

СИНТЕЗ ЕСТЕСТВЕННО - ЯЗЫКОВЫХ КОНСТРУКЦИЙ НА ЭТАПЕ СЕМАНТИЧЕСКОЙ ИНТЕРПРЕТАЦИИ

к.т.н. Б.Н. Судаков, к.т.н. И.В. Замаруева, к.т.н. В.Н. Шемаев
(представил д.т.н. профессор А.А. Рось)

В статье раскрывается актуальность и необходимость синтеза естественно - языковых (ЕЯ) конструкций при общении пользователя с ЭВМ, взаимосвязь этапов синтеза. Подробно рассмотрен этап семантической интерпретации ЕЯ выражений.

Широкое применение экспертных систем (ЭС), использующих формализованное представление знаний о предметной области, тормозится из-за отсутствия ЕЯ интерфейса, что затрудняет работу с системой. Это обусловлено, в первую очередь, трудностями, связанными с анализом и синтезом ЕЯ конструкций.

Если относительно задач анализа ЕЯ сообщений имеется достаточно большое число публикаций [1-2], освещающих различные подходы к формализации ЕЯ конструкций, то задачи синтеза в плане системных исследований остаются вне поля зрения разработчиков. Анализ работ в области разработки прикладных ЭС показал, что синтез реализуется, как правило, через лингвистический синтез (т.е. производится выбор нужной синтаксической конструкции по заданному шаблону и грамматическое согласование внутри нее) [2]. Такой подход применяется, например, в системах машинного перевода: синтез заменяется "семантической компонентой", которая априорно задается каждой вершине дерева (сети) [3]. Такой подход эффективен в ЭС реального времени. Однако в ряде систем, например, планирования, прогнозирования и т.д., данный подход не позволяет решить ряд важных задач, в частности, выдачи пользователю пояснительной компоненты о необходимых условиях достижения целевой установки на понятном ему языке, проверки фрагментов базы знаний на семантическую корректность при ее пополнении, изменении и т.д.

В то же время знание - ориентированные технологии, на которых основаны ЭС, позволяют решить эти задачи. Для решения этих приклад-

© к.т.н. Б.Н. Судаков, к.т.н. И.В. Замаруева, к.т.н. В.Н. Шемаев, 1998

ных задач необходимо разработать методы синтеза ЕЯ сообщений по формализованному фрагменту базы знаний.

Предлагаемая модель синтеза ЕЯ конструкций по логическому представлению базы знаний представляет собой концептуальное единство трех моделей:

- модель синтеза логической структуры сообщения;
- модель синтеза понятийной структуры сообщения;
- модель синтеза грамматической структуры сообщения.

На основе модели логической структуры сообщения решается задача выбора необходимого фрагмента (или фрагментов) знаний о предметной области (ПО). Решение данной задачи актуально при синтезе пояснительной компоненты. На основе модели синтеза понятийной структуры сообщения решается задача семантической интерпретации заданного (или выбранного на предыдущем этапе) фрагмента базы знаний.

На основе модели синтеза грамматической структуры сообщения решается задача построения грамматически и синтаксически правильного сообщения (в соответствии с лингвистическими правилами выходного языка) по данным семантической интерпретации.

Рассмотрим этап семантической интерпретации на примере решения задачи проверки фрагментов базы знаний на семантическую корректность при ее пополнении и изменении. Синтез ЕЯ конструкций в данном случае начинается непосредственно с этого этапа, так как фрагмент базы знаний там уже задан в этом случае. Задача синтеза на этапе семантической интерпретации включает решение следующих задач:

- 1) распознавание конструкций внутреннего языка представления фрагмента базы знаний;
- 2) интерпретация элементов конструкций (понятий и отношений) внутреннего языка средствами естественного языка;
- 3) построение (синтез) понятийной структуры выходного сообщения.

В качестве примера рассмотрим запись на внутреннем языке.

Пример 1.

ЗАП < Разгрузка, объект - процесс;

(Средство Разгрузки : порталный кран;

Объект Разгрузки : баржа;

Предмет Разгрузки : цемент;

Место Разгрузки : причал № 2)>.

Эта запись соответствует ЕЯ фразе: Портальный кран разгружает баржу с цементом на причале № 2.

В качестве исходных данных алгоритма распознавания будем рассматривать отдельные знаки внутреннего языка представления.

Знак := символ / ____ .

Символ := буква / цифра / специальный символ.

Буква := Латинская / Русская.

Латинская := Большая буква (Бб) / Малая буква (Мб).

Латинская Большая буква (ЛБб) := A / B / C / D

Латинская малая буква (Лмб) := a / b / c / d

Русская := Большая буква (Бб) / Малая буква (Мб).

Русская Большая буква (РБб) := A / Б / В

Русская малая буква (Рмб) := a / б / в

Число := цифра / цифра .

Цифра := 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 0.

Специальный символ := ; / , / : / . / - /) / (/ < / > .

Выходные данные алгоритма распознавания - информация о семантических классах объектов. Выходные данные: функция, формула. Функция - это команда, которая определяет, что надо сделать с данной формулой записи. Формула - формализованное описание (или описание на внутреннем языке того, что известно об объекте). Объект объединяет подобъекты.

