

УДК 004.822 004.5

*І. А. Козак*, канд. екон. наук, доцент,  
ДВНЗ «КНЕУ імені Вадима Гетьмана»

### **АВТОМАТИЗАЦІЯ ОНТОЛОГІЧНОГО ІНЖИНІРИНГУ В СИСТЕМАХ КЕРУВАННЯ ЗНАННЯМИ ВІРТУАЛЬНИХ ОРГАНІЗАЦІЙ**

В статті проаналізовано етапи, методи та підходи до автоматизації побудови онтологій з метою їх використання в системах керування знаннями віртуальних організацій.

**КЛЮЧОВІ СЛОВА:** онтологія, системи керування знаннями, методи автоматичної побудови онтологій.

В статье анализируются этапы, методы и подходы к автоматизации построения онтологий с целью их использования в системах управления знаниями виртуальных организаций.

**КЛЮЧЕВЫЕ СЛОВА:** онтология, системы управления знаниями, методы автоматического построения онтологий.

The stages, methods and approaches to automations of the building of ontology for their use in virtual organization knowledges management systems are analysed in the article.

**THE KEYWORDS:** ontologies, knowledges management systems, methods of the automatic building of ontology (Ontology Learning from Text).

**Вступ.** Віртуальні організації різних типів та колаборативні мережі є результатом еволюційного розвитку організаційних структур та можливістю отримання для підприємств конкурентних переваг в епоху метакapіталізму. Однак, створення віртуальних організацій висуває певні вимоги до використання інформаційних технологій. Можна навіть стверджувати, що інформаційні технології (поєднання деяких їх видів) є основою таких організацій.

Серед ключових для віртуальних організацій технологій, що виділяються в роботах [1, 2], особливу роль відіграють системи

керування знаннями. Вони мають забезпечувати управління інтегрованими процесами віртуальних підприємств (поряд із системами керування потоками робіт) на основі системного виявлення, організації та використання знань. Проте, у більш пізніх роботах, наприклад [3], системи керування знаннями взагалі не включаються вже до переліку інформаційних технологій для віртуальних організацій — перевага віддається більш апробованим і простим технологіям. А системи керування знаннями залишилися «поза увагою» через досить тривіальні причини, серед яких на першому місці тривалий час і складність упровадження та перевищення очікувань над результатами впровадження подібних систем.

Однак, інтерес до систем керування знаннями не згасає надовго — через значне зростання даних та неструктурованої інформації у сучасних організаціях. Так, серед світових лідерів у галузі розробки спеціалізованих систем керування знаннями можна назвати компанії: Hummingbird (<http://www.hummingbird.com>), Convera (<http://www.convera.com>, <http://www.convera.su/ru/>), Cognitive Technologies (<http://www.cognitive.ru>). Серед користувачів цих систем — уряди, банки, величезні фінансові компанії та підприємства по всьому світу.

Сподівання щодо керування знаннями у віртуальних організаціях пов'язані із акумулюванням та формалізацією знань; виявленням і поширенням наявної інформації і досвіду серед учасників організації, що змінюються. Ефективна система керування знаннями для віртуальної організації повинна охоплювати всі процеси перетворення знань — від їх виявлення, організації та зберігання до використання. Проте виявлення знань є на сьогодні найменш інтегрованим у системи керування знаннями процесом. Традиційні підходи до виявлення знань із використанням методів Data Mining передбачають попереднє структурування інформації і використання спеціалізованих програмних засобів.

У той же час, виявлення знань тісно пов'язане із побудовою онтологій, на основі яких працюватиме система керування знаннями (під онтологіями нами розуміється опис понять предметної області та відношень між цими поняттями). Так, відомий фахівець у галузі систем керування знаннями Т. А. Гаврилова відзначає: «Онтологический инжиниринг — ядро концепции “управления знаниями”» [4].

Але побудова онтологій на сьогоднішній день є процесом складним, трудовитратним і тривалим, оскільки здебільшого

здійснюється фахівцями «вручну» для кожного окремого проекту. Крім того, розробка онтологій вимагає від фахівців певної предметної області володіння методами виділення, структурування та формалізації знань (відповідно до правил певної формальної мови представлення онтологій). І якщо з формалізацією можуть допомогти спеціалізовані інструментальні засоби на зразок PROTEGE, WebOnto, OntoSaurus, ВИКОНТ та ін., то решта процесів цілком і повністю залежать від кваліфікації експерта.

