

Вербализация реальности и виртуальности. Ассоциативная семантика

Аннотация. Рассматривается одна из возможных технологий системного анализа, структурирования и частичной формализации описаний сущностей и проблемных ситуаций как попытка уменьшить влияние субъективизма при проектировании систем поддержки принятия решений. Предлагается разграничить понятия семантических и ассоциативных сетей, описываются операции с образами ситуаций для поиска аналогий. Предложена концептуальная схема ассоциативного узнавания текста.

Ключевые слова: вербализация, инженерия знаний, отношения виртуальности, семантическая сеть, ассоциативная семантика.

Введение

Идеи создания онтологий, которые позволили бы организовать знания о реальном мире в виде иерархии сущностей и отношений между ними и представить их в форме, приемлемой для компьютерной обработки, с момента своей актуализации в 80-х годах прошлого столетия привели к появлению новых разделов искусственного интеллекта, называемых онтологической инженерией и инженерией знаний, и послужили мощным толчком в развитии смежных разделов [1-3]

В широком смысле понятие *вербализация* означает словесное описание мира: предметов, процессов, явлений, в том числе и нашего восприятия мира – чувств, переживаний, поведения, мыслей с целью обмена информацией. В информатике вербализация реальности необходима для сжатия (реферирования) потоков данных (измерения, изображения, тексты) при хранении и передаче, а также при решении задач на знаниях для описания входных и выходных ситуаций в форме "понятной" компьютеру. Наверное, каждый человек воспринимает внешний мир по-своему, имеет свою "точку зрения", может быть, за исключением тех вещей, которые можно измерить, т.е. на его восприятие

реальности накладывается его индивидуальный вектор виртуальности.

В работе предлагается при оценке достоверности высказываний учитывать эту компоненту в естественно-языковых текстах, т.е. уже вербализованных сюжетах, которые содержат словосочетания, отражающие ощущения респондента. Осуществляется попытка частичной формализации при описании проблемных ситуаций с целью сближения анализаторов естественно-языковых текстов с языками решателей задач в системах поддержки принятия решений, а также при проектировании информационных систем в технических областях.

1. Этапы вербализации

В процессах вербализации выделим следующие основные этапы.

1. Перевод параметров сущностей (предметов, процессов, явлений и т.д.) в лингвистические переменные путем разбиения диапазонов их значений на поддиапазоны (в частности, с нечеткими границами) и присвоение поддиапазонам имен (идентификаторов).

Наиболее популярный инструментарий – статистический анализ и теория вероятностей, а также логические алгоритмы классификации [4-5].

2. Темпоральная индексация пространства параметров – присвоение качественных (дискретных) значений точкам и интервалам на оси времени (например, *раньше, до, в течение, позже* и т.п.).

3. Детализация и кластеризация (классификация) предметов и процессов. Детализация осуществляется, когда объект физически состоит из подобъектов (деталей) и необходимо их анализировать в отдельности. В некоторых случаях подобъект физически не отделен, но рассматривается как часть, обладающая специфическими свойствами (например, полуостров, мыс, коса, аппендикс, выступ, яма).

Кластеризация реальных субъектов (денотатов) осуществляется с помощью аппарата кластерного анализа значений их параметров (свойств, признаков, характеристик, атрибутов) и составов. Границы кластеров в n -мерном пространстве признаков, описывающих сущности, также могут быть нечеткими (мягкими), но никакой конкретный субъект не должен попадать одновременно в два кластера при решении одной задачи.

Гипотеза. Если два кластера пересекаются, то существует какой-то неизвестный параметр.

4. Таксономизация – построение иерархий сущностей предметов, процессов, ситуаций, сюжетов, отношений. Всякий объект может изменить свой статус в зависимости от той функциональности, которая в данный момент нас интересует. Например, человек на работе и дома. Например, одна деталь (при отсутствии номерного учета) может входить в разные сборочные узлы, которые сами находятся на разных уровнях дерева входимости. Поэтому онтология предметной области содержит историю развития, составляющую опыт системы.

5. Выявление и установление семантических и ассоциативных связей: по свойствам, по отношениям различного типа, по сходству образов и моделей.

2. Онтология предметной области и семантические модели сущностей

Что считается знаниями в СУБЗ КГ [6-7].

1. Декларативные знания:

- свойства, предметы, процессы – базовые типы и конструкции из них;
- сюжеты и ситуации на них – образцы из предметной области (ПрО).

2. Императивные (процедурные) знания:

- справочники, в т.ч. по функциям и программам, как форма систематизации правил, представляющих однотипные зависимости между параметрами;
- логические операции выявления и установления отношений;
- правила предметной области;
- правила вывода;
- планы решения задач – это разбиения задач на подзадачи, в том числе, подготовительные и заключительные действия.

Например, высказывание "*Каждый студент вуза знает два языка*" может быть представлено как вопрос - утверждение, требующее доказательства, или как правило ПрО: "*Если студент учится в вузе, то он владеет двумя языками*".

Онтология предметной области, как правило, содержит шесть семантических категорий концептов: свойство, предмет, процесс, отношение, сюжет, ситуация.

Концепт-свойство имеет вид:

$\langle \text{Свойство} \rangle ::= \langle \text{Имя} \rangle, [\langle \text{Комментарий} \rangle,] \langle \text{Тип значения} \rangle, [\langle \text{Метод} \rangle,]$ где

Тип значения – непрерывный, дискретный, денежный, качественный, текстовый, ссылочный;

Метод – способ вычисления атрибута.

Свойства используются в качестве атрибутов при описании всех других категорий концептов. Имеется пять семантических разновидностей атрибутов (параметров):

- имена (напр., ФИО, адрес);
- брэнды, шифры (код, шифр, ключ);
- свойства собственные (по областям науки);
- характеристики относительные (по видам отношений, см. ниже);
- время.

