

ПРИНЦИПЫ МОДЕЛИРОВАНИЯ ЛОГИКО-СЕМАНТИЧЕСКОГО АНАЛИЗА ЕСТЕСТВЕННОГО ЯЗЫКА

Ивашенко В. П.

Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: ivashenko@bsuir.by

Рассматриваются принципы и подход для построения модели логико-семантического анализа естественного языка. Выделены уровни анализа естественного языка и выделены этапы логико-диалектического анализа. Указаны основные упорядочивающие отношения для разных уровней логико-семантического анализа, организующих компоненты лингвистической базы знаний в рамках смыслового пространства.

ВВЕДЕНИЕ

Предложено множество подходов к анализу естественного языка [1]. В настоящее время развиваются и зачастую успешно используются подходы, использующие большие языковые модели [2]. Однако их нейросетевая реализация не позволяет в полной мере избежать ошибок и обеспечить их коррекцию путём их объяснения.

Более ранним подходами, в рамках логико-символьного подхода обработки знаний, являются подходы на основе сочетания лексико-грамматического, морфологического, синтаксического и семантического анализа [3]. Однако при этом количество этапов анализа не фиксировано и может варьироваться от подхода к подходу [3].

Известны подходы на основе модели «Смысл-текст» Ю. Мельчука [4]. Формальная модель семантики Р. Монтегю [5], Л. Кальмара [6].

Для анализа формальных языков может быть использован на стадиях лексико-грамматического и синтаксического разборов алгоритм Эрли [7] вместе с наборами правил.

I. ЭТАПЫ И УРОВНИ ЛОГИКО-СЕМАНТИЧЕСКОГО АНАЛИЗА ЕСТЕСТВЕННОГО ЯЗЫКА

Предлагается подход, который основывается на следующих принципах:

- использование унифицированного семантического представления для лингвистической информации [8];
- использование словарей синонимических наборов правил (продукций), $\text{supprodset}'\text{ов}$;
- использование в качестве основных этапов этапов лексического (лексико-грамматического) и семантического анализа;
- использование в качестве вспомогательных этапов этапов морфологического и синтаксического анализа;
- использование трёх уровней анализа в рамках семантического анализа: сущностно-систематического,

реляционно-феноменологического, логико-диалектического.

Этап лексического анализа сводится к поиску лексем в словаре и соответствующего набора правил.

Семантический анализ сводится к применению правил из найдённых наборов, при этом правила включают обработку онтологических структур на трёх уровнях. Кроме этого, правила в наборах правил ориентированы на работу с реляционными множествами [8], таким образом естественным образом поддерживается динамический характер процесса анализа естественного языка [9].

На уровне сущностно-систематического анализа учитываются и различаются такие языковые категории как число, род, имена собственные и нарицательные.

На уровне реляционно-феноменологического анализа учитываются такие категории как модальность, переходность и непереходность, пространственно-временные отношения [8,9], обстоятельства места, времени, причин и целей.

На уровне логико-диалектического анализа различаются утверждения, повествования, вопросы, императивы, при этом учитываются союзы, выражающие логические взаимосвязи, квантификаторы, обстоятельства условий и формируются гипотезы о предметных связях, высказываниях и интенциях.

Морфологический и синтаксический анализ инициируются, когда на уровне лексического анализа не обнаруживается лексема или на уровне логико-диалектического анализа не формулируется ни одна гипотеза.

На основных этапах и уровнях анализа учитывается порядок и топологические свойства онтологических структур языковой базы знаний [8].

На этапе лексического анализа [3] учитывается лексикографический порядок лексем.

На этапе семантического анализа на уровне сущностно-систематического анализа учитывается порядок, связанный с конденсацией транзитивного замыкания отношений `member-of`, `subclass-of`.

На уровне реляционно-феноменологического анализа учитывается порядок, связанный с конденсацией замыкания отношений становления и раньше-позже.

На уровне логико-диалектического анализа учитывается порядок, связанный с конденсацией транзитивного замыканий отношений под-формул, логического следования, выводимости, сообщений-реплик.

Перечисленные отношения порядка обеспечивают шкалирование лингвистических знаний и задаёт ключевые структуры в смысловом пространстве, необходимые для его организации, управления занятиями [8] и осуществления логического вывода в его рамках.

Для реализации правил используются специальная онтология, которая содержит понятия текущего корневого и атомарного высказываний (выражений), текущего феномена, сущности и др. Для описания правил может быть использован универсальный язык описания, к которому может быть применён алгоритм наподобие алгоритма Эрли. В этом языке грамматические правила дополняются пост-заклЮчениями для преобразования состояния лингвистической базы знаний, в соответствие со схемой «условие : заключение > пост-заклЮчение».

ЗАКЛЮЧЕНИЕ

Рассмотрены принципы и подход для построения модели логико-семантического анализа естественного языка. Выделены уровни анализа естественного языка и выделены этапы логико-диалектического анализа. Указаны основные упорядочивающие отношения для разных уровней логико-семантического анализа, организующих компоненты лингвистической базы знаний в рамках смыслового пространства.

Компоненты лексического анализатора прошли апробацию в интеграционной системе [10]. Предложенный подход и принципы ориентированы на создание интеллектуальных вопросно-ответных и обучающих систем [10, 11].

1. Большакова, Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. / Е. И. Большакова [и др.] . – М.: МИЭМ, 2011. – 272 с.
2. Jurafsky, D. Martin, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd Edition draft [Electronic resource]. – Mode of access: https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf. – Дата доступа: 28.10.2024.
3. Крапивин, Ю. Б. Естественнo-языковой интерфейс интеллектуальных систем. Лабораторный практикум : пособие / Ю. Б. Крапивин. – Минск : БГУИР, 2023. – 64 с.
4. Mel'cuk, I. Language: from Meaning to Text. Ed. by D. Beck. Moscow & Boston. 2016.
5. Montague, R. English as a Formal Language. In: Bruno Visentini (ed.): Linguaggi nella società e nella tecnica. – Mailand, 1970. – 189–223 pp.
6. Маркус, С. Теоретико-множественные модели языков // Перевод с англ. М.В. Арапова; Под ред. Ю. А. Шрейдера. – Москва, 1970. – 332 с.
7. Earley, J. An efficient context-free parsing algorithm. Communications of the Association for Computing Machinery. – 1970. –13, no. 2. – 94–102 pp.
8. Ivashenko, V. Semantic space integration of logical knowledge representation and knowledge processing models / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS); ed.: V. V. Golenkov [et al.]. – Minsk, 2023. – Iss. 7. – С. 95–114.
9. Ivashenko, V. Semantic Logging of Repeating Events in a Forward Branching Time Model / Ivashenko V., Zotov N., Orlov M. // Pattern Recognition and Information Processing (PRIP'2021) = Распознавание образов и обработка информации (2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus / United Institute of Informatics Problems of the National Academy of Sciences of Belarus. – Minsk, 2021. – 149–152 pp.
10. Ivashenko, V. Application of an integration platform for ontological model-based problem solving using an unified semantic knowledge representation / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS); ed.: V. V. Golenkov [et al.]. – Minsk, 2021. – Iss. 5. – С. 179–186.
11. Голенков В.В. Интеллектуальные обучающие системы и виртуальные учебные организации: Монография / В.В. Голенков, В.Б. Тарасов, О.Е. Елисеева и др.; под ред. В.В. Голекова, В.Б. Тарасова – Мн.:БГУИР, 2001. – 486 с.