

Б.Н. СУДАКОВ, канд. техн. наук, проф., НТУ "ХПИ", Харьков,
Ю.М. МИХАЙЛОВА, магистр, НТУ "ХПИ", Харьков

ЛИНГВИСТИЧЕСКИЙ ПРЕПРОЦЕССОР ОБРАБОТКИ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ ДЛЯ ЭКСПЕРТНЫХ СИСТЕМ

Анализ особенностей природных и формальных языков показал, что с точки зрения решаемых задач экспертными системами целесообразно иметь язык взаимодействия пользователя с системой, состоящий из двух уровней. Один уровень – ограниченный естественный язык. Другой – формальный язык, являющийся внутренним языком системы, на котором осуществляется логическая обработка запросов пользователей. Предложена структура лингвистического процессора, осуществляющего преобразование текстов языка взаимодействия с одного уровня языка на другой. Ил.: 2. Библиогр.: 10 назв.

Ключевые слова: лингвистический процессор, обработка естественно-языковых текстов, экспертные системы.

Постановка проблемы и анализ литературы. Одним из элементов экспертной системы для принятия решений является подсистема взаимодействия с пользователем. Основу взаимодействия составляют языковые средства, поскольку только с помощью языка (формального или естественного) можно достичь определенных целей в процессе общения коммуникантов.

Существующие языки взаимодействия и языки представления знаний в интеллектуальных системах, как правило, имеют узкоспециализированную направленность. Они базируются на известных логико-математических моделях (исчислении предикатов, системах продукций, фреймах, семантических сетях) и не позволяют учитывать такие особенности предметной области как неполнота и противоречивость знаний и данных, динамичность и неопределенность объектов предметной области.

В [1, 2] изложены основные принципы проектирования экспертных систем. В [3 – 7] освещены теоретические концепции и практические методы обработки естественно-языковых текстов на всех уровнях лингвистического анализа. В [8 – 10] приведены базовые понятия компьютерного анализа текстов. Проведенный обзор литературы показал, что отсутствует единый подход к формализации знаний о предметной области. Старение моделей, заложенных в систему, требует постоянного их совершенствования. Это приводит к необходимости проведения постоянных доработок программного обеспечения, что является весьма проблематичным в системах, алгоритмы которых построены по жесткому принципу.

Цель статьи – повышение эффективности процесса взаимодействия пользователей с экспертной системой на ограниченном естественном языке на основе разработки новых языковых средств и лингвистического процессора.

Модель внутреннего языка экспертной системы медицинской диагностики. В основу построения внутреннего языка положена категорная модель знаний, являющаяся более универсальным средством, по сравнению с известными моделями, для формализации знаний о процессах реального мира.

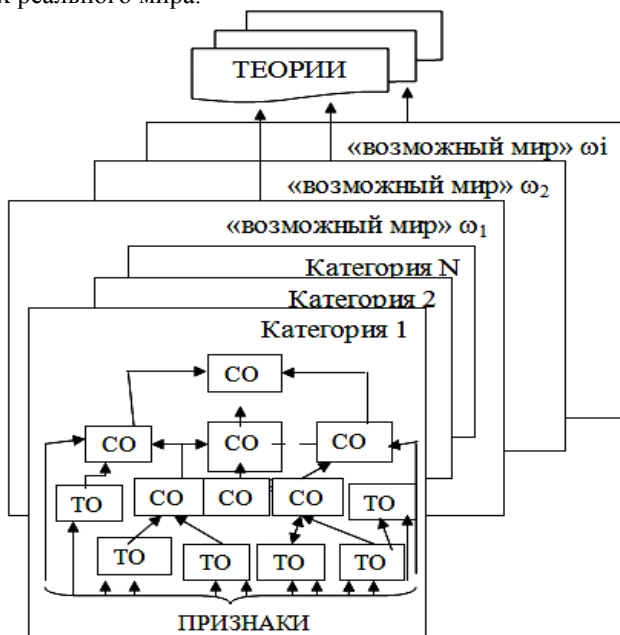


Рис. 1. Общая структура категорной модели знаний

Категория используется для описания концептов (понятий), которые ставятся в соответствие реальным (абстрактным) классам понятий или описывают частные примеры объектов реальности. Категория отражает определенный аспект знаний о концептах (состав, структура, состояние, пространство, время, параметры и т.д.). Состав категории меняется в зависимости от "возможных миров", в которых может использоваться эта категория. Категорная модель знаний носит двойственный характер. С одной стороны она описывает структуру концептов предметной области. С другой – определяет динамику взаимодействия (теорию) объектов

предметной области. Теория строится над каждым объектом и связывает по присутствию входящие и исходящие из него морфизмы. Это позволяет объединять теории объектов между собой и строить иерархические теории реальности. Использование для построения теорий операций топологической булевой алгебры позволяет описывать реальность в условиях неопределенности.