Предмет имеет состав (детали), схему соединения и геометрические модели (кинематическая схема, чертеж, карта, 3D-модель):

$\langle \text{Предмет} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Комментарий} \rangle, \langle \text{Список атрибутов} \rangle, [\langle \text{Список компонент} \rangle, [\langle \text{Список отношений между компонентами} \rangle,]] \langle \text{Геометрич. модель} \rangle]$

$\langle \text{Атрибут} \rangle ::= [\langle \text{Имя} \rangle], \langle \text{Концепт-свойство} \rangle, \langle \text{Диапазон изменения значений} \rangle, \langle \text{Ожидаемое значение} \rangle, [\langle \text{Единица измерения} \rangle,] [\langle \text{Экземпляр-метод} \rangle]$

Список отношений между компонентами определяет относительное положение, типы соединений, степени свободы и т.п.

Состав предмета – это дерево входимости деталей в сборочные узлы.

Граф-схема предмета: $GSh = G(V, E)$, где:

V – множество деталей предмета,

E – множество соединений (отношений предметной области).

Процессы имеют состав и алгоритм, т.е. их описание содержит вычислительные модели:

$\langle \text{Процесс} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Комментарий} \rangle, \langle \text{Список атрибутов} \rangle,$

$[\langle \text{Длительность} \rangle,] \langle \text{Список входных концептов} \rangle, \langle \text{Список выходных концептов} \rangle, [\langle \text{Вычислительная модель} \rangle,] [\langle \text{Состав подпроцессов} \rangle, \langle \text{Граф-схема} \rangle].$

Роли участников действия (входных/выходных концептов): *агент* (исполнитель); *бенефициант* (заказчик); *реципиент* (приемник действия); *предмет воздействия*; *сцена действия*; *инструмент*; *коагент*; *эффект* и т.д.

Отношение:

$\langle \text{Отношение} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Тип} \rangle, [\langle \text{Комментарий} \rangle,] \langle \text{Список атрибутов} \rangle,]$

$\langle \text{Имя объекта 1} \rangle, \langle \text{Имя объекта 2} \rangle, \langle \text{Тип значения} \rangle, \langle \text{Ожидаемое значение} \rangle,$

$[\langle \text{Единица измерения} \rangle,] \langle \text{Экземпляр-метод} \rangle.$

Здесь *экземпляр-метод* – это конкретный метод из некоторого класса процессов или запрос на выборку из БД.

Семантически непересекающиеся категории отношений (значения которых принадлежат разным множествам):

1) *принадлежности* предмету, процессу, классу, множеству;

2) *родства* – степень удаленности друг от друга двух субъектов в родовидовом дереве;

3) *положения* субъекта в пространстве предмета, сцены относительно другого субъекта;

4) *следования* во времени относительно события, предмета, процесса;

5) *каузальные* – причинно-следственные связи;

6) *сравнения/сопоставления* значений свойств или математических моделей, в частности, графов;

7) *коммуникации* (толерантности) субъекта к другому субъекту или классу;

8) *деятельностные* – ролевые отношения между участниками процесса;

9) *системные* – отношения на множестве субъектов, взаимодействующих между собой и оказывающих прямое или косвенное влияние друг на друга, например, в экологической системе.

Отношения 1 - 7 бинарные, 8 - 9 n – арные.

Отношения двузначные (принимаяющие булевские значения):

1) сравнение: $\langle \langle, \leq, =, \geq, \rangle, \rangle \rangle$;

2) принадлежность предмету, процессу: *Part of*;

3) принадлежность классу, множеству: *ISA* (*is an instance*);

4) родство: *AKO* (*a Kind of*);

5) следование (за): (*A Follow B*);

6) сопутствие (*accompanies*): $x @ Z$, в отличие от $x \in Z$ – множеству однотипных объектов, означает, что x является атрибутом (свойством, параметром, признаком) объекта Z , т.е. является одним из множества *различных* атрибутов всех субъектов класса Z . Запись $a @ b$, где $a \in A, b \in B$ можно также применять, когда A, B являются разными сущностями и имеют разные наборы атрибутов, например *стул @ стол*. Тогда это отношение многозначно.

7) *деятельностные*:

- исполнитель: ($x \text{ Actor } P$);

- бенефициант (заказчик): ($x \text{ Customer } P$);

- реципиент (приемник действия): ($x \text{ Recipient } P$);

- субъект воздействия: ($x \text{ Subject } P$);

- место действия: ($x \text{ PlaceOf } P$).

Пример. *Преподаватель читает лекции студентам*: ($x \text{ Actor } P$) & ($y \text{ Recipient } P$).

В конкретной ПрО роли могут быть более определенными и вводиться экспертами. Например, охрана труда – несчастный случай. Роли: *пострадавший, свидетель, член комиссии*.

Отношения многозначные (принимаяющие значения соответствующих лингвистических переменных):

• положение: ($X \text{ LocationOf } Y$);

• сопоставление (графов): ($x \text{ Pattern } y$);

• толерантность: ($x \text{ ToleranceTo } y$);

• системные (*зависит от*): ($x \text{ DependsOn } Y$).

Ситуация – это совокупность состояний предметов, процессов и отношений между ними, где состояние – это совокупность значений параметров предмета, процесса, отношения в некоторый момент или отрезок времени.

Сюжет – это временная диаграмма заданной совокупности процессов. Всякий сюжет может быть раскадрован на ситуации с произвольным шагом по времени, в частности путем логического деления в моменты событий начала или конца любого из рассматриваемых процессов.

С точки зрения сюжета предметы подразделяются на:

- исполнителей;
- других участников;
- статистов.

