



Лекция №6

МОДЕЛИРОВАНИЕ ЗНАНИЙ О ПРЕДМЕТНОЙ ОБЛАСТИ

Понятие «Знание»



В основе исследований в области ИИ лежит подход, связанный со знаниями.

Понятие «знание» относится к интуитивно определяемым.

В БСЭ дается следующее его толкование: **«Знание — проверенный практикой результат познания действительности, верное её отражение в сознании человека.»**

Знания бывают житейскими, донаучными, художественными, научными (теоретическими и эмпирическими)».

Разновидности и трактовки знаний



Разновидности знаний:

- декларативные;
- прагматические;
- процедурные;
- эвристические;
- экспертные;
- знания о Про.

Выделяют две характеристики знаний: **объектность** и **личность**.

Трактовки знаний могут быть объединены в четыре группы:

- психологическую;
- интеллектуальную;
- формально-логическую;
- информационно-технологическую.

КЛАССИФИКАЦИИ ЗНАНИЙ



В зависимости от источника знания можно разделить на две категории:

- **априорные знания** (определяются и закладываются в БЗ до начала функционирования ИС);
- **накапливаемые знания** (формируются в процессе использования БЗ):
 - **экспертные знания;**
 - **наблюдаемые знания;**
 - **выводимые знания.**

По характеру использования при решении задач в определенной ПрО выделяют следующие виды знаний:

- **декларативные (факты);**
- **процедурные;**
- **метазнания** («знания о свойствах знаний»), которые содержат общие сведения о принципах использования знаний.

К уровню метазнаний также относят стратегии управления выбором и применением процедурных знаний.

КЛАССИФИКАЦИИ ЗНАНИЙ



В зависимости от степени достоверности выделяют следующие категории знаний:

- **знания, имеющие определенную достоверность;**
- **знания с нечеткой степенью достоверности.**

В зависимости от глубины выделяют следующие виды знаний:

- **знания-копии;**
- **знания-знакомства;**
- **умения;**
- **навыки.**

НЕ-ФАКТОРЫ ЗНАНИЙ

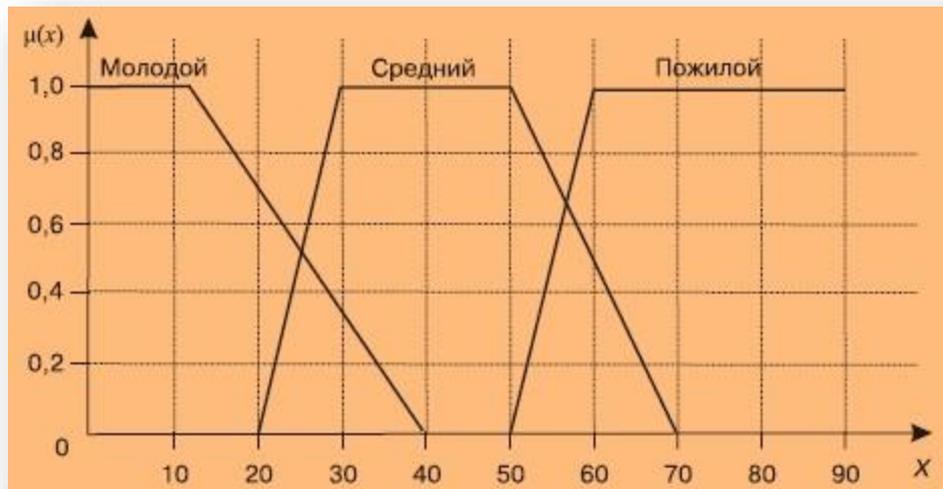


В основе деления знаний в зависимости от степени их достоверности лежат так называемые **«не-факторы»**, присущие знаниям:

- **неполнота информации о рассматриваемом фрагменте Про;**
- **неточность количественных оценок;**
- **размытость качественных оценок;**
- **неоднозначность ряда правил вывода новых знаний;**
- **несогласованность некоторых положений в БЗ;**
- **противоречивость.**

НЕ-ФАКТОРЫ ЗНАНИЙ

Один из способов учета подобных не-факторов при формализации знаний состоит в использовании **аппарата теории нечетких множеств**.



