

Когнитивные Модели Языков Программирования

Бессарабов Н.В.

Кубанский государственный университет, ул. Ставропольская, д. 149, г. Краснодар, 350040, Россия.

bes@fpm.kubsu.ru

Аннотация. *Исследован когнитивный аспект моделей языков программирования в процессах создания, изучения и использования этих языков. Предложены структуры концептов, позволяющие работать с высокоуровневыми языками, моделирующими специфику предметных областей. Рассмотрены проблемы, возникающие при измерении значений атрибутов в не изотропных вмещающих пространствах со сложной структурой. Для выявления когнитивных семантических структур и задания их связей с синтаксисом предложено использовать расслоенные семантические сети. Разработана расслоенная семантическая сеть для современных версий языка SQL*

Ключевые слова: когнитивный аспект, структура концепта, вмещающее пространство, измерение атрибутов, расслоенные семантические сети, когнитивная модель SQL

1 Введение

За последние два десятилетия усложнение синтаксиса и семантики языков программирования вышло на качественно новый уровень. Широко используются иноязычные вставки, например, в серверных страницах и языках, допускающих встроенный SQL. В SQL последних версий, используются вставки на языке регулярных выражений, аналитические функции и сложные конструкции типа *model*, предполагающие использование многомерной модели в дополнение к расширенной реляционной модели. В Oracle 11g можно работать с семантическими сетями и онтологиями.

Появились языки, определяемые предметной областью и программы управляемые моделями. В них, по крайней мере, часть задач по разработке синтаксиса и определению семантики языков высокого уровня может быть передана специалистам предметной области. И, наконец, в администрировании баз и, в частности, в процессах настройки приложений было бы полезно разобраться с системой концептов и используемой логикой.

Большинство появившихся проблем, связано с неадекватным описанием создаваемых артефактов. Необходимую ясность можно внести, вводя дополнительные описания в рамках направления, которое можно назвать когнитивной лингвистикой искусственных языков.

Мы рассмотрим когнитивный аспект процессов создания, изучения и использования языков программирования. Будет предложен ряд структур используемых концептов, установлены отличия измерений атрибутов и, возможность распада измерения величины в некоторое семейство величин, определяемое вмещающим пространством. Учёт когнитивного аспекта позволит предложить использование в качестве основы расслоенных семантических сетей с концептами специальной

структуры, зависящей от моделируемой предметной области. В сети выделяется база, расширяемая слоями, определяемыми каждый своим набором однопорядковых концептов. Сложные слои, такие, как описание модели в SQL, могут детализироваться иерархически. Поэтому сама система расширений образует семантическую сеть со связями двух типов: “расширяет” и “детализирует”.

2 Когнитивный аспект в создании языков программирования

Процессы создания языков и технологий программирования можно представлять неформализованными сценариями, выполняемыми коллективами программистов и определяемыми, в первую очередь, осознанными потребностями обслуживаемых предметных областей, особенностями знаний, мышления и восприятия человека-программиста.

Такая “социализация” заставляет вспомнить о замечании Н.Н. Моисеева [1] о том, что “законы в общественных науках имеют характер тенденций”. А тенденции могут и не реализоваться. Поэтому существует множество путей развития и выбор создаваемых моделей неоднозначен.

Языки могут быть специфичными не только для моделируемой предметной области, но и для собственных задач, таких, как разработка интерфейсов пользователя, администрирование баз данных и т.д. Переход к моделированию всех этих областей потребовал детального анализа используемых концептов – сущностей, связей, атрибутов, действий и сценариев. В [2] такой анализ выполнен для сущностей, связей и атрибутов. Используются и концепты без атрибутов, и концепты со сложно структурированными системами атрибутов, классифицируемыми по семантическим признакам. Связи рассматриваются как концепты с двумя основными видами атрибутов. Первые определяют стороны связи, а вторые – эмерджентные свойства, появляющиеся только при наличии связи.