Процессы бывают:

- внешние, например, *шел дождь*;
- внутренние, например, *Иванов gripпует*;
- деятельностные, например, *Петров пишет программу*.

Соответственно отношения:

- внешние;
- внутренние;
- межсубъектные.

Сюжет представляет собой полихромный граф, описывающий динамику отношений между субъектами (как способ запоминания опыта). Сюжет отличается от составного процесса отсутствием общей цели.

Граф сюжета – можно представить как микрофильм, содержащий последовательность кадров – ситуаций, связанных между собой процессами и участниками (Рис. 1):

- каждый кадр – это плоский относительно времени граф, множество вершин которого отображает процессы и участников, а множество ребер – связи между ними в заданный момент или интервал времени;

- каждые два соседних кадра связаны ребрами, соединяющими одноименные процессы и одноименных участников (двудольный граф).

Промежутки между так называемыми "ключевыми" кадрами могут быть различными. Если, например, граф сюжета отображает некий видеоклип, то промежуточные кадры генерируются автоматически с заданным шагом по времени.

3. Соотношения между высказываниями и действительностью

Высказывания о свойствах объекта или отношениях его атрибутов с атрибутами других объектов называют *предикатами*. Поскольку они формулируются экспертом или конечным пользователем, при решении задач необходимо учитывать степень достоверности высказываний, тем более, что большинство людей при этом употребляют ментальные (чувственные) глаголы.



Рис. 1. Временная диаграмма сюжета

В таких случаях предлагается рассматривать очередное задаваемое отношение в совокупности с *постотношением* между высказыванием как таковым и действительностью. Значение предотношения зависит от способа и средств (глаголов, словосочетаний, вспомогательных слов) его выражения и составляет степень достоверности основного отношения, т.е. служит коэффициентом, ослабляющим или усиливающим значение последнего.

Рассмотрим высказывание $A \rightarrow B$ (из A следует B). Логической связке "следует" в различных ПрО могут быть поставлены в соответствие одно или несколько свойств (параметров), принимающих одно из двух или более значений и выражающих различные аспекты отношения высказывания к реальности:

- вероятность истинности в действительности;
- модальность – степень необходимости выполнения высказывания (желательно, возможно, необходимо);
- временную зависимость достоверности высказывания, например, вчера оно было истинным, а сегодня – не совсем;
- сослагательность – гипотетичность высказывания ("Если бы...");
- возвратность и притяжательность – отношение принадлежности высказывания автору или другому субъекту;
- толерантность – отношение субъекта к высказыванию ("Мне нравится, что я больна не Вами"; "Он верит, что...");
- источник утверждения – органы чувств (зрение, слух, обоняние, вкус), интуиция, мыслительные, речевые, письменные и другие процессы, выражаемые в естественном языке

ментальными глаголами. Например, "*Лучше один раз увидеть, чем сто раз услышать*" выражает степень достоверности обсуждаемой характеристики некоего предмета.

Виртуальность высказывания запишем следующим образом:

$$\mathfrak{R}\left\{\begin{matrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{matrix}\right\}(A) \text{ или } \mathfrak{R}\{\alpha_1, \alpha_2, \dots, \alpha_n\}(A),$$

читается как "постольку, поскольку $\alpha_1, \alpha_2, \dots$ ", то A , где $\alpha_1, \dots, \alpha_n$ – лингвистические переменные [9], принимающие нечеткие значения соответствующих параметров отношения высказывания к действительности. Иначе говоря, виртуальность – это *вектор* в пространстве лингвистических переменных, характеризующий отношение между высказыванием и действительностью.

Вектор виртуальности умножается на значение истинности высказывания и, тем самым, корректирует влияние высказываний эксперта на принятие решений. Естественно-языковое предложение, как правило, описывает несколько фактов. Виртуальность может относиться к отдельному факту, предложению или рассказу в целом.

4. Семантические и ассоциативные сети в задачах на знаниях

После того как построены иерархии сущностей, устанавливаются связи между сущностями, в результате чего узлы деревьев переплетаются сетями.

Обычно при описании знаний строят *семантические сети*, которые устанавливают атрибутивные связи между понятиями, описывающими некоторую сущность, и ролевые связи между сущностями.

Атрибутивные связи (кто – какой, кто – чей) описывают свойства объектов, принадлежность, родство и другие их характеристики.

Ролевые связи (кто – кому – что – как – для чего – где – когда и т.д.) устанавливаются при выявлении отношений деятельностных, системных, коммуникации, положения, следования, в результате чего строится *концептуальная модель* некоторого процесса.

Помимо вышеперечисленных, в ходе обучения и накопления опыта возникают *ассоциативные связи* – по сходству сущностей или их

атрибутов или их отношений. Сходство может оцениваться по самым разнообразным критериям, которые устанавливаются жизненным опытом индивидуума, зависят от его психофизических характеристик. Как, например, в рассказе "Поэт" Карела Чапека: "*О шея лебедя, о грудь, о барабан и эти палочки – трагедии знаменье*" – таким запечатлелся в памяти поэта номер 235 автомобиля, сбившего женщину.

Ассоциативная связь фиксируется как факт наличия не нулевого значения при проверке отношения сопоставления. Вес этой связи учитывается при распознавании и оценке сходства образов ситуаций (см. ниже).

Ассоциативные связи динамичны. При решении задач, возникающих в ходе жизненного цикла информационной системы, они *модифицируются, развиваются, отмирают*. В общем случае, при поиске решения ассоциативный ряд может привести куда угодно, если нет критериев целевого отбора – ранжирования параметров по степени важности для каждого класса задач. Длина пути в ассоциативной сети не существенна для качества решения, т.к. после его нахождения этот путь, как правило, может быть оптимизирован, и в сети построятся новые связи.

