



## МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ ПОСРЕДСТВОМ ЕСТЕСТВЕННОГО ЯЗЫКА В ИНФОРМАЦИОННЫХ СИСТЕМАХ

БОСКЕБЕЕВ К.ДЖ., САИТОВ Н.

[izvestiya@ktu.aknet.kg](mailto:izvestiya@ktu.aknet.kg)

*Статья представляет собой создание базы знаний в виде набора предикатов первого порядка и семантическую структуру запроса на их обработку с помощью языка VisualProlog7.*

**Цель исследования.** Первый уровень машинного понимания предполагает использование при ответе на вопросы содержания, отражаемого текстом. Модели, с помощью которых воссоздаются знания специалистов-экспертов в интеллектуальных компьютерных системах, характеризуются различным уровнем синтаксичности. Этот уровень зависит от количества синтаксических правил, находящихся вне модели, которые, однако, следует знать, чтобы правильно понять назначение каждого из элементов модели.

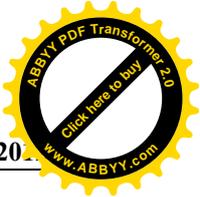
**Модель исследования.** Данная статья посвящена средствам, позволяющим облегчить подготовку базы знаний. Для целей анализа и систематизации знаний используется модель представления знаний, семантика и логика предикатов [1,2].

Высокий уровень синтаксичности приводит к ошибочной автоматической внутренней или внешней интерпретации модели. Например, невозможно восстановить содержание фрагмента семантической сети, заданного в виде такого отношения: подготовил (специалистов, бишкек, университет). Синтаксичность этого фрагмента настолько высока, что довольно затруднительно установить, подготовил ли «университет» специалистов, находясь в Бишкеке, или «университет» подготовил специалистов, обучаемых в Бишкеке. Если в процессе использования моделей человеком для выяснения их содержания можно привлечь авторов этих моделей или соответствующие письменные комментарии, то компьютерные системы такими возможностями не обладают. Все это требует максимального снижения уровня синтаксичности моделей представления знаний за счет их насыщения семантическими маркерами (падежами).

Одним из способов «борьбы» с синтаксичностью моделей представления знаний является явное указывание роли, выполняемой тем или иным элементом модели. Роли, или семантические отношения — это соотношение между обобщенными категориальными значениями связанных слов, отражающих объективно действующее соотношение между денотатами. Совокупность семантических отношений, отражающих связь между денотатами, представляет собой отношение отношений и называется семантической структурой высказывания, фразы, текста и др. Семантические структуры изображаются в виде дерева зависимостей, в узлах которого находятся понятия (имена объектов, процессов, состояний), а дуги символизируют семантические отношения. В корне дерева всегда находится имя предиката, охватывающего объекты и семантические отношения. В отличие от семантических сетей, отражающих отношения между элементами модели с помощью  $n$  арных предикатов, семантические структуры указывают на роль каждого элемента модели. Поэтому логичным является синтез семантических сетей и семантических структур, что, с одной стороны, обеспечит снижение синтаксичности модели, а с другой, — применение для обработки языка VisualProlog 7, известного своим высоким уровнем. А это значит, что будут обеспечены распознавание ролей каждого элемента модели запроса пользователя, сформулированного на языке деловой прозы, и сопоставление их с ролями модели базы знаний. Кроме того, можно будет воспользоваться стандартными средствами для обработки запроса путем его трансляции на язык VisualProlog 7.

Для иллюстрации приведем высказывания: «вуз «технический» находится в городе Бишкек. Вуз имеет факультеты. Факультет состоит из кафедр. Город входит в регион». Для представления этих высказываний в виде семантической сети, семантических структур и результата синтеза — расширенной семантической сети, применимы такие обозначения:

В — вуз, Т — технический, Б — Бишкек, Ф — факультет, К — кафедр, Г — город, Р — регион, Obj — семантическое отношение объекта, Loc — семантическое отношение места, 3 —



значение, Zn — семантическое отношение значения.

Соответствие вопросительных слов и семантических отношений

Таблица 1.