Тому, на сьогодні, можна говорити про необхідність розробки ефективних методів та засобів автоматичного створення онтологій, з метою використання їх як у системах керування знаннями, так і в інших типах інформаційних систем.

**Аналіз етапів та методів автоматичної побудови онтологій.** Дослідженням проблем автоматизації процесів побудови онтологій присвячено багато робіт, серед яких [5–9].

У роботі [5] як основні етапи процесу автоматичної побудови онтології (прошарки пирога побудови онтології) виділяються:

- визначення термів;
- визначення синонімів;
- визначення концептів;
- побудова таксономії концептів;
- визначення відношень;
- визначення правил.

Розглянемо, які методи на сьогоднішній день пропонуються для вирішення кожної з цих задач.

Виділення термів (як лінгвістичних реалізацій проблемно-залежних понять) є основою для вирішення подальших складніших задач. Для автоматичного визначення термів має використовуватися лінгвістична обробка тексту з тим, щоб визначити складені синтаксичні структури, які можна розглядати, як самостійні терми. Іноземні дослідники обмежуються тут побудовою зразків [6] для пошуку синтаксичних структур та використанням статистичної обробки [7] для включення лише «важливих» термів.

Визначення синонімів означає визначення семантичних варіантів терміну як у рамках однієї мови, так і між мовами (тобто, переклад терміну). Більшість робіт у цій області сфокусувалося на інтеграції з системою WordNet7, для отримання англійських синонімів, та EuroWordNet8, для двомовних і багатомовних перекладів синонімів і терміну. Однак, крім використання готових наборів синонімів, дослідники, наприклад [8], також працювали

над алгоритмами визначення синонімів на основі кластеризації та пов'язаних методів. Окремим напрямком досліджень стало також використання статистичних оцінок для веб-інформації — як у [9].

Визначення концептів, на думку авторів роботи [5], має забезпечувати:

- формулювання змісту поняття;
- набір прикладів поняття;
- набір лінгвістичних реалізацій.

Найбільше досліджень щодо визначення концептів розглядають концепти як групи зв'язаних термінів. Альтернативними напрямками є дослідження концептів з точки зору прикладів, як у [10], або з точки зору формальних і неформальних визначень [11].

Загалом, виділення термів та синонімів можна вважати підготовкою до визначення концептів у цілому.

Щодо автоматичного отримання таксономій із текстових даних, існує три основні парадигми. Перша полягає у використанні лексико-синтаксичних зразків, запропонованих у [12]. Друга полягає у використанні алгоритмів ієрархічної кластеризації [13]. А третя походить з інформаційно-пошукового суспільства і базується на категоризації термінів як запропоновано, наприклад, у [14].

Для визначення неієрархічних відношень між концептами використовується комбінування статистичного аналізу із лінгвістичним [7; 13].

У цілому, виділення в окремий етап визначення відношень одного виду (таксономічних) нам видається нелогічним і обумовленим лише більшою кількістю досліджень у даному напрямі.

Виділення правил виводу є найменш дослідженим питанням побудови онтологій. Деякі основні питання, пов'язані з автоматичним виділенням правил із текстів, сформульовані в роботі [15]. Хоча, на нашу думку, правила виведення на онтологіях мають бути загальними і універсальними, а не виводитися з текстів.

Крім цього, побудова онтології має логічно завершуватися її формальним представленням. Формальне представлення онтології передбачає запис виявлених концептів та відношень на одній із мов формального опису (OWL, KIF та ін.). Можливість автоматизації такого представлення неодноразово підтверджена чисельними редакторами онтологій, що дозволяють генерувати формальні описи онтологій на основі їх таблично-графічних представлень.

На основі вищесказаного ми виділимо наступні етапи процесу автоматичного створення онтологій:

1. визначення концептів.
2. визначення відношень між концептами;
3. формальне представлення онтології.

Узагальнено систему методів автоматичного формування онтологій з текстів предметної області зображено на рис. 1.

