

М. Л. Бойчук

# ОНТОЛОГІЯ СИНТАКСИЧНОГО АНАЛІЗУ УКРАЇНСЬКОЇ МОВИ

У статті описано результати побудови синтаксичного модуля лінгвістичного процесора української мови та його призначення.

**Ключові слова:** онтологія, синтаксичний аналіз, автоматична розбудова.

## 1. Вступ

Дослідження, про які йдеться у доповіді, відносяться до галузі автоматичної розбудови онтологій на базі джерел інформації, поданих українською мовою. Для того, щоб вручну побудувати повну зв'язану онтологію для певної предметної області (ПО) необхідно затратити достатньо багато часу та ресурсів. Це пояснюється тим, що прикладні онтології повинні містити десятки тисяч елементів, щоб бути придатними для розв'язування широкого кола задач, які виникають у цих ПО. Ручна побудова онтології — це довгий рутинний процес, який до того ж, вимагає ґрунтовних знань ПО та розуміння принципів побудови онтологій. Тому активно розвиваються методи та алгоритми автоматизованої побудови онтологій. В першу чергу це системи зорієнтовані на опрацювання природномовних англійських текстів, наприклад Link Grammar Parser. Однак, знання, які передбачаються додавати в онтологію, часто знаходяться в україномовних джерелах. Тому дослідження, про які йдеться в доповіді є актуальними.

## 2. Постановка проблеми

Опрацювати текст поданий українською мовою з метою видобування онтологічних знань неможливо без лінгвістичного процесора, який включає в себе морфологічний, синтаксичний та семантичний аналізи. Проблемою є побудова такого процесора для текстів, поданих українською мовою.

## 3. Основна частина

**3.1. Аналіз літературних джерел по темі дослідження.** Інтелектуальні системи, ядром баз знань яких є онтології, показали на практиці свою ефективність [1], зокрема під час побудови віртуального робочого місця наукового працівника [2]. У роботах [3–6] наведено функціонування таких систем для задач класифікації, а у [7, 8] для задач планування. Однак основною проблемою ефективного функціонування таких систем є автоматизація процесу розбудови онтології бази знань. Такий процес розглянуто, зокрема, у роботі [9].

**3.2. Результати досліджень.** У рамках проведених досліджень було побудовано онтологію синтаксичного аналізу української мови як окремих модуль лінгвістичного процесора. Під час побудови онтології виділяють такі блоки: терміносистема; номенклатура; засоби та правила формування понятійного апарату і термінів. Отже, щоб побудувати онтологію, необхідно побудувати її терміносистему  $O_T$  та номенклатуру  $O_N$ . Базова онтологія обов'язково містить частину терміносистеми, тобто  $O_{base} \cap O_T \neq \emptyset$ . Енциклопедії, термінологічні та тлумачні словники, на підставі яких будують терміносистему ПО, як правило, мають чітку структуру і складаються із словникових статей. Досліджено їх структуру з метою розпізнавання понять і відношень між ними. Процес побудови номенклатури складніший. Якщо в словниках терміни вже виокремлені, то в наукових текстах (підручники, монографії тощо) їх необхідно виділяти, вести пошук властивостей понять і відношень між поняттями. Отже, потрібна технологія природномовного опрацювання наукового тексту.

Нехай задана множина назв відношень  $V = \{v_1, v_2, \dots, v_s\}$ . Тоді відношення задається як відображення із  $C$  в  $C$ , використовуючи елемент множини  $V$ :  $R: C \xrightarrow{V} C$ . Тобто відношення  $r_i$  — триплет вигляду:  $r_i = \langle C_{i_1}, v_{i_j}, C_{i_2} \rangle$ . Оскільки онтологія формує таксономію понять, то, використовуючи, термінологію об'єктно-орієнтованого підходу, кожне поняття являє собою клас. Запропоновано поняття визначити як клас з такою структурою:  $C = \langle N, R^X, R^Y, S, D, A, Ob \rangle$ , де  $N$  — ім'я поняття;  $R^X$  — множина відношень, в яких клас  $C$  є доменом (областю визначення);  $R^Y$  — множина відношень, в яких клас  $C$  є множиною значень;  $S$  — суперкласи  $C$ ;  $D$  — підкласи  $C$ ;  $A$  — аксіоми визначення  $C$ ,  $Ob$  — екземпляри  $C$ .

Отже, розбудовуючи базову онтологію  $O_{base}$ , необхідно будувати триплети  $r_i$  та нові поняття  $C$ , які задаються вище наведеною структурою.

Приклад ієрархії понять розробленої онтології наведено на рис. 1. Окрім того побудовано множину SWRL-правил та перевірено якість онтології на основі мови запитів SPARQL. На черзі наповнення онтології окремими екземплярами.

