

ОНТОЛОГІЧНИЙ ПІДХІД ДО ПРОБЛЕМ ПІДВИЩЕННЯ ЯКОСТІ РОЗРОБЛЕННЯ НАЦІОНАЛЬНИХ СТАНДАРТІВ УКРАЇНИ



А. Я. Гладун

А. Я. Гладун, кандидат технічних наук, доцент, старший науковий співробітник, голова ПК-6 ТК-20 «Телекомунікації та обмін інформацією між системами», Міжнародний науково-навчальний центр інформаційних технологій та систем НАНУ,
Ю. В. Рогушина, кандидат фізико-математичних наук, доцент, старший науковий співробітник, Інститут програмних систем НАНУ, м. Київ

ОНТОЛОГИЧЕСКИЙ ПОДХОД К ПРОБЛЕМАМ ПОВЫШЕНИЯ КАЧЕСТВА РАЗРАБОТКИ НАЦИОНАЛЬНЫХ СТАНДАРТОВ УКРАИНЫ

А. Я. Гладун, кандидат технических наук, доцент, старший научный сотрудник, председатель ПК-6 ТК-20 «Телекоммуникации и обмен информацией между системами», Международный научно-учебный центр информационных технологий и систем НАНУ,
Ю. В. Рогушина, кандидат физико-математических наук, доцент, старший научный сотрудник, Институт программных систем НАНУ, г. Киев

ONTOLOGICAL APPROACH TO THE PROBLEMS OF IMPROVING THE QUALITY OF DEVELOPMENT OF NATIONAL STANDARDS OF UKRAINE

A. J. Gladun, Candidate of Technical Sciences, Associate Professor, Senior Researcher, Chairman of PC-6 TC-20 «Telecommunications and Information Exchange Between Systems», International Research and Training Center of Information Technologies and Systems, NASU,
Y. V. Ragulina, Candidate of Physico-mathematical Sciences, Associate Professor, Senior Researcher, Software Systems Institute, NASU, Kiev



Ю. В. Рогушина

Онтологічний аналіз використовується як основа для розроблення спільної та узгодженої термінологічної бази національних стандартів України та створення методичних рекомендацій, спрямованих на підвищення їхньої якості. Розглянуто задачі, для вирішення яких застосовані онтології та тезауруси стандартів: семантичний пошук, вибір та оцінювання компетентності експертів та розробників, порівняння стандартів, використання відображених у стандартах знань у відкритих ресурсах Web. Запропоновано методи застосування онтологій та тезаурусів для вирішення цих задач.

Ключові слова: стандартизація, онтологія, компетентність, семантичний пошук, тезаурус.

Онтологический анализ используется как основа для разработки совместной и согласованной терминологической базы национальных стандартов Украины и создание методических рекомендаций, направленных на повышение их качества. Рассмотрены задачи, для решения которых применены онтологии и тезаурусы стандартов: семантический поиск, выбор и оценка компетентности экспертов и разработчиков, сопоставление стандартов, использование отраженных в стандартах знаний в открытых ресурсах Web. Предложены методы применения онтологии и тезаурусов для решения этих задач.

Ключевые слова: стандартизация, онтология, компетентность, семантический поиск, тезаурус.

In this work an ontological analysis is used for development and harmonization of the terminological base of national standards of Ukraine and for creation of methodological recommendations deal with their quality. Tasks of usage of standard ontologies and thesauri such as semantic search, selection of experts and developers, matching of standards, standard knowledge discovery from the Web open sources are analyzed. Methods of solution of these problems are proposed.

Keywords: standardization, ontology, competence, semantic search, thesaurus.

Вступ

Стандартизація — це складна, міжгалузева, комплексна й багатоаспектна задача, рішення якої пов'язане із розробленням сучасних методів та технологій. Практично усі сфери діяльності людини важко представити без використання стандартів, які акумулюють передовий науково-технічний досвід багатьох країн, спрямований на забезпечення єдності вимог до продукції, яка є предметом міжнародного товарообміну, враховуючи взаємозамінність комплектувальних виробів, єдині методи випробувань та оцінювання якості виробів. Якість стандартів, що розробляються, залежить від багатьох чинників, включаючи компетентність розробників та узгодженість термінологічної бази, що безпосередньо впливає на ефективність їх використання.

Для створення об'єктивних методів для вирішення проблем, пов'язаних з розробленням державних стандартів України (оцінювання компетентності розробників, гармонізація термінологічної бази, реалізація семантичного пошуку тощо), доцільно застосовувати сучасні технології подання та оброблення знань, які базуються на онтологічному аналізі. Це потребує розроблення структури відповідних онтологій, методів їх поповнення та застосування у рішенні прикладних задач.

Етапи розроблення національного ДСТУ

В Україні діяльність зі стандартизації ґрунтується на правових нормах Закону України «Про стандартизацію» [1], Декрету КМУ «Про стандартизацію та сертифікацію», інших нормативних актах у цій сфері, з урахуванням принципів і положень міжнародних організацій зі стандартизації [2, 3]. Згідно з затвердженим планом стандартизації організація, яка повинна розробляти проект стандарту, призначає відповідальних виконавців та визначає термін виконання окремих етапів роботи. Розробляється *технічне завдання* (ТЗ) на розроблення стандарту. З цією метою здійснюється збір, вивчення та аналіз матеріалів щодо об'єкта стандартизації з урахуванням досягнень вітчизняного та закордонного досвіду, зокрема виконуються патентні дослідження [4].

