

ОНТОЛОГИИ:

формальное

и

неформальное

Валькман Ю. Р.

yur@valkman.kiev.ua

СОДЕРЖАНИЕ

- 1. Введение и некоторые определения**
- 2. Классификации онтологий**
- 3. Онтологии для обработки текстов на ЕЯ**
- 4. Об эволюции структур данных и знаний**
- 5. Применение онтологий**
- 6. Примеры**
- 7. Разработка онтологий 101**
- 8. Анализ формальных понятий**
- 9. Модели представления знаний в БЗ**

«Онтология - философское учение о бытии, его основах, принципах, структуре и закономерностях.

В этом определении еще не нашел отражения тот факт, что в начале 90-х годов этот термин был заново "осмыслен" ИИ. Для данной области такая склонность к постоянному заимствованию ярких этикеток очень типична.

С одной стороны, она обусловлена генетически, поскольку началась с выбора не самого скромного названия для самой области.

С другой, это постоянное подбадривание себя и контекста ИИ частым введением броских "лейблов", переименовывающих нечто начинающее приедаться специалистам и публике из-за слишком медленного прогресса по существу.

Что-то вроде очередного перекрашивания того же автомобиля.

В этом есть свой резон: в случае удачи автомобиль действительно начинает смотреться по-новому.

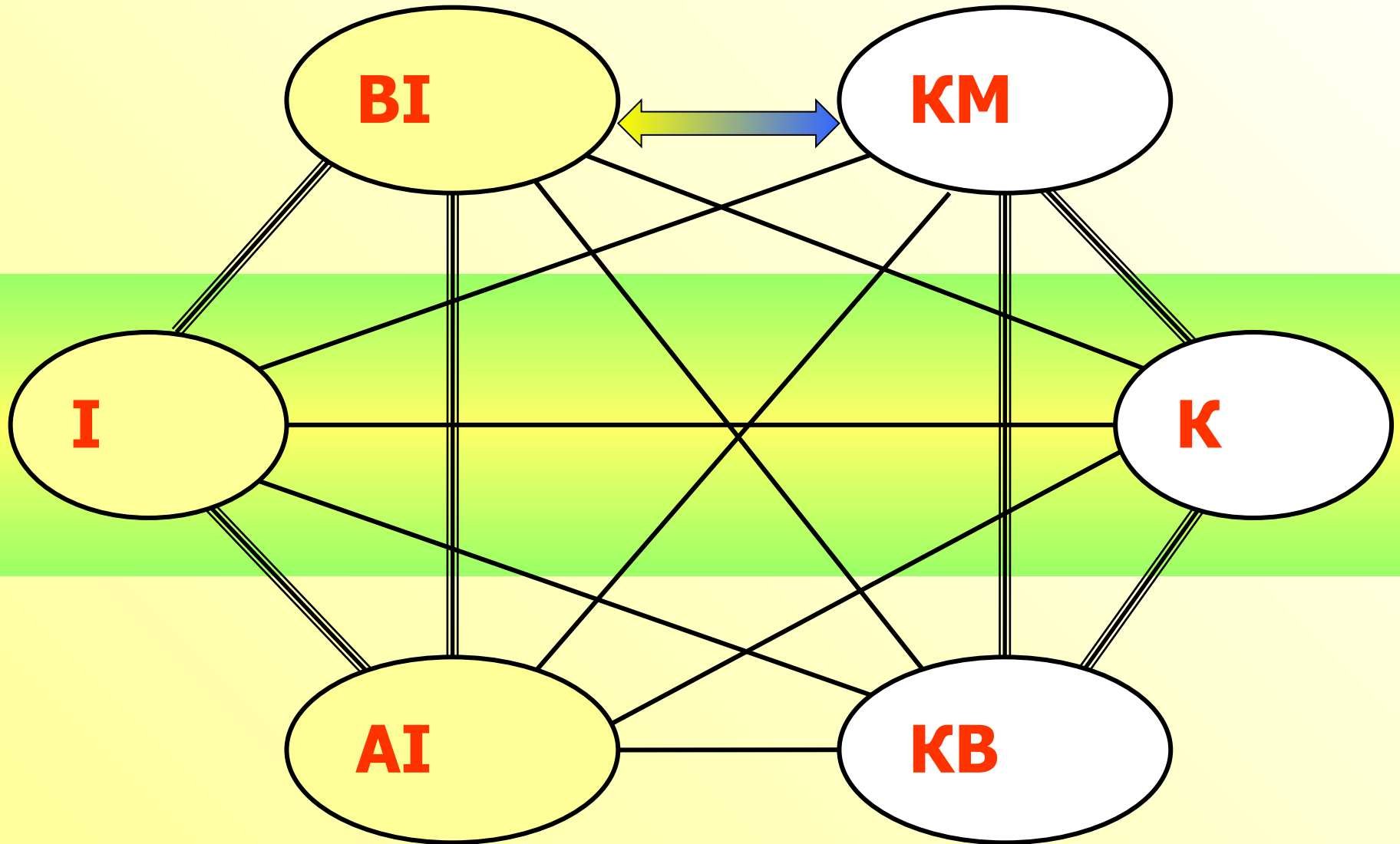
Итак, ИИ решил расширить горизонты и поднять до "основ бытия" уровень своего направления, занимавшегося **формальными моделями (описаниями) знаний о какой-либо области, переименовав их в Онтологии.»**

