

ЛИНГВИСТИЧЕСКИЙ ПРОЦЕССОР В СИСТЕМЕ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЯ С ЭКСПЕРТНОЙ СИСТЕМОЙ

проф. Б.Н. Судаков, Омар А.Х. Авадала

Предлагается ограниченный естественный язык для взаимодействия пользователя с экспертной системой и принципы обработки текстов этого языка. Обосновывается структура лингвистического процессора, реализующего анализ и синтез естественно-языковых конструкций. Дается описание основных этапов анализа и синтеза текстов.

Введение. Для поддержки принятия решений в различных сферах человеческой деятельности возникает необходимость разработки экспертных систем (ЭС). Естественно считать, что взаимодействие лиц, принимающих решения (ЛПР) – пользователей с ЭС, должно быть организовано на привычном для ЛПР профессиональном языке, близком к естественному [1]. Будем считать, что знания о предметной области представлены в ЕС в виде предложений некоторого формального внутреннего языка (ВЯ) (языка внутреннего представления) [2]. Полагая, что экспертная система «понимает» предложения ВЯ, представим взаимодействие пользователя с системой как процесс преобразования естественно-языковых профессиональных текстов, грамматика и семантика которых ограничена рамками решаемой задачи, на внутренний язык ЭС и обратно [3, 4]. При такой постановке задачи для реализации взаимодействия необходимо *построить ограниченный естественный язык общения и разработать лингвистический процессор, обеспечивающий обработку текстов этого языка.*

Ограниченный естественный язык для взаимодействия пользователя с ЭС. Язык, используемый пользователями, отличается от обычного разговорного языка. Он обогащен дополнительной терминологией и освобожден от малоупотребительных конструкций. Вместе с тем, этот язык недостаточно точен, а в ряде случаев многозначен. Вследствие этого он подвергается нормализации. Нормализованным считается язык, созданный на базе естественного и ограниченный с точки зрения грамматики и семантики. Ограничения определяются рамками предметной области и множеством решаемых в ней задач. На основе нормализованного языка строится модель, позволяющая формально описывать множе-

ство входных текстов. Для формализации поверхностной структуры текста необходимо установить связи (отношения синтаксического подчинения) между словоформами, входящими в его состав. Изображение синтаксических связей с помощью отношений синтаксического подчинения в основном совпадает с принятыми представлениями о грамматической структуре предложения (дополнения зависят от сказуемого, определение от определяемого и т.д.). Вследствие этого связи между словоформами любых текстов русского языка могут быть представлены в виде деревьев синтаксического подчинения. В качестве примера на рис. 1 показано дерево синтаксического подчинения, построенное на словоформах фразы «Какие мероприятия повседневной деятельности проводит директор предприятия?»

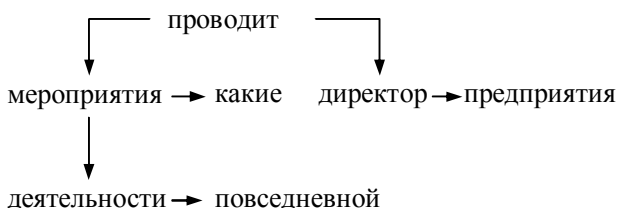


Рис. 1. Пример построения дерева синтаксического подчинения

Множество входных текстов для оперативности взаимодействия целесообразно ограничить простыми предложениями русского языка повествовательного, вопросительного и директивного типов, структура которых в нотации Бекуса-Наура имеет вид:

1. <Словоформа> ::= <С> | <П> | <ОС> | <Г> | <Н> | <А> | <МС> | <К> | <ПР> | <НЕ> | <ВС> | <'> | <.> | <?> | <!> | <Число> | <Ч> | <М>, <Число> ::= <Последовательность цифр> <Цифра> ::= <0\1\2\3\4\5\6\7\8\9>, <Ч> ::= <час>, <М> ::= <Минута>
2. <ГГ> ::= <Г> | <НеГ> | <МСГ> | <Не МС Г>
3. <Именная группа> ::= <С> | <П С> | <Число П С> | < Число С> | <К П С> | <К С> | <Пр К П С> | < Пр К С> | < Пр число П С> | <Пр число С> | <Пр П С> | <Пр С> | <Не Пр К П С> | <Не К П С> | <Не Пр К С> | <Не П С> | <Не К С> | < Не Пр П С> | <ВС П С> | <ВС С> | <Пр ВС П С> | < ВС Не П С> | <Пр ВС Не П С> | <Пр ВС С>, <ВС> ::= <Какой x,y,z> <сколько z>, <Пр ВС С>

где С – существительное; П – прилагательное; ОС – отглагольное существительное; Г – глагол; Н – наречие; А – аббревиатура; МС – модальное слово; Пр – предлог; НЕ – отрицательная частица; ВС – вопросительное

слово; ГГ – группа глагола; х – род; у – число; z – падеж.

