

УДК 681.3

DOI 10.33244/2617-5932.3.2019.20-29

В. М. Антоненко,*к.т.н., доцент,**Університет ДФС України,**e-mail: beta7733@ukr.net,***ORCID ID 0000-0002-5513-4423;****Ю. В. Рогушина,***к.ф.-м.н., доцент, с.н.с.,**Інститут програмних систем**НАН України,**e-mail: ladamandraka2010@gmail.com,***ORCID ID 0000-0001-7958-2557;****Н. К. Сьомка,***Університет ДФС України,**e-mail: n_semka@ukr.net,***ORCID ID 0000-0001-8893-450X**

ПРОБЛЕМИ СЕМАНТИЗАЦІЇ ІНФОРМАЦІЙНИХ РЕСУРСІВ WEB ДЛЯ ПІДТРИМКИ УПРАВЛІННЯ ПРОЦЕСАМИ ОПОДАТКУВАННЯ

Проаналізовано проблеми, пов'язані з пошуком у Web інформації, що стосується оподаткуванням. Обґрунтовано доцільність використання семантичного пошуку, в якому застосовуються знання цієї предметної області. Для формалізації цих знань пропонується використовувати онтологію, яка містить основні терміни зі сфери оподаткування та зв'язки між ними. Ці знання використовуються для того, щоб визначити семантичну близькість між знайденими ресурсами та інформаційними потребами користувачів.

Терміни онтології використовуються як основа для семантичної розмітки знайдених у процесі пошуку інформаційних ресурсів, що спрощує сприйняття інформації користувачами.

Надаються рекомендації щодо вибору джерел для поповнення цієї онтології та метод здобуття онтологічних знань з природномовних ресурсів.

Ключові слова: *семантична розмітка, семантизація, онтологічний аналіз, Web, оподаткування.*

В. М. Антоненко, Ю. В. Рогушина, Н. К. Сьомка. Проблемы семантизации информационных ресурсов Web для поддержки управления процессами налогообложения

Проанализированы проблемы, связанные с поиском в Web информации, касающейся налогообложения. Обоснована целесообразность использования семантического поиска,

в котором применяются знания данной предметной области. Для формализации этих знаний предлагается использовать онтологию, которая содержит основные термины из сферы налогообложения и связи между ними. Эти знания используются для того, чтобы определить семантическую близость между найденными ресурсами и информационными потребностями пользователей.

Термины онтологии используются как основа для семантической разметки найденных в процессе поиска информационных ресурсов, которая упрощает восприятие информации пользователем.

Предоставляются рекомендации относительно выбора источников для пополнения этой онтологии и метод получения онтологических знаний из естественных языковых ресурсов.

Ключевые слова: *семантическая разметка, семантизация, онтологический анализ, Web, налогообложение.*

Метою цієї статті є розробка методів семантичної розмітки інформаційних ресурсів Web, що враховують специфіку даних в області оподаткування, використовують сучасні технології подання знань (Wiki, онтології) та забезпечують ефективний доступ до відомостей, необхідних користувачам.

Постановка проблеми. Значна частина сучасних ІС орієнтована на розподілену обробку інформації та функціонування у відкритому інформаційному просторі Web. Це стосується значною мірою і сферою оподаткування, де актуальність і достовірність знайденої інформації є одними з найважливіших чинників прийняття рішень. Використання Web-ресурсів передбачає необхідність обробки великих обсягів гетерогенної інформації, що викликає потребу в автоматизації процедур обробки на рівні семантики. Потрібно відмітити, що сьогодні значна частка цієї інформації – це неструктуровані природномовні тексти. Безпосередній аналіз такої інформації є складним та ресурсоємним завданням. Це викликає потребу у розробці методів та засобів їхньої семантизації, які дозволяють використати знання предметної області (ПрО) для того, щоб перейти від обробки неструктурованих даних до аналізу семантично розміченої інформації. Тому основні напрями розвитку інформаційних технологій (ІТ) пов'язують зі створенням інтелектуальних інформаційних систем (ІС), які орієнтовані на аналіз інформації на рівні знань.