Распознавание образов отыскивает похожее известное, а интуиция – на основе ассоциаций из известных деталей синтезирует новое, неизвестное. Например, можно предположить (утрируя), что с И.Ньютоном именно это и произошло, когда он, обратив внимание на упавшее яблоко, вывел закон всемирного тяготения.

5. Операции с образами пространственно-временных ситуаций

Интеллектуальные системы поддержки принятия решений (ИСППР) предназначены для синтеза новых сущностей и новых решений. Существует два основных подхода (совместимых между собой) к автоматическому поиску новых решений:

- генерация допустимых вариантов конструкций;
- поиск аналогий для новых ситуаций и логический синтез (вывод) новых решений с использованием аналогов.

Для реализации второго подхода используется (с вариациями и расширениями) инструментарий,

работающий с семантическими моделями и конструкциями, составленными из них [8]:

- сравнение параметров;
- распознавание образов;
- сопоставление (сравнение) образов – выявление сходства и отличий;
- поиск цепочек ассоциаций;
- абстрагирование – отключение параметров, не существенных для задачи, например, когда кластер в пространстве трех параметров имеет вид приплюснутого эллипсоида, то одним из параметров можно пренебречь;
- идеализация образа – замена нескольких атрибутов одним обобщенным;
- аморфное преобразование образа (например, размерная параметризация геометрических моделей в САПР);
- сшивание образов – стыковка подграфов известных образов для получения графа нового образа.
- визуализация – перевод вербального в образное.

Заметим, что распознавание образов при неполном наборе видимых параметров – это как идентификация личности по обрывку фотографии – очень "тонкая" операция, но результат ее существенно влияет на планирование решений. В отличие от вероятностного распознавания здесь надежность определяется не только весомостью (уникальностью) имеющихся в нашем распоряжении параметров, но и пространственными связями между ними.

Будем называть *образом объекта* ту часть информации о нем, которая необходима и достаточна для решения поставленной задачи. Граф образа объекта Q^u порядка u – это граф $G^u = G(V, E, A)$, где:

V – множество вершин g , отображающих подобъекты q^{u-1} ;

E – семейство ребер $e \in E = (g_i^{u-1} R_{ij}^{u-1} g_j^{u-1})$, $g_i^{u-1}, g_j^{u-1} \in V$, отображающих отношения между образами подобъектов;

$A = (AH, EH)$ – гиперграф атрибутов $a_{il}(g_i), \dots, a_{im}(g_i) \in AH$ вершин и ребер $eh \in EH$, которые показывают принадлежность атрибутов из множества A некоторой вершине $g_i \in V$.

Сравнение образов. Пусть $G_1 = G(V_1, E_1, A_1)$ – граф известного (эталонного) образа;

$G_2 = G(V_2, E_2, A_2)$ – граф анализируемого образа;

$r = \langle A', A'' \rangle$ – гребенка (набор) анализируемых признаков, где

A' – множество имен признаков, определяющее *степень абстракции*,

A'' – множество коэффициентов, определяющих *точность сравнения* признаков как

$$da_i = a_{li} * a''_k.$$

Будем называть *общей гомоморфной частью* $\tau(G'_1)$ графов G_1 и G_2 часть G'_2 , определенную на множестве V'_2 и состоящую из всех ребер $\tau(E_1) = E_2 = (g_{2i}, g_{2j})$, $g_{2i}, g_{2j} \in V'_2 \subseteq V_2$, для которых существуют соответствующие ребра

$$E_1 = (g_{1i}, g_{1j}) = \tau^{-1}(E_2) = (\tau^{-1}(g_{2i}), \tau^{-1}(g_{2j})), \\ g_{1i}, g_{1j} \in V_1,$$

в графе G_1 , при этом однозначное отображение τ существует тогда и только тогда, когда вершины g_{2i}, g_{2j} и соответствующие им прообразы g_{1i}, g_{1j} совпадают в n -мерном пространстве A' анализируемых признаков с точностью заданной A'' .

Оценка сходства образов, взвешенная по k -тому свойству, есть

$$C = \frac{\sum_{j=1}^n L_k(D'_{2j})}{\sum_{i=1}^n L_k(D_{1i})}, \quad (1)$$

где $L_k(D_{1i})$ и $L_k(D'_{2j})$ – протяженность всех несоприкасающихся цепочек в эталоне и совпавших в анализируемом графе (соответственно), вычисленная с помощью суммируемых по цепочке значений k -того свойства вершин, что автоматически придает больший вес совпадению тех элементов объектов, которые в соответствии с заданной гребенкой признаков идентификации считаются более существенными.

Перечисленные операции с образами, представленными в виде нечетких моделей, дают широкие возможности в части имитации взаимодействия образного и логического мышления при решении задач в ИС ППР.

6. Схема ассоциативного узнавания текста

Ассоциативную семантику определим как нечеткое, выражаемое степенью сходства сопоставление (референцию) субъектов языкового высказывания объектам базы знаний. Неточность смысла (meaning imprecision) частично компенсируется анализом ассоциаций как системы отношений между субъектами и их параметрами.