НЕ-ФАКТОРЫ ЗНАНИЙ



К классу процедурных знаний с нечеткой степенью достоверности относятся **эвристики, описывающие приемы решения задач, базирующиеся на опыте экспертов в данной Про.**

С точки зрения меры возможной формализации различают **три группы эвристических методов:**

- полностью формализованные — **алгоритмы;**
- неформализованные на данном уровне развития науки — **эврисмы;**
- частично формализованные, частично неформализованные — **эвроритмы.**

Концептуальные свойства знаний



- 1. Внутренняя интерпретация**
- 2. Наличие внутренней структуры связей**
- 3. Наличие внешней структуры связей**
- 4. Шкалирование**
- 5. Погружение в пространство с семантической метрикой**
- 6. Наличие активности**

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ (МПЗ) ДЛЯ ИС



Среди МПЗ можно выделить следующие основные модели:

- **логические;**
- **продукционные (основанные на правилах);**
- **фреймовые;**
- **сетевые;**
- **объектно-ориентированные;**
- **специальные;**
- **комплексные.**

ЛОГИЧЕСКИЕ МОДЕЛИ ЗНАНИЙ



В **логических моделях** знания представляются в виде совокупности правильно построенных формул какой-либо **формальной системы (ФС)**.

Простейшей логической моделью является **исчисление высказываний**. Развитие логики высказываний нашло отражение в **исчислении предикатов первого порядка**. Представление знаний в рамках логики предикатов служит основой **логического программирования**.

Формальная система (ФС) задается четверкой (T, P, A, R) , где T — множество базовых (терминальных) элементов, из которых формируются все выражения ФС; P — множество синтаксических правил, определяющих синтаксически правильные выражения из терминальных элементов ФС; A — множество аксиом ФС, соответствующих синтаксически правильным выражениям, которые в рамках данной ФС априорно считаются истинными; R — конечное множество правил вывода, позволяющих получать из одних синтаксически правильных выражений другие.

ЛОГИЧЕСКИЕ МОДЕЛИ ЗНАНИЙ



Положительные черты логических моделей знаний:

- высокий уровень формализации;
- согласованность знаний как единого целого;
- единые средства описания как знаний о ПрО, так и способов решения задач в этой ПрО.

Недостатки логических моделей:

- представление знаний в таких моделях ненаглядно;
- ограничения исчисления предикатов первого порядка не допускают квантификации предикатов и использования их в качестве переменных;
- описание знаний в виде логических формул не позволяет проявиться преимуществам, которые имеются при автоматизированной обработке структур данных.

Пути повышения эффективности логических моделей знаний



Пути повышения эффективности логических моделей знаний связаны с использованием **многоуровневых и специальных логик**.

Модели знаний с открытыми БЗ и немонотонными механизмами выводов основываются на понятии **расширенной ФС (семиотической системы)**, задаваемой кортежем $(T, P, A, R, mT, mP, mA, mR)$, где T, P, A, R – составляющие замкнутой ФС; mT – правила изменения базовых элементов ФС; mP – правила изменения синтаксиса ФС; mA – правила изменения аксиом ФС; mR – правила изменения правил вывода ФС.

Семиотическая система может содержать **противоречивые и несогласованные** сведения, так как они соотносятся с разными Про.

ПРОДУКЦИОННАЯ МОДЕЛЬ



Центральным звеном **продукционной модели** является **множество продукций** или **правил вывода**.

Каждая такая продукция в общем виде может быть представлена выражением: $(W_i, U_i, P_i, A_i \rightarrow B_i, C_i)$, где W_i — сфера применения i -й продукции; U_i — предусловие i -й продукции; P_i — условие i -й продукции, определяемое факторами, непосредственно не входящими в A_i ; $A_i \rightarrow B_i$ — ядро i -й продукции, соответствующее правилу «если..., то...»; C_i — постусловие i -й продукции.

