

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИНФОРМАЦИОННЫХ СЕМАНТИЧЕСКИХ СИСТЕМ

Введение

Для современного этапа научно-технической революции характерно широкое объединение и глубокое взаимопроникновение различных наук, что создает благоприятные условия для постановки и решения сложных научно-технических проблем. К числу таких проблем относятся и исследования по информационным семантическим системам, т. е. системам, перерабатывающим осмысленную информацию для достижения целей. Решение указанной проблемы требует привлечения методологии диалектического материализма, взаимодействия теории познания, коммуникации, психологии, лингвистики, знаковых систем, прикладных дисциплин математики и других наук.

Актуальность этой проблемы обусловлена как необходимостью решения множества задач сегодняшнего дня, так и перспективными потребностями, связанными с функционированием информационных семантических систем.

Постановка задачи

Ежеминутно человек перерабатывает семантическую информацию предметов, явлений и окружающей его действительности, познавая мир вещей и их отношений и существуя благодаря этому в мире. Эта сторона жизни человека естественная. Человек не задумывается над тем, как он перерабатывает эту информацию. Однако человек накопил в виде документов колоссальный запас знаний (семантическую информацию всех форм представления). Эта сторона жизни человека – результат его деятельности. Человек задумывается над тем, как более эффективно пользоваться этой семантической информацией.

Указанные стороны жизни человека есть не что иное, как непрерывная реализация человеком соответствующих целей, а переработка семантической информации — неотъемлемое свойство (атрибут) систем, стремящихся к реализации указанных целей [1]. Так можно объяснить появление понятия *информационные семантические системы*. При этом не маловажной остается задача подвести базовую основу под применение информационных семантических сетей на практике. Такой основой должна стать теория информационной семантики.

Семантическая информация

Информационными семантическими (ISS) называются такие системы, функционирование которых направлено на достижение цели. Спе-

© В.И. Костюк, А.А. Стенин, С.В. Проценко, 2003

цифической особенностью их функционирования является семантическая переработка семантической информации. Семантическая информация – это **выраженные знаками сведения о выделенной заданием стороне (сторонах) объекта. Знаком же называется материальный предмет (процесс, явление, событие и пр.), выступающий в качестве представителя некоторого другого предмета, свойства или отношения и используемый для приобретения, хранения, переработки и передачи сообщений (семантической информации, знаний).** Как правило, знак — это условное обозначение чего-либо. Он не является обязательным. Один и тот же денотат (условное обозначение знака) можно обозначать по-разному. Например, знак π — отношение длины окружности к диаметру — является условностью. Это отношение могло быть обозначено любой другой буквой. Но было бы очень трудно (и неразумно) пытаться изменить принятую традицию. Знак может не только обозначать конкретный денотат, но и указывать его место в мире вещей, определять какие-то свойства денотата[2].

Различают однородные и комплексные формы представления семантической информации. Множество однородных форм можно определить выражением

$$N1 = \{t, S, g, C\}$$

где t — текстовая форма (t-форма) представления; S — аудиальная (речь, звуки) форма (S-форма); g — визуальная (жесты, пластика) форма (g-форма); C — изобразительная, графическая форма (C-форма). Посредством t-формы выражается содержание печатных и рукописных материалов: книг, газет, журналов, статей, отчетов, диссертаций, патентов, учебных планов, медицинских карт пациентов и др.; посредством S-формы — содержание механических, магнитных и других фонограмм: радиопередач, звукозаписей и др.; посредством g-формы — содержание художественных образов (пантомима, мультипликационное кино и пр.), бытовое общение, обряды, культовые запреты и т. п.; посредством C-формы — содержание графических и изобразительных материалов: чертежей, схем, графиков, карт, диаграмм, рисунков, картин и др.

Комплексная форма (и-форма) представления семантической информации объединяет одновременно несколько однородных форм, например текстовую и аудиальную, изобразительную и визуальную и др[3].

По видам семантическую информацию об объекте классифицируют на первичную и вторичную. Первичной считается семантическая информация, отражающая посредством знаков независимо от формы представления результаты обобщений, исследований, разработок, наблюдений, сочинений и пр., имеющие завершенный характер.

Вторичная семантическая информация, естественно, имеет меньшее содержание, чем первичная, так как при ее получении неизбежны потери. Для получения вторичной семантической информации необходим предварительный логический анализ первоисточника, переосмысливание и сопоставление его содержания с ранее накопленными знаниями

об объекте. С этой точки зрения получение вторичной семантической информации можно рассматривать как продолжение процесса познания объекта, заключающееся в выявлении наиболее устойчивых и характерных признаков.