КЕНТАВР ПО ИМЕНИ ТЕОН:ТЕЗАУРУС+ОНТОЛОГИЯ

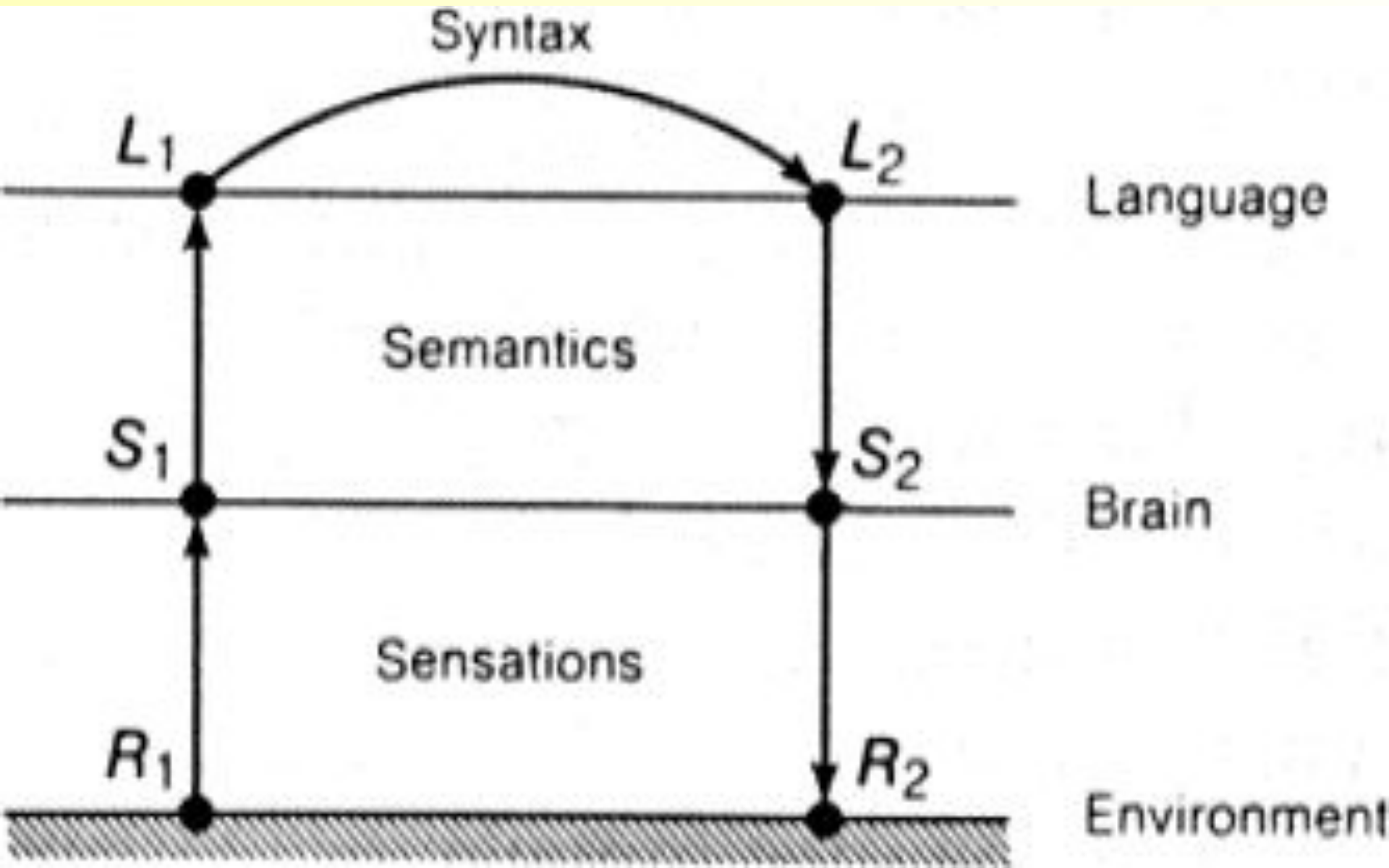
А.С. Нариньяни

Отношения «ВІ - КМ»

в контексте «интеллектуальности»



Двухэтажная языковая модель действительности В. Турчина



В зависимости от того, какого типа язык используется, можно говорить о неформальном и формальном мышлении.

При **неформальном мышлении** языковые объекты важны главным образом постольку, поскольку они вызывают у нас определенные комплексы представлений. Слова здесь — веревочки, дергая за которые мы извлекаем из памяти частицы нашего жизненного опыта, переживаем их вновь, сопоставляем, сортируем и т. п. Результатом этой внутренней работы является преобразование представлений **S1** → **S2**, которое моделирует переменны в окружающей среде **R1** → **R2**.

Это не значит, однако, что неформальное мышление тождественно с безъязыковым.

Во-первых, уже само расчленение потока ощущений зависит от системы понятий, фиксированной в языке.

Во-вторых, и в процессе преобразования $S1 \rightarrow S2$ «натуральный вид» языкового объекта — слово — играет немаловажную роль: мы очень часто пользуемся ассоциацией именно между словами, а не представлениями. Поэтому формулу неформального мышления можно изобразить в виде $(S1, L1) \rightarrow (S2, L2)$.

При формальном мышлении мы оперируем с языковыми объектами как с некоторыми самостоятельными и самодовлеющими сущностями, забывая на время об их семантике и вспоминая о ней лишь тогда, когда надо интерпретировать полученный результат или уточнить исходные посылки. Формула формального мышления такова:

$$S1 \rightarrow L1 \rightarrow L2 \rightarrow S2.$$

Для того чтобы формальное мышление приводило к правильным результатам, семантика языка должна обладать определенными свойствами, которые мы характеризуем такими терминами, как *точность, определенность, однозначность.*

