

О.В. Лозинська, М.В. Давидов

Предметно-орієнтована мова опису граматично доповненої онтології

На основе разработанной математической модели грамматически дополненной онтологии украинского словесного и жестового языков для построения системы перевода украинского жестового языка разработан предметно-ориентированный язык описания для наполнения данной грамматически дополненной онтологии.

На основі розробленої математичної моделі граматично доповненої онтології української словесної та жестової мов для побудови системи перекладу української жестової мови розроблено предметно-орієнтовану мову опису для наповнення даної граматично доповненої онтології.

Вступ. Українська жестова мова (УЖМ) є природним способом спілкування людей з вадами слуху, тому необхідність отримання інформації жестовою мовою для такої категорії громадян є важливою. Для вирішення цього завдання треба розробити комп'ютерну систему перекладу УЖМ, коли виникають такі ж проблеми, як і при перекладі з однієї мови на іншу. Задача перекладу УЖМ на українську словесну мову (УСМ) ускладнюється й тим, що кількість жестів значно менша за кількість слів УСМ [1]. Саме з цим пов'язана проблема усунення багатозначності слів *WSD (Word Sense Disambigulation)*, тобто вибору значення слова відповідно до контексту. Відомими методами вирішення цієї проблеми є методи навчання з вчителем, коли використовують розмічені корпуси, та без вчителя [2], методи на основі знань [3]. Для правильного визначення сенсу слова у контексті використовуються різноманітні джерела знань, до яких відносять електронні словники, тезауруси, онтології (наприклад, *WordNet*), корпуси текстів.

Використання онтології як лінгвістичного ресурсу для усунення багатозначності слів зумовлене їх формальною структурою, яка спрощує комп'ютерну обробку [4]. Онтологія визначає поняття (концепти) в певній галузі знань, а також співвідношення між цими поняттями.

Українськими вченими [3] досліджено проблему вирішення лексичної багатозначності та проаналізовано основні методи її розв'язання. З цією метою обрано методи на основі знань, досліджено сучасні інформаційні ресурси та встановлено, що мережа *WordNet* надає найбільше можливостей для автоматичного розрі-

знення значень слів. *WordNet* – це велика електронна база даних, організована в семантичну мережу, яка складається з різних співвідношень між словами, що включає синонімію, антонімію, узагальнення та деталізацію [5].

Постановка проблеми

Суттєвою задачею є інтегрування онтології в системи машинного перекладу. Для цього вона має містити більше інформації про граматичні зв'язки. Для ефективного усунення багатозначності слів при машинному перекладі, пропонується використовувати граматично доповнену онтологію, яка міститиме граматичну інформацію (граматичне значення слова, його граматичну форму).

Для практичного втілення розробленої математичної моделі граматично доповненої онтології необхідно створити інформаційну технологію, яка б давала змогу описувати цю онтологію, і перетворювала її у форму, придатну для використання у системах машинного перекладу.

Аналіз досліджень

З розвитком комп'ютерних технологій і техніки з'явилася можливість створення баз знань для автоматичного опрацювання інформації. Зростає інтерес до формальних онтологій як ефективного засобу подання знань. Онтології широко використовуються для досліджень в галузі комп'ютерної лінгвістики, побудови систем семантичного аналізу тексту та в інших галузях опрацювання природної мови. Онтології розроблено для багатьох мов, а саме *WordNet* для англійської [5], *plWordNet* для польської [6], *CWN (Chinese Wordnet)* для китайської [7], *WOLF (WordNet Libre du Français)* для французької [8], *MultiWordNet project* для італійської

[9], *The BalkaNet project* для шести європейських мов (болгарської, чеської, грецької, румунської, турецької та сербської) [10], *GermaNet* для німецької мови [11], проект *RussNet* [12] і тезаурус *RuThes* [13] – для російської.

Науковцями Київського національного університету імені Тараса Шевченка розроблено багаторівневий програмний комплекс для поповнення, редагування даних української онтологічної лексико-семантичної бази знань *UkrWordNet (UWN)* та модерзації роботи редакторів у мережі [15]. На даний момент онтологія містить близько 80 тис. концептів.

Українські вчені займаються розробленням *WordNet*-подібного словника для частини іменників української мови [14]. У результаті дослідження розроблено фрагмент *WordNet*-подібного словника української мови, в якому реалізовано 194 синсети, пов'язані між собою зв'язками гіпо-/гіперонімії, антонімії, а також додатково зв'язками метонімії/голонімії.

Вибір системи побудови онтології залежить від практичного завдання (мети розробника), галузі знань, в межах якої будується онтологія.

Створення онтології складається з таких етапів:

- виділення понять;
- формулювання відношень між поняттями;
- наповнення онтології конкретними прикладами;
- формулювання аксіом (правил).