Аналіз публікацій. Онтологічний аналіз є потужним засобом для формалізації знань у різних ПрО [6, 12]. Менеджмент знань на основі онтологій [14] нині широко застосовують в ІС у різних сферах. Для цього в рамках проекту Semantic Web [13] розроблено велику кількість стандартів, мов [8, 9] та інструментальних засобів [11]. Також вони дозволяють формалізувати знання про суб'єкти економічної діяльності, які входять до складу економічної системи [10], що забезпечує їхнє інтероперабельне використання в різних інформаційних системах, тобто знання, згенеровані внаслідок аналізу інформації однією ІС, можуть використовуватися як вхідні дані іншою ІС. Це дозволяє не тільки позбутися додаткових складних операцій зі здобуття знань (такі операції потребують багато часу й участі експертів), але й отримувати більш об'єктивну та узгоджену інформацію, в процесі створення якої бере участь більше спеціалістів.

Формалізація знань про суб'єкти економічної діяльності у вигляді онтологій спрощує пошук потрібних знань та дозволяє узгоджувати інформацію з різних джерел на

семантичному рівні, використовуючи інформаційні джерела з відкритого інформаційного середовища – Web, локальних та корпоративних мереж тощо [15]. Але це викликає потребу у використанні засобів та методів керування знаннями, що враховують специфіку середовища Web.

Керування знаннями (інша назва – менеджмент знань, Knowledge management – КМ) – це систематично керована організаційна діяльність, спрямована на аналіз явних і неявних знань як ключові стратегічні ресурси організації і призначена для поліпшення обробки цих знань на індивідуальному, колективному, організаційному і міжорганізаційному рівнях для досягнення більш ефективної діяльності податкової служби. У суспільстві знань (knowledge-based society), де саме знання стають вирішальним ресурсом, застосування методів керування знаннями є необхідною умовою успішної податкової діяльності.

Термін «керування знаннями» («knowledge management») вперше було використане 1986 р. американським вченим Карлом Віггом на конференції у Швейцарії, яка проводилася Міжнародною організацією праці під егідою ООН, за аналогією з такими поняттями, як «керування даними» та «керування інформацією». На сьогодні існує багато досить суперечних визначень цього терміна, що відображають різні аспекти, пов'язані з ефективним використанням знань у різних областях [1].

Основні проблеми керування знаннями у Web пов'язані з:

- *інтеграцією* знань, отриманих від різних інформаційних ресурсів (ІР);
- пошуком *протиріч* між знаннями, що відображені в контенті різних ІР, оцінкою їхньої достовірності та надійності;
- *здобуттям нових знань* з уже наявних та їхнім поданням у формі, зрозумілій користувачеві;
- *пошуком* знань, потрібних конкретному користувачеві для вирішення певних задач;
- автоматизацією *створення метаданих*, що коректно описують контент ІР на рівні змісту, та ефективним пошуком у таких метаописах.

Технології Semantic Web. Значні кроки до встановлення єдиного простору знань зроблені у проекті Semantic Web [7], спрямованому на перетворення Web на єдину базу знань на основі існуючих інформаційних комп'ютерних технологій (далі – ІКТ) та міжнародних стандартів, базуючись на онтологічному представленні знань ПрО [3].

Виклад основного матеріалу дослідження

Використання онтологій для створення спільного інформаційного середовища для суб'єктів оподаткування. У сучасних умовах створення та поширення інформаційно-комунікаційних технологій (ІКТ) настала можливість об'єднання всіх суб'єктів економіки (підприємства й організації) в загальне інформаційне середовище на основі онтологічного аналізу [5]. Знання про це середовище можуть бути формалізовані за допомогою ієрархії онтологій, що містять інформацію різного рівня узагальнення (рис. 1).

Вирішити цю проблему дозволяють методи Data Mining, призначені для здобуття знань з різних типів ІР. Чим більше структурованими та семантизованими є такі ресурси, тим більш ефективно та легко здобувати з них податкову інформацію.

