової діяльності, зокрема, індекс Гірша та імпакт-фактор.

У 2005 р. Х. Гірш запропонував новий показник — індекс Гірша (*h*-індекс) [14], який є максимальним цілим числом *h* і який означає, що автор опублікував *h* статей, кожна з яких наведено хоча б *h* раз. Простота розрахунків та нечутливість до типових прийомів штучного покращення розглянутих вище показників миттєво зробили індекс Гірша популярним наукометричним індикатором. Індекс Гірша можна обчислювати, користуючись як безкоштовними загальнодоступними НМБД в Інтернеті (наприклад, Google Scholar, Elibrary.ru, ADS NASA), так і БД із платною передплатою (наприклад, SCOPUS або ISI Web of Science); однак платні БД часто теж подають *h*-індекс учених у вільному доступі. Слід зазначити, що індекс Гірша має різні значення одного й того самого вченого залежно від проіндексованих IP. Індекс Гірша дає об'єктивніші результати в разі вилучення посилань авторів на власні статті. Наприклад, у рейтингу вчених України відповідно до індексу Гірша виконано обчислення з БД SCOPUS з вилученням посилань авторів на власні статті.

Імпакт-фактор показує, скільки разів у середньому наведено кожному опубліковану в журналі статтю протягом *x* наступних років після її виходу. Це числовий показник значущості наукового журналу, який щорічно розраховує Інститут наукової інформації (Institute for Scientific Information — ISI) і опубліковує в журналі Journal Citation Report. Імпакт-фактор дає змогу за формальними ознаками порівнювати різні журнали та дослідницькі групи.

Зазвичай обчислення імпакт-фактора ґрунтується на трирічному періоді. Імпакт-фактор журналу *A* за рік *x* обчислюють за формулою

$$\text{Im } p(A, x) = \frac{\text{Cit}(A, x - 2, x) + \text{Cit}(A, x - 1, x)}{\text{Pub}(A, x - 2) + \text{Pub}(A, x - 1)},$$

де $Cit(A, y, z)$ — кількість посилань протягом року z статей, опублікованих у журналі A в році y в публікаціях журналів, які відслідковує Інститут наукової інформації, а $Pub(A, y)$ — кількість публікацій в журналі A за рік y .

Індекс наведення — це прийнятий в науковому світі ступінь значущості наукової праці будь-якого вченого або наукового колективу, що є сумарною кількістю посилань у проіндексованих працях на розглянуті публікації. Зазвичай слабкі і вторинні роботи не наводять. Наведеність залежить не тільки від рівня наукових результатів, але й від інших факторів, наприклад, своєчасності.

Для отримання цих оцінок користуються НМБД — бібліографічними та реферативними БД з інструментами для відстеження наведеної статей, опублікованих у наукових виданнях.

Найвідоміша НМБД SCOPUS видавничої корпорації Elsevier. У SCOPUS не застосовують поняття імпаکت-фактора, але широко користуються індексом Гірша. База даних є доступною на умовах передплати через Web-інтерфейс (<http://www.SCOPUS.com>).

Не менш популярною є НМБД Web of Science (WoS) компанії Thomson Reuters. Вона містить посилання на повні тексти в першоджерелах і переліки всіх бібліографічних посилань, які трапляються в кожній публікації.

Менш відома міжнародна НМБД Index Copernicus (Польща) (<http://www.indexcopernicus.com>), яка охоплює індексування, ранжування і реферування журналів, а також є платформою для наукової співпраці та виконання спільних наукових проектів.

Найповніше представлено у Web наукові праці, проіндексовані в Google Scholar (<http://scholar.google.com/>) — вільно доступній пошуковій системі, що індексує повний текст наукових публікацій усіх форматів і дисциплін. Основний наукометричний показник, що генерує ця НМБД, — індекс Гірша (як загальний, так і за останні п'ять років).

