

ИННОВАЦИОННЫЕ РЕШЕНИЯ В ФОРМИРОВАНИИ ФУНКЦИОНАЛЬНОЙ СТРУКТУРЫ ПРЕДМЕТНОЙ ОБЛАСТИ

THE NEW APPROACHES IN CONSTRUCTING OF FUNCTIONAL STRUCTURES OF PROBLEMS DOMAIN

В статье описываются подходы выделения знаний о предметной области на основе анализа текстов естественного языка. Для формализации знаний о предметной области и построения структуры бизнес-процессов активно используются различные методы представления знаний. В данной статье рассматривается организация знаний на основе семантических сетей.

Ключевые слова: знания, семантические сети, фрагменты знаний, объекты, отношения.

The article is devoted to the development of new approaches in natural language analysis based on problems domain investigation. The article describes the structure of knowledge with use of the extended semantic networks. Models of knowledge are constructed with fragments sets of multirelational objects.

Keywords: knowledge, semantic networks, fragments of knowledge, objects, relations.

Одним из способов представления знаний являются семантические сети. Они были первоначально разработаны для использования в качестве психологических моделей человеческой памяти, затем их стали использовать для представления и обработки естественного языка; теперь это стандартный метод представления в системах обработки знаний. Структура семантической сети представляется в виде узлов и связывающих их дуг, описывающих отношения между узлами. Узлы в семантической сети соответствуют объектам, концепциям или событиям. Дуги могут быть определены разными методами, зависящими от вида представляемых знаний. Обычно дуги, используемые для представления иерархии, включают дуги типа *is a* (является) и *has part* (имеет часть).

Существует большое разнообразие инструментальных сред для обработки семантических сетей. Одна из них – система Деклар с языком ДЕКЛ. Это язык продукционного типа, который служит для преобразования семантических сетей. Язык ДЕКЛ предназначен для ввода в систему правил, или продукций, которые записываются в форме «ЕСЛИ... ТО...» Семантика языка реализуется программным ядром, работающим

как интерпретатор. При считывании с диска в ОП продукции, записанные на ДЕКЛ, транслируются в язык РСС, где работает программное ядро, которое осуществляет применение продукций.

Это язык продукционного типа. По некоторым своим компонентам ДЕКЛ близок к типовому языку логического программирования – ПРОЛОГ. Однако формы его несколько другие, богаче по своим возможностям. В ДЕКЛ условная часть записывается первой, т.е. слева, а следствие – справа (в языке ПРОЛОГ – наоборот). Соответственно, выбрана другая процедура применения правил. Такое применение управляется индикаторами и сводится к проверке условия с порождением следствий. С помощью продукций реализуется как прямой, так и обратный вывод. При этом в нем отсутствуют многие ограничения, характерные для языка ПРОЛОГ. Во-первых, формы языка ДЕКЛ не ограничиваются правилами, у которых в части, соответствующей следствию, – только один предикат. В продукциях таких предикатов может быть множество. Во-вторых, правила, или продукции (без дополнительных списковых операций), могут применяться к правилам, что весьма необходимо, например, когда нужно анализировать правила, менять их. Более того, одни правила могут строить другие и тут

¹ Кандидат технических наук, доцент, доцент НОУ ВПО «Российский новый университет».

же включать их в работу, что необходимо при металолической обработке. В-третьих, процедура вычисления не носит «жесткого» характера, как в языке ПРОЛОГ. Правила при применении осматривают все знания, находящиеся в ОП, а не только предикаты запроса. Такой процедурой легко управлять, заставляя правила осматривать только нужную часть знаний.

С помощью продукций языка ДЕКЛ удается поддерживать настраиваемые пользовательские представления – языки представления знаний различных предметных областей, и для этого не требуется программирования работы над списками. Поддержка языков сводится к обработке структур расширенной семантической сети (РСС), что осуществляется на металолическом уровне только с помощью продукций. Специальных средств работы со списками не требуется. Поэтому программное ядро – проще. Термин РСС характеризуется понятием фрагмента, которое определяется в контексте описания структурного объекта.

Определим некоторые понятия. Структурным объектом В называется такой объект, который может быть описан с помощью других объектов (назовем их элементарными) и отношений (связей) между ними. При этом, элементарные объекты не являются сами структурными объектами. Например:

$R1(D2, B2/D1) R2(B3, B4/D2) R3(B5, B6/D3)$.

Здесь: $R1, R2, R3$ – имена отношений;

$D1, D2, D3$ – имена фрагментов ($D1$ – имя всего фрагмента $R1(D2, B2)$; при этом оно может опускаться, если этого не требуется для обработки специально);

$B2-B6$ – имена объектов; при этом $D2$ является именем фрагмента $R2(B3, B4/D2)$ и одновременно именем объекта для $R1(D2, B2/D1)$.

Фрагмент $D1$ в данном случае включает в себя объект $D2$.

$D1$ называется структурным объектом. $D2$ и $D3$ – элементарные объекты.

На множестве элементарных объектов $\{P_i\}$ определяется множество отношений $\{R_j\}$, характерных для данного структурного объекта (объекты связываются между собой отношениями). Выражение $R_i (P_1, \dots, P_k)$ назовем фрагментом знаний.

Понятие фрагмента шире, чем предиката, по следующим причинам:

– во-первых, во фрагментах имя отношения равнозначно имени объекта, т.е. на месте отношения может стоять имя объекта, а также константы или переменные;

– во-вторых, в предикатах отсутствуют средства их именования, которые есть во фрагментах (в системе это константа или переменная, которая равнозначна всему фрагменту в целом, вместе с отношением и объектами);

– в-третьих, понятие фрагмента, в отличие от предиката, никак не связано с логическими значениями «истинности» или «ложности», хотя такие значения всегда могут быть введены во фрагмент – для них может быть выделено свое аргументное место.