Связи между представленными фрагментами входного текста устанавливаются естественным образом в соответствии с грамматикой русского языка. В соответствии с изложенным, для построения дерева синтаксического подчинения на словоформах S_i входного текста создается база знаний о языке, включающая: множество синтаксической Sin_i , семантической Sem_i информации о S_i ; информацию о моделях управления (МУ) S_i :

$$MU_k(A_1(Sin_1, Sem_1), \dots, A_l(Sin_l, Sem_l)),$$

где A_l – актанты, характеризующие ситуацию ПО и содержащие множество семантической и синтаксической информации.

Для перехода к внутреннему представлению используются следующие правила вывода.

1. Устанавливаются связи для S_i с МУ:

$$\{H; S_i (A_l (Sem_l, Sin_l); S_j; D) \rightarrow S_i R^{ij} S_j,$$

где S_i – словоформа, имеющая МУ; S_j – словоформа, подходящая на роль актанта; D – информация о расположении этих словоформ в анализируемой фразе; H – условия применимости правила (условие согласованности Sem_i и Sem_j , Sin_i и Sin_j); R^{ij} – тип связи между словоформами S_i и S_j .

2. Для поиска связи между словоформами, не имеющими МУ, используются правила

$$\{H; S_i; S_j; D\} \rightarrow S_i R^{ij} S_j.$$

3. Для формирования типа запроса (уточнить, определить количество, перечислить, изменить содержание базы знаний и др.), связанного с множеством решаемых задач, используются правила

$$\{H; Sem_i; Sem_j; R^{ij}; T\} \rightarrow F; Sem_{i(j)},$$

где T – тип вопроса, F – тип запроса.

В результате применения правил формируется внутреннее представление:

<Внутреннее представление> ::= <Описание 1> [<Описание 2>];

<Описание 1> ::= {<Понятие><Имя>}, {<Тип запроса>};

<Описание 2> ::= [Атрибут: {<Имя>, <Тип запроса>}].

Квадратные скобки указывают на возможность многократного повторения того, что в них заключено. Пример разбора предложения «Какие мероприятия повседневной деятельности проводит директор предприятия?» приведен на рис. 2.

Модель управления «Проводит»

Информация	Актанты	
	A ₁ (Кто?)	A ₂ (Что?)
Синтаксическая	S им.	S вин.
Семантическая	Руководитель	Понятия, выражающие действия

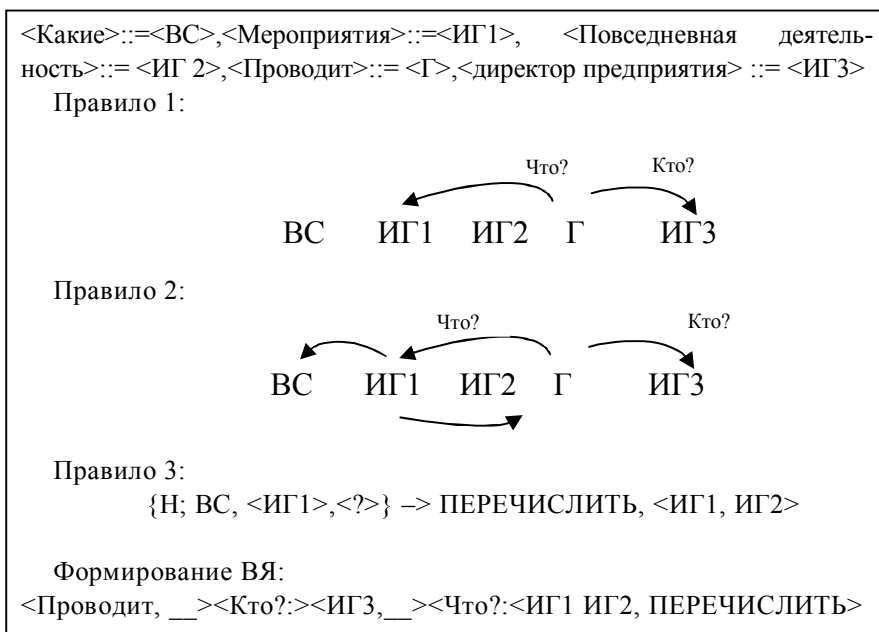


Рис. 2. Пример разбора текста

Переход от внутреннего представления к поверхностному уровню осуществляется путем синтеза естественно-языковых конструкций с применением аналогичных правил.

Лингвистический процессор. С использованием рассмотренного подхода разработан лингвистический процессор (ЛП), обеспечивающий трансляцию входных текстов, лексика и семантика которых ограничена рамками предметной области. Структура лингвистического процессора ЛП представлена на рис. 3.

В его составе можно выделить процедуры анализа и синтеза текстов ограниченного естественного языка (ОЕЯ), а также базу знаний о языке (словарь), включающую совокупность морфологической, синтаксической

и семантической информации о словоформах ограниченного естественного языка.