Є також національні НМБД, орієнтовані на індексацію й оцінку публікацій мовами, відмінними від англійської. Наприклад, Web-сайт «Український індекс наукового цитування» (<http://uincit.uran.ua>). Передбачається, що сайт дасть змогу з показників публікаційної активності як окремих учених, так і наукових установ України, довідатися про показники наведеної їхніх робіт і ключові наукометричні показники.

ОНТОЛОГІЧНИЙ АНАЛІЗ У СТАНДАРТИЗАЦІЇ

Онтології ПрО, що характеризують певні стандарти, можна застосовувати не лише для оцінювання компетентності їх розробників та експертів. Вони є потужним інструментом для оброблення, аналізу та застосування знань, що містяться в цих стандартах, забезпечують їх автоматизоване оброблення та інтегрування [15].

Важливим питанням, пов'язаним з розробленням стандартів, є їх інтероперабельність і забезпечення їх автоматизованого пошуку та порівняння. Тому в процесі розроблення стандартів виникає потреба в побудові їх семантичних метаописів. Такими метаописами можуть стати онтології ПрО стандарту, у яких формалізовано основні поняття ПрО, що їх відображено в стан-

дарті, та зв'язки між ними. Онтологічний підхід полегшує коректний переклад стандартів і дає змогу визначити, якими саме близькими за змістом термінами треба користуватись у кожному окремому випадку.

Щоб виявити семантичну близькість між стандартами, потрібно зіставити їх онтології, знайти близькі за змістом поняття і на підставі цього обчислити семантичну відстань між стандартами. Саме онтології ПрО можуть стати джерелом знань для пошуку компетентних фахівців для розроблення національних стандартів, якщо за чинним стандартом побудувати онтологію та порівняти її з онтологіями фахівців, які можна отримати з аналізу результатів їх науково-технічної та навчально-методичної діяльності, що відображається в їх публікаціях.

Ще одна сфера застосування онтологій стандартів — автоматизована побудова галузей компетенції підкомітетів зі стандартизації, формалізованих через поєднання онтологій уже розроблених стандартів. Аналіз відношень між онтологіям стандартів дасть змогу також визначити порядок розроблення стандартів та з'ясування не лише формальних, але й семантичних зв'язків між ними. Наприклад, стандарт, у якому наведено певні терміни, потрібно розробляти раніше, ніж той, у якому ці терміни вже застосовано.

Розглянемо це на прикладі побудови онтології для стандарту [16] з таксономії профілів (рис. 1). У цьому стандарті застосовують терміни, наведені в розділі стандарту «Терміни та визначення понять». Усі ці терміни мають бути класами онтології відповідної ПрО, яка описує цей стандарт, і є підкласами класу «терміни стандарту». Спочатку в онтології створюють клас «стандарт», який описує основні характеристики стандарту, такі як назва, код, тематика, рік прийняття, розробники, обсяг тощо.