Достаточно широкое определение концепта можно получить [3] выделив четыре класса атрибутов – обязательные, необязательные, состояния (с переходами между ними) и ресурсы. Дополнительно предполагается, что в моделях понятий-сущностей допускаются любые классификационные структуры на атрибутах, а пути сужения/обобщения понятий задаются явно.

Обобщение в предикатных формулировках понятий, как известно, может производиться через обобщение по именам, по квантификаторам, по модификаторам и по императивам [4]. Допускается задание понятий, для которых не имеется предикатной формулировки, с помощью набора типичных образцов. Используются явно выраженные ограничения на значения атрибутов и их комбинаций [5].

Понятия-связи уточняются добавлением атрибутов, определяющих возможность задействования связи, в частности, задающих нелинейности, и атрибутов, определяющих состояние и историю связи.

Выделяется явно класс сущностей-ресурсов. Желательно определить для них характеристики доступности и, если это возможно, законы сохранения, хотя бы локальные. Используются ограничения на ресурсы и атрибуты любых концептов.

Остаётся добавить действия, которые определим как связи особого вида с тремя группами атрибутов, определяющих то, что действует (субъект), то, над чем выполняется действие (объект) и, наконец, при каких условиях, с какими ограничениями и свойствами выполняется действие.

Под сценарием будем понимать связный и взаимно обусловленный набор действий, который может описываться одним концептом. Введение сценариев позволяет рассматривать динамические модели на семантических сетях.

Было бы интересно реализовать характерные для гуманитарных наук и естественные для фреймового представления знаний контекст и подтекст. Программные реализации контекста и подтекста возможны, но, безусловно, будут весьма ресурсоёмки.

3 Об измерении атрибутов

В работе [6] исследованы расширения, особенности и ограничения применения предложенной в [7] парадигмы единой геометрической теории управления. В соответствии с ней система, как естественно существующая, так и искусственно созданная, рассматривается в пространстве представляющем последовательность расслоений со своими геометриями слоев.

Попробуем перенести понятие “измеряемая величина” на признаки сложных концептов, используемых для моделирования предметных областей в общественных и компьютерных науках, в частности, в администрировании программных продуктов и в моделях интерфейсов пользователя. Особенности таких измерений зачастую определяются невозможностью адекватного моделирования описываемого объекта вне контекста или вмещающей среды. Результат измерения может зависеть от времени, от состояния объекта и системы в целом, от ресурсов, имеющихся в распоряжении измеряемого объекта и, наконец, от ограниченности ресурсов, совместно используемых объектом измерения, другими объектами системы и измерителем. Опыт социологических измерений показал, что возможны интерпретации результата измерения в моделях отличных от модели измерения. И, наконец, измеряемая величина может распадаться в семейство измеряемых величин. Не исключено, что не существует простого глобально описания, но существует система локальных моделей.

Выпишем обычно неявно предполагаемые постулаты классической теории измерений, сделав упор на нюансы важные для рассматриваемого класса измеряемых величин:

- Измеряемый объект есть замкнутая система. Он не взаимодействует с другими объектами и, в частности, с измерителем. Отсюда следует, что его можно без ущерба извлечь из контекста (системы) или вмещающего пространства, и что влиянием измерителя можно пренебречь.
- Измеряемый объект не меняется вообще и, тем более, во время измерения. Это означает, что результаты измерений выполненных в разные моменты времени совпадают.
- Измеряемая величина не имеет ограничений на применимость.
- Результат измерения интерпретируется непосредственно в рамках модели измерения.
- Измеритель и алгоритм измерения не рассматриваются в рамках теории, то есть считаются всегда существующими или, по крайней мере, не заслуживающими внимания.

Заметим, что в социологических измерениях невыполнение последних трёх пунктов учитывается давно [8]. У нас возможны нарушения всех перечисленных постулатов.