Технічний комітет зі стандартизації (ТК), який розробляє стандарт, для виконання робіт визначає відповідний *підкомітет* (ПК), який формує робочу групу (РГ) або залучає фахівців, компетентніших у цій предметній області (ПрО). ПК готує проект першої редакції стандарту згідно з договором і ТЗ на розроблення стандарту. Для підготовки проекту стандарту виконують науково-дослідні, проектно-конструкторські роботи, випробування тощо.

Онтологічний аналіз

У семантичних інформаційних технологіях під *онтологією* розуміється детальний опис деякої про-

блемної області, що використовується для формального і декларативного визначення її концептуалізації [5]. Часто онтологією називають базу знань спеціального виду, яку можна розділяти, відчувувати і самостійно використовувати в рамках розглянутої ПрО¹. Онтологія — це точна специфікація певної області, що містить словник термінів цієї області та множину зв'язків, що описують, як ці терміни співвідносяться між собою. Аналіз публікацій показує, що саме онтології є адекватними та ефективними засобами для моделювання уявлень щодо різноманітних ПрО, які дозволяють формально подавати їх семантику.

У різних джерелах пропонуються різні формальні моделі представлення онтологій [6]. Проте схожим у всіх є наступне:

- множина термінів (понять, концептів), яка може бути поділеною на множину класів і множину екземплярів;
- множина відношень між поняттями, у якій можуть явно виділятися відношення «клас-підклас», ієрархічні (таксономічні) відношення і відношення синонімії (подоби), а також функції — спеціальний випадок відношень;
- аксіоми і функції інтерпретації понять і відношень.

Формально онтологія представляється трійкою $\langle X, R, F \rangle$, де X — множина концептів, R — множина відношень між концептами, F — функції інтерпретації концептів з множини X та відношень з R .

Поняття онтології перетинається з уже давно прийнятим в інформатиці та лінгвістиці поняттями *тезауруса* та *терміносистеми*. Онтології дозволяють представити поняття в такому вигляді, що вони стають придатними для машинного опрацювання.

Онтології та семантична розмітка стандарту

Текст стандарту, розміщений у Web, може бути простим документом без спеціальної розмітки для інформаційно-пошукових систем (ІПС). Звичайна людина, прочитавши його, може скласти коротке резюме в одній-двох фразах, виділивши найнеобхідніше. У випадку ІПС їй доведеться аналізувати документ традиційними методами, видобуваючи і структуруючи необхідну інформацію. Якими б досконало не були алгоритми ІПС, якість пошуку завжди буде недосконалою, особливо якщо потрібний не просто пошук за ключовими словами, а встановлення основного змісту (суті контенту). У випадку ж застосування семантичної розмітки, де інформація чітко структурована, і тегами виділені потрібні категорії, опрацювання буде під силу алгоритмам

¹ Предметна область (або домен) — це частина реального чи віртуального (абстрактного) світу, теорії концепції, явища, події тощо, що розглядається в межах певного контексту. Під контекстом розуміється область дослідження або область, яка є об'єктом певної діяльності.

ІПС. Таким чином ІПС можна підказати, в якій частині тексту розміщена інформація певної категорії. Використання онтологій ПрО дозволяє здійснювати семантичну розмітку стандартів, що забезпечує виконання семантичного пошуку в їх контенті.

ПрО, для якої здійснюється семантична розмітка, характеризується набором термінів із скінченої множини термінів $T, T = \{t_1, \dots, t_m\}$. Ці терміни можуть використовуватися як теги семантичної розмітки. Довільний фрагмент тексту $\langle t_p, \dots, t_q \rangle, p, q = 1, n, p < q$ може бути пов'язаний з одним чи декількома поняттями з T . Для цього відповідним тегом відзначають початок такого фрагмента, а парним йому, що закриває, — кінець фрагмента.

Семантична розмітка текстів дозволяє автоматизовано аналізувати їх надалі на семантичному рівні, виконувати над ними різні логічні операції, видобувати з них нові знання тощо. При формуванні семантичної розмітки потрібно використовувати не лише знання ПрО (чи хоча б її термінологічну базу), але і правила тієї конкретної природної мови, якою написано текст. На жаль, створення такої розмітки є нетривіальною і досить трудомісткою задачею. Семантична розмітка залежить і від того, які саме засоби використовуються для опису ПрО [7].

Методи оцінювання компетенції експертів та розробників стандартів

Оцінки, побудовані на основі аналізу формальних характеристик розробників стандарту (такі як дипломи, кількість публікацій у зазначеній ПрО, стаж роботи тощо), є занадто приблизними й не враховують семантику розв'язуваного завдання.

Тому створення об'єктивних методів оцінювання компетентності претендентів на основі онтологічного аналізу з метою прогнозування успішності реалізації розроблення стандартів є актуальною і важливою сьогодні науковою проблемою, рішення якої повинне базуватися на використанні й опрацюванні знань щодо тієї ПрО, в якій планується проект розроблення нового стандарту.

Серед кількісних методів оцінювання наукової діяльності можна виділити *методи аналізу кількісних характеристик первинних документів* (контент-аналіз, метод підрахунку кількості публікацій, тезаурусний метод тощо) та *методи кількісного аналізу вторинних джерел інформації*, зокрема метод наукового цитування.

Для забезпечення об'єктивних методів оцінювання компетентності учасників виконання проекту зі стандартизації (розробників, наукових редакторів, експертів тощо) необхідно розробити засоби зіставлення опису стандарту і відомостей щодо його учасників, що характеризує їхні знання й досвід у ПрО стандарту та враховує їхні знання в цій ПрО. Таке зіставлення зводиться до аналізу двох множин при-

родномовних документів і пов'язаних з ними метада-них і знань. Розглянутий підхід дозволить інтегрувати персоніфіковані онтологічні знання щодо експертів у ПрО і логічні методи їхнього опрацювання, використовуючи онтологічний аналіз і методи семантичного пошуку.