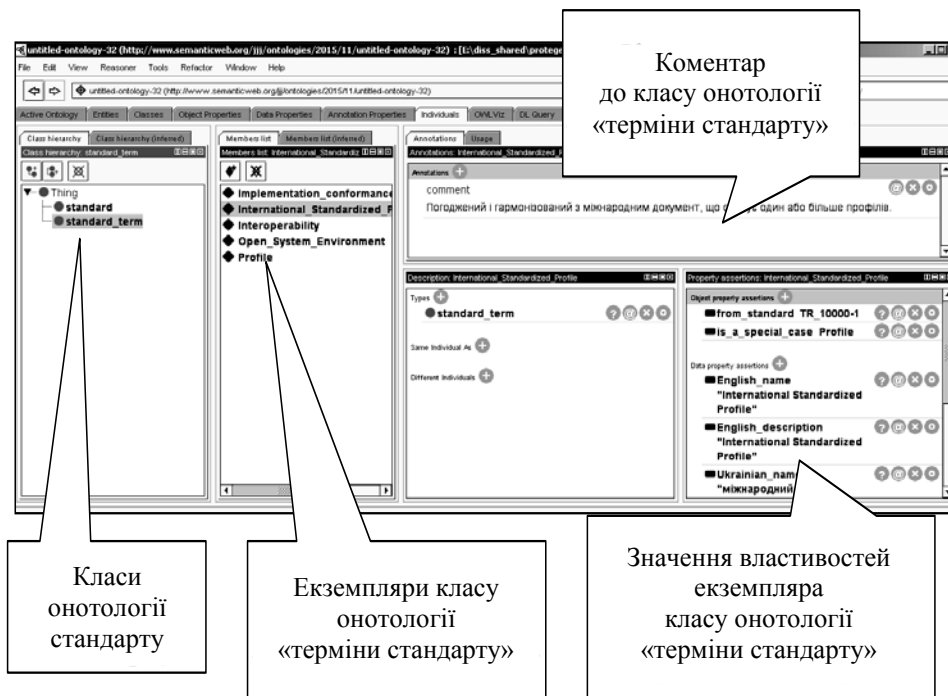


Рис. 1. Онтологія стандарту

Для інтегрованого використання знань, які подано в стандартах, доцільно також додати до опису класу «стандарт» зв'язки цього стандарту з іншими. Крім того, можна зазначити, які саме терміни належать до цього стандарту за допомогою властивості об'єктів «стандарт використовує термін».

Клас «терміни стандарту» має властивості даних (Data Properties): «назва стандарту», «назва терміна українською», «назва терміна англійською», «опис терміна українською», «опис терміна англійською» і «примітка» типу «рядок символів» та властивості об'єктів (Object Properties) «є підкласом» та «є синонімом», які дають змогу встановлювати семантичні зв'язки між елементами термінів стандарту. Онтологію ПрО для ДСТУ ISO/IEC TR 10000-1:2006 показано на рис. 2.

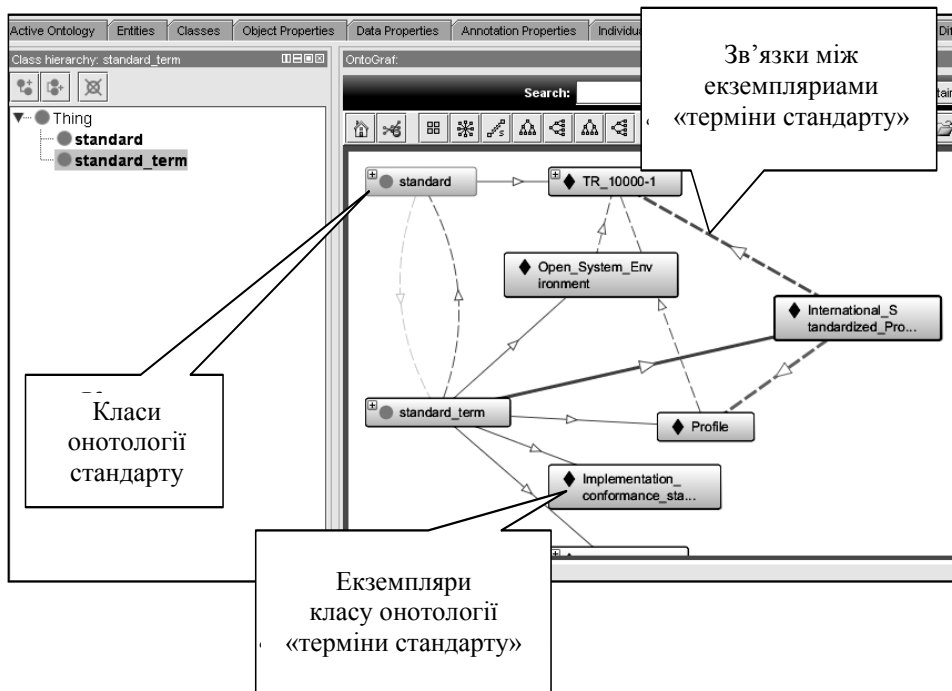


Рис. 2. Візуалізація зв'язків між класами, екземплярами класів та їх властивостями в онтології стандарту

Зв'язки між класами, екземплярами класів та їх властивостями в онтології стандарту відображають візуально так, як показано на рис. 2.