Например, измеряемая величина “размер таблицы” для системы управления базами данных Oracle распадается в семейство из шести величин (размер сегмента как место, которое не могут занять другие таблицы, размер как количество экстендов занятых данными таблицы, размер как количество символов, получаемых при распечатке таблицы и т.д.). Заметим, что величины этого семейства, за исключением размера определяемого через число символов, вне базы смысла не имеют. В других системах пространство, вмещающее таблицы, может быть устроено по-другому. Так, в СУБД Cache таблицы хранятся в структуре B*-деревьев.

Не менее сложно определение измеряемой величины, определяющей потребность в использовании пространства в интерфейсе пользователя. В адаптивной системе, позволяющей оптимизировать интерфейс пользователя [9] структура интерфейса может быть представлена деревом прямоугольников (блоков), может быть, вложенных. Прямоугольники одного уровня конкурируют за пространство вмещающего блока. Величина “потребность в пространстве” должна учитывать особенности сложившейся конфигурации.

Классическая измеряемая величина это концепт с двумя обязательными и одним необязательным атрибутом:

имя_измеряемой_величины(модель_измерения, результат_измерения, погрешность).

Измеряемая величина рассматриваемого класса имеет более сложную спецификацию:

имя_измеряемой_величины (вмещающее_пространство, область_применения, модель_измерения, модель_интерпретации, результат_измерения, погрешность, параметры_измерения).

В ней атрибут “область_применения” характеризует ограничения на применимость величины. Указание моделей измерения и интерпретации отражает различия между измерением и интерпретацией. Атрибут “параметры_измерения” характеризует условия, при которых выполнялось измерение свойства объекта.

Ограничения областей существования измеряемых величин -- одна из основных причин ограниченности описываемых моделей. В отличие от известной практики, когда кусочные описания вызываются техническими причинами, в основном, желанием упрощения используемых моделей, в

нашем случае причины явления лежат глубже -- в свойствах самой изучаемой системы, в большом объеме плохо обобщаемых локальных знаний, в невозможности построения простой глобальной модели. Существенные ограничения на применение моделей делают получаемые знания менее универсальными и выдвигают на первый план аспекты, связанные с классификациями и выделением локальных свойств.

4 Расслоенные семантические сети

Перейдём к рассмотрению когнитивных структур. Любой язык можно определить на семантической сети с концептами, типы которых определяются предметной областью, или следуют из идеологии языка. Для исследования структуры языка лингвистическими методами достаточно оперировать явно заданными фрагментами этой сети. В расслоенной сети выделяется база – начальная система связанных концептов, которая разворачивается путём добавления слоёв. В языках, разработанных давно, база определяется, как правило, на основе некоторой математической модели. В SQL это реляционное исчисление на кортежах, в LISP -- λ -исчисление, и т. д. В предметно-ориентированных языках база может быть не связана с какими-либо математическими построениями.

Слой (и база) -- это фрагменты семантической сети, распознаваемые человеком как наборы равноположных концептов, сильно связанных между собой и слабо связанных с остальными равноположными концептами сети. Предполагается, что существует ядро слоя, такое, что удаление хотя бы одного концепта ядра разрушает слой. Для того, чтобы обеспечить хорошее восприятие слоя человеком, слои должны содержать, как правило, небольшое количество концептов. Существуют вертикальные структуры, развиваемые вглубь, но и в них начальный слой устроен достаточно просто.

При введении нового слоя нежелательно изменять предыдущие слои. Это нарушает совместимость версий языка. Введение слоя может вызвать следующие эффекты в существующих слоях:

- изменение прагматики, например, появление запретов или ограничений на использование отдельных концептов или слоя в целом в контексте нового слоя;
- изменение семантики концептов или слоя в целом.

Отмеченные выше изменения в использовании других слоёв, по-видимому, появляются, когда введённый слой использует концепты других слоёв или два слоя используют общие равноположные основным концепты.

Выделим четыре способа добавления слоёв:

- расширение языка за счёт добавления слоя; в синтаксисе добавляется одна или несколько фраз;
- агрегирование -- новый слой вводит элементы в старый; необходимо указать условия соединения и, может быть, возникающие ограничения;
- “добавление вариантов” -- это создание агрегатов с несколькими близкими слоями;
- “рекурсия” – слой есть компонента себя.