Системы, основанные на продукционной модели, состоят из **трех типовых компонентов**:

- **базы правил (продукций);**
- **базы фактов, содержащей декларативные знания о ПрО;**
- **интерпретатора продукций.**

Существуют **два типа механизмов вывода** в продукционных системах: **прямой** и **обратный вывод**.

ПРОДУКЦИОННАЯ МОДЕЛЬ



Положительные стороны:

- ясность и наглядность интерпретации отдельных правил;
- простота механизмов вывода и модификации БЗ.

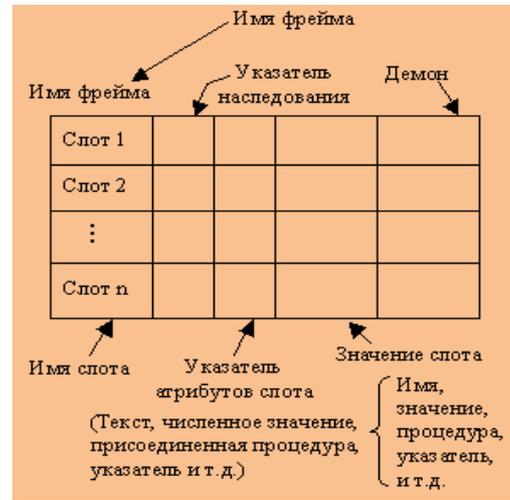
Недостатками продукционной модели являются:

- сложность управления выводом, неоднозначность выбора конкурирующих правил;
- низкая эффективность вывода в целом, негибкость механизмов вывода;
- неоднозначность учета взаимосвязи отдельных продукций;
- несоответствие психологическим аспектам представления и обработки знаний человеком;
- сложность оценки целостного представления ПрО.

ФРЕЙМОВАЯ МОДЕЛЬ

Фундаментом **фреймовой модели знаний** служит понятие **фрейма** (теория фреймов М. Минского) — структуры данных, представляющей некоторый концептуальный объект или типовую ситуацию.

Фрейм – форма описания знаний, очерчивающая рамки рассматриваемого фрагмента Про.

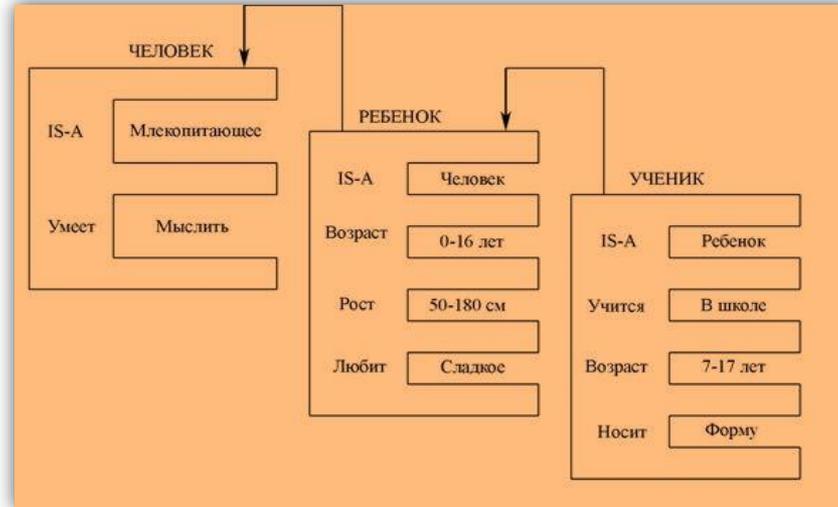


Структура фрейма

ФРЕЙМОВАЯ МОДЕЛЬ

Организация вывода во фреймовой системе базируется на **обмене сообщениями между фреймами, активации и выполнении присоединенных процедур.**

Реализация фреймовой модели знаний базируется на **языках линии *LISP, FRL, KRL.***



Пример фрейма с описанием «УЧЕНИКА»

ФРЕЙМОВАЯ МОДЕЛЬ



Положительными чертами в целом являются:

- наглядность;
- гибкость;
- однородность;
- высокая степень структуризации знаний;
- соответствие принципам представления знаний человеком в долговременной памяти;
- интеграция декларативных и процедурных знаний.