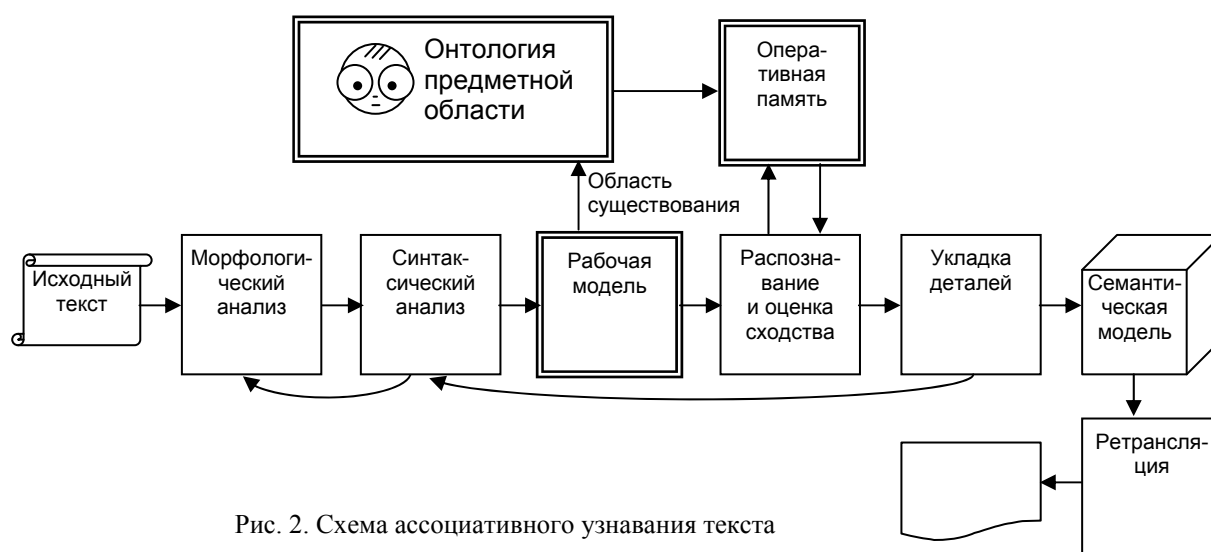


Рис. 2. Схема ассоциативного узнавания текста

Описываемый здесь инструментарий следует рассматривать как один из возможных подходов к конкретизации операции разъяснения [10], являющейся первым этапом вычислений со словами (CW – Computing with Words).

Сущность предлагаемого подхода заключается в преобразовании пространства, которое образует субъект (фрагмент) исходного текста (слово, словосочетание, предложение, текст) в пространство некоторого концепта (понятия, сущности) из базы знаний, сопоставлению векторов и синтезу соответствующей ассоциативной связи (АС).

Имеется *атрибутивная онтология* предметной области (ОПрО), в которой концепты и экземпляры (образы) описаны таблицами, а отношения между ними и их атрибутами – формулами.

Параметры и атрибуты образов:

1. *Область существования* – это узел дерева концептов в ОПрО и множество осей (параметров), в общем случае – нечеткое.
2. *Область определения субъекта текста* – это информация, вычлененная в ходе морфологического и синтаксического анализа из текста.
3. *Физический (механический) смысл* – это назначение объекта ОПрО, определенное как нечеткое множество возможных применений объекта.

Предмет, как правило, может употребляться для нескольких целей, а процесс всегда выполняет одну функцию, хотя результатов может давать несколько. Например, в результате *строгания* получается *доска* и *стружка*.

Несколько терминальных языковых выражений могут иметь одно механическое значение. Например, глаголы: *идти, бежать, ползти, ехать, лететь, плыть, двигаться* имеют один физический смысл – предмет перемещается с одного места на другое.

4. *Время существования.*

5. *Основные параметры* описывают собственные (внутренние) свойства объекта. В задаче узнавания они участвуют как лингвистические переменные, значения которых определяют нечеткий вектор образа объекта.

Добавление осей к пространству образа и наиболее вероятных значений соответствующих параметров из альтернативной области существования в случае многозначности языкового выражения как бы обогащает палитру восприятия. Нарушением такого механизма можно объяснить, например, дальтонизм или отсутствие чувства юмора у человека.

6. *Атрибуты* – внешние, относительные характеристики объекта. В частности, для процесса важными являются атрибуты, определяющие назначение каждого его участника и, тем самым, ролевые отношения между предметами в контексте данного процесса.

Схематично процесс узнавания текста показан на Рис. 2.

По результатам морфологического и синтаксического анализа текста строятся рабочие модели словосочетаний, предложений, связанных групп.

Определяются области существования моделей в ОПрО по совокупности имен их параметров – лингвистических переменных.

Выполняется выборка и загрузка в оперативную память соответствующих концептов и принадлежащих им экземпляров.

Распознавание и оценка сходства. Как видим, пространство сопоставления не строгое (переменное). Его оси – это имена свойств, найденные в исходном тексте. Оценка сходства (1) нового субъекта с известным является результатом пошагового рекурсивного процесса узнавания его деталей и выражает степень принадлежности известному образу. В результате возникает нечеткая ассоциативная связь $AC(X, Rout(A_1, \dots, A_n), Y)$, где AC – функция близости векторов образов X, Y в пространстве $Rout$ параметров $A_1, \dots, A_n$.

Укладка (сборка) найденных в базе знаний вариантов деталей в результирующую модель сюжета позволяет, помимо основной своей функции, сократить множественность возникших ассоциаций. Она выполняется с помощью операции *Распознавание Места* и продукционных правил предметной области. Операция распознавания места основывается на той же процедуре сравнения графов образов (см. выше), но в данном случае ищется место, куда можно вставить очередную деталь так, чтобы она не противоречила текущему состоянию сборки.

Семантическая модель текста представляет нагруженным (атрибутивным) графом.

Ретрансляция семантической модели, выполняемая путем лингвистической аппроксимации [10], служит для визуализации результата в языковой форме.

Пример 1. "Я выхожу на следующей остановке" (Рис. 3).

1. Для каждого субъекта (слова, словосочетания, группы), выделенного из текста в ходе морфологического и синтаксического анализа, находим в онтологии область существования:

Я – человек (предмет одушевленный, местоимение).