Слои образуют сеть с многосортными, в общем случае, связями. Название слоя должно быть термином концепта, представляющего подсеть семантической сети, которая вводится как слой. Если такой термин, может быть многословный, не находится, значит невозможно распознать эту подсеть как единое целое. Скорее всего, не стоит выделять эту подсеть в слой.

5 Когнитивная модель запросов SQL

Нами создан прототип учебника по языку SQL для СУБД Oracle, основанный на расслоенных моделях, в котором используются образные представления слоёв, и фрагменты семантических сетей создаваемые в инструментальном средстве CmapTools (www.cmap.ihmc.us). Опыт преподавания показал, что использование этих компонентов существенно облегчает освоение материала, особенно таких трудных разделов, как аналитические функции, модель и семантически поддерживаемые запросы.

База – запрос исчисления на кортежах -- и её семантическая сеть приведена на рисунках 1 и 2.

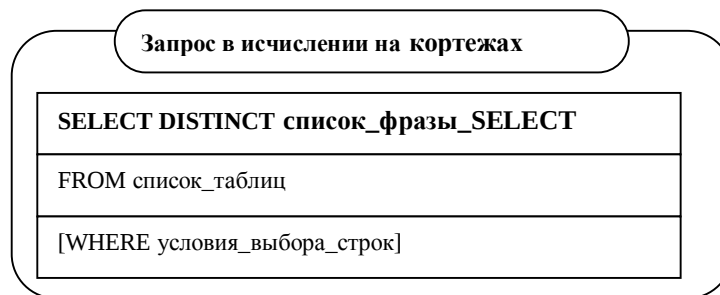


Рис. 1. База сети запросов SQL

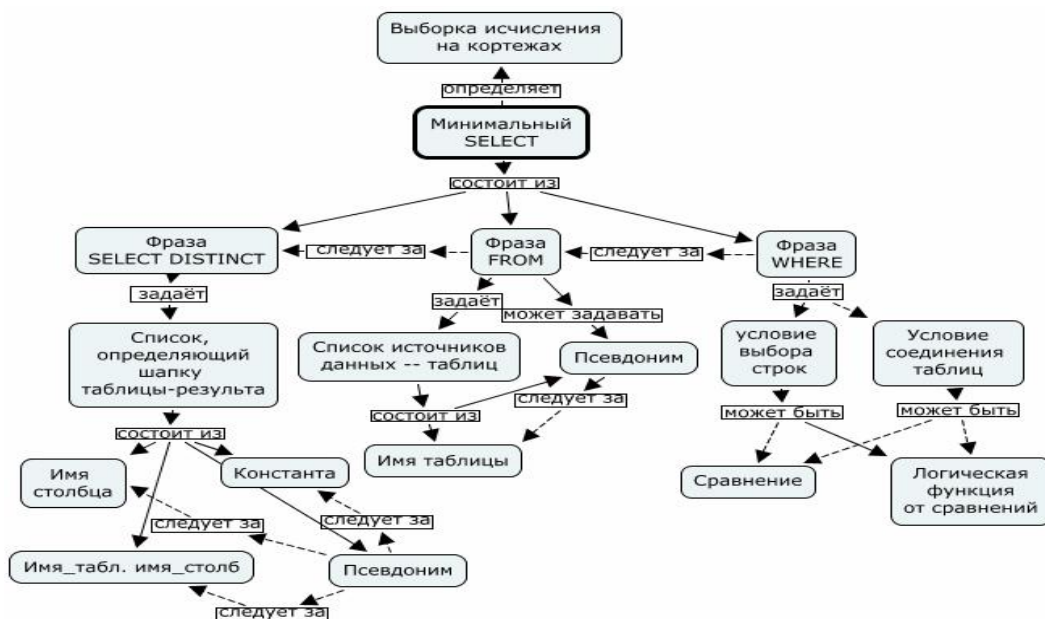


Рис.2. Семантическая сеть для базы.