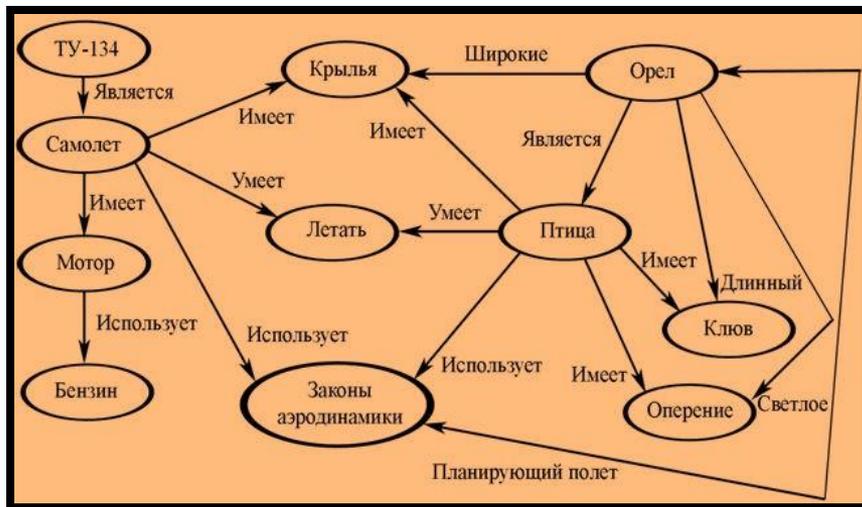
Недостатки фреймовой модели:

- сложность управления выводом;
- низкая эффективность его процедур.

СЕТЕВЫЕ МОДЕЛИ

Наиболее общий способ представления знаний, при котором ПрО рассматривается как совокупность объектов и связывающих их отношений, реализован в **сетевой модели знаний**.

В качестве носителя знаний в этой модели выступает **семантическая сеть (СС)**, вершины которой соответствуют объектам (понятиям), а дуги — отношениям между понятиями.



Фрагмент СС

СЕТЕВЫЕ МОДЕЛИ



В общем случае под СС понимается структура: $S=(O, R) = (\{o_i | i= 1,2,\dots,k\}, \{r_j | j= 1,2,\dots,l\})$, где O – множество объектов ПрО ($|O|=k$); R – множество отношений между объектами ПрО ($|R|=l$); o_i – i -й объект ПрО; r_j – j -е отношение между объектами ПрО.

Типизация семантических сетей обуславливается смысловым содержанием образующих их отношений.

Например, если дуги сети выражают **родовидовые отношения**, то такая сеть определяет **классификацию объектов ПрО**. Аналогично, наличие в сети **причинно-следственных (каузальных) отношений** позволяет интерпретировать ее как **сценарий**.

СЕТЕВЫЕ МОДЕЛИ



Очевидные достоинства сетевой модели:

- высокая общность;
- наглядность отображения системы знаний о ПрО;
- легкость понимания подобного представления.

Недостатки сетевой модели:

- в семантической сети имеет место смешение групп знаний, что усложняет интерпретацию знаний;
- трудность унификации процедур вывода и механизмов управления выводами на сети.

Семантические сети



Наиболее известные отечественные **модели СС.**

- 1. Расширенные СС И.П. Кузнецова;**
- 2. Неоднородные СС Г.С. Осипова;**
- 3. Нечеткие СС И.А. Перминова;**
- 4. Обобщенная модель представления знаний о Про А.И. Башмакова.**

Объектно-ориентированная модель знаний



Объектно-ориентированная модель знаний получила широкое применение в современных технологиях проектирования разнообразных программных и информационных систем.