Носителями первичной семантической информации являются первичные документы.

Примеры первичных документов для различных форм представления семантической информации:

t-форма — книги, журналы, статьи, авторские свидетельства, стандарты и др.;

S-форма — магнитные кассеты, содержащие записи речей, музыкальных произведений, радиопередач, а также диски грамзаписей музыкальных произведений, стихотворений и пр.;

g-форма — видеозаписи мультимедиа, балета, пантомимы, мимики и жеста;

C-форма — структурные, функциональные схемы; карты, чертежи, картины, рисунки, графики и пр.

Примеры вторичных документов для различных форм представления семантической информации:

t-форма — рефераты, аннотации, каталожные карточки библиотек и пр.;

S-форма — звукозаписи изложения речей, фрагмент музыкальных произведений и пр.;

g-форма — видеозаписи фрагментов мультимедиа пантомимы, мимики и жестов и пр.;

C-форма — упрощенные структурные, функциональные схемы; контурные карты, приближенные графики и др[4].

Семантические операции

Понятие “семантическая операция” является синонимом понятия “семантическая переработка” семантической информации.

Семантические операции различаются степенью сложности. Приведем примеры семантических операций. Запишем выражения вида

$$P = \{SI_T, SI_S, SI_G, SI_C\};$$

$$P = \{SI_t, SI_s, SI_g, SI_c\},$$

где SI_T, SI_S, SI_G, SI_C — первичная текстовая, аудиальная, визуальная, изобразительная семантическая информация; SI_t, SI_s, SI_g, SI_c — вторичная семантическая информация соответственно.

Рассмотрим некоторые варианты семантических операций, связанных с преобразованием семантической информации.

Вариант 1. Преобразование “первичное в первичное”. Формально указанное преобразование записывается в виде

$$SI_T \text{Sēm} SI_S; SI_T \text{Sēm} SI_G; SI_C \text{Sēm} SI_S; SI_C \text{Sēm} SI_T; \dots,$$

Запись вида $SI_T S\bar{e}mSI_S$ соответствует выполнению семантической операции, связанной с преобразованием первичной семантической информации текстовой формы представления в первичную семантическую информацию аудиальной формы представления. Аналогично можно пояснить и другие приведенные семантические операции.

Вариант 2. Преобразование “вторичное во вторичное”. Формально преобразование записывается в вид

$$SI_t S\bar{e}mSI_g; SI_g S\bar{e}mSI_s; SI_c S\bar{e}mSI_g; \dots,$$

Запись вида $SI_g S\bar{e}mSI_s$ соответствует выполнению семантической операции, связанной с преобразованием вторичной семантической информации визуальной формы представления в вторичную семантическую информацию аудиальной формы представления. Аналогично можно пояснить и другие семантические операции.

Вариант 3. Преобразование “первичное во вторичное”. Формально преобразование записывается в виде

$$SI_T S\bar{e}mSI_t; SI_S S\bar{e}mSI_s; SI_G S\bar{e}mSI_g; \dots,$$

Запись вида $SI_T S\bar{e}mSI_t$ соответствует выполнению семантической операции, связанной с преобразованием первичной семантической информации текстовой формы представления во вторичную семантическую информацию той же формы представления. Аналогично можно пояснить и другие семантические операции[4].

Методы сжатия семантической информации. Проведенные исследования позволяют выделить два основных метода сжатия: эвристический и модельный.

В процессе сжатия уменьшается содержание первичной семантической информации, выявляются существенные признаки объекта познания, содержание которого представляется первичной семантической информацией; признаки выражаются соответствующими знаками по определенным правилам.

Эвристический метод сжатия. Для этого метода характерно нестрогое задание, которое не определено конкретным содержанием. Оно выражает лишь плохо сформулированную потребность человека уменьшить содержание первичной семантической информации. Человек располагает определенными знаниями о предмете, приспосабливается к содержанию первичной семантической информации и на этой основе создает вторичную семантическую информацию, являющуюся моделью первичной.

В процессе эвристического сжатия логическая структура первичной семантической информации разрушается и создается новая, с меньшим содержанием. Это достигается либо отбрасыванием несущественных признаков, либо заменой некоторых существенных признаков более общими понятиями, отличными от имеющихся в первичной семантической информации.

Для эвристического сжатия характерен субъективизм, т. е. разное понимание людьми существенных и несущественных признаков, а следовательно, разная степень сжатия и разный уровень обобщения понятий. Это приводит к неоднообразию плана выражения семантической информации, получаемой в результате сжатия одной и той же первичной семантической информации разными людьми[5].