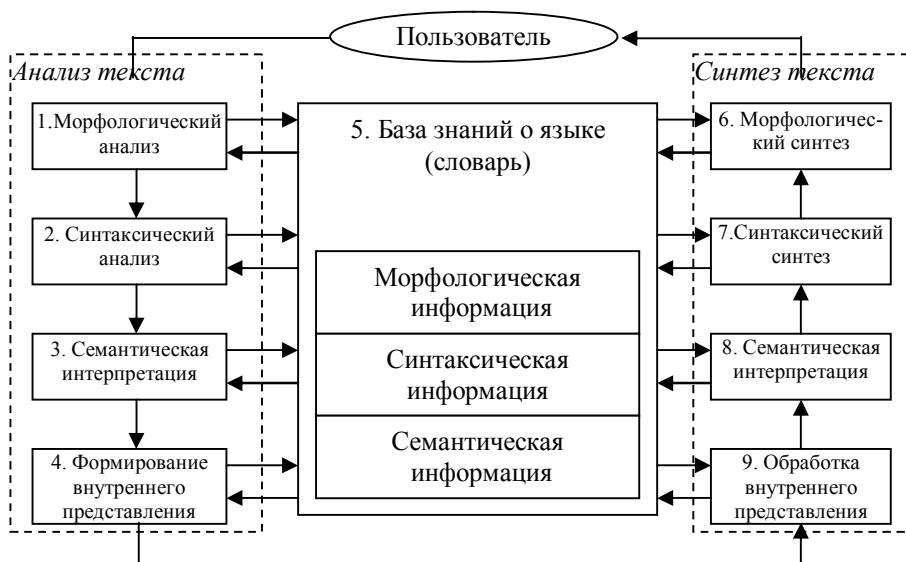


Рис. 3. Структура лингвистического процессора

На этапе морфологического анализа [3] осуществляется отождествление словоформ входного текста с информацией словаря и приписывание им характеризующей их синтаксической и семантической информации. В зависимости от принципов построения словаря (словоформы либо основы слов) эта процедура осуществляется либо непосредственно (словарь словоформ), либо более сложным путем после предварительного выделения основы слова и окончания. Информационные массивы, соответствующие каждой входной словоформе, приводятся к виду, удобному для дальнейшей обработки.

В соответствии с операциями блоков 2 – 4 производится установление направленных связей между словами (построение дерева синтаксического подчинения) подобно тому, как это показано на рис. 1. Для этих целей используются как данные о роли отдельных слов в предложении, информация моделей управления, так и другие сведения о синтаксической структуре фразы. Завершается анализ текста формированием внутреннего представления, которое является «понятным» для машины и которое в зависимости от потребностей пользователя после соответствующей обработки выдает ответ на его запрос, корректирует базу знаний и т.п.

Для получения реакции экспертной системы в естественной для пользователя форме производится синтез текста (блоки 6 – 9) с использованием процедур обратной трансляции, аналогичных описанным выше.

Следует заметить, что в процессе анализа и синтеза текстов ОЕЯ предусматриваются меры по исключению неоднозначности слов и отдельных выражений языка взаимодействия, поэтому эти процессы реализуются многократно путем анализа и отбрасывания неверных структур до тех пор, пока не будет построено синтаксически и семантически правильное выражение.

Выводы. 1. Процесс взаимодействия пользователя с экспертной системой представлен как преобразование естественно-языковых профессиональных текстов на внутренний язык ЭС и обратно. Разработан ограниченный естественный язык для взаимодействия пользователя с экспертной системой. Для формализации поверхностной структуры текста предложен аппарат деревьев синтаксического подчинения.

2. Обоснована и приведена структура лингвистического процессора, реализующего анализ и синтез естественно-языковых конструкций. Дано описание основных этапов обработки текста.

ЛИТЕРАТУРА

1. Ярушек В.Е., Прохоров В.П. *Применение логики присутствия в интеллектуальных информационных управляющих системах // II всесоюзная конференция по «ИИ-90».* – Минск. – 1990. – С. 99 – 105.
2. Гаврилова Т.А. *Базы знаний интеллектуальных систем.* – С.-Пб.: Питер, 2000. – 384 с.
3. Шлепаков Л.Н. *Зональная морфологическая обработка текстов на естественных языках // Кибернетика и системный анализ.* – 2001. – Вып. 3. – С. 28 – 34.
4. Бондарев В.Н., Аде Ф.Г. *Искусственный интеллект.* – Севастополь: изд-во Сев. НТУ, 2002. – 615 с.

Поступила 20.02.2003

СУДАКОВ Борис Николаевич, канд. техн. наук, профессор, профессор кафедры НТУ «ХПИ». В 1970 году окончил ВИРТА ПВО. Область научных интересов – системы взаимодействия «человек – машина».

ОМАР А.Х. Авадала, аспирант НТУ «ХПИ». В 1996 году окончил НТУ «ХПИ». Область научных интересов – обработка языковой информации.