Выходить [из] – процесс:

- движение – перемещение с места на место (откуда/куда: комната, дом, транспорт, лес);
- комплектование – исключение/включение в состав участников процесса (откуда/куда: игра, бизнес);
- событие – смена состояния (кризис).

Остановка (следующая):

- сооружение – место, специально оборудованное для кратковременной приоста-

новки движения транспорта с целью входа/выхода пассажиров и (или) груза;

- движение – фаза процесса.

2. Каждый из вариантов толкования субъекта текста добавляет в рабочую модель свою информацию – объект или атрибут или отношение:

2.1. Прилагательное '*следующая*' – означает, что надо рассматривать по меньшей мере две остановки. В таблице предметов появляется строка 3 – остановка следующая.

2.2. '*Остановка*' принадлежит двум областям (предмет и фаза движения).

Остановке как предмету *сопутствует* предмет '*транспорт*', а транспорту – '*груз*'. В таблицу предметов добавляется транспорт (груз уже есть).

В то же время, остановка является частью процесса '*Движение*', поэтому в таблицу процессов добавляется пятая строка – движение и шестая строка – '*останавливаться*', связанная со следующей остановкой.

3. Установка отношений между процессами и участниками. Формирование списка отношений.

4. Уточнение параметров времени. Поскольку некто '*Я*' выходит на второй остановке, то длительность действия '*выходить*' можно приравнять к длительности остановки. Соответственно, совпадает и их время. Остаются неизвестными время и длительность движения, но это можно уточнить, если в базе знаний есть информация о маршрутах транспорта. Кроме того, если считать, что анализируемое высказывание (как правило) произносится где-то в начале движения, то, очевидно, время глагола '*выхожу*' (настоящее) должно быть скорректировано на будущее.

Таким образом, явная или скрытая в естественно-языковом тексте виртуальность раскрывается с помощью ассоциативных связей базы знаний, благодаря чему уточняется семантическая модель текста. В нашем примере в модели сюжета (Рис. 3):

1) появились "подразумеваемые" предметы *остановка* (предыдущая) и *транспорт* и процессы *остановка* – *движение* – *остановка*;

2) установилась *система отношений*, связывающих между собой все объекты, кроме второго и третьего вариантов толкования процесса '*выходить*'.

Пример 2. Классическая задача Эйнштейна.

В пяти домах разного цвета живут люди пяти национальностей.

Они курят пять разных сортов сигарет, пьют пять разных напитков и содержат пять разных видов животных.

1. Англичанин живет в красном доме.
2. Швед держит собак.
3. Датчанин пьет чай.
4. Зеленый дом слева от белого.
5. Хозяин зеленого дома любит пить кофе.
6. Любитель Pall Mall содержит птиц.
7. Хозяин желтого дома курит Dunhill.
8. В центральном доме предпочитают молоко.
9. Норвежец живет в первом доме.

10. Курящий Blend живет по соседству с хозяином кошек.

11. Хозяин лошадей живет рядом с тем, кто курит Dunhill.

12. Любитель пива курит Bluemasters.

13. Немец курит Prince.

14. Норвежец живет рядом с синим домом.

15. Курящий Blend живет по соседству с пьющим воду.

16. Вопрос: Кто держит рыбок?

Определим отношения:

- Рядом, ПоСоседству: $|i - j| = 1$;

Предметы (Subject)

№ п/п	Имя	Область существования	Назначение	Время	Длительность	Параметры (собственные)			Атрибуты (относительные)		
						P 1	P 2	P 3			
1	Я	Человек	Деятельность								
2	Остановка	Сооружение	Место выгрузки/погрузки							Транспорт	
3	Остановка	Сооружение	Место выгрузки/погрузки							Следующая	
4		Транспорт	Перевозка груза								Груз

Процессы

№ п/п	Имя	Область существования	Назначение	Время	Длительность	Параметры			Атрибуты			
						v	a	...	кто/что	откуда	куда	для чего
1	Выходить	Движение	Перемещение							(место)		
2	Выходить	Комплектование	Исключение							список состава		
3	Выходить	Событие	смена состояния							ситуация		
4	Останавливаться	Движение	Прерывание		краткая				транспорт			погрузка/выгрузка
5		Движение	Перемещение						транспорт			груз
6	Останавливаться	Движение	Прерывание		краткая				транспорт			погрузка/выгрузка

Отношения

№ п/п	Имя	Субъекты			Формула
		S1	S2	S3	
1	Следующий	Sub _i	Sub _j		$j = i + 1$
2	Исполнитель	Sub _i	Proc _i		Sub _i Actor Proc _i
3	Исполнитель	Sub _j	Proc _j		Sub _j Actor Proc _j
4	Исполнитель	Sub _j	Proc ₅		Sub _j Actor Proc ₅
5	Исполнитель	Sub _j	Proc ₆		Sub _j Actor Proc ₆
6	Приемник действия	Sub _i	Proc ₅		Sub _i Recipient Proc ₅
7	Приемник действия	Sub _i	Proc ₆		Sub _i Recipient Proc ₆
8	Направление	Proc _i	Sub _j		Proc _i From Sub _j
9	Местоположение	Sub ₂	Proc ₄		Sub ₂ Place Proc ₄
10	Местоположение	Sub ₃	Proc ₄		Sub ₃ Place Proc ₄

Рис. 3. Граф сюжета "Поездка"

- СлеваОт: $j - i = 1$ и его инверсия СправаОт $\equiv$ СлеваОт $^{-1}$ как $i - j = 1$.

Все остальные связи между сущностями в тексте данной задачи обозначим символом @ - "сопутствует" (см. выше). Например 'Швед' @ 'Собаки'.

Сформируем графы деталей для нашей ситуации (Рис. 4).