Анализ существующих моделей языков показал, что для описания внутреннего языка системы наиболее целесообразно использовать аппарат формальных грамматик. Дополнение КС-грамматики семантическими продукционными правилами позволило обеспечить взаимнооднозначное соответствие между элементами категорной модели знаний и языком представления знаний. Кроме того, описание языка с помощью правил-продукций (декларативных знаний) позволяет использовать их как для распознавания (порождения) терминальных цепочек языка, так и для синтеза меню, пополнения словаря, что обеспечивает дополнительные возможности для организации работы непрограммируемых пользователей. Предложенные прагматические функции позволяют осуществить манипулирование знаниями и формировать многоуровневые иерархические теории реальности.

Модель ограниченного естественного языка. Для перехода от текста ограниченного естественного языка к внутреннему представлению необходимо использовать модель языка, основанную на многоуровневом описании. Тексты ограниченного естественного языка (ОЕЯ) представляют поверхностный уровень языка. Поверхностно-синтаксический уровень языка задается системой синтаксических групп (ССГ), которая является обобщением деревьев синтаксического подчинения и системы составляющих. Глубинный уровень – структурой семантико-синтаксических групп (СССГ), что соответствует размеченной системе синтаксических групп (РССГ). При этом связи между словоформами (синтаксическими группами) устанавливаются не только с использованием знаний о синтаксисе языка, но и знаний о проблемной области и семантике внутреннего языка системы. Уровень смыслов представлен выражениями внутреннего языка системы (С-языка).

Для реализации языка в соответствии с моделью необходимо разработать следующие алгоритмы:

- построения системы синтаксических групп;
- построения структуры семантико-синтаксических групп;
- преобразования данной структуры в выражения внутреннего языка системы.

Совокупность процедур, реализующих указанные алгоритмы, будем называть лингвистическим процессором.

Структура лингвистического процессора. В соответствии с выбранной моделью языка рассмотрим общую структуру ЛП, изображенную на рис. 2.