РОЛЬ ОНТОЛОГІЇ В ОЦІНЮВАННІ КОМПЕТЕНЦІЇ ЕКСПЕРТІВ

Крім загального професійного рівня, потрібно оцінювати компетентність розробників та експертів саме в тій ПрО, якої стосується конкретний стандарт. Формальною моделлю такої ПрО є її онтологія. Саме онтології є натеper найуживанішим засобом подання знань про ПрО, що дає змогу аналізувати і зіставляти компетенції експертів та розробників у нових дослідницьких галузях. Сам факт наявності в розробників стандарту онтології, що відповідає ПрО, зазвичай свідчить про ґрунтовніші пізнання у вибраній галузі (особливо якщо йдеться про інформаційні технології).

Разом з онтологіями ПрО доцільно застосовувати загальну онтологію наукової діяльності, яка дає змогу однозначно встановлювати термінологію, пов'язану з рейтингом публікацій, науковими ступенями та вченими званнями, типами організацій тощо (рис. 3). Таку онтологію спеціально розроблюють для визначення компетентності авторів проекту на основі організаційних онтологій наукових організацій, класифікатора УДК та паспортів спеціальностей.

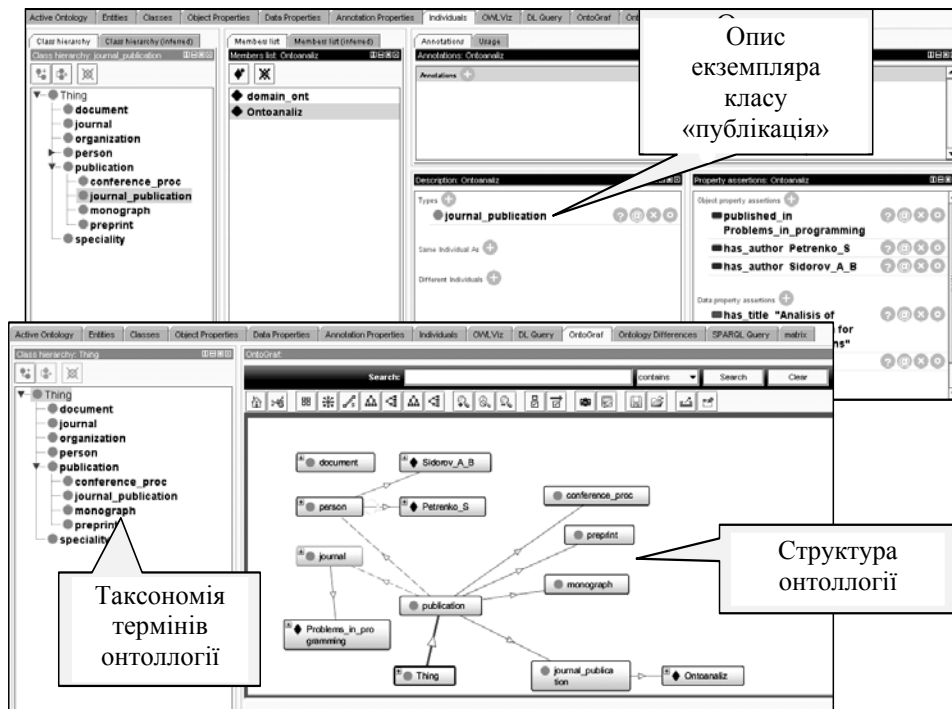


Рис. 3. Онтологія наукової діяльності для аналізу компетентності розробників

У такій онтології відображають, наприклад, такі взаємозв'язки між класами, як «бути співавтором», «працювати в організації», «бути автором публікації», «мати вчений ступінь за фахом» і властивості «бути публікацією», «мати індекс Гірша».

Цю онтологію запропоновано розробникам стандарту як зразок для опису та класифікації поданих документів, які мають засвідчити рівень їх компетентності в науковій діяльності загалом та в ПрО стандартизації зокрема.