Термінологія для розроблення стандартів

Особлива увага під час розроблення стандартів ДСТУ приділяється фахівцями ТК термінологічній науковій базі (термінології), яка була накопичена експертами певної предметної області протягом діяльності комітету зі стандартизації. *Термінологія* — сукупність термінів, тобто слів або словосполучень, що висловлюють специфічні поняття з ПрО (галузі), а також сукупність усіх термінів, наявних у тій чи тій мові. Від звичайних слів терміни відрізняються точністю семантичних меж [8].

Термін — це слово або словосполучення, яке порівнюється з чітко окресленим поняттям ПрО і вступає в системні відношення з іншими одиницями мови, утворюючи разом з ними особливу систему — термінологію.

Сучасні дослідники виділяють такі основні ознаки терміну: 1) має чітке визначення, зафіксоване в словнику; 2) однозначний в межах певної термінологічної системи (або має тенденцію до однозначності); 3) точний та не залежить від контексту; 4) стилістично нейтральний; 5) не має синонімів у межах обраної терміносистеми; 6) системний; 7) виражений коротко. Принциповим у відборі термінів є цілісність системи понять та термінів, що змістовно підпорядковуються певній теорії або концепції.

Галузеві термінології (тобто сукупності термінів конкретних галузей) називають *терміносистемами*, або *термінологічними системами*. Характеризуючи сучасну термінологію в сфері інформаційних технологій, необхідно відзначити динаміку її виникнення та інтенсивність збагачення новими лексичними одиницями.

Незалежно від ПрО, кожна терміносистема має характерні спільні риси: 1) сукупність термінів, що виконує комунікативно-тематичне чи комунікативно-ситуативне завдання; 2) елементи структури пов'язані між собою (безпосередньо або опосередковано); 3) кожен термін має належати хоча б до однієї терміносистеми; 4) має чіткі межі; 5) терміносистеми можуть перетинатися; 6) терміносистему можуть використовувати одна або кілька ПрО (міжгалузєва термінологія) [9].

Терміносистема є основою для побудови онтології стандарту, яка описує його семантику та дозволяє як здійснювати пошук в ньому, зіставляти його як з іншими стандартами, так і з різноманітними документами, наприклад, тими, що характеризують компетенції експертів та розробників стандартів. ▶

Проблеми термінології в стандартизації

Основна складність перекладу тексту у сфері інформаційних технологій, а саме переклад термінологічних одиниць, полягає у розкритті та переданні засобами української мови іншомовних реалій. Інтенсивний розвиток науки і техніки разом із процесами глобалізації у сучасному суспільстві приводять до безперервного збагачення термінологічної системи словами і словосполученнями, пов'язаними зі сферою інформаційних технологій. Комп'ютерна термінологія є найбільш динамічною із термінологічних систем, і в ній навряд чи колись можна буде поставити крапку та вважати дослідженою, оскільки інновації в комп'ютерній техніці розвиваються досить швидко [10].

На жаль, сучасна українська науково-технічна мова не завжди містить адекватні терміни для перекладу нових понять або понять, що набули нового значення в іншій предметній області. 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 перекручування 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, наприклад, «access code» — «код доступу»; «autorepeat» — «автоматичний повтор»; «absolute disk read» — «абсолютне зчитування з диску»; «abstract semantic network» — «абстрактна семантична мережа». Такий вид перекладу досить часто призводить до створення незрозумілих та нехарактерних для української мови слів, які ускладнюють розуміння тексту.

Транскодування (транскрипція) — це передання звукової або графічної форми слова вихідної мови засобами абетки мови перекладу. Наприклад, «Semantic Web» перекладається як «Семантичний Веб» (а не «змістовне 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 та не мають безпосередніх аналогів в українській мові, а калькування не відображає особливості термінів. Нерідко в процесі перекладу транскрипцію і калькування використовують одночасно: «control panel» — «панель керування»; «hot keys» — «гарячі кнопки».

Описовий переклад, або експлікація (від лат. explicatio — роз'яснення) — це лексико-граматична трансформація, за якої лексична одиниця мови-

оригіналу замінюється словосполученням, яке пояснює або визначає цю одиницю. За допомогою експлікації перекладають багатокomпонентні термінологічні словосполучення, наприклад, «native mode» — «режим роботи у власній системі команд»; «processor-specific code» — «програма, прив'язана до певного процесора»; «policy module» — «модуль керування використанням ресурсів».

Якщо значення англійського терміна повністю та однозначно відповідає значенню українського, то це — *еквівалентний переклад*. Цей вид перекладу досить поширений, хоча в українській мові існує не достатній обсяг термінології, приміром, у галузі ІТ. Можна навести такі приклади еквівалентного перекладу: «keyboard» — «клавіатура»; «network» — «мережа»; «memory» — «пам'ять»; «error» — «помилка».

Вибір способу перекладу досить часто залежить від компетентності перекладача у предметній області тексту, що перекладається: найбільш кваліфіковані здатні обрати найбільш відповідні терміни української мови, і лише за їх відсутністю звертатися до калькування та транскрипції.

Практичні рекомендації

стосовно визначення термінів тезаурусу

Для представлення знань про певну ПрО, документ, об'єкт чи суб'єкт широко використовують моделі знань про ПрО: тезауруси, таксономії, онтології. При створенні тезаурусу, визначення його термінів можуть формуватися автоматично (шляхом аналізу повнотекстових документів та інших інформаційних джерел), видобуватися з інших баз знань (тезаурусів, онтологій тощо) або створюватися вручну експертом ПрО. В усіх випадках доцільно дотримуватися певних правил та удосконалювати отримані визначення відповідно до наведених далі принципів [11].