Об онтологиях

Экспликация концептуализации, специальная БЗ, семантическая сеть (в Дубне) и т. д.???

Ответ Мизогучи: Позвольте мне процитировать фразу, найденную в моем архиве электронной почты по онтологии:

- (1)** Представляет ли **нечто** - консенсус-знание некоторого сообщества людей?
- (2)** Используют ли **его** при ссылках на точно определенные термины?
- (3)** Представляет ли **оно** консенсус-знание некоторого сообщества агентов?
- (4)** Является ли **используемый язык** достаточно выразительным для людей, чтобы они могли сказать то, что они хотят?

(5) Может ли **оно** быть использовано повторно в нескольких эпизодах решения проблем?

(6) Является ли **оно** стабильным?

(7) Может ли **оно** использоваться для решения множества проблем различного сорта?

(8) Может ли **оно** служить отправной точкой для конструирования множественных (в некотором смысле) приложений, включая новую базу знаний, схему базы данных, объектно-ориентированную программу?

Чем увереннее звучит ответ «да» на эти вопросы, тем это НЕЧТО «онтологичнее».

Онтология (артефакт)

- Неформально, онтология представляет собой некоторое описание взгляда на мир применительно к конкретной области интересов.
- Это описание состоит из терминов и правил использования этих терминов, ограничивающих их значения в рамках конкретной области

**Типичная - одна из многих -характеристика
читается так:**

**Онтология - это набор определений (на
формальном языке) фрагмента
декларативных знаний, ориентированный
на совместное многократное
использование различными
пользователями в своих приложениях.**

**В онтологии вводятся термины, типы и
соотношения (аксиомы), описывающие фрагмент
знания.**

Одно из самых известных определений онтологии дал Том Грубер, звучит оно следующим образом: Онтология — это точная спецификация концептуализации.