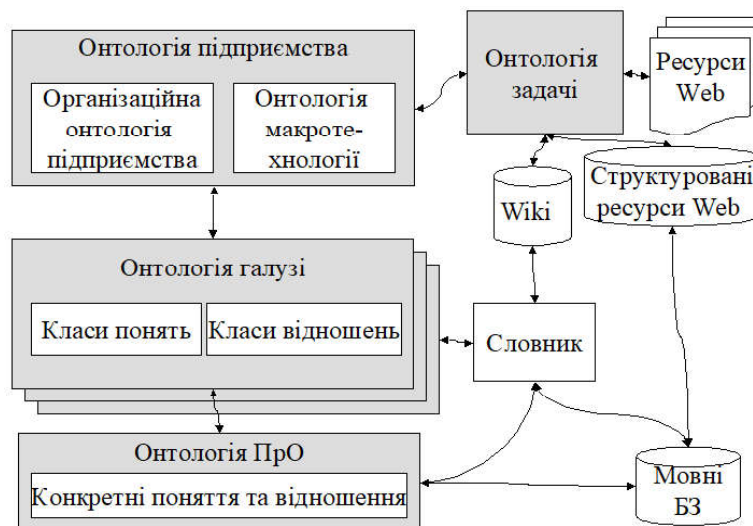


Рис. 1. Ієрархія знань щодо суб'єктів економічної діяльності

Проблеми семантизації ІР у сфері оподаткування. Семантизацією у мовознавстві називають виявлення змісту мовних одиниць, наприклад появу нових семантичних одиниць у структурі лексеми, тобто набуття словом нового значення. У більш загальному випадку *семантизація* – це встановлення зв'язку між певним інформаційним об'єктом (ІО) та його змістом. Під *семантизацією* інформаційних ресурсів розуміють встановлення зв'язків між цими ІР (або їхніми елементами) та певними формалізованими знаннями. Основні задачі, які при цьому вирішуються, – це пошук ІР, пертинентних потребам користувачів, та здобуття з цих ІР необхідної інформації. Окремими випадками семантизації ІР є семантичний пошук, семантичне анотування (побудова семантичних метаописів) та семантична розмітка.

Але виникає проблема, пов'язана з тим, що інформація про суб'єкти оподаткування постійно змінюється, а онтології є недостатньо динамічними – їхнє оновлення потребує значних зусиль та досвіду інженерів зі знань.

Семантичний пошук [4] – це пошук інформації, що задовольняє інформаційні потреби користувача в процесі пошуку інформації з оподаткування із застосуванням знань щодо суб'єктів і об'єктів пошукової процедури й методів аналізу цих знань [2]. Ці знання можуть стосуватися як користувача та його інформаційних потреб (персоніфікація пошуку), так і структури та типів ІР, серед яких здійснюється пошукова процедура. Знання можуть використовуватися приховано від користувача (внутрішня база знань пошукової системи, до якої користувач не має доступу), що ускладнює роботу такої системи та зменшує довіру до неї, або безпосередньо за його вибором (база знань користувача), що вимагає від користувача додаткових зусиль.

Аналіз існуючих на сьогодні засобів семантичного пошуку показує, що для переважної кількості задач для опису ПрО достатньо використовувати «легковагові» онтології, в яких не застосовуються аксіоми, а множина відношень між класами, що описуються додатковими властивостями, зазвичай є порожніми. *Легковагові* онтології – це стосовно

прості таксономічні структури термів з відповідними визначеннями. Вони менш аксіоматизовані. Такі онтології мають менші виразові можливості, але їх значно легше та швидше обробляти. Для великовагової онтології можна побудувати її легковагову версію. Для таких онтологій характерна наявність тільки таких відношень:

- відношення «клас-підклас», значення якого пов'язані транзитивно;
- об'єктне відношення синонімії, значення якого пов'язані симетрично;
- об'єктні відношення, специфічні для ПрО, значення яких не мають бінарних властивостей.