Відомі системи побудови онтологій *DOE (Differential Ontology Editor)* [16], *Ontolingua* [17], *OntoEdit* [18], *WebOnto* [19], *Protégé* [20] не мають засобів зв'язування понять і можливих граматичних конструкцій для їх вираження. Для введення в онтологію граматичних конструкцій української словесної та жестової мов було розроблено спеціальну мову опису граматично доповненої онтології.

Мова для опису граматично доповненої онтології

Основою для побудови мови опису граматично доповненої онтології стала математична модель такої онтології [21].

Наповнення онтології відбувається за використання розробленої предметно-орієнтованої

мови (*Domain Specific Language, DSL*). Цей термін позначає мову програмування або моделювання, яка застосовується для розв'язання конкретних задач в термінах, максимально наближених до даної предметної галузі [22].

Предметно-орієнтовану мову *UkrNet* реалізовано з використанням додатку *Xtext* середовища розробки програмного забезпечення *Eclipse*, який дозволяє створити для граматики мову *DSL* [23]. Платформа *Xtext* на основі граматики *DSL*-мови забезпечує створення набору засобів для наповнення, редагування та верифікації граматично доповненої онтології, що включає в себе аналізатор коду мови, інструмент форматування коду, компілятор, редактор коду та ін. Вікно текстового редактора для мови *UkrNet* зображено на рисунку.

Розроблення предметно-орієнтованої мови *UkrNet* здійснено з урахуванням граматики даної мови в *Xtext*. Потім, використовуючи засоби *Xtext*, створено синтаксичний аналізатор і текстовий редактор для мови *UkrNet*.

Розглянемо поняття граматично доповненої онтології.

Для позначення граматичних значень введено поняття ***grammar_attribute*** <назва категорії> [= опис]. Граматичні категорії можуть бути спільними для декількох частин мови. Категорія числа охоплює багато частин мови (іменник, прикметник, займенник, дієслово), так само й категорія роду, а категорії особи, часу, виду, способу властиві тільки дієсловом, що пояснюється самим виявом предметів і явищ. Граматична категорія власних показників не має, вона виявляється через граматичні значення, засобами вираження яких є граматичні форми. Для опису онтології з граматичною компонентою введено такі граматичні значення:

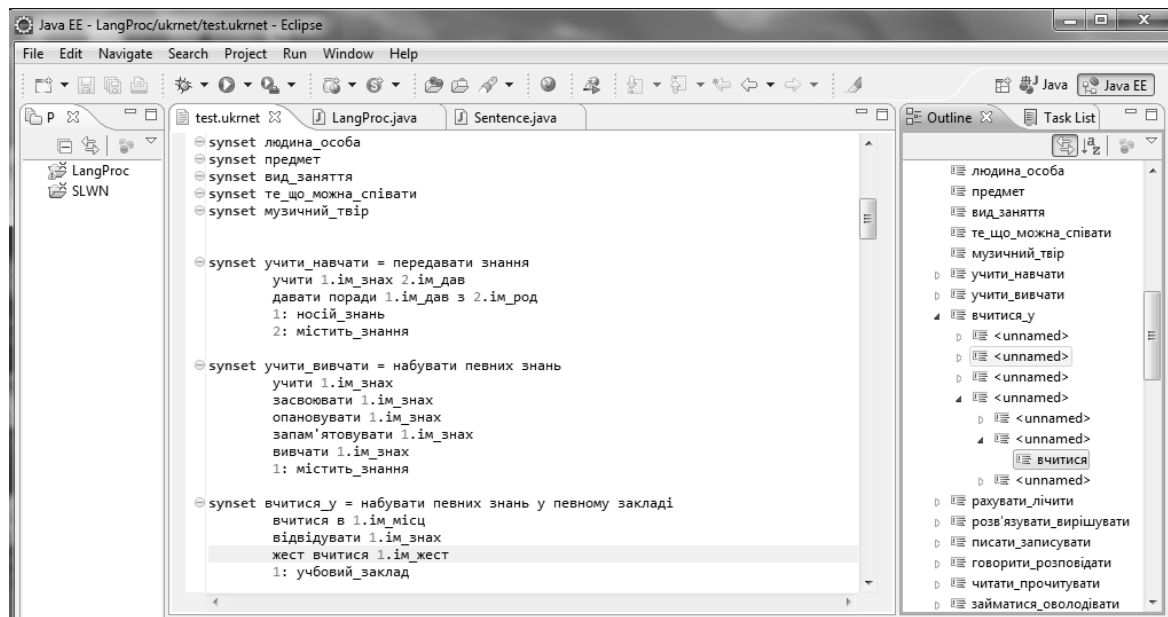
grammar_attribute ім_наз // іменник у називному відмінку

grammar_attribute ім_род // іменник у родовому відмінку

grammar_attribute ім_дав // іменник у давальному відмінку

grammar_attribute ім_знах // іменник у знахідному відмінку

grammar_attribute ім_оруд // іменник у орудному відмінку



grammar_attribute ім_місц // іменник у місцевому відмінку

grammar_attribute д_інф // дієслово (інфінітив)

grammar_attribute д_жест // дієслово (жест)

grammar_attribute ім_жест // іменник (жест)