Концептуализация — это структура реальности, рассматриваемая независимо от словаря предметной области и конкретной ситуации.

Например, если мы рассматриваем простую предметную область, описывающую кубики на столе, то концептуализацией является набор возможных положений кубиков, а не конкретное их расположение в текущий момент времени.

Более поздней модификацией определения Грубера является такое определение:

Онтология — это формальная спецификация согласованной концептуализации.

Под согласованной концептуализацией подразумевается, что данная концептуализация не есть частное мнение, а является общей для некоторой группы людей.

Сформулировано еще достаточно много разных определений онтологии.

Например, Никола Гуарино определяет онтологию следующим образом: **Онтология** — это формальная теория, ограничивающая возможные концептуализации мира.

Некоторые определения отражают способы, которыми авторы строят и используют онтологии, например: **Онтология** — это иерархически структурированное множество терминов, описывающих предметную область, которое может быть использовано как исходная структура для базы знаний.

НЕФОРМАЛЬНО ОНТОЛОГИЯ представляет

собой некоторое описание взгляда на мир применительно к конкретной области интересов.

Это описание состоит из терминов и правил использования этих терминов, ограничивающих их значения в рамках конкретной области.

НА ФОРМАЛЬНОМ УРОВНЕ ОНТОЛОГИЯ — это система, состоящая из набора понятий и набора утверждений об этих понятиях, на основе которых можно описывать классы, отношения, функции и индивиды.

Онтологии связывают два существенных аспекта:

определяют формальную семантику информации, позволяя обработку информации компьютером



определяют семантику реального мира, позволяя, основываясь на общей терминологии, связывать информацию, представленную в виде, требуемом для компьютерной обработки, с информацией, представленной в удобной для восприятия человеком форме



Основной областью применения онтологии является интеграция информации

Онтология в модели Ontolingua определяется кортежем:

$\langle C, R, I, A \rangle,$

где C – множество классов как унарных отношений над объектами,

$R: C \times \dots \times C$ – множество отношений, введённых как множества кортежей объектов, и функций как особой разновидности отношений,

I – множество индивидов,

A – утверждения, определяющие классы и отношения.

Определения включают вербальные описания объектов и формальные описания на языке **KIF** (Knowledge Interchange Format). В связи с выразительностью языка **KIF**, основанного на логике первого порядка, спецификации на языке Ontolingua не поддаются автоматическому логическому выводу для многих задач. Разработка онтологий в этой модели подчиняется специальной технике дополнения непротиворечивых спецификаций.

Онтологии в моделях, основанных на дескриптивной логике, составляют кортеж

$$\langle C, R, H_C, H_R, I, A \rangle,$$

где C – множество классов,

$R: C \times C$ – множество свойств классов,

$H_C: C \times C$ – иерархия классов,

$H_R: R \times R$ – иерархия отношений,

I – множество индивидов,

A – множество аксиом над классами и ограничений отношений.

В некоторых моделях могут отсутствовать H_R и/или I , а A составляют только конструкции определённых видов, предусмотренные в данной модели. Этот набор определяет сложность вывода в задачах непротиворечивости и поглощения спецификаций.

$$O = \{C, R, A\}$$

где **O** — онтология,

C — совокупность концептов предметной области,

R — совокупность отношений между ними,

A — набор аксиом (законов и правил, которые описывают законы и принципы существования концептов).

По глубине проработки все онтологии делятся на:

«весомые» онтологии (Heavy-weighted),

содержащие аксиомы $\{C, R, A\}$

«легкие» (Light-weighted), их не

содержащие $\{C, R\}$

«Соответствие между сообщением и информацией (информация передается посредством сообщения) не является взаимно-однозначным. Решающим для связи между *сообщением S* и *информацией I* является некоторое отображение *int* (правило интерпретации):

$$S \xrightarrow{\text{int}} I$$

Пусть $\{t_1, \dots, t_n\}$ – множество терминов, а V – множество их возможных значений.

Будем называть **вербальным представлением информации** такое сообщение, которое имеет вид некоторого отображения множества терминов $\{t_1, \dots, t_n\}$ в множество значений V .

Смысл терминов t_1, \dots, t_n может быть представлен **концептуализацией Conc** таких вербальных представлений информации, которые имеют интерпретацию.