Визначення терміну повинне:

- 1) бути викладене в однині (виключення складають поняття, які самі є множинними);
- 2) визначати, чим є наведене поняття, а не тільки чим воно не є;
- 3) мати вигляд описової фрази або речення;
- 4) містити лише широко відомі скорочення;
- 5) бути викладеним без використання визначень інших даних або базових понять;
- 6) відображати суттєвий зміст поняття;
- 7) бути точним та однозначним;
- 8) бути коротким;
- 9) мати можливість використовуватися окремо;
- 10) бути поданим без використання пояснювальної інформації, функціонального використання або процедурної інформації (якщо вони необхідні, то їх слід не включати до визначень, а розмішувати в інших атрибутах метаданих (ISO/IEC 11179-3:2013), але припустимо додавати декілька прикладів після визначення);

11) запобігати циклічних посилань — визначення не повинні визначатися одне через інше та використовувати визначення іншого поняття як своє;

12) використовувати ту ж саму термінологію та логічну структуру для пов'язаних визначень — для близьких або пов'язаних визначень слід використовувати однакову термінологію та синтаксис, тому що інакше користувачі можуть не зрозуміти, чому для пов'язаних визначень використовуються різні терміни.

Професійні компетенції та компетентність фахівців зі стандартизації

Проблемам розроблення стандартів присвячено багато методичних розроблень, а аналіз їхньої актуальності й прогнозування ефективності є завданням експертів відповідних предметних областей [13].

Експерти, що оцінюють розроблення, повинні мати досить глибокі знання в тій сфері, до якої належать усі стандарти в цілому, але можуть і не бути експертами в усіх аналізованих підобластях кожного окремого стандарту. У той же час потрібно, щоб розробники мали високу кваліфікацію саме в тій ПрО, якої стосується стандарт. Тому виникає необхідність у розробленні засобів автоматизованого оцінювання компетентності потенційних учасників проекту (розроблення стандарту) на основі зіставлення матеріалів, які описують їх знання, кваліфікацію й навички, з описом самого проекту. Необхідно, щоб у такому зіставленні використовувалися знання щодо ПрО у сфері стандартизації.

Варто розрізнати *компетенцію* й *компетентність*. *Компетенція* — це поняття, що не пов'язане з конкретною особою. У загальному значенні компетенція — здатність успішно діяти на основі наявних знань і практичного досвіду під час їх вирішення. Тобто компетенція — це властивість проекту зі стандартизації, а компетентність — властивість розробника. Елементи компетенції — знання й навички, життєвий досвід, здатності, риси характеру, інтелект тощо, об'єднані в різних конфігураціях для вирішення людиною конкретних проблем.

Компетентність — рівень досягнень (досвіду, знань, звичок) особи в певній ПрО. Компетентність може бути визначена на основі аналізу діяльності фахівця, його поінформованості щодо досягнень науки й техніки, його розуміння досліджуваних проблем і можливих шляхів їхнього рішення. Для кількісного оцінювання рівня компетентності використовується коефіцієнт компетентності.

Експерт — це компетентний фахівець, який здатен творчо розв'язувати складні проблеми, знає останні досягнення в оцінюваній галузі, поєднує вузьку спеціалізацію й загальний кругозір. У сучасних умовах, коли відбувається експонентне зростан-

ня обсягів наукової інформації, виконувати експертну оцінку стає дедалі важче, і тому важливо якнайчіткіше визначати область компетентності кожного потенційного експерта.

У сфері розроблення стандартів існують як формальні, так і неформальні вимоги до експертів. Потреба у фахівцях, спроможних систематично використовувати і накопичувати знання щодо розроблення стандартів та термінології певної ПрО, стає все актуальнішою. Професійні компетенції та вимоги фахівців і експертів зі стандартизації регламентовані Законом України «Про стандартизацію», де описуються вимоги до освіти, стажу в галузі діяльності стандартизації, знань у сфері сертифікації, стандартизації й метрології, термінології, міжнародної стандартизації, знання етапів і правил розроблення тощо. Крім того, велика увага приділяється особистим якостям експерта: здатність до логічного мислення, міркування; аналітичний склад розуму; об'єктивність; забезпечення конфіденційності інформації; наявність постійних наукових публікацій стосовно цієї ПрО у фахових виданнях тощо.

У науковій сфері компетентність експертів мають підтверджувати наукові публікації, патенти, звіти, описи прикладних розробок та інші матеріали. У більшості випадків всі ці матеріали подані в електронній формі, що припускає створення засобів їхнього автоматизованого аналізу. Значна частина наукових публікацій є доступною через Web.

Відкриті Web-джерела відомостей стосовно ПрО та фахівців

Передбачається, що основним джерелом відомостей стосовно проекту розроблення стандарту є його природномовний опис (заявка, ТЗ, публікація міжнародного стандарту, план стандартизації тощо), а також додаткові зовнішні інформаційні ресурси — онтології, Wiki-сторінки тощо, які містять структуровані й семантично значимі відомості щодо ПрО.

Складнішою задачею є аналіз інформації щодо учасників проекту з розроблення стандартів. Частина відомостей про них є чітко формалізованою й може бути однозначно оціненою без урахування семантики ПрО й додаткових знань про стандарт. Зокрема, це рівень освіти учасників, їхній стаж роботи у релевантній області, досвід роботи. Але часто цих відомостей буває недостатньо, щоб визначити їхню компетентність для розроблення конкретного стандарту, особливо для стандартів у нових і динамічно змінюваних ПрО, де ще не сформована остаточно термінологія. Проблема в тому, що наукові співробітники з однаковою освітою й захищеними дисертаціями з однієї спеціальності, у процесі своїх досліджень спеціалізуються в різних питаннях.