Модельный метод сжатия. Для этого метода характерно строгое формализованное задание. Основой его является фиксированная модель. Таким образом, содержание первичной семантической информации приписывается к наперед заданной, логической структуре — формализованной модели. Узловыми аспектами ее служат общие и устойчивые понятия соответствующих областей знаний, позволяющие упорядочить последовательность логических шагов при сжатии, облегчить выбор существенных признаков и таким образом найти типовой алгоритм сжатия, пригодный для обработки разнообразной первичной семантической информации. Процесс сжатия при этом сводится к тому, что выраженные общими понятиями аспекты модели раскрываются конкретным содержанием первичной семантической информации, находящейся в первичных документах. Субъективизм в выборе существенных признаков (степени сжатия уровней обобщения понятий и др.) при модельном способе сжатия сведен к минимуму[6].

Семантические объекты

Объект считается семантическим, если он является элементом информационной семантической системы. В качестве элемента информационной семантической системы семантический объект может выполнять только три функции: источника семантической информации, приемника семантической информации, посредника между семантическими объектами.

Являясь источником, семантический объект генерирует семантическую информацию. В функции приемника семантический объект понимает семантическую информацию. В качестве посредника семантической объект выполняет семантические операции над семантической информацией и, кроме того, может бы источником и/или приемником. В логике понятие, выражающие совокупность объектов, удовлетворяющих какому-либо условию (условиям) или свойству (свойствам, признакам), называется *классом*. Предполагается, что с каждым свойством можно рассматривать класс объектов, обладающих этим свойством. Класс, соответствующий некоторому свойству, может состоять из любого конечного числа объектов.

Понятие “класс” (классы) применительно к информационным семантическим системам рассматривается трех аспектах.

В первом аспекте в основу класса положен признак, определяющий число уровней системы. По данному признаку различаются семантические системы одно-, двух-, трех- и многоуровневые. На каждом уровне сосредоточено (определено) конечное число семантических объектов.

Во втором аспекте в основу класса положен признак, характеризующий тип семантического объекта (человек, документ, машина) в информационной семантической системе.

В третьем аспекте в основу класса положен индекс (первый, второй и т. д.), определяющий группу *ISS*. Различают *ISS* первой группы (одноуровневой), второй (двухуровневой), третьей (трехуровневой), *n*-й (многоуровневой)[7].

Необходимо отметить, что в настоящее время для автоматизированной обработки семантической информации в прикладном плане наибольшее применение находят двух- и трехуровневые *ISS*. Однако существует большая необходимость создания многоуровневых *ISS*, например, многоуровневых экспертных систем для решения практических задач народного хозяйства.

Переход информационных семантических систем из класса в класс

Переход из класса в класс — объективная закономерность диалектического развития информационных семантических систем. В основе такого развития находятся противоречия, как внутренние (внутри *ISS*), так и внешние (*ISS* с окружающей средой). Следует различать два вида перехода *ISS* из класса в класс: прогрессивный и регрессивный.

В процессе прогрессивного перехода *ISS* повышает эффективность функционирования по сравнению с ранее существующей; при этом имеет место адекватная замена (в функциональном отношении) семантического объекта одного типа другим. Различают переход *ISS* из класса в класс как в пределах одной группы, так и между группами. Так, возможен прогрессивный переход во второй группе (двухуровневая *ISS*) из класса документ — человек в класс документ — машина. В данном случае человек заменен машиной, так как машина адекватна человеку в функциональном плане. Такой переход в пределах одной группы очевиден. Повышение эффективности функционирования *ISS* в новом классе обусловлено тем, что машина (ЭВМ) обладает лучшими (по сравнению с человеком в прежнем классе) параметрами, например производительностью выполнения логических операций в единицу времени (считаем, что логические операции являются определяющими операциями семантической переработки семантической информации в данном случае).

В процессе регрессивного перехода *ISS* ухудшает эффективность функционирования. В основе такого перехода могут быть различные причины, например снижение эффективности функционирования *ISS* за счет неправильного научного обоснования, положенного в основу разработки *ISS*, или отсутствие опыта и др.

Причиной регрессивного перехода может быть и деградация *ISS*. Например, функционировала *ISS* класса человек— машина, реализующая поиск семантической информации по специальности “Вычислительная техника”. Затем ЭВМ устарела и была списана (уничтожена), а новую модель ЭВМ не удалось получить. Естественно, что имеет место переход

к классу *ISS* человек — документ (архив), т. е. к традиционной системе ручного поиска информации[8].