Вопросительно е слово	Семантическое отношение
Какого	Zn
Какой	Zn
Каким	Zn
Какому	Zn
Что	Obj (Zn)
Чего	Obj (Zn)
Где	Loc (Zn)
Кем	Obj (Zn)
Для чего	Gol (Zn)
В чем	To (Zn), Loc (Zn)
За что	Obj (Zn)
Кому	Obj (Zn)
С какой целью	Gol (Zn)
Сколько(какоекоп	Obj (Zn)

В семантической сети базы фактов отсутствуют роли, исполняемые каждым из ее элементов. Однако можно составить перечень семантическое отношение отражает содержание вопросительных слов. Наиболее употребляемые вопросительные слова и соответствующие им семантические отношения приведены в табл. 1. Очень часто в запросах пропускаются имена объектов (предметов, процессов, состояний), значения которых интересуют пользователя. Например, в запросе «Где находится...» не указываются конкретно ни город, ни местность. Более полно этот запрос должен был бы звучать так: «В каком городе находится...». Если же запрос появился в сокращенной форме, то вопросительное слово «Где» трансформируется в структуру Loc(Zn), что означает: Loc— отношение местонахождения (например, город), aZn— отношение значения (например, Бишкек).

Всякую иерархическую структуру можно представить записью, отделив соответствующим количеством скобок (или другим знаком) одни элементы в иерархии от других.

Например, если в дереве два уровня, как это показано на рис. 1, а, то выражение будет таким: **a(b,c)**. А если три (рис.1.б.), то таким : **a(b(f),c)**.

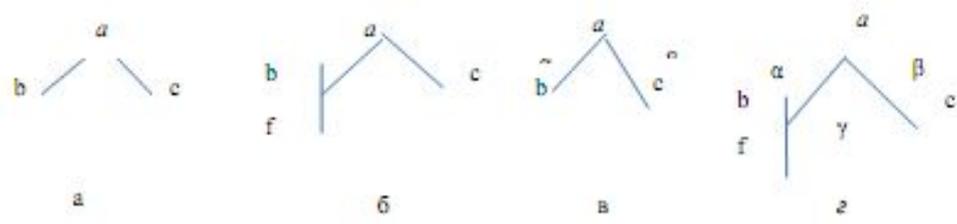


Рис.1. Два и три уровня в иерархической структуре



Варианты  $\beta$  и  $\gamma$  представляют не просто иерархические отношения, но и имена этих отношений ( $\beta, \gamma$ ). Выразить их можно так:  $a(\alpha_b, \beta_c); a(\alpha_b(\gamma_f), \beta_c)$ . Имена отношений ( $\alpha, \beta, \gamma$ ) записывают перед элементом структуры, который связан с элементом, более старшим по иерархии. Предикатное представление семантической структуры имеет две формы: одна для отображения семантики базы знаний, а вторая – для семантики запроса пользователя. Разница между ними в том, что первая содержит лишь семантические отношения, а вторая – еще и понятия, указанные пользователем в запросе. База знаний состоит из следующих фактов вышеуказанного высказывания *находится* ( $B, T, G, B$ ); *имеет* ( $B, \Phi$ ); *состоит* ( $\Phi, K$ ); *входит* ( $G, P$ ). Семантические структуры базы знаний, в предикатной форме имеют следующий вид: *находится* ( $Obj(Zn), Loc(Zn)$ ); *имеет* ( $Obj(Zn), Obj(Zn)$ ); *состоит* ( $Obj(Zn), Obj(Zn)$ ); *входит* ( $Obj, Loc$ ). Предикатная форма семантической структуры запроса включает и понятия, между которыми устанавливаются отношения семантических структур, охватывающих факты, и хранение его в качестве справочного материала. Тогда при появлении запроса с его семантической структурой возникает возможность выделения в базе фактов тех элементов, которые соответствуют семантической структуре запроса. Такая процедура следующая: *находится* ( $Obj_B(Zn_T), Loc_G(Zn_B)$ ). Прежде чем объяснить семантическую структуру запроса, необходимо проанализировать семантические отношения, которые будут использоваться в иллюстрациях и формальных способах записи структур:

$Obj$  — семантическое отношение объекта: существительное или группа существительных, на которые направлено действие, выраженное глаголом, деепричастием;

$Loc$  — семантическое отношение места: группа существительных, отражающая место действия или нахождения;

$Goal$  — семантическое отношение цели: группа слов, отражающая цель, ради которой осуществляется действие;

$Zn$  — семантическое отношение значения: слово или группа слов, приписываемая иным словам.