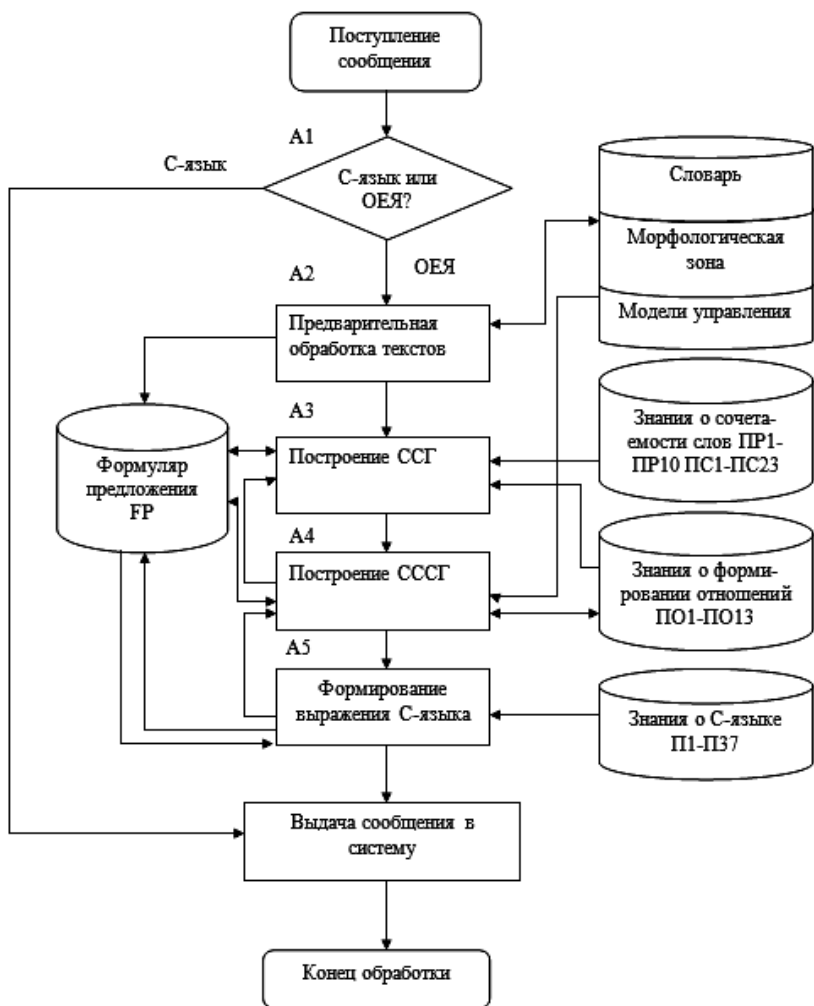


Рис. 2. Общая структура ЛП

В его составе можно выделить элементы, связанные с реализацией алгоритмов каждого из отмеченных этапов обработки текстов ограниченного естественного языка, базу знаний, включающую знания о предметной области и знания о языке.

Выводы. Выполнены исследования по разработке естественно-языкового интерфейса интеллектуальной экспертной системы на базе лингвистического процессора, обеспечивающего обработку естественно-языковых текстов.

Список литературы: 1. *Джексон П.* Введение в экспертные системы / *П. Джексон.* – М.: Основа, 2001. – 624 с. 2. *Поспелова Д.А.* Искусственный интеллект / *Д.А. Поспелова.* – М.: Наука, 1990. – 246 с. 3. *Карпова Г.Д.* Компьютерный синтаксический анализ: описание моделей и направлений разработок / *Г.Д. Карпова, Ю.К. Пирогова, Т.Ю. Кобзарева, Е.В. Микаэлян.* – М.: ВИНТИ, 1991. – 304 с. 4. *Романенко В.Н.* Сетевой информационный поиск / *В.Н. Романенко.* – М.: Профессия, 2005. – 290 с. 5. *Попов Э.В.* Общение с ЭВМ на естественном языке / *Э.В. Попов.* – К.: Наука, 1992. – 254 с. 6. *Анно Е.И.* Типологии алгоритмов синтаксического анализа (для формальных моделей естественного языка) / *Е.И. Анно.* – СПб.: Питер, 1989. – 152 с. 7. *Омар А.Х. Авадала* Подход к синтезу естественно-языковых сообщений по формализованному представлению базы знаний: автореф. дис. на соискание ученой степени канд. техн. наук / *Омар А.Х. Авадала.* – Х., 2001. – 20 с. 8. *Фостер Дж.* Автоматический синтаксический анализ / *Дж. Фостер.* – М.: Мир, 1975. – 71 с. 9. *Русских И.В.* Инкрементальный синтаксический анализ в средах разработки и текстовых редакторах // Нижегородский университет. – 2007. – С. 277. 10. Автоматическая обработка текста [Электронный ресурс] / *Н.Н. Леонтьев.* – 2003. – С. 5. – Режим доступа к статье: <http://www.aot.ru/technology/html>.

Статью представил д.т.н., проф. НТУ "ХПИ" Серков А.А.

УДК 004.031.42

Лінгвістичний препроцесор обробки природно-мовних текстів для експертних систем / Судаков Б.М., Михайлова Ю.М. // Вісник НТУ "ХПИ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПИ". – 2012. – № 38. – С. 178 – 183.

Аналіз особливостей природних і формальних мов показав, що з точки зору розв'язуваних завдань в експертних системах доцільно мати мову взаємодії користувача з системою, що складається з двох рівнів. Один рівень – обмежена природна мова. Інший – формальна мова, яка є внутрішньою мовою системи та на якій здійснюється логічна обробка запитів користувачів. Запропоновано структуру лінгвістичного процесора, що здійснює перетворення текстів мови взаємодії з одного рівня мови на іншу. Іл.: 2. Бібліогр.: 10 назв.

Ключові слова: лінгвістичний процесор, обробка природно-язикових текстів, експертні системи.

UDC 004.031.42

The linguistic preprocessor of natural language texts for expert systems processing / Sudakov B.N., Mykhailova Y.M. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modeling. – Kharkov: NTU "KhPI". – 2012. – №. 38. – P. 178 – 183.

Analysis of the characteristics of natural and formal languages has shown that in terms of tasks of expert systems it's useful to have a language for interaction of user and system, which consists of two levels. One level – limited natural language. Another – a formal language, which is an internal language of the system and which is used for logical processing of user requests. The structure of the language processor which carries out the transformation language texts interacts with one level of language to another was proposed. Figs.: 2. Refs.: 10 titles.

Keywords: linguistic processor, natural language texts, expert systems.

Поступила в редакцию 10.05.2012