Структуры информационных семантических систем

В информационных семантических системах понятие структуры используется в двух основных значениях. В первом значении структура используется как форма, т. е. наружный вид или внешнее очертание. При этом понятие структуры представляется важным сточки зрения классификации существующих и вновь создаваемых форм семантических систем. Во втором значении структура рассматривается как атрибут *ISS*, необходимый (наравне с шестью другими атрибутами) для формального описания *ISS*. Иначе говоря, структура, являясь обязательным аспектом *ISS* (наряду с другими аспектами), служит достижению цели, которую реализует *ISS*[9].

Рассмотрим понятие структуры в первом ее значении. Данной книге классификация форм *ISS* основывается на аналогиях с другими понятиями. Аналогами являются либо геометрические фигуры (плоские, объемные), либо другие понятия. Следует заранее оговориться, что в данном изложении отсутствуют строгие аналитические выводы, а само изложение носит в основном качественный характер.

Плоские структуры *ISS* (модели). Существует большое разнообразие структур *ISS*. Еще большее количество структур можно формально построить. Пример плоской структуры *ISS* приведен на рисунке 1.

Объемные структуры *ISS* (модели). Объемные структуры *ISS* являются чаще всего аналогами объемных фигур. Не представляет особого труда показать множество объемных структур *ISS*, в том числе и иерархических. Пример объемной структуры *ISS* приведен на рисунке 2.

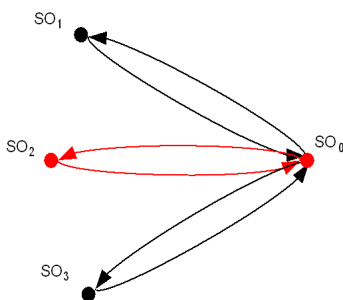


Рис. 1 – Пример плоской структуры *ISS*

Анализируя модели плоских и объемных структур *ISS* можно сделать вывод о том, что основу всех структур составляет дуальная. Этот вывод является важным, так как, применяя метод декомпозиции, можно проводить исследования структур *ISS* любой сложности, опираясь на знание дуальной структуры.

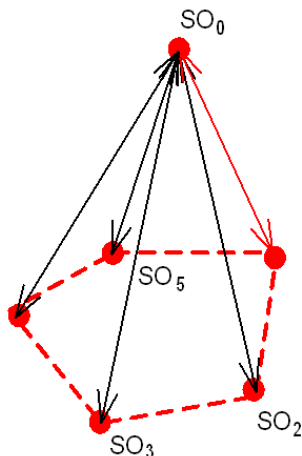


Рис. 2 – Пример объемной структуры *ISS*

Заключение

“Мир ЭВМ” все больше и больше становится семантическим — вот главный вывод по перспективам развития ЭВМ. Это не только реализация сокровенной мечты человека сделать “себе подобного”, но и насущная необходимость обрести “умного партнера” по решению жизненно важных задач, сущность которых составляет переработка смысла.

В этой книге сделана попытка сформулировать основы теории информационных семантических систем. Одним из важных направлений применения сформулированных положений теории информационных семантических систем является проектирование перспективных ЭВМ. В качестве примера можно указать на классифицированные формы представления и преобразования семантической информации; формирование первичной семантической информации об объекте, переработка семантической информации путем выполнения семантических операций, осуществляемых семантическим процессором; проектирование баз знаний по предметным областям методом модельной семантики; место и функции ЭВМ как семантического объекта в системах, стремящихся к достижению цели и др.

Литература

1. Аткинсон Р. Человеческая память и процесс обучения: Пер. с англ. — М.: Прогресс, 1980.
2. Бехтерева Н. Новое в изучении мозга человека. — Коммунист, 1975, 13.
3. Бирюков Б. В., Гутчин И. Б. Машина и творчество. — М.: Радио и связь, 1982.

4. Вилкас Э. Я., Майлинас Е. З. Решения, информация, моделирование. — М.: Радио и связь, 1981.
5. Ершов А. П. Человек и машина. — М.: Знание, 1985.
6. Кибернетика. Дела практические/Под ред. В. Д. Пекелис. — М.:Наука, 1984.
7. Криницкий Н. А. Алгоритмы и роботы. — М.: Радио и связь, 1983.
8. Кулик А. И. Информационные сети и языковая совместимость дескрипторных ИПС. — М.: Сов. радио, 1977.
9. Ложе И. Информационные системы: Пер. с англ. — М.: Мир, 1979.