У більш складних випадках деякі відношення ПрО з оподаткування можуть також бути визначені як транзитивні або симетричні, і це дозволяє реалізувати більш складне логічне виведення на знаннях ПрО. Але частіше для семантичного пошуку податкової інформації достатньо самої наявності відношення між екземплярами класів, щоб оцінити пертинентність області пошуку. Це пов'язано з тим, що аналіз, наприклад, природномовного тексту з урахуванням усіх знань щодо ПрО для визначення його семантики займає досить багато часу, що неприйнятно для аналізу великої кількості IP, релевантних запиту. Залежно від того, наскільки складно описує користувач потрібну ПрО, залежить як час пошуку, так і його якість. Обирати компроміс між цими протилежними критеріями має сам користувач.

Така «легковагова» онтологія може базуватися на онтології більш широкої ПрО (з більшою кількістю класів, відношень та екземплярів), яка має значно більшу виразну потужність (наприклад, подана на OWL Full або OWL-DL). Але безпосередньо побудувати підмножину онтології для переважної більшості користувачів є досить складною задачею. Значно легше обробляти IP, що містять відповідну семантичну розмітку, обираючи ті, що належать до поточної задачі.

На сьогодні переважна частина ресурсів Web не є семантизованими, але кількість IP, що містять семантичну розмітку та різноманітні метаописи інформації з податків, постійно зростає. Але для користувачів все ще важко відокремлювати такі ресурси від загальної маси IP, незважаючи на те, що існує велика кількість пошукових систем, орієнтованих саме на пошук структурованих IP, наприклад для виявлення, індексації та запитування документів з оподаткування у форматі RDF або OWL.

Використовуючи такі системи, можна знайти конкретний документ, але досить складно побудувати множину документів, що відповідають будь-якій поточній задачі, тому що безпосередні посилання між такими IP, як правило, відсутні або не є наочними. Тому доцільніше використовувати бази знань, побудовані на Wiki-платформі. По-перше, такі системи застосовують стандартизовані засоби для подання семантичної розмітки (за допомогою системи категорій та властивостей). Ці елементи можна легко розпізнавати навіть у тих випадках, коли різні інформаційні об'єкти (IO) отримані з різних ресурсів [2]. Інша важлива перевага – у семантичних Wiki-ресурсах завжди є в наявності різноманітні засоби для внутрішньої навігації та пошуку, і це дозволяє користувачу досить швидко визначити набір Wiki-сторінок, пов'язаних з його задачею. Третій важливий фактор – на сьогодні вже створено багато різноманітних семантичних Wiki-ресурсів, і їхня кількість, обсяг та якість постійно збільшуються. По-четверте, у випадку, якщо інформації в семантичних Wiki недостатньо, їх досить легко доповнити відомостями з несемантизованих Wiki (наприклад, з Вікіпедії різними мовами або Wiki-довідників) – з таких ресурсів можна здобути менше семантичної інформація, але у поєднанні із семантизованими вони дозволяють досить коректно описати проблему.

Основні етапи побудови онтології задачі в оподаткуванні.

Етап 1. Знайти семантичний Wiki-ресурс, що за тематикою співвідноситься із задачею оподаткування. Рекомендується використовувати ті Wiki-довідники та Wiki-енциклопедії, які надають достовірну інформацію (реферовані видання) та пов'язані із країною, для якої вирішуються задачі. Якщо користувач має відомості до більш спеціалізованих ресурсів, то їхнє застосування може збільшити ефективність роботи. Припускається використовувати й несемантизовані Wiki-ресурси, але це надає менше інформації.

Етап 2. У цьому Wiki-ресурсі відібрати множину IP, що безпосередньо пов'язані із вирішенням задач з оподаткування, – Wiki-сторінки відповідних термінів. Почати цей відбір можна з пошуку Wiki-сторінок, назви яких співпадають із ключовими словами задачі, а надалі за допомогою вбудованих засобів навігації Wiki-ресурсом переходити до сторінок, сполучених із цими сторінками семантичними властивостями (усіма або тільки тими, що цікавлять користувача) або належать до категорій оподаткування. На цьому етапі користувач має виконати певну кількість роботи самостійно, щоб охарактеризувати ту інформацію, що йому потрібна, та відкинути ту, що не стосується його поточної задачі (це може бути цінна інформація, важлива для Про в цілому, але не потрібна саме для вирішення поточної проблеми). Від того, наскільки точно буде виконано відбір, залежить ефективність виконання семантичного пошуку: відсутність потрібної інформації не дозволить знаходити відповідні ресурси, а наявність зайвих сторінок збільшить час обробки.