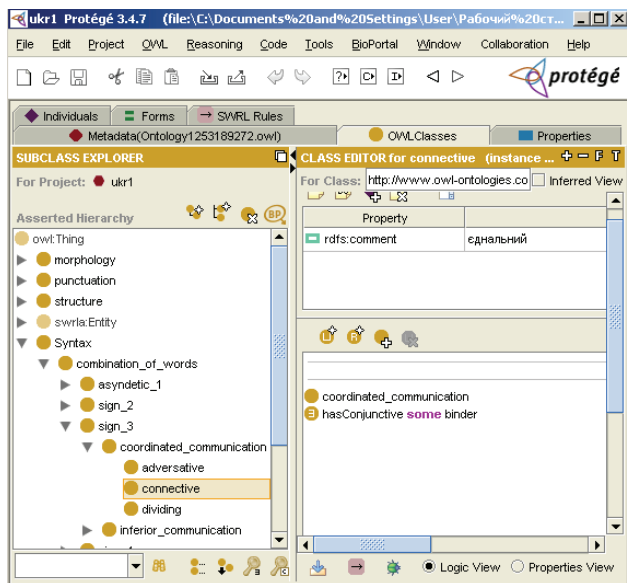


Рис. 1. Ієрархія понять синтаксичного аналізу української мови

В результаті виконання роботи розроблено окремий модуль лінгвістичного процесора опрацювання природномовних текстів, поданих українською мовою, який містить правила про синтаксис рідної мови. Створена онтологія є достатньо розробленою і дозволяє виконувати такі важливі функції: створення ієрархії класів та підкласів понять предметної області; введення представників класів та підкласів, що розширює можливості розуміння і використання бази знань; створення системи зв'язків між класами та підкласами; створення і виконання запитів різного характеру; побудову правил для опрацювання даних.

**Література**

1. Литвин В. В. Бази знань інтелектуальних систем підтримки прийняття рішень [Текст] : монографія / **В. В. Литвин** // Міністерство освіти і науки, молоді та спорту України, Національний університет «Львівська політехніка». — Львів : Видавництво Львівської політехніки, 2011. — 240 с.
2. Литвин В. В. Застосування інформаційних технологій для координації наукових досліджень [Текст] : монографія / Р. Р. Даревич, Д. Г. Досин, **В. В. Литвин**, Л. С. Мельничок // Національна академія наук України, Фізико-механічний інститут ім. Г. В. Карпенка. — Львів : «СПОЛОМ», 2008. — 240 с.
3. Литвин В. В. Моделювання інтелектуальних систем підтримки прийняття рішень з використанням онтологічного підходу [Текст] / **В. В. Литвин** // Радіоелектроніка, інформатика, управління. — Запорізький національний технічний університет. — Запоріжжя, 2011. — № 2(25). — С. 93–101.
4. Литвин В. В. Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології [Текст] / **В. В. Литвин** // Радіоелектроніка, інформатика, управління. — Запорізький національний технічний університет. — Запоріжжя, 2009. — № 2(21). — С. 120–126.

5. Литвин В. В. Інтелектуальні агенти пошуку релевантних прецедентів на основі адаптивних онтологій [Текст] / **В. В. Литвин** // Математичні машини і системи. — Національна академія наук України; Інститут проблем математичних машин і систем. — Київ, 2011. — № 3. — С. 66–72.
6. Литвин В. В. Проектування інтелектуальних агентів прийняття рішень в просторі ознак з використанням онтологічного підходу [Текст] / **В. В. Литвин**, Р. Р. Даревич, Д. Г. Досин, Н. В. Шкутяк // Штучний інтелект. — Національна академія наук України; Інститут проблем штучного інтелекту. — Донецьк, 2010. — № 4. — С. 398–403.
7. Литвин В. В. Інтелектуальні системи підтримки прийняття рішень на основі адаптивних онтологій [Текст] / **В. В. Литвин**, Р. Р. Даревич, Д. Г. Досин, А. С. Мельник, А. О. Мазур // Штучний інтелект. — Національна академія наук України; Інститут проблем штучного інтелекту. — Донецьк, 2011. — № 2. — С. 35–44.
8. Литвин В. В. Моделювання плану поведінки інтелектуального агента на основі мереж Петрі та онтологічного підходу [Текст] / **В. В. Литвин** // Вісн. Нац. ун-ту «Львівська політехніка». Серія : Інформаційні системи та мережі. — 2009. — № 653. — С. 170–175.
9. Литвин В. В. Метод автоматичної розбудови адаптивної онтології [Текст] / **В. В. Литвин**, Д. І. Угрин // Вісник Національного технічного університету «ХПІ». — Харків, 2011. — № 10. — С. 75–82.

**ОНТОЛОГИЯ СИНТАКСИЧЕСКОГО АНАЛИЗА УКРАИНСКОГО ЯЗЫКА**

**М. Л. Бойчук**

В статье описаны результаты построения синтаксического модуля лингвистического процессора украинского языка и его назначение.

**Ключевые слова:** онтология, синтаксический анализ, автоматическое построение.

*Марьяна Любомировна Бойчук, магистр кафедры информационных систем и сетей Национального университета «Львовская политехника», тел.: (032) 258-25-38, e-mail: m.bojchuk@mail.ru.*

**ONTOLOGY PARSING UKRAINIAN LANGUAGE**

**M. Wojchuk**

The article describes the results of constructing syntax module linguistic processor Ukrainian language and its purpose.

**Keywords:** ontology, syntactic analysis, automatic building.

*Maryana Wojchuk, student of Department of Information Systems and Nets, National University Lviv Polytechnic, tel.: (032) 258-25-38, e-mail: m.bojchuk@mail.ru.*