Для позначення синсетів використовуються наступні описи:

synset <назва синсету> [= опис] // синсет без відповідних словесних реалізацій (просто поняття);

synset <назва синсету> [= опис]

учити 1.ім знах 2.ім дав // певне словесне вираження, напр. учити учнів плаванню

давати поради 1.ім_дав з 2.ім_род // інша словесна реалізація, напр. давати поради учням з плавання

1: носій_знань // вимоги до слова 1 (*optional*)

2: містить_знання // вимоги до слова 2 (*optional*)

Кожен синсет представляє деяке значення, поняття мови. Синсет можна розглядати як множину значень слів, які виражають одне поняття. Усі іменники, дієслова, прикметники і прислівники української словесної та української жестової мов утворюють синонімічні набори (синсети), кожен з яких представляє один семантичний концепт.

Для систематизації іменників введено загальне поняття до синсету1: **common**(<назва синсету1>) = <назва синсету2>, що дає змогу побудувати родовидове відношення, відоме як гіперонімія, та обернене йому відношення гіпонімії. Наприклад, синсет *книга* є гіпонімом синсету *носії_знань*:

common(книга) = носій_знань

Введене поняття **parts**(<назва синсету1>) визначає відношення *частина-ціле*, або меронімію. Опис даного поняття має такий вигляд:

parts(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN є невід'ємною частиною синсету1;

Наприклад, синсет *книга* завжди містить синсет *сторінка*:

parts(книга) = сторінка

Поняття **optional_parts**(<назва синсету1>) має наступний опис:

optional_parts(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN є можливою частиною синсету1 (у звичному розумінні)

Наведене поняття подібне до співвідношення **parts**(<назва синсету1>) та використовується у випадку, коли синсет2,..., синсетN є можливою частиною синсету1. Наприклад, синсет *сторінка* не завжди має синсет *номер_сторінки*:

optional_parts(сторінка) = номер_сторінки

Для позначення співвідношення однозначної асоціації між синсетами використовується поняття **associations**(<назва синсету1>), що має наступний опис:

associations(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN завжди асоціюються з синсет1 ;

Для прикладу, синсет *школяр* завжди асоціюється з синсетом *школа*:

associations(школяр) = школа // школяр завжди ходить в школу

Для позначення співвідношення можливої асоціації між синсетами введено поняття *optional_associations*(<назва синсету1>). Це поняття має наступний опис в онтології:

optional_associations(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN можуть асоціюватися з синсет1;

Це співвідношення використовується тоді, коли не завжди синсет1 асоціюється з синсетом2,..., синсетомN. Наприклад, синсет *заліковка* не завжди є у студента:

optional_associations(студент) = заліковка // часто студенти мають заліковки

Крім того, поняття може входити у різні класифікації:

common(учень) = людина // учень має всі ознаки людини

common(людина) = людина_біологічний_вид // людина має всі ознаки людина_біологічний_вид

Отже, учень має всі ознаки людина_біологічний_вид.

При цьому зрозуміло, що учень не є більш конкретним представником людина_біологічний_вид:

parts(учень) = людина_біологічний_вид

Приклад неправильного співвідношення – школяр не є частиною учня, а є більш конкретним об'єктом.

optional_parts(учень) = школяр // неправильно!!!

common(школяр) = учень // правильно

Аналіз результатів дослідження

Для побудови граматично доповненої онтології вибрано такі предметні галузі: *Навчання, Природа, Подорож, Держава, Сім'я, Виробництво, Професії, Військо, Театр, Культура, Лікарня*. Для цього зібрано 1200 слів, пов'язаних з даними галузями, та описано пояснення до кожного слова з використанням тлумачного словника української мови [24] та словника жестів [25]. Також до кожного слова зібрано всі його синоніми. Після цього в редакторі онтологій занесено ці слова у вигляді синсетів (синонімічних рядів) та описано граматичні конструкції, в яких використовується це слово.

У синсет також включено граматичні конструкції жестової мови. Отже, можна отримати однозначний переклад певного словосполучення.