Етап 3. Для кожної Wiki-сторінки виконуються такі дії:

– інформація про класи сторінки (усі або відібрані користувачем) дозволяє поповнити класи онтології задач з оподаткування. Ієрархічні відношення між цими класами можна визначити за допомогою сторінок цих категорій;

– ім'я самої сторінки заноситься до множини екземплярів класів онтології задачі;

– імена тих семантичних властивостей сторінок, що використані на цій сторінці, та область, визначення яких належить до типу «Сторінка» (усі або відібрані користувачем), заносяться до множини об'єктних властивостей класів онтології задачі;

– імена сторінок, на які ця сторінка посилається за допомогою гіперпосилань (усі або відібрані користувачем), також заносяться до множини екземплярів класів онтології задачі;

– якщо ця сторінка відсилає до іншої сторінки, тоді ім'я такої сторінки розглядається як синонім поточної сторінки, заноситься до множини екземплярів класів онтології задачі та пов'язується відношенням синонімії із іменем поточної сторінки.

Ці дії дозволяють побудувати онтологію задачі користувача для отримання інформації про оподаткування. Хоча існує можливість їхньої програмної реалізації через те, що більшість операцій потребує втручання користувача та ухвалення рішення щодо кожного поняття і відношення, доцільніше виконувати ці дії безпосередньо за допомогою редактора онтологій (наприклад, Protégé). Таким чином, побудована онтологія представлена мовою OWL та може використовуватися семантичними пошуковими системами для фільтрації результатів запитів. Основна ідея фільтрації інформації з оподаткування полягає у тому, що контент знайдених IP (або певних їхніх елементів, визначених користувачем) порівнюється із елементами онтології. За результатами порівняння знайдені IP впорядковуються відповідно до кількісного значення коефіцієнта близькості. Алгоритм порівняння та міри близькості залежить від специфіки конкретних пошукових

систем та обраних користувачем налаштувань. Приклад такої системи семантичного пошуку, що використовує онтології користувачів, – метапошукова система МАПС, що призначена для пошуку інформації, пов'язаної з податковою інформацією. Формалізація інформаційних потреб користувача МАПС базується на представленні знань через онтології Про. Персоналізація семантичного пошуку в МАПС базується на тому, що користувач чітко визначає, які саме знання (онтології та базовані на них тезауруси задач) використовуються для формалізації його інформаційних потреб. Це є характерною рисою цієї ССП та її основною відмінністю від інших систем, що підтримують пошук податкової інформації на семантичному рівні.

Етап 4. Побудована онтологія задачі використовується для семантизації довільних природномовних IP, пов'язаних з оподаткуванням. Надалі такі IP самі стають об'єктом аналізу і дозволяють оновлювати онтологію задачі, що використовується у різних застосунках.

Висновки. На сьогодні збільшення ефективності роботи інформаційних систем з отриманням інформації з оподаткування пов'язують не тільки з алгоритмами обробки, але й з тим, наскільки актуальні та пертинентні дані вони обробляють. Це збільшує важливість підсистеми пошуку інформації. Персоніфікація дозволяє з великої кількості IP, релевантних запиту користувача, відібрати ті IP, які стосуються його поточних інтересів, позбавивши самого користувача від рутинного аналізу запропонованої інформації. Основна тенденція семантизації пошукових засобів пов'язана з переходом від виявлення документів, що містять певні ключові слова, до пошуку знань, необхідних для виконання поставленого перед користувачем завдання. Це особливо важливо до динамічних областей знань, до яких належить оподаткування. Використання онтології для отримання податкової інформації дозволить знаходити та обробляти найновіші документи – закони та інструкції, аналітичні звіти, рекомендації тощо. Це дозволить користувачам, що не є спеціалістами в інформаційних технологіях, легко та успішно застосовувати методи керування знаннями.