Одним із об'єктивних критеріїв оцінювання сфери компетентності наукових співробітників є аналіз ▶

їхніх публікацій, представлених в Web, — наукових статей, доповідей, звітів і презентацій.

Пертинентність документа (публікації) залежить від таких параметрів, як кількість згадувань основних термінів проекту в документі й кількості основних термінів, згаданих у документі. Існує багато робіт щодо автоматичного визначення компетентностей на основі документів. Крім того, важливо враховувати оцінку діяльності дослідників науковим співтовариством — через наявність посилань на їхні роботи, згадуванням їхніх робіт у навчальній та довідковій літературі. Крім того, джерела відомостей щодо дослідників можуть бути імпортовані з баз знань інтелектуальних застосувань, що забезпечують персоніфіковане інформаційне обслуговування користувачів, наприклад, із семантичних пошукових і рекомендувальних систем [12].

Таким чином, як джерела відомостей щодо сфери компетентності авторів проекту пропонується розглядати такі інформаційні ресурси (ІР):

- офіційні документи, що підтверджують освіту та досвід роботи експертів та розробників, які є необхідною умовою для подальшого аналізу ресурсів (диплом про вищу освіту, науковий ступінь, сертифікати та нагороди);
- інформаційні ресурси, що розкривають семантику офіційних документів розробників (паспорти спеціальностей та дисциплін, вимоги до отримання наукового ступеню та вченого звання, посадові інструкції, таксономія національних вчених ступенів тощо);
- тексти опублікованих статей, тез доповідей, монографій, підручників, технічних звітів, авторських свідоцтв, патентів та інших об'єктів інтелектуальної власності, представлених у вигляді природномовних документів, опублікованих у Web, з урахуванням їх рейтингів у наукометричних базах даних;
- тексти інших об'єктів інтелектуальної власності, представлених у вигляді природномовних документів, представлених самими авторами (тексти статей із платних журналів, проміжні звіти тощо);
- Wiki-сторінки авторів розроблення стандартів, що містяться у різних Вікіпедіях і Wikidовідниках, які забезпечують структуроване подання інформації [14];
- Wiki-сторінки організацій та установ, з якими пов'язані автори розроблення (членство в ТК зі стандартизації, співробітництво з Національною академією наук України, навчальними закладами);
- звіти про раніше розроблені стандарти та дослідження інших типів у відповідній ПрО, з урахуванням успішності їхнього виконання (така інформація дозволяє прогнозувати успішність виконання нової роботи).

Одним з найважливіших факторів, які визначають кваліфікацію наукового співробітника, є його диплом і, за наявності вченого звання, спеціальність, за якою була захищена дисертація. Кожна спеціальність має свій паспорт — документ, що закріплює область досліджень, за які може бути присуджено вчене звання кандидата або доктора наук з відповідної спеціальності. У паспорті спеціальності вкладено також визначення спеціальності, науки, до яких належить ця спеціальність.

Значною мірою компетентність співробітника визначається місцем його роботи. Тому доцільно враховувати відомості про організації (колективи), у яких працюють автори розроблення, і зіставляти їх із ПрО розроблення. Іноді корисніше використовувати вже структуровані знання щодо організації. Організаційна онтологія — це онтологія, що відображає знання щодо організаційної та функціональної структури певної організації. Вона містить інформацію стосовно працівників підприємства, ієрархію виробничих відносин між ними [11]. Для об'єктивної оцінки компетентності розробників та експертів доцільно користуватися зовнішніми кількісними параметрами, що відображають загальну ефективність та інтенсивність їх наукової діяльності.

Використання онтологічного аналізу в стандартизації

Онтології ПрО, що характеризують певні стандарти, є потужним інструментом для опрацювання, аналізування та застосування знань, що містяться в цих стандартах, забезпечують їх автоматизоване опрацювання та інтеграцію.

Важливим питанням, пов'язаним з розробленням стандартів, є їх інтегрований характер, забезпечення автоматизованого пошуку та порівняння. Тому в процесі розроблення стандартів виникає необхідність у побудові їх семантичних метаописів. Такими метаописами можуть стати онтології ПрО стандарту, в яких формалізуються основні поняття ПрО, що відображаються в стандарті, та зв'язки між ними. Онтологічний підхід полегшує коректний переклад стандартів та дозволяє визначити, які саме близькі за значенням терміни потрібно використовувати у кожному окремому випадку.

Щоб виявити семантичну близькість між стандартами, потрібно порівняти їх онтології, знайти близькі за значенням поняття та на основі цього обчислити семантичну відстань між стандартами. Саме онтології ПрО можуть стати джерелом знань у пошуку компетентних фахівців для розроблення національних стандартів: за існуючим стандартом треба 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. Крім

того, наявність онтології стандарту дозволяє надалі знаходити потрібні стандарти, здобувати з них необхідні користувачам знання та аналізувати їх вміст на семантичному рівні.

Ще одна сфера застосування онтологій стандартів — автоматизована побудова сфер компетенцій ПК зі стандартизації, формалізованих через поєднання онтологій вже розроблених стандартів та наявність об'єктивних автоматизованих методів для класифікування нових стандартів до найбільш релевантного ПК. Аналіз відношень між онтологіями стандартів дозволить визначити порядок розроблення стандартів та визначення не лише формальних, але й семантичних зв'язків між ними: наприклад, стандарт, в якому визначаються певні терміни, потрібно розробляти раніше, ніж той, в якому ці терміни вже використовуються.