Наприклад, розглянемо слово *учити*. У словосполученні слово *учити* може означати *навчати когось* (учити учнів математиці) або може означати *вивчати самому* (учити математику). Тому для цього слова у граматичній онтології записано два синсети:

synset *учити_навчати* = передавати знання
учити 1.ім_знах 2.ім_дав
давати поради 1.ім_дав з 2.ім_род
жест навчати 1.ім_жест 2.ім_жест
1: носій_знань
2: містить_знання

synset *учити_вивчати* = набувати певних знань
учити 1.ім_знах
засвоювати 1.ім_знах
опанувати 1.ім_знах
запам'ятовувати 1.ім_знах
вивчати 1.ім_знах
жест вивчати 1.ім_жест
1: містить_знання

Також записано синсет *учитись у*, що означає *учитись у певному навчальному закладі*:

synset *вчитися_у* = набувати певних знань у певному закладі

вчитися в 1.ім_місц
відвідувати 1.ім_знах
жест вчитися 1.ім_жест

1: учбовий_заклад

Висновки. На основі аналізу граматики української словесної та української жестової мов розроблено математичну модель граматично доповненої онтології. Для того, щоб забезпечити можливість введення в онтологію граматичних категорій, розроблено спеціалізовану предметно-орієнтовану мову опису *UkrNet* граматично доповненої онтології. Предметно-орієнтовану мову *UkrNet* реалізовано з використанням додатку *Xtext* середовища розробки програмного забезпечення *Eclipse*, який дає змогу для граматики створити мову *DSL*. Граматично доповнена онтологія містить 1200 слів з різних предметних галузей. Ці слова занесено у вигляді синсетів (синонімічних рядів) та описано граматичні конструкції, в яких використовуються дані слова. У синсет також включено граматичну конструкцію жестової мови. Таким чином, можна отримати однозначний переклад певного словосполучення.

ня. Отже, граматично доповнена онтологія дозволяє покращити результат перекладу і може використовуватись у системах машинного перекладу.

1. Бармак О., Романишин С. Комп'ютерна технологія для перекладу з української на жестову мову // Вісн. ТНТУ. – 2012. – 65, № 1. – С. 195–208.
2. Navigli R. Word Sense Disambiguation: A Survey // ACM Computing Surveys, ACM Press. – 2009. – N 41(2). – P. 1–69.
3. Романюк А.Б., Сундутова І.А., Романишин М.М. Методи вирішення лексичної багатозначності. Використання WORDNET для вирішення проблем // Вісн. Нац. ун-ту «Львівська політехніка» Комп'ютерні системи проектування. Теорія і практика. – 2011. – № 711. – С. 147–157.
4. Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Снятие лексико-семантической омонимии в новостных и газетно-журнальных текстах: поверхностные фильтры и статистическая оценка // Интернет-математика. – 2005. – С. 38–57.
5. Al-Halimi R., Berwick R.C., Burg J.F.M. An Electronic Lexical Database / Ed. by Christiane Fellbaum. – Cambridge, MA: MIT Press, 1998. – 422 p.
6. Проект PlWordNet. – <http://plwordnet.pwr.wroc.pl/wordnet/>
7. Chinese WordNet: Design and Implementation of a Cross-Lingual Knowledge Processing Infrastructure / H. Chu-Ren, H. Shu-Kai, H. Jia-Fei et al. // J. of Chinese Inform. Processing. – 2010. – N 24(2). – P. 14–23.
8. Benoît S., Darja F. Building a free French wordnet from multilingual resources // Proc. of Ontolex 2008.
9. Pianta E., Bentivogli L., Girardi C. MultiWordNet: Developing an aligned multilingual database // Proc. of the 1st Int. Conf. on Global WordNet, Mysore, India. – 2002. – P. 21–25.
10. Tufis D., Cristea D., Stamou S. Balkanet: Aims, methods, results and perspectives. A general overview // Romanian J. Sci. Tech. Inform. (Special Issue on Balkanet). – 2004. – N 7(1–2). – P. 9–43.
11. Henrich V., Hinrichs E., Vodolazova T. Semi-Automatic Extension of GermaNet with Sense Definitions from Wiktionary // Proc. of 5th Language & Technology Conf. (LTC 2011), Poznań, Poland, Nov. 2011. – P. 126–130.
12. Проект RussNet. – http://www.phil.pu.ru/depts/12/RN/index_ru.shtml
13. Портал УИС «Россия». – <http://www.cir.ru/>
14. Кульчицький І.М., Романюк А.Б., Харів Х.Б. Розроблення wordnet-подібного словника української мови // Вісн. Нац. ун-ту «Львівська політехніка»: Інформ. сист. та мережі. – 2011. – № 673. – С. 306–319.
15. Анісімов А.В., Марченко О.О., Никоненко А.О. UWN: Універсальна онтологічна база знань української мови // Проблеми програмування. – 2012. – № 2–3. – С. 348–355.
16. Проект DOE, 2006. – <http://opales.ina.fr/public/index.htmlFarquhar>
17. Fickas A., Rice J. The Ontolingua Server: a tool for collaborative ontology construction // Proc. of the 10th Banff KnowledgeAcquisition for Knowledge based system workshop. – 1995.
18. Sure Y., Erdmann M., Studer R. OntoEdit: Collaborative Engineering of Ontologies // Proc. of the First Int. Semantic Web Conf. 2002, Sardinia, Italia, 2002. – P. 221–235.
19. Domingue J., Motta E., Garcia C.O. Knowledge modelling in WebOnto and OCML: A user guide. – 1999.
20. Creating Semantic Web Contents with Protege-2000 / N.F. Noy, M. Sintek, S. Decker et al. / IEEE Intel. Syst., 2011. – N 16. – P. 60–71.
21. Лозинська О.В., Давидов М.В. Математична модель граматично-доповненої онтології // Вісн. Нац. техн. ун-ту «ХП». – 2015. – № 11(1120). – С. 102–108.
22. Казакова А.С. Методы и инструменты реализации предметно-ориентированных языков программирования. – М: Вильямс, 2009. – С. 1–79.
23. Жуков М.В. Реализация предметно-ориентированных языков средствами фреймворка Xtext // Изв. Пензенского гос. пед. ун-та. – 2009. – № 13 (17). – С. 98–102.
24. Білодід І.К. Словник української мови: В 11 т. – Київ: Наук. думка, 1970–1980.
25. Програмний тренажер навчання сурдоперекладу / М.В. Давидов, В.В. Пасічник, Ю.В. Нікольський та ін. / Проблеми освіти : Наук.-метод. зб. – 2009. – № 60. – С. 108–113.