Онтология есть явное описание (на некотором языке L) смысла терминов, неявно определенных концептуализацией *Conc*:

$$O = \langle \text{Conc}, L \rangle.$$

Е.А. Шалфеева

Институт автоматизации и процессов управления ДВО РАН, г. Владивосток, Россия

Классификация структурных свойств онтологий

В проектировании онтологии условно можно выделить два направления, до некоторого времени развивавшихся отдельно.

Первое связано с представлением онтологии как формальной системы, основанной на математически точных аксиомах.

Второе направление развивалось в рамках компьютерной лингвистики и когнитивной науки.

Там онтология понималась как система абстрактных понятий, существующих только в сознании человека, которая может быть выражена на ЕЯ (или средствами какой-то другой системы символов). При этом обычно не делается предположений о точности или непротиворечивости такой системы.

Таким образом, существует два альтернативных подхода к созданию и исследованию онтологий.

Первый (формальный) основан на логике (предикатов первого порядка, дескриптивной, модальной и т.п.).

Второй (лингвистический) основан на изучении естественного языка (в частности, семантики) и построении онтологий на больших текстовых массивах, так называемых корпусах.

В настоящее время данные подходы тесно взаимодействуют. Идет поиск связей, позволяющих комбинировать соответствующие методы.

Поэтому иногда бывает сложно отделить лексические онтологии с элементами формальных аксиоматик от логических систем с включениями лингвистических знаний.

Независимо от различных подходов можно выделить 3 основных принципа классификации онтологий:

- по степени формальности;**
- по наполнению, содержимому;**
- по цели создания.**

2. КЛАССИФИКАЦИИ ОНТОЛОГИЙ

Спектр онтологий. Косая черта разделяет системы, предоставляющие "человеко-понятные" (выше черты) и "машино-понятные" (ниже черты) описания



В целом с необходимостью описывать сложные факты выразительные средства онтологии (и ее структура) усложняются. Например, может потребоваться заполнить значение какого-либо свойства экземпляра, используя математическое выражение, основанное на значениях других свойств данного экземпляра или значениях свойств других экземпляров.

Многие онтологии позволяют объявлять два и более класса дизъюнктивными (непересекающимися). Это означает, что у данных классов не существует общих экземпляров.

Некоторые языки описания онтологий позволяют делать произвольные логические утверждения о концептах — аксиомы.

Языки описания онтологий, такие как СусL и Ontolingua, позволяют фиксировать утверждения на языке логики предикатов первого порядка (FOL)

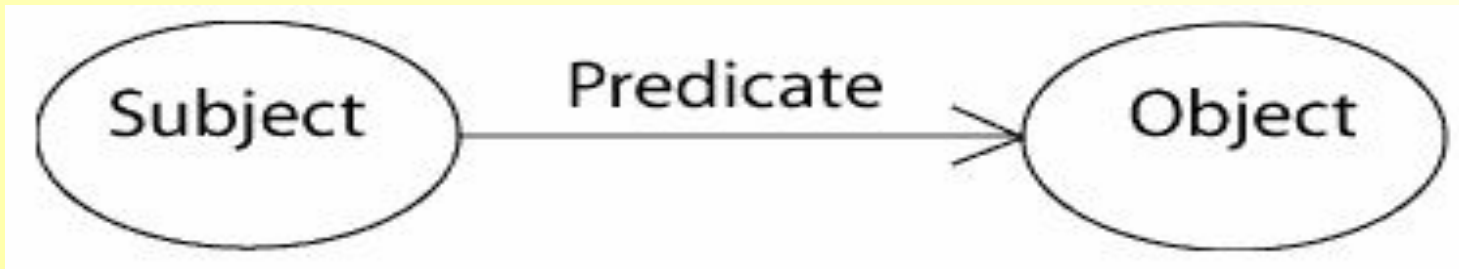
Классификация по цели создания



Онтология представления

Цель ее создания — описать область представления знаний, создать язык для спецификации других онтологий более низких уровней. Пример: описание понятий языка **OWL (Web Ontology Language)** средствами **RDF/RDFS**. В данном описании определяются такие понятия, как "класс", "отношение", "ограничение на значение свойства", "домен", "диапазон" и т.п.