Оказалось, что слои модели SQL образуют довольно сложную сеть. Небольшой фрагмент сети слоёв для запроса SELECT представлен на рис. 3. Заметим, что в случае, когда слой имеет более одного предшественника, возможны коллизии в реализации сходных или зависимых концептов, введенных в слоях-предшественниках. В описание слоя кроме образных представления слоёв, оформляемые в стиле, рис.1, и фрагментов семантических сетей входит шаблон синтаксиса на XML.

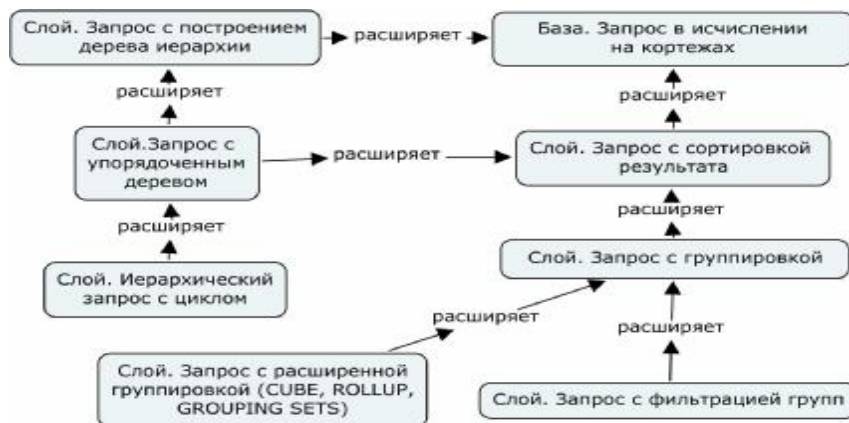


Рис. 3. Фрагмент сети слоёв для запросов SQL

6 Заключение

Использование расслоенных семантических сетей, позволяет выявить когнитивные семантические структуры языка и связать их с синтаксисом. Выделение этого, лингвистического по сути, аспекта позволяет внести в язык структуры удобные для человека, облегчает изучение и использование языка, а также может прояснить перспективы его развития. Получение значимых результатов при лингвистическом подходе возможно при допущении многих вольностей, например, удаётся ограничиться фрагментами семантической сети, оформленными в свободном стиле.

Это не означает, однако, что мы пренебрегаем более строгим подходом. В частности, предполагается решить задачу выделения слоёв как нечётко связанных компонент семантической сети.

Литература

- [1] Моисеев Н.Н. Математика в социальных науках // Математические методы в социологическом исследовании. - М., 1981. С. 10-24.
- [2] Бессарабов Н.В. О моделях понятий, связей, терминов и онтологиях // Труды восьмого международного симпозиума «Интеллектуальные системы». – М.: РУСАКИ, – 2008. с. 396-399.
- [3] Бессарабов Н. В., Бессарабова К. Н. Онтологии и динамические системы / Труды седьмого международного симпозиума «Интеллектуальные системы». – Краснодар, КИИЗ. – 2006. с. 282-286.
- [4] Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.
- [5] Шалфеева В. А., Грибова В. В. Внутренние свойства онтологий // Труды IV Международной конф. «Идентификация систем и задачи управления» SICPRO'05 Москва 25-28 января 2005. – с. 1109-1128.
- [6] Бессарабов Н.В. Об измерении признаков сложных открытых объектов. (в печати)
- [7] Бабичев А.В., Бутковский А.Г., Похьолайнен Сеппо. К единой геометрической теории управления.- М.: Наука, -- 2001.—352 с.
- [8] Толстова Ю.Н. Измерение в социологии. М.: Инфра-М, 2003.
- [9] Бессарабов Н.В., Коблов А.Г., Муса-Оглы Е.С., Торовец Н.Г. Морфизмы в анализе и проектировании информационных систем. Обзорные прикл. и промышл. матем. 2008, т. 15, в. 1, С.115-117.