Количество семантических отношений может быть значительным (сотни) и зависит от целей создания системы и специфики предметной области.

Поскольку, как правило, обращение к системе оформляется с помощью вопросительных предложений, то необходимо знать, как. Предикатную запись семантических структур следует читать так:

- для структур базы фактов: с глаголом «находится» объект состоит в отношении  $Obj$ , который в свою очередь имеет значение (состоит в отношении  $Zn$ ); с глаголом «находится» некоторое понятие состоит в отношении  $Loc$ , имеющее также некоторое значение;
- для структуры запроса: с глаголом «находится» «вуз» состоит в отношении  $Obj$ , которое состоит в отношении  $Zn$  с понятием «кгту»; с глаголом «находится» «город» состоит в отношении  $Loc$ , который состоит в отношении  $Zn$  с понятием «бишкек».

Каким образом семантическая структура базы фактов (знаний) и запроса используются для ответа, показано в предикате *находится* ( $B, T, G, B$ );

Описание предметной области представляется вначале в виде предикатной формы, а затем в семантической сети: *находится* (*вуз, технический, город, бишкек*); *имеет* (*вуз, факультет*); *состоит* (*факультет, кафедра*); *входит* (*город, регион*).

Семантическая структура этих фактов такая:

*находится* ( $Obj(Zn), Loc(Zn)$ ); *имеет* ( $Obj(Zn), Obj(Zn)$ ); *входит* ( $Obj, Obj$ ); *состоит* ( $Obj, Obj$ ).

Если запрос подтверждающего характера, например «Находится ли вуз «технический» в городе Бишкеке?, то его семантическая структура будет такой:

*находится* ( $Obj_{вуз}(Zn_{технический}), Loc_{город}(Zn_{бишкек})$ );

Если же запрос содержит неизвестные, например: «Где находится ВУЗ технический?», то структура его такова:

*Находится* ( $Obj_{вуз}(Zn_{технический}), Loc_X(Zn_Y)$ );

Неизвестные  $X$  и  $Y$  появились в результате обработки вопросительного слова «Где», которому соответствует семантическая структура  $Loc(Zn)$ .

Далее запросы преобразуются в предикаты языка Пролог, имеющие такую форму:

имя предиката (аргумент1, аргумент2,..., аргумент N), где аргумент может быть:

- абстрактным ( $X, Y, Z$ );
- состоит (энергетический, электромеханика);



- анонимным ( $\_$ ).

Для последнего примера запрос примет вид:

**goal: находится (вуз, технический, X, Y).**

Служебное слово «goal» здесь и далее обозначает, что данная фраза написана в соответствии с синтаксисом языка VisualProlog 7, и может быть использована его компилятором для выдачи ответа.

#### **Выводы:**

1. Процесс преобразования естественно-языкового запроса в Пролого ориентированный синтаксис предполагает выполнение операций, с помощью которых отыскивают семантическую структуру факта, наиболее соответствующего содержанию запроса.
2. Представление о предметах различных классов дает возможность установить смысловую связь между конкретными элементами этих классов. Например, понятия «город» и «вуз» совершенно необходимы для распознавания смысловой связи между элементами этих классов, а именно Бишкек и КГТУ. Понятия явно или неявно, но всегда принимают участие в процессе выявления смысла высказывания или запроса.

#### **Литература**

1. Соломатин Н.М. Перспективы развития вычислительной техники: В 11 кн. Справ. пособие / под ред. Ю.М. Смирнова. Кн. 1: Информационные семантические системы. - М.: Высш. шк., 1989. – 127 с.
2. Кузнецов О.П., Адельсон – Вельский Г.М. Дискретная математика для инженеров. – М.: Энергоатомиздат, 1988. – 480 с.