Розроблення онтології стандарту

Розглянемо це на прикладі побудови онтології для ДСТУ ISO/IEC TR 10000-1:2006 «Основи і таксономія міжнародних стандартизованих профілів. Частина 1. Загальні принципи та основи документування» [13]. У цьому стандарті використовують терміни, що визначаються у розділі стандарту «Визначення». Усі ці терміни мають бути класами онтології відповідної ПрО, яка описує цей стандарт та є підкласами класу «Термін стандарту». Спочатку в онтології створюється клас «Стандарт», який описує основні характеристики стандарту, такі як назва, код, тематика, рік прийняття, розробники, обсяг тощо (рис. 1).

Для інтеграбельного використання знань, які відображаються у стандартах, доцільно також додати до опису класу «Стандарт» зв'язки даного стандарту з іншими. Крім того, можна вказати, які саме терміни належать до цього стандарту за допомогою властивості об'єктів «Стандарт використовує термін».

Клас «Термін стандарту» має властивості даних (Data Properties) «Назва стандарту», «Назва терміну українською», «Назва терміну англійською», «Опис терміну українською», «Опис терміну англійською» та «Примітка» типу «рядок

символів» та властивості об'єктів (Object Properties) «Є підкласом» та «Є синонімом», які дозволяють встановлювати семантичні зв'язки між елементами термінами стандарту.

На рис. 2 подано онтологію ПрО, до якої належить розроблений ДСТУ ISO/IEC TR 10000-1:2006 — «Телекомунікації та обмін інформацією між системами». Зв'язки між класами, екземплярами класів та їх властивостями в онтології стандарту відображаються візуально наступним чином.

Роль онтології в оцінюванні компетенцій експертів

Крім загального професійного рівня, потрібно оцінювати компетентність розробників та експертів саме в тій ПрО, до якої належить конкретний стандарт.

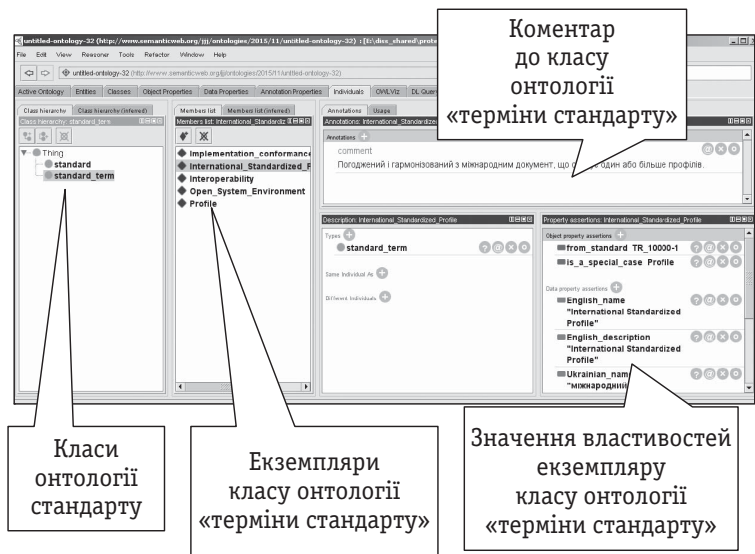


Рис. 1. Онтологія стандарту

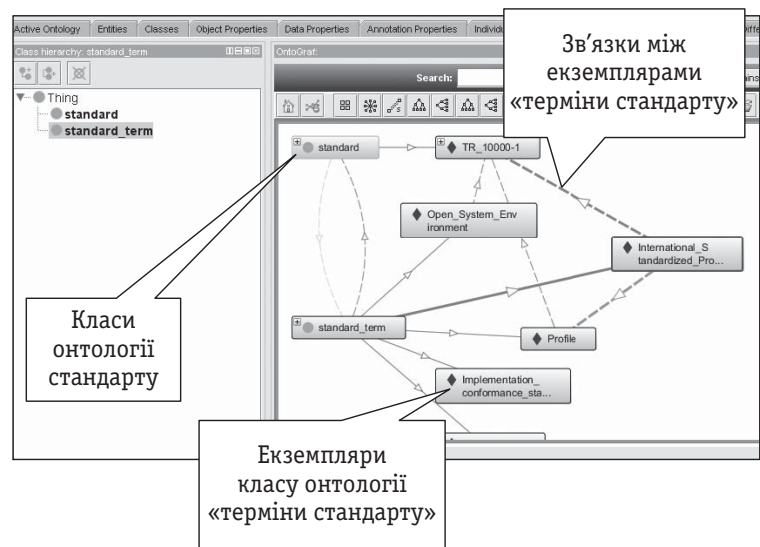


Рис. 2. Візуалізація зв'язків між класами, екземплярами класів та їх властивостями в онтології стандарту

Формальною моделлю такої ПрО є її онтологія. Саме онтології є на сьогодні широкорозповсюдженим засобом подання знань стосовно ПрО, що дозволяє аналізувати й зіставляти компетенції експертів і розробників у нових дослідницьких областях.

Щоб використовувати знання ПрО, необхідно використовувати її онтологію. Пошук (або розроблення) такої онтології є завданням учасників розроблення стандартів, бо саме вони повинні визначити, у якій сфері вони планують вести розроблення. Якщо ж учасники не надають таку онтологію, то використовується загальніша онтологія тієї області знань, до якої належить розроблюваний стандарт. Використання загальніших онтологій призводить до двох проблем: необхідно обробляти більший обсяг термінів; у проекті стандарту можуть зустрічатися загальніші, не істотні для суті проекту терміни, опрацювання яких трохи знизить точність результативної оцінки. Крім того, сам факт наявності в розробників стандарту онтології, що відповідає ПрО, як правило, засвідчує глибокі пізнання в обраній області (особливо якщо мова йде про інформаційні технології).