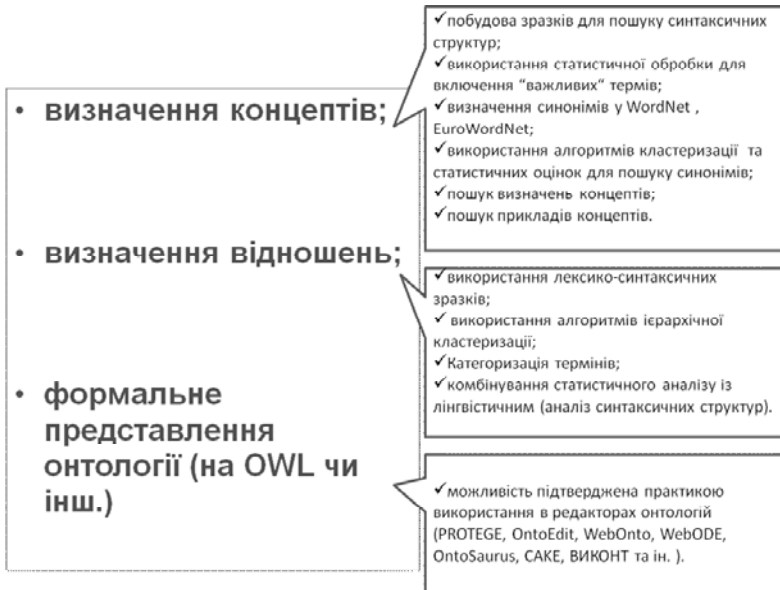


Рис. 1. Методи автоматичної побудови онтологій

Звичайно, на етапі визначення концептів для мов слов'янської групи, необхідне використання морфологічного та синтаксичного аналізу, а також теорії синтаксичних груп [16] тощо.

**Використання зовнішнього онтологічного інжинірингу в системах керування знаннями.** Висока трудоемність розробки систем автоматизованого виділення онтологій, їх універсальність та затребуваність вимагають уточнення їх місця в системах керування знаннями організацій.

Якщо традиційні системи керування знаннями зазвичай мають у своєму складі інтернет-портали та сервери знань, що забезпе-

чують збір, аналіз, класифікацію та структурування інформації, то використання зовнішніх систем онтологічного інжинірингу (на основі сервісної архітектури) дозволить здійснити «аутсорсинг» багатьох функцій серверів знань (рис. 2).

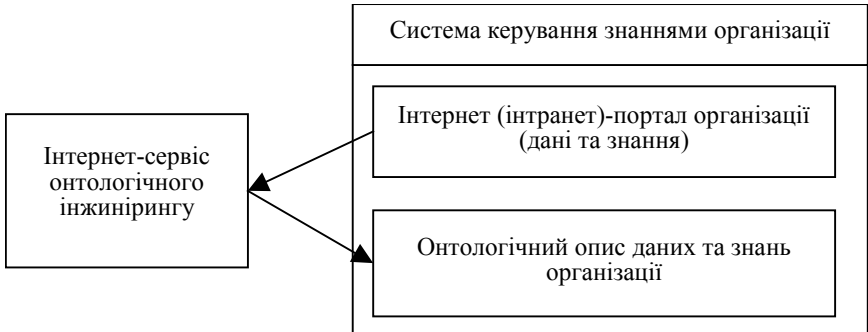


Рис. 2. Використання зовнішнього онтологічного інжинірингу

Основними вимогами до сервісу онтологічного інжинірингу будуть:

- можливості аналізу текстів із файлів різного формату, розміщених на порталі організації;
- формальне представлення онтологій у кількох стандартних форматах;
- онтологічна розмітка (індексація) даних та знань організації.

На системи керування знаннями покладатимуться функції щодо пошуку необхідних даних і знань організації на основі здійснених онтологічних описів, логічного виведення та забезпечення інтерфейсу користувача при доступі до даних і знань.

**Висновки.** В роботі виділено основні етапи автоматичного онтологічного інжинірингу, а також узагальнено основні методи, що можуть використовуватися на кожному з етапів онтологічного інжинірингу.

Запропоновано використання сервісу зовнішнього онтологічного інжинірингу в системах керування знаннями. Такий підхід дозволить забезпечити роботу з даними і знаннями не лише традиційних організацій, але й віртуальних, оскільки дозволить у короткі терміни здійснювати онтологічні описи будь-яких нових ресурсів та ефективно працювати з усією експліцитною інформацією віртуальної організації.