ТЕЗАУРУС І ОНТОЛОГІЯ ПрО ЯК ЗАСІБ ЗІСТАВЛЕННЯ ПРИРОДОМОВНИХ ДОКУМЕНТІВ

Для зіставлення стандарту та компетенції розробників стандарту запропоновано побудувати тезаурус стандарту та тезауруси всіх ІР, які описують учасників розроблення стандарту. Тезаурус природомовного ІР є проєкцією онтології, аналізованої ПрО на документ, який досліджують тепер. Тезаурус учасників розроблення визначають як об'єднання тезаурусів окремих ІР.

При цьому слід зважати на вагу окремих ІР, у якій має бути враховано як значущість самого документа для опису компетенції, так і рівень довіри до самого ІР. Наприклад, більшу вагу має автореферат дисертації, ніж опис диплома, а під час оцінювання публікацій доцільно враховувати рейтинг різних журналів і конференцій у НМБД.

Припускають, що кожен з розробників стандарту сам формує набір документів, які є найбільш пертинентними запропонованому проекту стандарту. Наприклад, якщо у автора є n наукових публікацій, то він сам відбирає m з них, які пов'язані з тією проблемою, якої стосується розроблення стандарту, причому не обов'язково це повинні бути статті з найбільшим індексом цитування або ті, що входять до SCOPUS. Однак автор повинен прагнути до того, щоб для всіх понять онтології ПрО, для яких виявлено лінгвістичні аналоги в тексті стандарту, було знайдено такі відповідності і в його роботах (вагу кожного окремого зіставлення визначають вагою ІР, що обумовлений як функція від статусу документа і його рейтингу). Статус документа характеризує рівень його документального підтвердження, а рейтинг — його оцінку в НМБД [17].

Спочатку будують тезаурус стандарту Th_{proj} , який залежить від опису проекту розроблення стандарту та від вибраної онтології ПрО. Він являє собою множину пар (t_i, q_i) , де $t_i \in T$, T — множина термінів онтології ПрО $O_{ПрО} = \langle T, R, A \rangle$, а q_i — кількість знайдених відповідностей, що обумовлює вагу цього терміна. (Якщо певний термін трапився в описі проекту стандарту 10 разів, то припускають, що він важливіший, ніж той, який трапився 2 рази.) Для кожного терміна онтології ПрО пробують знайти фрагменти тексту.

Загальну оцінку компетентності колективу авторів визначають за допомогою масиву $\left(t_i, \sum_{j=1}^m p_{IR_{j_i}} * v_{IR_j} \right)$, де t_i — множина термінів онтології ПрО $O_{ПрО} = \langle T, R, A \rangle$; $p_{IR_{j_i}}$ — кількість знайдених відповідностей із цим терміном в j -му ІР; v_{IR_j} — вага j -го ІР.

Щоб визначити вагу j -го ІР, кожному документу попередньо призначають за замовчуванням вагу 1, яка може збільшуватися за такими умовами (оцінки можна підсумовувати):

- для наукових публікацій — якщо стаття є публікацією у виданнях з імпаکت-фактором понад 0,5, то +5, якщо матеріалами конференції, то +1, якщо її опубліковано в закордонному виданні, то +3, якщо проіндексовано в Google Scholar, то +2, а якщо в SCOPUS, то +10;
- для паспорта спеціальності, диплома +5; для опису профілю діяльності організації +3; для описів раніше виконаних проектів +3; для автореферату дисертації +3.

Важливим є те, що в цьому масиві є не всі терміни онтології ПрО (яка загалом може бути значно ширшою, ніж та, що її безпосередньо досліджують у проекті стандарту), а лише ті, для яких знайдено відповідності в цьому проекті стандарту.

Очевидно, що різні ІР мають різну вагу під час оцінювання компетенції фахівців. Наприклад, публікація в реферованому журналі є більш пріоритетною, ніж технічний звіт, викладений у Web. Тому під час оцінювання публікацій доцільно враховувати імпакт-фактор видання, яке опублікувало статтю.