Поступила 08.04.2015

E-mail: kanmirg@gmail.com, maks.davydov@gmail.com

© О.В. Лозинская, М.В. Давыдов, 2015

О.В. Лозинская, М.В. Давыдов

Предметно-ориентированный язык описания грамматически дополненной онтологии

Введение. Украинский жестовый язык (УЖЯ) – естественное средство общения людей с недостатками слуха, поэтому необходимость получения информации на жестовом языке для такой категории граждан важна. Для решения

этой задачи необходимо разработать компьютерную систему перевода УЖЯ, когда возникают такие же проблемы, как и при переводе с одного языка на другой. Задача перевода УЖЯ на украинский словесный язык (УСЯ) осложняет

ется и тем, что количество жестов значительно меньше количества слов УСЯ [1]. Именно с этим связана проблема устранения многозначности слов *WSD (Word Sense Disambiguation)*, т.е. выбора значения слова в соответствии с контекстом. Известные методы решения этой проблемы – методы обучения с учителем, с использованием размеченных корпусов, и без учителя [2], методы на основе знаний [3]. Для правильного определения значения слова в контексте используются различные источники знаний, к которым относятся электронные словари, тезаурусы, онтологии (например, *WordNet*), корпуса текстов.

Использование онтологии как лингвистического ресурса для устранения многозначности слов обусловлено их формальной структурой, упрощающей компьютерную обработку [4]. Онтология определяет понятия (концепты) в определенной области знаний, а также связи между этими понятиями.

Украинскими учеными [3] исследована проблема решения лексической многозначности и проанализированы основные методы ее решения. С этой целью избраны методы на основе знаний, исследованы современные информационные ресурсы и установлено, что сеть *WordNet* предоставляет больше возможностей для автоматического различения значений слов. *WordNet* – это большая электронная база данных, организованная в семантическую сеть, состоящую из различных связей между словами, включая синонимы, антонимы, обобщения и детализацию [5].

Постановка проблемы

Существенной задачей есть интеграция онтологии в системы машинного перевода. Для этого она должна содержать больше информации о грамматических связях. Для эффективного устранения многозначности слов при машинном переводе предлагается использовать грамматически дополненную онтологию, которая будет содержать грамматическую информацию (грамматическое значение слова, его грамматическую форму).

Для практического воплощения разработанной тематической модели грамматически дополненной онтологии необходимо создать информационную технологию, которая даст возможность описывать эту онтологию, а также превратит ее в форму, пригодную для использования в системах машинного перевода.

Анализ исследований

С развитием компьютерных технологий и техники появилась возможность создания баз знаний для автоматической обработки информации. Растет интерес к формальным онтологиям как эффективного средства представления знаний. Онтологии широко используются для исследований в области компьютерной лингвистики, построения систем семантического анализа текста и в других отраслях обработки естественного языка. Онтологии разработаны для многих языков, а именно онтология *WordNet* для английского [5], *plWordNet* для польского [6], *CWN (Chinese Wordnet)* для китайского [7], *WOLF (WordNet Libre du Français)* для французского [8], *MultiWordNet project* для итальянского [9], *The*

BalkaNet project для шести европейских языков (болгарского, чешского, греческого, румынского, турецкого и сербского) [10], *GermaNet* для немецкого [11], проект *RussNet* [12] и тезаурус *RuThes* [13] для русского языка.