### Література

1. Тарасов В. Б., Шильников П. С. Виртуальные предприятия: свойства, технология создания, компоненты инфраструктуры // Информационные технологии. — 2000. — № 9. — С. 2–7.
2. National Industrial Information Infrastructure Protocols (NIIP), December 31, 1998. [www.niip.org](http://www.niip.org)
3. Camarinha-Matos L. M., Afsarmanesh H. Collaborative networked organization in manufacturing // IFAC 2007, 16 p
4. Гаврилова Т. А. Логико-лингвистическое управление как введение в инженерию знаний // Новости искусственного интеллекта. — № 6. — 2002. — С. 28–33.
5. Buitelaar P., Cimiano P., Magnini B. editors. *Ontology Learning from Text: Methods, Evaluation and Applications*, volume 123 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, Nieuwe Hemweg 6B, 1013 BG Amsterdam, The Netherlands, July 2005.
6. Cimiano P. *Ontology Learning and Population: Algorithms, Evaluation and Applications*. PhD thesis, University of Karlsruhe, 2005.
7. Buitelaar P., Olejnik D., Sintek M. A prot'eg'e plug-in for ontology extraction from text based on linguistic analysis. In Davies J., Fensel D., Bussler C., Studer R., editors, *Proceedings of the 1st European Semantic Web symposium(ESWS)*, Heraklion, Greece, May 2004.
8. Hindle D.. Noun classification from predicate-argument structures. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 268–275, 1990.
9. Baroni M., Bisi S. Using cooccurrence statistics & the web to discover synonyms in a technical language. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 5, pages 1725–1728, 2004.
10. Etzioni O., Cafarella M., Downey D., Kok S., Popescu A.-M., Shaked T., Soderland S., Weld D.S., Yates A. Web-scale information extraction in KnowItAll (preliminary results). In *Proceedings of the 13th World Wide Web Conference*, pages 100–109, 2004.
11. Navigli R., Velardi P., Cucchiarelli A., Neri F. Extending and Enriching WordNet with OntoLearn. GWC 2004, *Proceedings*, pp. 279.284.c Masaryk University, Brno, 2003.
12. Hearst M.A. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 539–545, 1992.
13. Faure D., Nedellec C. A corpus-based conceptual clustering method for verb frames and ontology. In P. Velardi, editor, *Proceedings of the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications*, pages 5–12, 1998

14. Sanderson M., Croft B. Deriving concept hierarchies from text. In Research and Development in Information Retrieval, pages 206–213. 1999.

15. Lin D., Pantel P. Dirt — discovery of inference rules from text. In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 323–328, 2001.

16. Гладкий А. В. Синтаксические структуры естественного языка в автоматизированных системах общения. — М.: Наука. Главная редакция физико-математической литературы, 1985. — 144 с. — (Серия «Проблемы искусственного интеллекта»).

Статтю подано до редакції 09.02.10 р.

УДК 338.47

О. В. Стець, доцент  
кафедри математичного  
моделювання економічних систем,  
С. В. Михайлович, студентка групи УК-41,  
факультет менеджменту та маркетингу,  
Національний технічний університет України «КПІ»

## МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ СИСТЕМИ МІСЬКИХ ПАСАЖИРСЬКИХ АВТОПЕРЕВЕЗЕНЬ ТА ЇХ ЕКОНОМІЧНИЙ АНАЛІЗ

Сучасні темпи розвитку міського пасажирського транспорту не відповідають темпам розвитку міської інфраструктури, сучасним потребам населення, рівню урбанізації міста. Тому необхідно уточнення методологічної основи аналізу та удосконалення систем пасажирського транспорту.

Стаття присвячена пошуку оптимального (за часовими характеристиками) маршруту міського пасажирського транспорту при пересуванні з одного транспортного району в інший. В якості критерію оптимізації обрано час, що витрачається пасажиром на дорогу.

При пошуку найраціональнішого маршруту враховується інтервал руху засобів пасажирських перевезень, час у дорозі, пасажиропотоки, можливі ситуації на пунктах зупинки транспортних одиниць та ймовірність їх виникнення.