У такому випадку загальну оцінку компетентності авторів можна подати таким чином:

$$C = \sum_{i=1}^n q_i \left(\sum_{j=1}^m p_{IR_{j_i}} v_{IR_j} \text{Im } p(IR_{j_i}) \right), \quad (1)$$

де $\text{Im } p(IR_{j_i})$ — імпакт-фактор видання, що опублікувало ІР.

Надалі доцільно ввести різні нормалізовані коефіцієнти, які дають змогу зменшити вплив документів великого обсягу, що є слабонасиченими термінами ПрО. Однак це потребує детальнішого дослідження вмісту документів, надаваних авторами на експертизу, і значною мірою залежить від специфіки проведеного конкурсу.

Оцінка (1) не використовує семантику ПрО і не враховує зв'язки між термінами в онтології ПрО. Для коректнішого врахування цих зв'язків запропоновано використовувати таку оцінку:

$$C = \sum_{i=1}^n q_i \left(\sum_{j=1}^m p_{IR_{j_i}} v_{IR_j} \right) s_i,$$

де параметр s_i визначає значущість i -го терміна онтології через кількість його зв'язків з іншими термінами онтології, що також увійшли до тезаурусу проекту стандарту, з урахуванням семантичної відстані між ними.

Щоб урахувати, крім спеціалізації у вибраній ПрО, загальну кваліфікацію кожного з розробників проекту стандарту, доцільно враховувати в оцінюванні і їх рейтинг як дослідників, отриманий з НМБД. Зокрема, запропоновано користуватись відомостями з Google Scholar і SCOPUS, оскільки ця інформація є відкритою й доступною для всіх користувачів Web. Крім того, це дає змогу диференціювати кваліфікацію окремих авторів, а не просто підсумувати їх результати. При тому, що оцінювання якості формування колективу розробників не є завданням експертів, однак наявність у ньому людей із кваліфікаціями, що значно розрізняються, може викликати здивування. Запропоновано користуватись таким критерієм оцінювання x -го автора проекту:

$$C_x = \sum_{i=1}^n q_i \left(\sum_{j=1}^m p_{IP-x_{j_i}} v_{IP-x_j} \right) s_i h_x,$$

де h_x — сума індекса Гірша вченого з Google Scholar та SCOPUS.

Загальну оцінку кваліфікації розробників проекту стандарту можна оцінювати як суму оцінок учасників або як їх нормалізовану суму. Перший підхід має переваги, оскільки знання й досвід кожного можна застосовувати незалежно від кількості учасників. Тому нормалізовану оцінку може бути застосовано лише як додаткову.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Запропоновано новий підхід до проблеми об'єктивного оцінювання компетентності в контексті нових інформаційно-комунікаційних технологій, що характеризуються високою динамікою розвитку наукових досліджень, використання й оброблення інформаційних ресурсів про експертів Про, яка нас цікавить. Метою розроблення поданих методів є об'єктивізація процесу оцінювання кваліфікації потенційних учасників з урахуванням значної кількості доступних знань про їх кваліфікацію в сфері розроблення та стандартизації.

Запропонований підхід дає змогу на підставі опису проекту та природомовної інформації про учасників проекту з розроблення стандартів (їх публікацій, дипломів, описів раніше виконаних проектів, відомостей про їхні організації тощо) будувати тезаурус проекту й тезауруси учасників, які базуються на онтології Про проекту з розроблення стандарту, зіставляти ці тезауруси і визначати на основі цього рівень компетентності учасників.

Надалі результатом цієї роботи має стати створення *глобальної семантичної мережі стандартів*, яка пов'язує окремі національні та міжнародні стандарти; об'єкти, що використовують ці стандарти та посилаються на них (як матеріальні, так і інформаційні об'єкти); фахівців, які є експертами в сфері розроблення стандартів, та організації різного рівня, що підтримують різні види діяльності, пов'язаної з розробленням та застосуванням стандартів. Інтероперабельне подання знань та застосування відповідних технологій і форматів дасть змогу інтегрувати цю мережу знань зі знаннями, поданими в Semantic Web.