*Функция := Рбб {Рбб} V Рбб {Рмб} Рбб {Рмб} V
V Рбб {Рмб} Рбб {Рбб} V Рбб {Рмб}.*

*Формула := < начальная строка;
(вторая строка; ...; строка ; ..., конечная строка) >.*

Начальная строка := <Роль : Имя Объекта (ИО), Тип Объекта (ТО)

*Код Области Знаний (КОЗ); V Роль : ИО, ТО; V Роль, КОЗ; V
V Роль; V ИО, ТО; V Роль: ИО, КОЗ.*

Вторая строка := (Роль : Имя Подобъекта (ИПО),

Служебная Информация (СИ); V Роль, СИ; V Роль: ИПО;.

Строка := Роль : ИПО, СИ; V Роль, СИ; V Роль : ИПО;.

Конечная строка := Роль: ИПО, СИ; V роль, СИ; V Роль : ИПО)>.

*ИО = ИПО := РБб {РБб} V РБб {РМб} V число Рмб {Рмб} V
V Рмб {Рмб} Лбб {ЛБб} число.*

ТО := объект – Рмб {Рмб}.

Роль: = Имя Роли (ИР) ИО V ИРИПО.

ИР := Рбб {Рмб} V Рбб.

КОЗ и СИ являются дополнительной информацией и в данной статье не рассматриваются. В соответствии с алгоритмом, на основе идентификации исходных данных и правил их объединения, производится распознавание имен объекта, подобъектов и приписывание им соответствующего семантического класса. В результате на выходе алгоритма распознавания, получаем выражение:

Разгрузка (ИО) объект - процесс (ТО);

Средство Разгрузки (ИРИО): портальный кран (ИПО);

Объект Разгрузки (ИРИО): баржа (ИПО);

Предмет Разгрузки (ИРИО): цемент (ИПО);

Место Разгрузки (ИРИО): причал № 2 (ИПО).

После решения задачи распознавания конструкций внутреннего языка решаются задачи семантической интерпретации распознанных классов и формирование понятийной структуры фрагмента базы знаний. В качестве исходных данных будем рассматривать объекты, подобъекты и их семантические классы. Подобъекты входят в структуру теории объекта.

ИО = ИПО - последовательность символов, расположенных между специальными символами: $_ / - / ; / : /$. Для нашего примера объек-

тами являются: разгрузка, подобъектами: порталный кран, баржа, цемент, причал № 2.

После распознавания и приписывания каждому понятию и отношению соответствующего класса решается задача семантической интерпретации. Для решения этой задачи используются априори заданные таблицы языковой интерпретации всех классов объектов и подобъектов, присутствующих в заданном внутреннем языке.

Фрагменты записей для интерпретации приведенного примера заданы в табл.1 и табл.2. Таблица 1 служит для интерпретации понятий, таблица 2 - для интерпретации отношений.

Таблица 1 - Семантическая интерпретация понятий

№	Тип объекта	Семантическая интерпретация
1.	Объект-процесс	производить ИО ___ Винительный падеж (ВП)

2.	Объект-класс	включить ИО ___ ВП

К	Объект-состояние	находиться ИО ___ Предложный падеж (П.П.)

В соответствии с алгоритмом на основе идентификации входных данных и информации табл. 1 и табл. 2, правил их преобразования, строится их понятийная структура выходного текстового сообщения. В результате на выходе алгоритма семантического синтеза, получаем следующее выражение:

*портальный кран * ИПО (ИП) __ производить __ Разгрузка **
** ИО (ВП) __ баржа * ИПО (РП) __ баржа **
** ИПО (ИП) __ Содержать __ цемент **
** ИПО (ВП) __ Место (ИП) __ Разгрузка **
** ИО (РП) __ причал __ № 2 * ИПО (ИП).*

Выходные данные этапа семантической интерпретации являются, в свою очередь, входными данными для модели синтеза грамматической структуры сообщения. На последнем этапе решаются следующие задачи:

- формирование границ предложения в соответствии с синтаксическими правилами выходного языка;

- синтез словоформ на основе сформированной грамматической информации;
- синтез выходного сообщения.

Таблица 2 - Семантическая интерпретация отношений

№	Имя роли	Семантическая интерпретация
1.	Средство - являться средством	ИПО ___ именительный падеж (ИП)
2.	Объект - являться объектом	ИПО ___ родительный падеж (РП)
3.	Предмет - являться предметом	ИПО (объект) ИП ___ Содержать ___ ИПО ___ ВП

К.	Место - являться местом	Место (ИП) _ ИО_ РП _ ИПО_ ИП

Таким образом, подход к решению задачи синтеза ЕЯ сообщения по формализованному представлению базы знаний позволяет не только значительно повысить инженерно-психологические показатели системы, но и контролировать машинное "понимание" выполнения прикладной задачи на принципиально новом качественном уровне.

Предложенный подход синтеза ЕЯ конструкций в ЭС допускает как независимое решение задачи (например, в системах, где отсутствует анализ ЕЯ сообщений), так и разработку ЕЯ интерфейса на единой основе решения задач анализа и синтеза.

В последнем случае структура и содержание моделей синтеза определяется методами представления и обработки знаний, выбранными при анализе ЕЯ текстовых сообщений.

ЛИТЕРАТУРА

1. Попов Э.В. Общение с ЭВМ на естественном языке. - М.: Наука, 1982. - 375 с.
2. Пospelов Д.А. Логико - лингвистические модели в системах управления. - М.: Энергоатомиздат, 1981. - 385 с.
3. Организация взаимодействия человека с техническими средствами АСУ. Кн.2. Языковые средства диалога человека с ЭВМ. - М.: Высшая школа, 1990. - 235 с.