Учеными Киевского национального университета имени Тараса Шевченко разработан многоуровневый программный комплекс для наполнения, редактирования данных украинской онтологической лексико-семантической базы знаний *UkrWordNet (UWN)* и модерации работы редакторов в сети [15]. На данный момент онтология содержит около 80 тыс. концептов.

Украинские ученые занимаются разработкой *WordNet*-подобного словаря для части существительных украинского языка [14]. В результате исследования разработан фрагмент *WordNet*-подобного словаря украинского языка, в котором реализовано 194 синсета, связанных между собой гипо-/гиперонимией, антонимией, а также дополнительно связями метонимии/голонимии.

Выбор системы построения онтологии зависит от практической задачи (целей разработчика), области знаний, в рамках которой строится онтология.

Создание онтологии состоит из следующих этапов:

- выделения понятий;
- формулировки отношений между понятиями;
- наполнения онтологии конкретными экземплярами;
- формулировки аксиом (правил).

Известные системы построения онтологий *DOE (Differential Ontology Editor)* [16], *Ontolingua* [17], *OntoEdit* [18], *WebOnto* [19], *Protégé* [20] не имеют средств связывания понятий и возможных грамматических конструкций для их выражения. Для введения в онтологию грамматических конструкций украинского словесного и жестового языков был разработан специальный язык описания грамматически дополненной онтологии.

Язык для описания грамматически дополненной онтологии

Основой для построения языка описания грамматически дополненной онтологии стала математическая модель такой онтологии [21].

Наполнение онтологии происходит при использовании разработанного предметно-ориентированного языка (*Domain Specific Language, DSL*). Этот термин обозначает язык программирования или моделирования, применяемый для решения конкретных задач в терминах, максимально приближенных к данной предметной области [22].

Предметно-ориентированный язык *UkrNet* реализован с использованием приложения *Xtext* среды разработки программного обеспечения *Eclipse*, позволяющим создать для грамматики язык *DSL* [23]. Платформа *Xtext* на основе грамматики *DSL*-языка обеспечивает создание набора средств для наполнения, редактирования и верификации грамматически дополненной онтологии, включает в себя анализатор кода языка, инструмент форматирования кода, компилятор, редактор кода и др. Окно

текстового редактора для языка *UkrNet* изображено на рисунке.

Разработка предметно-ориентированного языка *UkrNet* осуществлено с учетом грамматики данного языка в *Xtext*. Затем, используя средства *Xtext*, создан синтаксический анализатор и текстовый редактор для языка *UkrNet*.

Рассмотрим понятия грамматически дополненной онтологии.

Для обозначения грамматических значений введено понятие ***grammar_attribute*** <название категории> [= описание]. Грамматические категории могут быть общими для нескольких частей речи. Категория числа охватывает много частей речи (существительное, прилагательное, местоимение, глагол), так же и категория рода, а категории лица, времени, вида, способа свойственны только глаголам, что объясняется самим проявлением предметов и явлений. Грамматическая категория своих показателей не имеет, она проявляется через грамматические значения, средствами выражения которых являются грамматические формы. Для описания онтологии с грамматической компонентой введены следующие грамматические значения:

grammar_attribute *сущ_им* // существительное в именительном падеже

grammar_attribute *сущ_род* // существительное в родительном падеже

grammar_attribute *сущ_дат* // существительное в дательном падеже

grammar_attribute *сущ_вин* // существительное в винительном падеже

grammar_attribute *сущ_тв* // существительное в творительном падеже

grammar_attribute *сущ_пред* // существительное в предложном падеже

grammar_attribute *гл_инф* // глагол (инфинитив)

grammar_attribute *гл_жест* // глагол (жест)

grammar_attribute *сущ_жест* // существительное (жест)

Для обозначения синсетов используется следующее описание:

synset <название синсета> [= описание] // синсет без соответствующих словесных реализаций (просто понятия);

synset <название синсета> [= описание]

учить 1.сущ_вин 2.сущ_дат // определенное словесное выражение, например учить учеников плаванию

давать советы 1.сущ_дат с 2.сущ_род // другая словесная реализация, например давать советы ученикам по плаванию

1: носитель_знаний // требования к слову 1 (*optional*)

2: содержит_знания // требования к слову 2 (*optional*)

Каждый синсет представляет некоторое значение, понятие языка. Синсет можно рассматривать как множество значений слов, выражающих одно понятие. Все существительные, глаголы, прилагательные и наречия украинского словесного и украинского жестового языков образуют синонимические наборы (синсеты), каждый из которых представляет один семантический концепт.