В настоящее время существуют два основных подхода к моделированию знаний, базирующихся на объектной парадигме:

- четырехуровневая модель *MDA (Model Driven Architecture)* консорциума *Object Management Group (OMG)*;
- трехуровневая модель *ODP (Model of Open Distributed Processing)*, зафиксированная в стандарте *ISO/ITU*.

Объектно-ориентированная модель знаний



Стандарт *ODP* определяет пять возможных видов представлений:

- корпоративное (*enterprise*);
- информационное (*information*);
- вычислительное (*computational*);
- разработчиков (*engineering*);
- технологическое (*technology*).

Сравнение данных подходов показывает, что они **позволяют моделировать одну и ту же Про** с разных точек зрения. В *MDA* модели описывают Про непосредственно в отличие от *ODP*.

Специальные и комплексные модели знаний



Класс специальных моделей знаний объединяет модели, отражающие особенности представления знаний и решения задач в отдельных, относительно узких ПрО.

Смешанные или комплексные модели, интегрирующие преимущества рассмотренных выше базовых моделей представления ПрО.

Ключевые требования к моделям знаний



- 1) общность (универсальность);
- 2) «психологичность», наглядность представления знаний;
- 3) однородность;
- 4) реализация в модели свойства активности знаний;
- 5) открытость БЗ;
- 6) возможность отражения в БЗ структурных отношений объектов ПрО;
- 7) наличие механизма «проецирования» знаний на систему семантических шкал;
- 8) возможность оперирования нечеткими знаниями;
- 9) использование многоуровневых представлений (данные, модели, метамодел, метаметамодел и т.д.).

Ни одна из моделей знаний не удовлетворяет всем девяти требованиям.

К представлению знаний как направлению ИИ традиционно относят задачи **верификации знаний**, **пополнения БЗ** за счет логического вывода, **обобщения** и **классификации знаний** (систематизация знаний).

Онтологический подход. Понятие онтологии



Онтологии имеют непосредственное отношение к построению БЗ и частично к реализации интеллектуального интерфейса.

Они помогают обеспечить одинаковое понимание всеми пользователями смысла применяемых при решении терминов, их атрибутов и отношений между ними.

Термин «**онтология**» в ИИ употребляется в контексте с такими понятиями, как концептуализация, знания, модели знаний, системы, основанные на знаниях.

Онтология — это формально представленные на базе концептуализации знания о ПрО.

Под **концептуализацией** понимается процесс перехода от представления ПрО на ОЕЯ (или ЕЯ) к точной спецификации этого описания на некотором формальном языке, ориентированном на компьютерное представление.

Концептуализация также трактуется как результат подобного процесса, т.е. описание множества понятий (концептов) ПрО, знаний о них и связях (отношениях) между ними.

Самым распространенным на данный момент является определение, согласно которому **онтология есть точная (выраженная формальными средствами) спецификация концептуализации.**

Онтологический подход



Онтология является не абсолютной (единственной) спецификацией концептуализации Про, а зависит от целей ее создания. Независимо от вида **онтологии** она должна включать **словарь терминов** и некоторые **спецификации их значений**. При таком подходе **онтология** похожа на **тезаурус**.

Интерпретации понятия «онтология»:

- 1. неформальная концептуальная система (представление концептуализации);**
- 2. формальный взгляд на семантику;**
- 3. спецификация концептуализации;**
- 4. представление концептуальной системы через логическую теорию;**
- 5. словарь, используемый логической теорией;**
- 6. метауровневая спецификация логической теории.**

Онтологический подход



В неформальной трактовке онтология представляет собой описание некоторой ПрО.

На формальном уровне **онтология** — это система, состоящая из набора понятий и набора утверждений об этих понятиях, на основе которых можно строить классы, объекты, отношения, функции и теории.