Требуется восстановить ситуацию, т.е. построить граф ситуации, который можно представить таблицей (Рис. 5), где по строкам расположены вершины (дома), а по столбцам – их атрибуты (обитатели). Взаимосвязи между ними определяются отношениями, заданными в условии задачи.

Решение основывается на операции распознавания ситуаций, которая применяется для поиска места, куда можно вставить деталь, построенную из исходных высказываний:

1. Вставка адресных деталей, непосредственно привязанных к какому-либо субъекту (детали g и h , дома 3 и 1).

2. Вставка безальтернативных деталей таких, для которых отыскивается единственное подходящее место в сборке (в графе ситуации) в цикле, пока есть такие детали (детали d , a , f).

Например, при наложении детали (d) на дом 1 оценка сходства $C_1 = 2/3$, для дома 2 - $C_2 = 2/3$, соответственно, $C_3 = 2/3$, $C_4 = 1$, $C_5 = 2/3$ (в операции распознавания места, которая анализирует пересечения вершин, оценка сходства, т.е. совместимости (1), вычисляет длину цепочки как количество вершин при условии совпадения инцидентных ребер).

3. Распознавание ситуаций.

3.1. Рассортировать оставшиеся детали в порядке убывания их величины по количеству вершин в графе детали (i, b, c, e, j, k, l).

3.2. Найти все возможные положения для каждой из оставшихся деталей путем сравнения графа детали с графом текущей ситуации и выявления пересечений.

Например, для детали (i): $C_1 = 1/3$, $C_2 = C_3 = C_4 = 1$, $C_5 = 2/3$, т.е. на момент распознавания имеется три варианта размещения.

3.3. Взять такие вершины графа ситуации, которые совпали с единственной деталью из оставшихся ($i1$: $E1 = 'Вода'$).

3.4. Исключить другие варианты расположения этой детали ($i2, i3$).

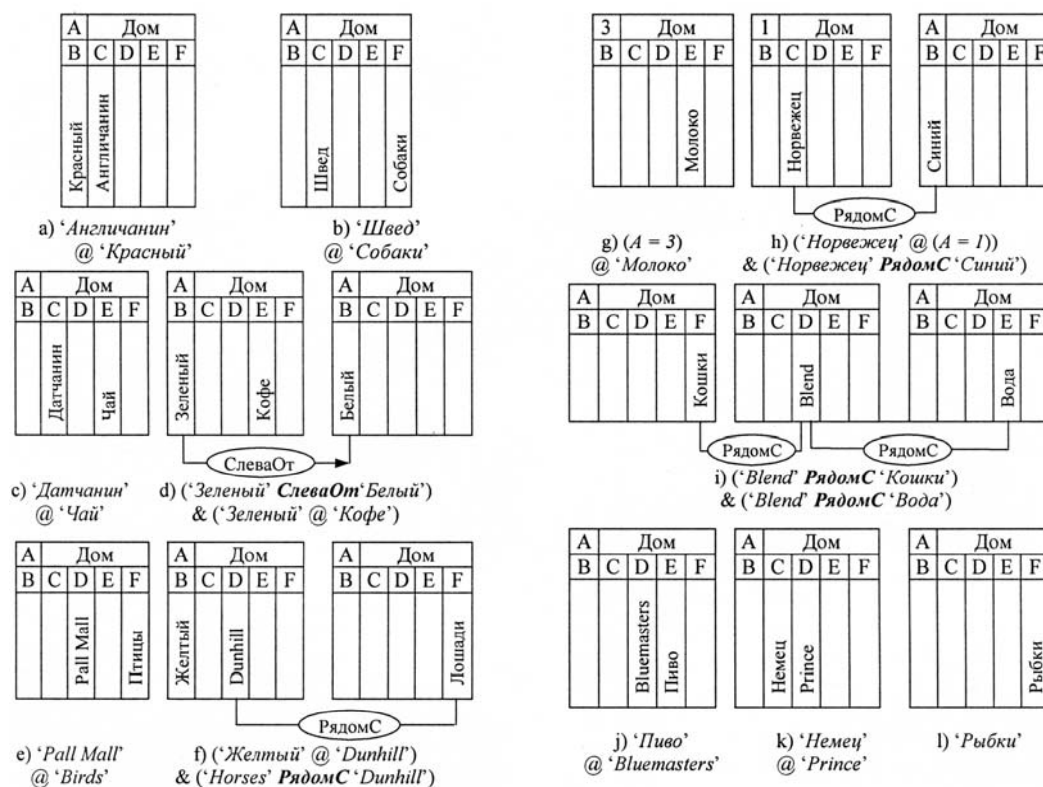


Рис. 4. Графы деталей ситуации (из условия задачи)