Перспективи подальших досліджень у цьому напрямі пов'язані зі створенням семантизованих Wiki-ресурсів з проблем оподаткування, побудовою методів та програмних засобів для здобуття з них онтологій, які можуть використовуватися в інших розподілених системах. Таким чином, здійснення переходу від неструктурованих природномовних документів (а саме у такому вигляді сьогодні представлена переважна частка інформації у сфері оподаткування – закони, нормативні акти, інструкції тощо) до таких частково структурованих даних має стати основою для застосування методів машинного навчання для їхнього автоматизованого аналізу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гладун А. Я. Data Mining: пошук знань в даних / А. Я. Гладун, Ю. В. Рогушина. – К. : ТОВ «ВД «АДЕФ-Україна», 2016. – 452 с.
2. Рогушина Ю. В. Семантичний пошук у Web на основі онтологій: розробка моделей, засобів і методів / Ю. В. Рогушина. – Мелітополь : МДПУ ім. Богдана Хмельницького, 2015. – 291 с.
3. Davies J., Fensel D., van Harmelen F. Towards the Semantic Web: Ontology-driven knowledge management. John Wiley & Sons Ltd, England, 2002. – 288 p.

4. Semantics-driven modelling of user preferences for information retrieval in the biomedical domain Informatics for health and social care / Gladun A., Rogushina J., Valencia-García R., Martínez-Béjar R. – 2013. – V. 38, № 2. – P. 150–170.
5. Gruber T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies. – 1995. – V. 43. – P. 907–928.
6. Guarino N. Formal Ontology in Information Systems. Formal Ontology in Information Systems. Proceedings of FOIS'98, 1998. – P. 3–15.
7. Introduction to Semantic Web [Електронний ресурс]. – Режим доступу : <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
8. Lassila O., Swick R. Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation [Електронний ресурс]. – Режим доступу : <http://www.w3.org/TR/REC-rdf-syntax>.
9. OWL Web Ontology Language Semantics and Abstract Syntax. Section 2. Abstract Syntax [Електронний ресурс]. – Режим доступу : <http://www.w3.org/TR/owl-semantics/syntax.html>.
10. Tang, X., & Yang, F. A study on dynamic ontology for information integration in E-governmental virtual organization. Wireless Communications, Networking and Mobile Computing, 2007. – P. 3600–3604.
11. Tudorache T., Nyulas C., Noy N. F., Musen M. A. WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. Semantic web. – 2013. – № 4(1). – P. 89–99 [Електронний ресурс]. – Режим доступу : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691821/>.
12. Uschold M., & Grüninger M. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review. – 1996. – 11(2). – P. 93–155.
13. W3C Semantic Web Activity [Електронний ресурс]. – Режим доступу : <http://www.w3.org/2001/sw/Activity/>.
14. Warren P. Knowledge Management and the Semantic Web: From Scenario to Technology. IEEE Intelligent Systems. 21, 1, 2006. – P. 53–59.
15. Yoo, D., & No, S. Ontology-based economics knowledge sharing system. Expert Systems with Applications. – 2014. – № 41(4). – P. 1331–1341.

REFERENCES

1. Hladun A. Y., Rohushina Y. V. Data Mining: poshuk znan v danyh [Data Mining: knowledge retrieval in data]. – К. : TOV VD «ADEF-Ukraine», 2016. – 452 s.
2. Rohushina Y. V. Semantychnyi poshuk u Web na osnovi ontologyi: rozrobra modelei, zasobiv i metodiv [Semantic retrieval into the Web: development of models, tools and methods]. – Melitopol : MDPU, 2015. – 291 s.
3. Davies J., Fensel D., van Harmelen F. Towards the Semantic Web: Ontology-driven knowledge management. John Wiley & Sons Ltd, England, 2002. – 288 p.
4. Gladun A., Rogushina J., Valencia-García R., Martínez-Béjar R. Semantics-driven modelling of user preferences for information retrieval in the biomedical domain. Informatics for health and social care. – 2013. – V. 38, № 2. – P. 150–170.
5. Gruber T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies. – 1995. – V. 43. – P. 907–928.