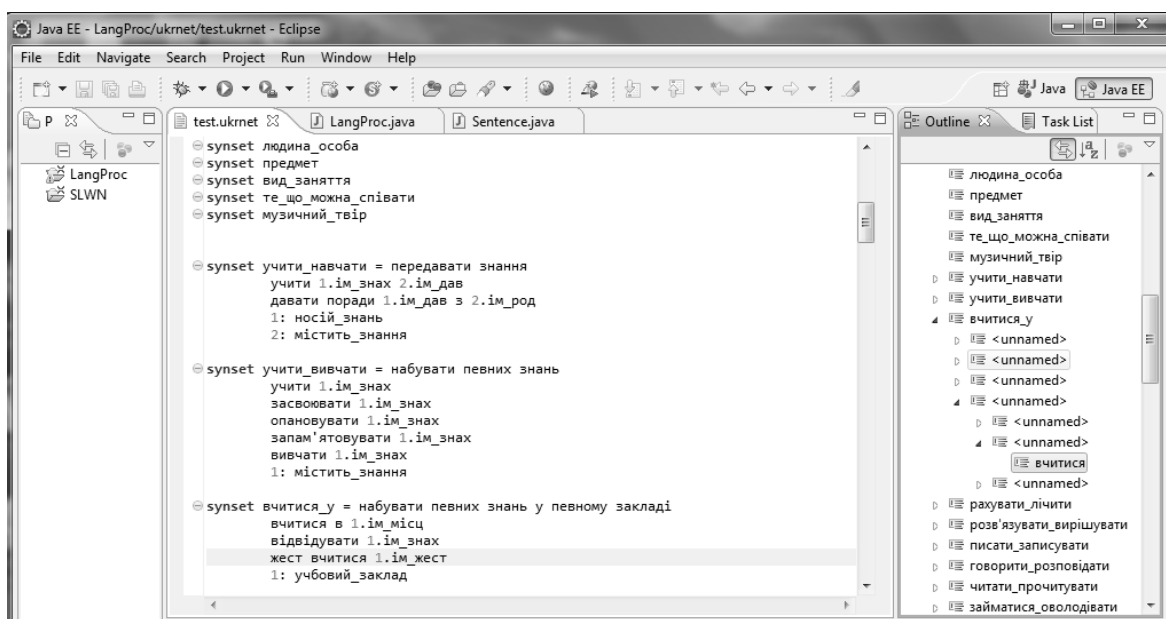
Для систематизации существительных введено общее понятие к синсету1: ***common*** (<название синсета1>) = <название синсета2>, что позволяет построить родо-видовое отношение, известное как гиперонимия, и обратное ему отношение гипонимии. Например, синсет *книга* есть гипонимом синсету *носитель_знаний*:

common (книга) = носитель_знаний

Введение понятия ***parts*** (<название синсета1>) определяет отношение «часть-целое», или меронимию. Описание данного понятия выглядит так:

parts (<название синсета1>) = <название синсета2>, ..., <название синсетаN> // синсет2, ..., синсетN является неотъемлемой частью синсета1;

Например, синсет *книга* всегда содержит синсет *страница*:



parts (книга) = страница

Понятие *optional_parts* (<название синсета1>) имеет следующее описание:

optional_parts (<название синсета1>) = <название синсета2>, ..., <название синсетаN> // синсет2, ..., синсетN может быть частью синсета1 (при обычном понимании);

Приведенное понятие подобно отношению *parts* (<название синсета1>) и используется в том случае, когда синсет2, ..., синсетN является возможной частью синсета1. Например, синсет *страница* не всегда имеет синсет *номер_страницы*:

optional_parts (страница) = номер_страницы

Для обозначения отношения однозначной ассоциации между синсетами используется понятие *associations* (<название синсета1>), имеющее следующее описание:

associations (<название синсета1>) = <название синсета2>, ... <название синсетаN> // синсет2, ..., синсетN всегда ассоциируются с синсет1;

Например, синсет *школьник* всегда ассоциируется с синсетом *школа*:

associations (школьник) = школа // школьник всегда ходит в школу

Для обозначения отношения возможной ассоциации между синсетами введено понятие *optional_associations* (<название синсета1>). Это понятие имеет следующее описание в онтологии:

optional_associations (<название синсета1>) = <название синсета2>, ..., <название синсетаN> // синсет2, ..., синсетN могут ассоциироваться с синсет1;

Это отношение используется тогда, когда не всегда синсет1 ассоциируется с синсетом2, ..., синсетомN. Например, синсет *зачетка* не всегда есть у студента:

optional_associations (студент) = зачетка // часто студенты имеют зачетки

Кроме того, понятие может входить в различные классификации:

common (ученик) = человек // ученик имеет все признаки человека

common (человек) = человек_биологический_вид // человек имеет все признаки человек_биологический_вид

Таким образом, ученик имеет все признаки человек_биологический_вид.