Условие связности фрагментов. Если:

$\{R\}$ – множество отношений;

$\{P\}$ – множество объектов;

$\{F\}$ – множество имен фрагментов.

Выражение R_i/F_j означает, что у фрагмента F нуль-местным аргументом является R_i , т.е. имеет место следующее выражение $R_i (.../F_j)$. Вместо «...» могут через запятую стоять какие-либо объекты. Выражение типа P_i^k/F_j – обозначает фрагмент с именем F_j , у которого на k -ом аргументном месте стоит P_i .

Имеем три условия. Если существует:

P_i/F_m и P_j/F_n , что $P_i = P_j$, при $F_m \# F_n$,

R_i/F_m и R_j/F_n , что $R_i = R_j$, при $F_m \# F_n$,

P_i/F_m и R_j/F_n , что $P_i = R_j$, при $F_m \# F_n$,

то такие два фрагмента являются связанными и обозначаются: $F_m \Leftarrow F_n$.

Итак, сформулируем условие связности сети.

Если для любого F_p , принадлежащего $\{F\}$, существует P^k , принадлежащее $\{P\}$, и F_j , принадлежащий $\{F\}$, а вместе с ним и P_m^1 , принадлежащее $\{P\}$, такие, для которых существуют фрагменты $R_p (P^k/F_i)$ и $R_q (P_m^1/F_j)$, причем

$P_n^k = P_m^1$, а также последовательность:

$F_i \Leftarrow F_s \Leftarrow \dots \Leftarrow F_t \Leftarrow F_j$,

то такая сеть является связанной.

Для разработки метода структуризации использовался метод обобщения, в основу действия которого положен окрестностный подход. Обобщение рассматривается как выделение множества вершин с одинаковыми («идентичными») окрестностями. Эти вершины объединяются во множество, которому сопоставляется собственная вершина, имеющая ту же окрестность. В описываемом ниже методе используется принцип обобщения вершин объектов через связанные пары отношений.

Можно видеть существенное отличие приведенного метода обобщения от предложенного метода структуризации. Здесь производится не только выделение окрестностей вершин с обобщением, но и формирование пар отношений, характерных для данных вершин. При этом предлагается после анализа вершин,

их отношений, выделения свойственных данному Z_b (раздел знаний предметной области) пар отношений с n -местными аргументами, которые получены в результате анализа связей объектов, выделять типы объектов, которые характеризуются парами n -местных отношений этих объектов. Для рассмотрения данного метода нам потребуются предыдущие определения структурного объекта B , множеств $\{R\}$, $\{P\}$ и $\{F\}$. Структура $R_i (P^k/F_j)$ будет означать, что объект P вступает в отношение R_i на k -ом аргументном месте в именованном фрагменте F_j . При этом существует структурный объект B , которому соответствует раздел знаний Z_b ; на знаниях Z_b определены все вышеприведенные множества.

При рассмотрении пары фрагментов $R_m (P^s/F_h)$ и $R_i (P^k / F_j)$ устанавливаем цель – выделение множеств пар связанных фрагментов. После выделения пары связанных фрагментов конкретные объекты в них заменяются на типы объектов с учетом их аргументных мест во фрагментах. На основе этого строятся новые знания в виде пар фрагментов, в которые входят данные отношения. Рассмотрим этот механизм подробнее.

В случае если существуют R_m/F_h и R_i/F_j , такие, что верно соотношение $P^s/F_h = P^k/F_j$, при том, чтобы $F_h \neq F_j$, то множество объектов P_i составляет тип, порожденный данной парой отношений, т.е. происходит типизация объектов для данной пары отношений (объекты P^s и P^k в двух фрагментах заменяются на типизированный объект P). В результате имеем пару отношений $[R_m (P^s), R_i (P^k)]$. Здесь s и k – аргументные места, определяющие тип объекта во фрагментах. После этого производится выделение связанных пар фрагментов по всем типам пар отношений. Получим полное множество пар отношений, характерных для данной ПО или данного

Z_b . Если у нас множество $\{R\}$ содержит h элементов, то необходимо провести анализ по всему множеству пар отношений от 1 до h . В результате получим анализ R_i^s при изменении i от 1 до $h - R^s$, общее множество пар отношений между объектами, заданных на Z_p . Выделяются пары отношений, свойственных объектам, вступающим в эти отношения. Это обобщенный подход, при котором отношения могут выступать как в роли объектов, так и в роли свойств.

Предлагаемый способ обобщения позволяет выделять наиболее характерные особенности предметной области и способствует формированию формализованной структуры бизнес-процессов. Использование данного метода позволяет в значительной степени ускорить анализ предметной области с целью построения бизнес-процессов предприятия, которое является одним из первых этапов на пути модернизации организации, улучшения управляемости и повышения эффективности.

Литература

1. Кузнецов И.П., Шарнин М.М., Крюков В.В., Пузанов В.В., Золотарёв О.В. Система обработки декларативных структур знаний ДЕ-КЛАР-2, препринт. – М. : АН СССР, ИПИАН, 1989.
2. Кузнецов И.П., Шарнин М.М., Золотарёв О.В. Экспертные решения на основе анализа взаимосвязанных объектов в системе СПРУТ-93. – М. : ИПИРАН, 1994.
3. Kuznetsov Igor, Matskevich Andrey. System for Extracting Semantic Information from Natural Language Text : труды Международного семинара Диалог-2002 по компьютерной лингвистике и ее приложениям. – Протвино : Наука, 2002.