

**М.М. ГЛИБОВЕЦЬ, О.В. БОБКО**

*Національний університет «Києво-Могилянська академія»,  
факультет інформатики, Київ, Україна  
glib@ukma.kiev.ua*

## **МЕТОДИ АВТОМАТИЧНОЇ ГЕНЕРАЦІЇ ОНТОЛОГІЙ**

*В даній роботі здійснено класифікацію методів автоматичної генерації онтологій, та проаналізовано можливості поєднання різних методів автоматичної генерації онтологій для покращення ефективності систем для керування знаннями.*

### **Вступ**

Метою даного дослідження є виявлення існуючих методів автоматичної генерації онтологій, побудова системи критеріїв їх оцінки й порівняння, та обрання оптимального рішення для створення засобу автоматичної генерації онтологій на основі набору текстових і/або XML - подібних документів з метою використання для керування знаннями в корпоративних інформаційних системах.

### **1. Процес автоматичної побудови онтологій**

Аналіз інструментів буде здійснюватись на кожній стадії роботи з онтологією, перелік яких включає такі стадії підготовки, аналізу, валідації, модифікації та серіалізації.

### **2. Критерії оцінки онтологій та методів їх генерації**

Ми можемо умовно поділити систему критеріїв на вимоги до інструменту та вимоги до онтології.

#### **2.1. Вимоги до інструменту**

- Рівень (можливості) автоматизації процесу генерації онтологій на кожному етапі.
- Розширюваність (масштабованість), цілісність та точність  
Незалежність від сторонніх джерел
- Мінімальна потреба адаптації при зміні мови тексту документів.

## **2.2. Вимоги до онтології**

- Придатність до подальшого розвитку, зміни чи доповнення
- Відповідність загальноприйнятим стандартам
- Відображення подібності між концептами, сили відношень між ними, залежності між екземплярами та концептами, тощо.

## **3. Методи генерації онтологій**

- Побудова онтологій шляхом перетворення XML-подібних документів (Напр. [1])
- Використання готових словників. (Напр. [2])
- Застосування лінгвістичного аналізу текстів написаних природною мовою. (Напр. [3])
- Застосування кластеризації та аналізу формальних концептів. (Напр. [4])

## **Висновки**

Розглядаючи можливість автоматизації різних стадій автоматичної генерації онтології ми дійшли до таких висновків.

Підготовчу стадію та стадію серіалізації у всіх випадках можливо повністю автоматизувати, оскільки дані процеси є повністю тривіальними і зводяться до примітивних операцій над рядковими даними або ж перетворення деревовидних структур даних у деякий XML-подібний формат.

Стадію аналізу також можна ефективно автоматизувати. Побудова концептів, таксономічних відношень між концептами та відношень належності екземплярів до класів автоматизується при застосуванні всіх описаних вище методів генерації онтологій. Можливість повної автоматизації побудови нетаксономічних відношень все ще лишається відкритим питанням, більше того наявні методи є залежними від мови оброблюваних текстових даних. Відкритим є також питання повної автоматизації найменування концептів, особливо це стосується методів автоматичної генерації онтологій заснованих на використанні ієрархічної кластеризації та формального аналізу концептів. Дану проблему можливо вирішити за рахунок використання зовнішніх джерел знань, наприклад тезаурусів чи словників на зразок WordNet.

Стадія валідації в тій чи іншій мірі потребує втручання експерта. Виключення складають випадки, коли онтологія генерується на основі набору XML-документів або ж підмножини записів деякого словника або тезауруса. В цьому контексті засоби на зразок WordNet без сумніву заслуговують на особливу увагу через великі можливості для автоматизації валідації. Попри те, що WordNet має занадто широке призначення, і не може бути адаптованим до певної предметної області людської діяльності, використання даного методу для валідації та узгодження онтологій є цікавою темою для подальшої розробки.

Стадія розширення також потребує особливої уваги при автоматизації. Особливо це стосується випадку її зведення до узгодження і злиття онтологій. Процес злиття онтологій тісно пов'язаний з узгодженням; на даний час існує ряд методів його реалізації для двох вхідних онтологій, однак одночасне злиття кількох онтологій залишається відкритим питанням. Ця проблема може бути вирішена шляхом послідовних злипів, але в такому випадку кінцева онтологія буде залежати від вибору порядку злиттів. В деяких випадках додання нових сутностей та зв'язків в онтологію може відбуватись методом відмінним від того, який використовувався на початковому етапі. Так чи інакше після стадії розширення онтології має здійснюватись повторна валідація, проблеми якої описані вище.

В цілому ми дійшли до висновку, що для досягнення максимальної ефективності обробки та систематизації інформації система керування знаннями, що використовує онтології має поєднувати різні методи автоматичної генерації. Генерація онтологій на основі XML документів становить важливість через те, що даний формат став ію-суті стандартом для обміну даними між прикладними програмами. Великий інтерес становлять методи засновані на аналізі формальних концептів та ієрархічній кластеризації зогляду на їх більшу незалежність від мов текстових документів та потужну математичну основу. Для автоматизації стадії валідації ефективним є використання сторонніх словників і тезаурусів, що відповідно приводить до потреби розробки україномовних відповідників WordNet. Ключовою складовою системи керування знаннями має бути засіб узгодження та злиття онтологій, що забезпечив би коректну інтеграцію знань.

### ВИКОРИСТАНА ЛІТЕРАТУРА

1. Xu J., Li W. Using Relational Database to Build OWL Ontology from XML Data Sources. [Електронний ресурс]. Xu, Li W. - Режим доступу: <http://bit.ly/pclrWE>
2. Moldovan D., Girju R. Domain-Specific Knowledge Acquisition and Classification using WordNet. [Електронний ресурс] / D. Moldovan, R Girju. - Режим доступу: <http://www.seas.smn.edu/~roxana/papers/nairs2000.ps>
3. Ермаков А.Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово. 4-8 июня 2008 г.). Вып. 7 (14).- М.: РГГУ. 2008. С. 154-158.
4. Obitko M., Snoel V. Ontology Design with Formal Concept Analysis. [Клектронний ресурс] / M. Obitko, V. Snoel, J. Smid - Режим доступу: <http://bit.ly/pfu5FO>