При этом понятно, что ученик не является более конкретным представителем человек_биологический_вид:

parts (ученик) = человек_биологический_вид

Пример неправильного отношения – школьник не является частью ученика, а более конкретным объектом.

optional_parts (ученик) = школьник // неправильно !!!

common (школьник) = ученик // правильно

Анализ результатов исследования

Для построения грамматически дополненной онтологии выбраны следующие предметные области: *Обучение, Природа, Путешествие, Государство, Семья, Производство, Профессии, Войско, Театр, Культура, Больница*. Для этого собрано 1200 слов, связанных с данными

отраслями, и описано объяснение к каждому слову с использованием толкового словаря украинского языка [24] и словаря жестов [25]. Также к каждому слову собраны все его синонимы. После этого в редакторе онтологий занесены эти слова в виде синсетов (синонимических рядов) и описаны грамматические конструкции, в которых используется это слово. В синсет также включены грамматические конструкции жестового языка. Таким образом, можно получить однозначный перевод определенного словосочетания.

Рассмотрим, например, слово *учить*. В словосочетании слово *учить* может означать *обучать кого-то* (учить учеников математике) или может означать *изучать самому* (учить математику). Поэтому для этого слова в грамматической онтологии записаны два синсета:

synset *учить_обучать* = передавать знания

учить 1.сущ_вин 2.сущ_дат

давать советы 1.сущ_дат с 2.сущ_род

жест обучать 1.сущ_жест 2.сущ_жест

1: носитель_знаний

2: содержит_знания

synset *учить_изучать* = приобретать определенные знания

учить 1.сущ_вин

усваивать 1.сущ_вин

овладевать 1.сущ_вин

запоминать 1.сущ_вин

изучать 1.сущ_вин

жест изучать 1.сущ_жест

1: носитель_знаний

Также записан синсет *учиться в*, что означает *учиться в определенном учебном заведении*:

synset *учиться_в* = приобретать определенные знания в определенном заведении

учиться в 1.сущ_предл

посещать 1.сущ_вин

жест учиться 1.сущ_жест

1: учебное_заведение

Заключение. На основе анализа грамматики украинского словесного и украинского жестового языков разработана математическая модель грамматически дополненной онтологии. Для того, чтобы обеспечить возможность введения в онтологию грамматических категорий, разработан специализированный предметно-ориентированный язык описания *UkrNet* грамматически дополненной онтологии. Предметно-ориентированный язык *UkrNet* реализован с использованием приложения *Xtext* среды разработки программного обеспечения *Eclipse*, позволяющий создать для грамматики язык *DSL*. Грамматически дополненная онтология содержит 1200 слов из разных предметных областей. Эти слова занесены в виде синсетов (синонимических рядов) и описаны грамматические конструкции, в которых они используются. В синсет также включены грамматические конструкции жестового языка. Таким образом, можно получить однозначный перевод определенного словосочетания. Итак, грамматически дополненная онтология позволяет улучшить результат перевода и может использоваться в системах машинного перевода.

Domain-Specific Language for Describing Grammatically Augmented Ontology

The problem of developing a machine translation system for sign language has been the object of scientific research for a long time. The modern society computerization, the use of multimedia and Internet technologies have created the conditions for development of computer communication systems for people with impaired hearing. For easy communication with deaf people it is enough to develop a system that translates sign language into the text and vice versa. Automatic translation can help these people to communicate and to overcome language and cultural barriers. The urgent task is to develop software tools that help in the study and translation of Ukrainian Sign Language (USL).

Ukrainian sign language is the natural way of the deaf people communication, so the need to receiving information on sign language for this category of people is important. To solve this problem it is necessary to develop a computer translation system of Ukrainian sign language. The problem of translation from Ukrainian sign language to Ukrainian spoken language (USpM) is complicated by the fact that the number of signs less than number of words of USpL. This is the problem of word sense disambiguation (WSD), which includes the selection of a word according to the context.

The formal structure of ontologies facilitates the development of word sense disambiguation algorithms. Ontologies define concepts in a particular field of knowledge, and the relationship between these concepts.

An important task is the integration of ontologies into machine translation systems. To do this, it should contain more information about grammatical attributes of words. To solve the problem of word sense disambiguation in machine translation, it is proposed to use a grammatically augmented ontology that will contain grammatical information (grammatical meaning, its grammatical form).

Based on the developed mathematical model grammatically augmented ontology of Ukrainian Spoken and Sign Languages for building Ukrainian Sign Language translation system developed domain specific language UkrNet for filling grammatically augmented ontology. UkrNet is implemented using Xtext extension for Eclipse, which allows for the grammar to create a DSL. Grammatically augmented ontology contains 1200 words in various subject areas. These words are listed in the form of synsets and described the grammatical constructions in which these words are used. Synset also includes grammatical constructions of the sign language. Thus, we can obtain monosemantic translation of a certain word combination. Therefore, grammatically augmented ontology can improve the translation result and can be used in machine translation systems.