ЛІТЕРАТУРА

1. Закон України «Про стандартизацію». Верховна Рада України, Закон від 05.06.2014, № 1315 – VII. — <http://zakon0.rada.gov.ua/laws/show/1315-18>
2. ДСТУ 1.1:2015 Національна стандартизація. Стандартизація та суміжні види діяльності. Словник термінів. — Чинний від 20.12.2015.
3. ДСТУ 1.2:2015 Національний стандарт України. Національна стандартизація. Правила проведення робіт з національної стандартизації. — Чинний від 20.12.2015.
4. ДСТУ 1.7:2015 Національна стандартизація. Правила та методи прийняття міжнародних і регіональних нормативних документів. — Чинний від 20.12.2015.
5. Гладун А.Я. Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу / А.Я. Гладун, Ю.В. Рогушина // Искусственный интеллект, 2008. — № 5. — С.112–124.
6. Грицик Н. Комп'ютерна термінологія та основні способи її перекладу / Н. Грицик // XVIII Міжнародна науково-практична Інтернет-конференція «Проблеми та перспективи розвитку науки на початку третього тисячоліття», Переяслав-Хмельницький, 2013. — С. 45–52.
7. Цициліано О. Креативність та стандартизація як основа конкурентоспроможності організації / О. Цициліано // Стандартизація. Сертифікація. Якість. — 2007. — № 1 (44). — С. 42–44.

8. *Доманцевич Р.І.* Основи стандартизації, метрології та управління якістю / Р.І. Доманцевич, І.С. Полікарпов, Б.П. Яцишин. — К.: НМЦ «Укоопосвіта», 1997. — 219 с.
9. *Костенко Л.Й.* Наукова періодика України та бібліометричні дослідження / Л.Й. Костенко, О.І. Жабін, Є.О. Копанєва, Т.В. Симоненко // НАН України, Нац. б-ка України ім. В. І. Вернадського. — К., 2014. — 173 с.
10. *Rogushina J.* Ontology-based competency analysis in new research domains / J. Rogushina, A. Gladun // Journal of Computing and Information Technology. — 2012. — 23, N 4. — P. 123–134.
11. *Налимов В.В.* Наукометрия: Изучение развития науки как информационного процесса / В.В. Налимов, З.М. Мульченко. — М.: Наука, 1969. — 192 с.
12. *Штовба С.Д.* Обзор наукометрических показателей для оценки публикационной деятельности ученого / С.Д. Штовба, Е.В. Штовба // Наукометрия и экспертиза в управлении наукой. — М.: ИПУ РАН, 2013. — С. 262–278.
13. *Бурков В.Н.* Параметры цитируемости научных публикаций в наукометрических базах данных / В.Н. Бурков, А.А. Белошицкий, В.Д. Гогунский // Управління розвитком складних систем. — 2013. — Вып. 15. — С. 134–139.
14. *Hirsch J.E.* An index to quantify an individual's scientific research output / J.E. Hirsch // Proc. Of the National academy of Sciences of the United States of America. — 2005. — 102(46). — P. 16569–16572.
15. *Гладун А.Я.* Семантичні технології: принципи та практики: моногр. / А.Я. Гладун, Ю.В. Рогушина. — К.: Універсаріум, 2016. — 314 с.
16. *Гладун А.Я.* ДСТУ ISO/IEC TR 10000-1:2004 «Інформаційні технології. Основи та таксономія міжнародних стандартизованих профілів. Частина 1. Загальні принципи та основи документування». — К.: Держспоживстандарт, 2004. — 20 с. — Чинний від 01.01.2006.
17. *Рогушина Ю.В.* Онтологічний аналіз у Web: моногр. / А.Я. Гладун, В.В. Осадчий, С.М. Прийма. — Мелітополь: МДУПУ ім. Богдана Хмельницького, 2015. — 407 с.

Надійшла 15.04.2016