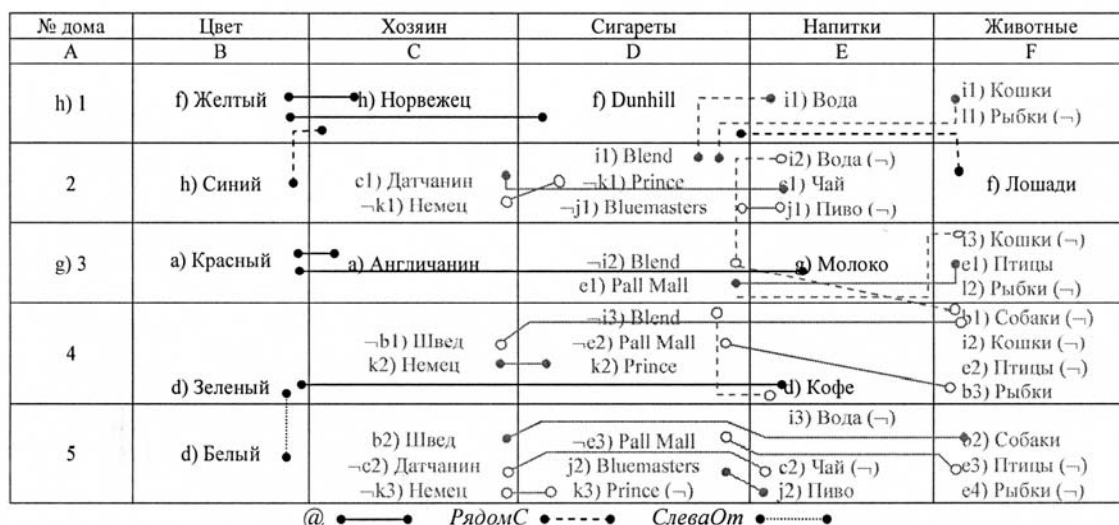


Рис. 5. Граф ситуации с вариантами вставки деталей

3.5. Исключить пересекающиеся с найденной деталью варианты расположения оставшихся деталей (i1, j1, k1).

3.6. Возврат к п. 3.3 пока есть новые однозначные установки.

Второй проход (п.п. 3.3 – 3.6). Непересекающиеся варианты: деталь c1: c2 = 'Датчанин'; e1: D3 = 'Pall Mall'. Исключаем другие варианты этих деталей: c2, e2, e3. Исключаем пересекающиеся с c1 и e1: k1, i2.

Третий проход (п.п. 3.3 – 3.6). Непересекающиеся варианты: деталь j2: E5 = 'Пиво'; k2: D4 = 'Prince'. Исключаем другие варианты этих деталей: k3. Исключаем пересекающиеся с деталями j2 и k2: b1.

Четвертый проход. Непересекающиеся варианты: деталь b2: C5 = 'Швед'. Исключим другие варианты этой детали: нет других. Исключим пересекающиеся с b2: i4.

Получили ситуацию, в которой все детали установлены однозначно, в том числе деталь i: 'Рыбки' – это дом, в котором живет 'Немец'.

Заключение

Таким образом, вербализация плавно переходит в механизацию (компьютеризацию). Полученный в результате тезаурус можно считать коэффициентом интеллекта системы, где объем тезауруса – это эрудиция, а талант – скорость и частота переходов гиперграницы между образным и логическим пространствами восприятия.

Онтологическая модель предметной области служит путеводителем по реальным пространствам, где есть один общий параметр – время, но его шкала в разных пространствах меняет частоту отсчетов. Переходы между однородными пространствами – это параметризация, т.е. установление зависимостей между имеющими одинаковое семантическое значение параметрами этих пространств, тогда как переходы между разнородными пространствами возможны только через обобщение или расщепление параметров, или не возможны, если нет никакой общности.

Очевидно, качество и надежность всей технологии сильно зависит от корректности процесса вербализации знаний о предметной области. Разные эксперты зачастую строят разные онтологии одной предметной области. И не только потому, что устанавливают приоритеты в зависимости от своих задач, но и по-разному понимают некоторые сущности. Семантические модели в виде графов, в которых сущности формализованы и связаны между собой ролевыми отношениями и ассоциативными связями, дают возможность "переводить" описания предметной области с одного языка на другой и пополнять друг друга.

Литература

1. Искусственный интеллект. Справочник в трех томах. /Под ред. В.Н. Захарова, Э.В. Попова, Д.А. Поспелова, В.Ф. Хорошевского – М.: Радио и связь. 1990.

2. Люгер Джордж Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание.: Пер. с англ. – М.: Издательский дом "Вильямс". 2005. – 864 с.
3. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. – М.: Издательский дом "Вильямс". 2006. – 1408 с.
4. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах / Под ред. В.Н.Вагина, Д.А.Поспелова. – М.: Физматлит. 2004. – 704 с.
5. Воронцов К.В. Лекции по логическим алгоритмам классификации, 2007 г.
<http://citeseer.ist.edu/zheng01real.html>.
6. Кучуганов В.Н. Вопросы корректности предметных онтологий / Интеллектуальные системы. Коллективная монография. Выпуск второй / Под ред. В.М.Курейчика. – М.: Физматлит. 2007.
7. Габдрахманов И.Н., Кучуганов В.Н. KNOWLEDGE'S GUIDE – учебная система для распределенной разработки онтологий. // Труды Междунар. научно-технич. конференций "Интеллектуальные системы" (AIS'08)" и "Интеллектуальные САПР" (CAD-2008). Научное издание в 4-х томах. – М.: Физматлит. 2008. Т.1. С. 328-336.
8. Кучуганов В.Н. Семантика графической информации. Известия ТРТУ. Тематич. вып. "Интеллектуальные САПР". Материалы междунар. научн.-техн. конф. "Интеллектуальные САПР". Таганрог: Изд-во ТРТУ. 2002. №3(26). – С. 157-166.
9. Заде Л.А. Основы нового подхода к анализу сложных систем и процессов принятия решений // Математика сегодня. – М.: Знание, 1974. – С. 5-49.
10. Заде Л.А. Роль мягких вычислений и нечеткой логики в понимании, конструировании и развитии информационных / интеллектуальных систем. – Новости Искусственного Интеллекта. № 2-3. 2001. С. 7-11.

Кучуганов Валерий Никонорович. Зав. кафедрой "Автоматизированные системы обработки информации и управления" ГОУ ВПО "Ижевский государственный технический университет". Окончил Ижевский механический институт в 1964 году. Д.т.н., профессор. Автор 110 статей и 4 монографий, имеет 1 авторское свидетельство. Область научных интересов: когнитивное моделирование, интеллектуальные информационные технологии. E-mail: kuchuganov@istu.ru.