На метауровне **онтология** является **разновидностью сетевой модели знаний о ПрО**.

Эта модель может быть **статической** или **динамической**. Во втором случае говорят об онтологии как о модели мира, которая может представлять состояния моделируемой ПрО во времени.

Web-онтологией называют онтологию, которая либо доступна на одном из web-узлов Internet, либо используется в рамках корпоративного портала.

ОСНОВНЫЕ ЗАДАЧИ, РЕШАЕМЫЕ С ПОМОЩЬЮ ОНТОЛОГИЙ



- 1. Создание и использование БЗ.**
- 2. Организация эффективного поиска в БД, информационных каталогах, БЗ.**
- 3. Создание систем, реализующих механизмы рассуждений (ЭС, системы управления, интеллектуальные роботы).**
- 4. Организация поиска по смыслу в текстовой информации.**
- 5. Семантический поиск в Internet.**
- 5. Представление смысла в метаданных об ИР.**
- 6. Построение и использование баз общих знаний для различных ИС.**
- 7. Обеспечение общей терминологии для множества специалистов и совместно используемых приложений.**
- 8. Многократное применение БЗ и информационных массивов, представляющих сведения о технических системах на различных стадиях их жизненного цикла.**

МОДЕЛЬ ОНТОЛОГИИ



Модель онтологии O задает тройка $O = (X, R, \Phi)$, где

X — конечное непустое множество концептов ПрО, которую представляет онтология;

R — конечное множество отношений между концептами;

Φ — конечное множество функций интерпретации, заданных на концептах и (или) отношениях.

Эта модель является разновидностью сетевой модели знаний. В начале развития онтологического подхода для представления онтологии использовались языки ИППП и другие языки математической логики. Сейчас применяются эквивалентные по семантике языки нелогического типа, специально предназначенные для описания онтологии.

МОДЕЛЬ ОНТОЛОГИИ



Направления обобщения частных случаев моделей онтологии:

- представление множества концептов в виде сетевой структуры;
- представление в R отношений, отражающих специфику конкретной ПрО, а также средств расширения R ;
- использование декларативных и процедурных представлений элементов Φ и R , включая возможность определения новых интерпретаций.

Для спецификации пространств знаний, охватывающих несколько взаимосвязанных ПрО, предложена модель расширенной онтологии:

$$O_{\text{расш}} = (O_m, \{(O_p, O_z)\}, MB),$$

где O_m — онтология верхнего уровня (метаонтология); O_p — предметная онтология; O_z — онтология задач ПрО; MB — модель машины вывода.

МОДЕЛЬ ОНТОЛОГИИ



Метаонтология O_m оперирует, не зависящими от ПрО концептами и отношениями (время, объект, свойство и т.д.). Она изменяется весьма незначительно и может считаться статической. Поэтому вывод на ее уровне достаточно эффективен.

Предметная онтология O_p содержит понятия, описывающие конкретную ПрО, семантически значимые для нее отношения, а также декларативные и процедурные интерпретации этих понятий и отношений.

В **онтологии задач O_z** в качестве понятий выступают типы решаемых задач, а отношения, как правило, специфицируют декомпозицию задач на подзадачи.

Машина вывода начинает работу при активации понятий или отношений, описывающих исходную ситуацию (задачу). Вывод на СС организуется как волновой процесс, использующий свойства отношений, выходящих из узлов, задающих исходную ситуацию.

Критерием останова процесса является достижение целевой ситуации, либо превышение длительности времени, отведенного для решения задачи.

МЕТОДИКИ ПОСТРОЕНИЯ ОНТОЛОГИИ И ТРЕБОВАНИЯ К СРЕДСТВАМ ИХ СПЕЦИФИКАЦИИ



Существует много предложений по методикам разработки онтологии.

Стандарт онтологического исследования IDEF5 подготовленный фирмой Knowledge Base Systems, Inc. в качестве проекта национального стандарта США (1994).