6. Guarino N. Formal Ontology in Information Systems. Formal Ontology in Information Systems. Proceedings of FOIS'98, 1998. – P. 3–15.
7. Introduction to Semantic Web. URL : <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
8. Lassila O., Swick R. Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation. URL : <http://www.w3.org/TR/REC-rdf-syntax>.
9. OWL Web Ontology Language Semantics and Abstract Syntax. Section 2. Abstract Syntax. URL : <http://www.w3.org/TR/owl-semantics/syntax.html>.
10. Tang, X., & Yang, F. A study on dynamic ontology for information integration in E-governmental virtual organization. Wireless Communications, Networking and Mobile Computing, 2007. – P. 3600–3604.
11. Tudorache T., Nyulas C., Noy N. F., Musen M. A. WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. Semantic web, 4(1), 2013. P. 89–99. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691821/>.
12. Uschold M., & Grüninger M. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2), 1996. – P. 93–155.
13. W3C Semantic Web Activity. URL : <http://www.w3.org/2001/sw/Activity/>.
14. Warren P. Knowledge Management and the Semantic Web: From Scenario to Technology. IEEE Intelligent Systems. 21, 1, 2006. – P. 53–59.
15. Yoo, D., & No, S. Ontology-based economics knowledge sharing system. Expert Systems with Applications, 41(4), 2014. – P. 1331–1341.

V. M. Antonenko, Yu. V. Rogushina, N. K. Siomka. Problems of semantization of the Web information resources for support the management of taxation processes

Introduction. Now information support of taxation processes is oriented on intelligent analysis of the Web resources. Such analysis requires the processing of large volumes of heterogeneous information at the semantic level. Ontological analysis is a powerful tool for formalizing knowledge in different domains. At present a large number of standards, languages and software tools are developed for these purposes within the of the Semantic Web project. The formalization of knowledge about the subjects of economic activity in the form of ontologies simplifies the search for the necessary knowledge and allows integration of data from different sources from the open information environment – the Web, local and corporate networks, etc.

The use of ontologies involves retrieval of law information deal with taxation from the Web at the semantic level (and this kind of task is often a component of modern intellectual applications that operate in an open information environment). Retrieval results not only documents or data, but also knowledge about the domain-specific information objects with complex and formalized structure.

Purpose. Efficiency of knowledge management methods used for taxation problems depends of pertinence of domain knowledge to particular user task. We propose to use ontologies for representation of taxation knowledge. This approach causes the necessity of methods providing retrieval of sources for creation and replenishment of such ontology and it's use for domain problems.

Methods. Ontological analysis is the basis for semantic processing and integration of Web resources pertaining to taxation. Now ontological representation of knowledge is prevailing for

Web-oriented intelligent applications, and we propose to use this approach modified for taxation data specifics.

Information on subjects of taxation is continually changing, and ontologies are not enough dynamic – their updating requires considerable effort and experience of knowledge engineers. This problem can be solved by Data Mining methods oriented on various types of information resources. Semantic structuring of these resources simplifies their analysis and makes obtaining of tax information from them more efficient. Recommendations about the choice of sources for ontology replenishing and the method for acquiring of ontological knowledge from natural language resources are given.

Results. We propose the method of construction of the taxation task ontology by processing of the pertinent semantic Wiki information resources. This method uses the specifics of the semantic Wiki markup (categories and semantic properties) and relates the standard elements of the Wiki-page with taxation concepts and relations formalized by the task ontology. This ontology can be used for semantic retrieval by filtering of information sources pertinent to particular user task. The terms of ontology are used as a basis for semantic marking of found information resources which simplifies the perception of information by the users.

Conclusion. Application of ontological approach for analysis of taxation information allows non-IT professionals successfully apply knowledge management techniques to find and process the most up-to-date documents – laws and regulations, analytical reports, recommendations, etc.

Key words: *semantic markup, semantization, ontological analysis, Web, taxation.*

Стаття надійшла до редколегії 11 березня 2019 року