Разом з онтологіями ПрО, доцільно застосовувати загальну онтологію наукової діяльності, яка дозволяє однозначно встановлювати термінологію, пов'язану з рейтингом публікацій, науковими ступенями й ученими званнями, типами організацій тощо. Така онтологія спеціально розробляється для цілей визначення компетентності авторів проекту на основі організаційних онтологій наукових організацій, класифікатора УДК та паспортів спеціальностей ВАК. У такій онтології відображаються, наприклад, такі відношення між класами, як «співавторство», «робота в організації», «авторство публікації», «наявність вченого звання за фахом» і власності «бути публікацією ВАК», «мати індекс Гірша» (рис. 3).

Ця онтологія пропонується розробникам стандарту як зразок для опису й класифікації представлених документів, що повинні засвідчити рівень їх компетентності в науковій діяльності у цілому та у предметній області стандартизації зокрема.

Тезаурус і онтологія ПрО як засіб зіставлення природномовних документів

Для зіставлення стандарту і компетенцій його розробників пропонується побудувати тезаурус стандарту і тезауруси усіх ІР, які описують учасників розроблення стандарту.

Тезаурус природномовного ІР є проекцією онтології, аналізованої ПрО на документ, який розглядається в цей час. Докладніше алгоритм побудови таких тезаурусів розглядається в [11].

Тезаурус учасників розроблення визначається як об'єднання тезаурусів окремих ІР. При цьому слід враховувати вагу окремих ІР, у якій повинно бути враховано як значимість самого документа для опису компетенції, так і рівень довіри до самого ІР. Наприклад, більшу вагу має автореферат дисертації, ніж опис диплому, а при оцінюванні публікацій корисно враховувати рейтинг різних журналів і конференцій у наукометричних базах даних [16].

Припускається, 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 із них, що пов'язані з тією проблемою, якій присвячено розроблення стандарту, причому не обов'язково це повинні бути статті з найбільшим індексом цитування або ті, що входять до Scopus. Однак автор повинен прагнути, щоб для всіх понять онтології ПрО, для яких виявлено лінгвістичні аналоги у тексті стандарту, були знайдені такі відповідності й у його роботах (вага кожного окремого зіставлення визначається вагою ІР, обумовленого як функція від статусу документа та його рейтингу). Статус документа характеризує рівень його документального підтвердження, а рейтинг — його оцінку в наукометричних БД.

Спочатку будується тезаурус стандарту Th_{proj} , який залежить від опису проекту розроблення стандарту та від обраної онтології ПрО. Він є множиною пар (t_i, q_i) , де $t_i \in T$, T — множина термінів онтології

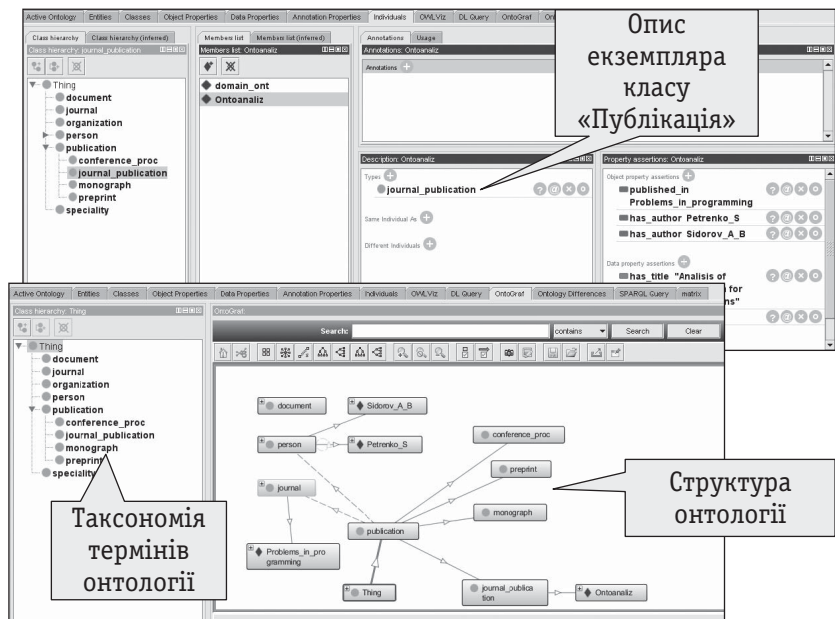


Рис. 3. Онтологія наукової діяльності, яка використовується для аналізу компетентності розробників стандарту

ПрО $O_{\text{ПрО}} = \langle T, R, A \rangle$, а q_i — кількість знайдених відповідностей, що обумовлює вагу цього терміна (якщо певний термін зустрівся в описі проекту стандарту 10 разів, то припускається, що він важливіший, ніж той, який зустрівся 2 рази). Для кожного терміна онтології ПрО виконується спроба знайти фрагменти тексту.