Процесс построения онтологии в рамках IDEF5 состоит из **пяти основных этапов**:

- 1. Изучение и систематизация начальных условий.** Этот этап устанавливает основные цели и контекст разработки онтологии, а также распределяет роли членов проекта.
- 2. Сбор и накопление данных для построения онтологии.**
- 3. Анализ и группировка собранных данных** для облегчения согласования терминологии.
- 4. Начальное развитие онтологии.** На этом этапе формируется предварительная онтология на основе систематизированных данных.
- 5. Уточнение и утверждение онтологии.**

Онтологии в IDEF5



Для поддержки процесса построения **онтологии в IDEF5** определены **специальные онтологические языки**:

- **схематический язык (Schematic Language — SL);**
- **язык доработок и уточнений (Elaboration Language — EL).**

Язык SL является наглядным графическим языком диаграммного типа, предназначенным для формирования начального представления онтологии, а также дополнения существующих онтологий.

Язык EL – структурированный текстовый язык, позволяющий детализировать элементы онтологии (структурировать элементы концептуализации).

Онтологии в IDEF5



В стандарте IDEF5 предусмотрены четыре вида схем, предназначенных для представления онтологической информации в наглядной графической форме:

Диаграммы классификации служат средством логической систематизации знаний, накопленных при изучении системы.

Существует два типа таких диаграмм:

- **диаграмма строгой классификации (Description Subsumption — DS);**
- **диаграмма естественной или видовой классификации (Natural Kind Classification — НКС).**

В DS определяющие свойства класса являются необходимым и достаточным признаком принадлежности объекта этому классу. С помощью диаграмм DS, как правило, классифицируются логические объекты.

В НКС, наоборот, свойства класса не являются необходимым и достаточным признаком принадлежности ему тех или иных объектов. В диаграммах этого типа интерпретация свойств класса является более общей.

Онтологии в IDEF5



Композиционные схемы (Composition Schematics) служат для графического представления состава классов онтологии. В частности, с помощью них можно наглядно отобразить состав объектов, относящихся к тому или иному классу.

Схемы взаимосвязей (Relation Schematics) позволяют визуализировать и изучать связи между различными классами объектов системы, а также представлять зависимости между взаимосвязями классов.

Диаграмма состояния объекта (Object State Schematic) позволяет описать процесс изменения состояния объекта (смена состояния или смена класса).

Стандарт IDEF5 отражает методологию, с помощью которой можно наглядно и эффективно разрабатывать онтологии, но данный стандарт охватывает не все этапы создания онтологии.

Инструментальная среда проектирования онтологии ODE



Существует еще одна методология построения онтологии. Для ее поддержки предназначена специальная **инструментальная среда проектирования онтологии (Ontology Design Environment – ODE)**.

Она включает **подсистемы управления проектом и поддержки разработки**. **Первая подсистема** обеспечивает решение задач планирования, контроля за ходом выполнения проекта и управления качеством.

Вторая ориентирована на задачи приобретения знаний, их оценки, интеграции, документирования и управления конфигурациями.

Процесс разработки онтологии включает четыре стадии:

- 1) спецификация;
- 2) концептуализация;
- 3) формализация;
- 4) реализация.

Инструментальная среда проектирования онтологии ODE



Наиболее сложной задачей является концептуализация. От успешности ее выполнения зависит эффективность всей разработки.

Концептуализация включает два этапа:

- построение глоссария терминов;
- построение классификационных деревьев концептов.

Вторая задача начинает решаться тогда, когда объем глоссария по мнению экспертов достигает существенного объема. Затем для каждого классификационного дерева формируются словарь концептов и совокупность таблиц, описывающих бинарные отношения между концептами, экземпляры, атрибуты экземпляров и классов, логические аксиомы, константы и формулы.

Существует много программных решений для создания ИС на основе онтологического подхода, которые поддерживают различные форматы и языки онтологического проектирования: **OWL, RDF (S), RIF, ObjectLogic**.