

## **НЕКОТОРЫЕ ВОПРОСЫ СОЗДАНИЯ ПРОГРАММЫ ПОСТРОЕНИЯ СЕМАНТИЧЕСКОЙ СЕТИ ПО ТЕКСТУ ЗАДАЧИ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ\***

*В.П. ЗУБКОВ*

**(Ивановский государственный энергетический университет)**

В текстильной промышленности есть много задач, постановка которых на естественном языке имеется, однако формализованное представление отсутствует в силу различных причин: отсутствие подходящего программного обеспечения, для которого необходимо представить в соответствующей форме сформулированную на естественном языке задачу; проблема вида формального представления для более эффективного решения; отсутствие формальной теории для данного класса задач и т.д.

С целью попытки устранения первых двух причин создается программная ин-

теллектуальная система автоматизированного программирования (ИнтСАПр), не являющаяся универсальной с точки зрения наполнения баз знаний и баз данных, но содержащая инвариантные программные компоненты, пригодные для различных приложений.

Аналогично тому как программные компоненты некоторой СУБД являются инвариантной частью, так и в ИнтСАПр имеются инвариантные компоненты более высокого уровня. ИнтСАПр содержит подсистему анализа текстов задач на естественном языке (АТнЕЯ), подсистему со-

---

\* Публикуется в порядке обсуждения.

ставления плана решения задачи (СПРЗ), подсистему порождения алгоритмов (СПА), подсистему порождения программ на целевом языке (СПП).

АТнЕЯ содержит такие компоненты, как программная компонента морфологического анализа текста на естественном языке (ПКМА); программная компонента синтаксического анализа текста на естественном языке (ПКСинА); программная компонента семантического анализа текста задачи на естественном языке (ПКСемА).

В статье рассматривается вопрос о способах перехода от синтаксических представлений с помощью морфологической информации и словарей к семантической сети первого уровня, под которой понимается разбиение текста на группы слов, соответствующие понятиям и действиям на уровне лексики и семантики словарей русского языка.

При решении задач конкретной предметной области в интеллектуальной системе автоматизированного программирования используются различные виды семантических сетей. Дуга семантической сети нагружена отношениями между объектами, процессами, явлениями, которые представляются вершинами.

Синтаксическая структура текста задачи на естественном языке содержит следующую информацию: [Метка части речи / Часть речи / Ссылка на морфологическую информацию, Ссылка на синтаксическую информацию, (Список ссылок на зависимые слова)].

Ссылка на морфологическую информацию для данной части речи имеет вид RMI#n, ссылка на синтаксическую информацию для данной части предложения имеет вид RSI#n, ссылка на n-ю часть речи в данном предложении или тексте имеет вид (n), ссылка на пунктуационную информацию для данного знака пунктуации имеет вид RPI#n.

В прямоугольных скобках выделена часть речи, которая помечается меткой MI (I – номер). Сама часть речи находится в косых скобках, далее следует ссылка на морфологическую информацию, а затем – на синтаксическую информацию. В круг-

лых скобках находится ссылка на зависимые слова либо в виде одной ссылки, либо в виде списка ссылок.

Структура семантической сети первого уровня для сформулированной задачи многое заимствует из синтаксической структуры. Появляются дуги, нагруженные отношениями. Если рассматривать семантическую сеть как граф, то он будет являться мультиграфом. При построении семантической сети второго уровня для задачи используются знания о предметной области, которые сами представлены в виде нескольких видов семантических сетей.

Семантическая структура первого уровня для данной задачи выглядит следующим образом: [Номер ветки / Часть предложения / Ссылка на семантическую информацию (Список отношений)]. В данной структуре в квадратных скобках находятся словосочетания, имеющие вначале метку вида SI (I – номер метки). Непосредственно само словосочетание находится в косых скобках, далее следует ссылка на семантическую информацию. В круглых скобках указывается дуга, которой соответствуют три элемента: начальная вершина, конечная вершина, ссылка на описание отношения. Дуг может быть несколько. Если в круглых скобках ничего нет, то это указывает на отсутствие дуги.

Возможен другой формат семантической сети первого уровня (ЗдчСемС1): (Номер дуги (Номер начальной вершины (Описание объекта, явления, процесса, соответствующего начальной вершине дуги)) (Номер отношения (Описание отношения)) (Номер конечной вершины (Описание объекта, явления, процесса, соответствующего конечной вершине дуги))).

В данном случае используется списочное представление, похожее на представление в функциональных языках программирования. В виде списка представляется дуга, которая содержит в виде списков начальный и конечный элементы дуги, помеченные метками вида PI (I – номер метки), и в виде списка наименование дуги, помеченное меткой RI (I – номер метки).

В семантической сети первого уровня выделяется словосочетания и образуется

семантическая сеть, учитывающая смыслы, изложенные в толковых словарях. Семантическая сеть второго уровня (ЗдчСемС2) привязывается к предметной области задачи, принимая во внимание классификационные (КлСемС), функциональные (ФСемС), понятийные (ПСемС), казуальные (КазСемС) семантические сети. При стыковке КлСемС, ФСемС, ПСемС, КазСемС и ЗдчСемС1 создается ЗдчСемС2.

Семантические сети в виде списков, структура которых такова: (Порядковый номер списка (Начальная вершина (Возможен неопределенный вложенный список дополнительной информации))(Конечная вершина (Возможен неопределенный вложенный список дополнительной информации))(Ссылка)Отношение[или связь]).