Загальна оцінка компетентності колективу авторів визначається за допомогою масиву $\left(t_i, \sum_{j=1}^m P_{IR_{ji}} v_{IR_j} \right)$, де $P_{IR_{ji}}$ — кількість знайдених відповідностей із цим терміном в j -му IP, а v_{IR_j} — вага j -го IP. Очевидно, що різні IP мають різну вагу під час оцінювання компетенцій фахівців. Наприклад, публікація в реферованому журналі є показовішою, ніж ТЗ, викладений у Web. Тому під час оцінювання публікацій доцільно враховувати імпаکت-фактор видання, яке опублікувало статтю. У такому випадку загальну оцінку компетентності авторів можна подати наступним чином:

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

де $\text{Imp}(IR_{ji})$ — імпакт-фактор видання, що опублікувало IP.

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

$$C = \sum_{i=1}^n q_i \left(\sum_{j=1}^m P_{IR_{ji}} v_{IR_j} \right) s_i, \quad (2)$$

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

Загальна оцінка кваліфікації розробників проекту стандарту може оцінюватися як сума оцінок учасників або як їхня нормалізована сума. Перший підхід має переваги, бо знання й досвід кожного можуть

використовуватися незалежно від кількості учасників. Тому нормалізована оцінка може застосовуватися лише як додаткова.

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 опрацювання інформаційних ресурсів про експертів ПрО, яка нас цікавить. Метою розроблення представлених методів є об'єктивізація процесу оцінювання кваліфікації потенційних учасників на основі урахування великої кількості доступних знань щодо їхньої кваліфікації в сфері розроблення та стандартизації.

Запропонований підхід дозволяє на основі опису проекту й природномовної інформації щодо учасників проекту з розроблення стандартів (їх публікацій, дипломів, описів раніше виконаних проектів, відомостей щодо їхньої організації тощо) будувати тезаурус проекту й тезауруси учасників, які базуються на онтології предметної області проекту з розроблення стандарту, зіставляти ці тезауруси й визначати на основі цього рівень компетентності учасників.

У майбутньому 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 й інформаційні); фахівців, що є експертами у сфері розроблення стандартів, та організації різного рівня, що підтримують різні види діяльності, пов'язаної із розробленням та використанням стандартів. Інтероперабельне подання знань та застосування відповідних технологій і форматів дозволить інтегрувати цю мережу знань із знаннями, поданими у Semantic Web.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Закон України «Про стандартизацію» від 05.06.2014, № 13-15. — VII. Верховна Рада України.
2. Національна стандартизація. Стандартизація та суміжні види діяльності. Словник термінів : ДСТУ 1.1:2015. — К. : ДП «УкрНДНЦ», 2015.
3. Національний стандарт України. Національна стандартизація. Правила проведення робіт з національної стандартизації : ДСТУ 1.2:2015. — [Чинний від 01-07-2003]. — (Національний стандарт України). — К. : ДП «УкрНДНЦ», 2015.
4. Національна стандартизація. Правила та методи прийняття міжнародних і регіональних нормативних документів : ДСТУ 1.7:2015. — [Чинний від 31-03-2004]. — (Національний стандарт України). — К. : ДП «УкрНДНЦ», 2015.
5. Gruber T. What is an Ontology? — [Електронний ресурс]. — Режим доступу : <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
6. Гладун А. Я. Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу / А. Я. Гладун, Ю. В. Рогушина // Штучний інтелект. — 2008. — № 5. — С. 112—124.
7. Лесько О. В. Анализ семантики естественно-языковых законодательных документов с использованием

- онтології предметної області / О. В. Лесько, Ю. В. Рогушина // Проблеми програмування. — 2015. — № 4. — С. 58—71.
8. Грицик Н. Комп'ютерна термінологія та основні способи її перекладу / Грицик Н. // XVIII та Міжнародна науково-практична Інтернет-конференція «Проблеми та перспективи розвитку науки на початку третього тисячоліття». — Переяслав-Хмельницький, 2013. — С. 45—52.
 9. Наукові засади та теоретико-методологічні принципи створення сучасних енциклопедій: колективна монографія / За ред. Киридон А. М. — К. : ДНУ «Енциклопедичне видавництво», 2015. — 160 с.
 10. Цициліано О. Креативність та стандартизація як основа конкурентоспроможності організації // Стандартизація, сертифікація, якість. — 2007. — № 1 (44). — С. 66—71.
 11. Гладун А. Я. Семантичні технології: принципи та практики / А. Я. Гладун, Ю. В. Рогушина. — К. : Універсаріум, 2016. — 314 с.
 12. Рогушина Ю. В. Онтологічний аналіз в Web / Ю. В. Рогушина, А. Я. Гладун, В. В. Осадчий, С. М. Прийма — Мелітополь: МДУПУ ім. Богдана Хмельницького, 2015. — 407 с.
 13. Rogushina J. Ontology-based competency analysis in new research domains/ J. Rogushina, A. Gladun // Journal of Computing and Information Technology. — 2012. — V. 23. — № 4. — P. 123—134.
 14. Рогушина Ю. В. Семантичний пошук у Web на основі онтологій: розробка моделей, засобів і методів / Ю. В. Рогушина. — Мелітополь : МДУПУ ім. Богдана Хмельницького, 2015. — 291 с.
 15. Гладун А. Я. ДСТУ ISO/IEC TR 10000-1:2006 «Інформаційні технології — Основи і таксономія міжнародних стандартизованих профілів — Частина 1 : Загальні принципи та основи документування» / А. Я. Гладун, А. В. Кулик. — К. : Держспоживстандарт, 2006. — 52 с.
 16. Штовба С. Д. Обзор наукометрических показателей для оценки публикационной деятельности ученого / С. Д. Штовба, Е. В. Штовба // Наукометрия и экспертиза в управлении наукой : Сб. статей. — М. : ИПУ РАН, 2013. — С. 262—278. ■

НОС ІНФОРМУЄ

ВЗАЄМОДІЯ З ТК