При сопоставлении классификационных семантических полей происходит выделение понятий и слов с помощью АСемС в каждой вершине и соответственно в каждой характеристике, если она присутствует, на основании морфологической, синтаксической и семантической информации, хранящихся для каждой вершины семантических сетей. Анализатор семантических сетей (АСемС) может вызываться из любого программного приложения с указанием семантической сети и режима работы.

После работы АСемС в ПСемС появляются неопределяемые понятия, которые определяются с помощью ИИн(Интеллектуальный Интерфейс) и ППрЗн(Программа Приобретения Знаний). С помощью компоненты КСемС (Корректировка Семантической Сети) можно удалить, изменить, вставить информацию в семантическую сеть. С помощью ССемС (Сопоставителя Семантических Сетей) для двух указанных сетей и определения базовой сети выделяются в зависимости от указанного режима работы понятия, отношения, характеристики и дополняются недостающими элементами в базовой семантической сети.

Приступать к решению задачи можно только в том случае, когда известна вся информация об условиях, ограничениях,

требованиях, входных данных, целях задачи, что представлено в семантической сети задачи.

В ЗдчСемС1 могут быть словосочетания, которые не все могут быть понятиями. Значит, такие словосочетания надо разбить на более мелкие словосочетания или слова. Процесс приобретения знаний состоит в выполнении пяти основных функций: анализа словосочетаний и шаблонов задач, сопоставлении, нахождении источника знаний, создании шаблонов задач, записи полученной информации в нужное место. При сопоставлении один объект должен являться базовым, а другой – входным; при этом должно быть указано, что сопоставляется: понятия, характеристики, отношения и т.д.

Операции сопоставления в математике соответствует операция соответствия. Семантические сети, являющиеся базами знаний, всегда являются базовыми по отношению к семантическим сетям задач. Кратко алгоритм сопоставления состоит в следующем. Выбирается очередная начальная вершина в базовой семантической сети и анализируется словосочетание, связанное с этой вершиной на предмет соответствия его указанному виду сопоставления. После выделения элемента сопоставления происходит контекстный поиск во входной семантической сети (ВхСемС) с целью выявления повторяемости.

В случае индивидуальной встречаемости выделенного элемента сопоставления во ВхСемС делается вывод о том, что элемент сопоставления определен в ВыхСемС. Если элемент сопоставления встречается только в контексте, то это означает, что он используется в определениях, но сам не определен. Если элемент сопоставления нигде не встречается, то это означает, что он не определен и сам не используется для определения. В последних двух случаях нужно находить источник знания. Сначала надо обратиться к функциональной семантической сети – нет ли там функционального определения элемента сопоставления. При наличии элемента сопоставления в функциональной сети посредством ИИн организуется диа-

лог для доопределения элемента сопоставления. При отсутствии элемента сопоставления рассматривается КлСемС с целью организации диалога с помощью ИИн и найденной информации в сети.

В этом случае ППрЗн работает с КлСемС следующим образом. Выбирается очередная начальная вершина с текстом «ТТТ...ТТТ». Допустим, элементу сопоставления соответствует текст «ССС...ССС». В этом случае ППрЗн формирует вопрос типа ««ССС..ССС» является видом «ТТТ..ТТТ» ?Ответ(Да,Нет)». В случае положительного ответа формируется строка в КлСемС, в которой начальной вершине соответствует текст «ТТТ..ТТТ», а конечной вершине – текст «ССС...ССС». При этом организуется дополнительный диалог для получения разного рода характеристик.

В случае отрицательного ответа происходит переход к следующей начальной вершине в КлСемС, не совпадающей с текущей, и так, пока не закончится информация в КлСемС. Если всегда был отрицательный ответ, то надо обратиться к казуальной семантической сети С помощью казуальной семантической сети образовывать новые понятия. Для этого берется словосочетание конечной вершины и добавляется словосочетание начальной вершины в родительном падеже. Это возможно не всегда.

Семантическая сеть шаблонов задач представляет собой следующую структуру:  
(1(Описание шаблона 1-й задачи)((Список списков ссылок на подобные задачи)(Ссылка на решение задачи)))  
(2(Описание шаблона 2-й задачи)((Список списков ссылок на подобные задачи)(Ссылка на решение задачи)))  
.....  
(n(Описание шаблона n-й задачи)((Список списков ссылок на подобные задачи)(Ссылка на решение задачи))).

Внутри текста может быть ссылка на задачи от понятия в виде /относительный номер начального слова понятия – относительный номер последнего слова понятия / @ [список номеров списков шаблонов задач]. Ссылка всегда находится после последнего слова понятия; таких ссылок может быть несколько. Подобность задачи отражает некоторую идентичность (схожесть) по структуре, по классу задач и т.д. Построение шаблона задачи напоминает построение схемы программы.

Для анализа шаблонов нужно уметь исходную задачу преобразовывать в шаблон и потом искать подходящий. Можно искать похожесть по конкретным сочетаниям и, найдя некоторый шаблон, попытаться с помощью такой подсказки преобразовать исходную задачу.

В первом формате представления семантической информации имеются ссылки на нужный вид информации, но отсутствует наглядность представления. Во втором случае при наличии наглядности представления более затруднительным является поиск описательной информации об элементах списка. В данном случае первое представление является внутренним для разрабатываемой системы, а второе – внешним для показа в интерфейсе. Существуют различные способы визуального изображения графа.

В данной статье была поставлена цель рассмотреть на содержательном уровне представление текста задачи на естественном языке в виде семантических сетей. Представление в виде семантической сети представляет собой некоторый формализм, с которым далее легче работать при создании программного комплекса «Интеллектуальная система автоматизированного программирования».

Рекомендована кафедрой программного обеспечения компьютерных систем. Поступила 06.12.04.