Базовой структурной единицей RDF (*Resource Description Framework*) является коллекция троек (или триплетов), каждая из которых состоит из **субъекта, предиката и объекта (S,P,O)**. Набор триплетов называется RDF-графом. В качестве вершин графа выступают субъекты и объекты, в качестве дуг — предикаты (или свойства). Направление дуги, соответствующей предикату в данной тройке (S,P,O), всегда выбирается так, чтобы дуга вела от субъекта к объекту.



Онтология представления для языка OWL



Онтология верхнего уровня

Ее назначение — в создании единой "правильной" онтологии, фиксирующей знания, общие для нескольких предметных областей, и в многократном использовании данной онтологии.

Существует несколько крупных онтологий верхнего уровня: Сус, DOLCE, SUMO, онтология Джона Совы (J.Sowa) и другие.

Многие онтологии верхнего уровня похожи друг на друга. Они содержат одни и те же концепты:

сущность, явление, процесс, объект, роль, пространство, время, материя, событие, действие и т.п.

Верхние уровни иерархии DOLCE



Онтология предметной области

Другое название — онтология домена.

Назначение схоже с назначением онтологии верхнего уровня, но область интереса ограничена одной предметной областью (т.н. доменом), например, *авиация, медицина, культура, дистанционное обучение, Интернет-технологии*.

Онтология предметной области обобщает понятия, использующиеся в некоторых задачах домена, абстрагируясь от самих задач (так, онтология автомобилей независима от любых особенностей конкретных марок машин).

Прикладная онтология

Назначение такой онтологий в том, чтобы описать концептуальную модель конкретной задачи или приложения.

Прикладные онтологии описывают концепты, которые зависят как от онтологии задач, так и от онтологии предметной области. Примером может служить онтология для автомобилей, строительных материалов, вычислительной техники.

Такие онтологии содержат наиболее специфичную информацию.

Классификация онтологий по содержанию



Сущности, События,
Пространство, Время...



Составление расписаний,
определение целей,
классификация



Множества предметов: скальпели,
сканеры

Общие онтологии описывают наиболее общие концепты (пространство, время, материя, объект, событие, действие и т.д.), которые независимы от конкретной проблемы или области.

Онтология, ориентированная на задачу — это онтология, используемая конкретной прикладной программой и содержащая термины, которые используются при разработке ПО, выполняющего конкретную задачу. Задачи могут быть самыми разнообразными: составления расписания, определение целей, диагностика, продажа, разработка ПО, построение классификации.

Предметная онтология (или онтология предметов) описывает реальные предметы, участвующие в какой-либо деятельности (производстве). Например, онтология всех частей и компонентов самолетов определенной марки (Boeing) и сведения об их поставщиках, характеристиках, способе соединения друг с другом и т.п.

3. Онтологии для обработки текстов на ЕЯ.

Лексические онтологии

Для того чтобы применить онтологию для автоматической обработки текстов, в частности, для решения задач информационного поиска, необходимо понятиям онтологии сопоставить набор языковых выражений (слов и словосочетаний), которыми понятия могут выражаться в тексте.

Процедура сопоставления понятий онтологий и языковых выражений может быть осуществлена различными способами.

1. Во-первых, онтология может быть сделана заранее, путем логической классификации, а затем к ее элементам могут быть приписаны языковые единицы.

Так, например, Дуг Ленат (Doug Lenat), руководитель известного проекта в области представления знаний **Сус**, в рамках которого предполагалось формализовать знания здравого смысла (common sense) и использовать их, в частности, для обработки текстов на ЕЯ, считает, что учет значений слов может только запутать ("words are often red herrings"), что значения слов делят мир неоднозначно, а линии деления происходят из самых различных причин: исторических, физиологических и т.п.

Предлагается создавать онтологию путем логического анализа, "сверху-вниз". При этом имена вводимых понятий (желательно) должны отражать те признаки, которые заложены в основу деления.

В результате получаются имена понятий достаточно громоздкие, неестественные, с ними трудно оперировать как разработчикам, так и возможным пользователям.

Другая проблема такого подхода: приписывая языковые выражения логически обоснованной системе понятий, мы получаем, что одно и то же слово может соответствовать слишком большому количеству таких "правильных" понятий в зависимости от контекста, а значит, возникает излишняя многозначность лексической единицы.

Кроме того, тогда как небольшие онтологии могут быть построены методом сверху-вниз, разработка подробных онтологий для реальных приложений — задача нетривиальная.

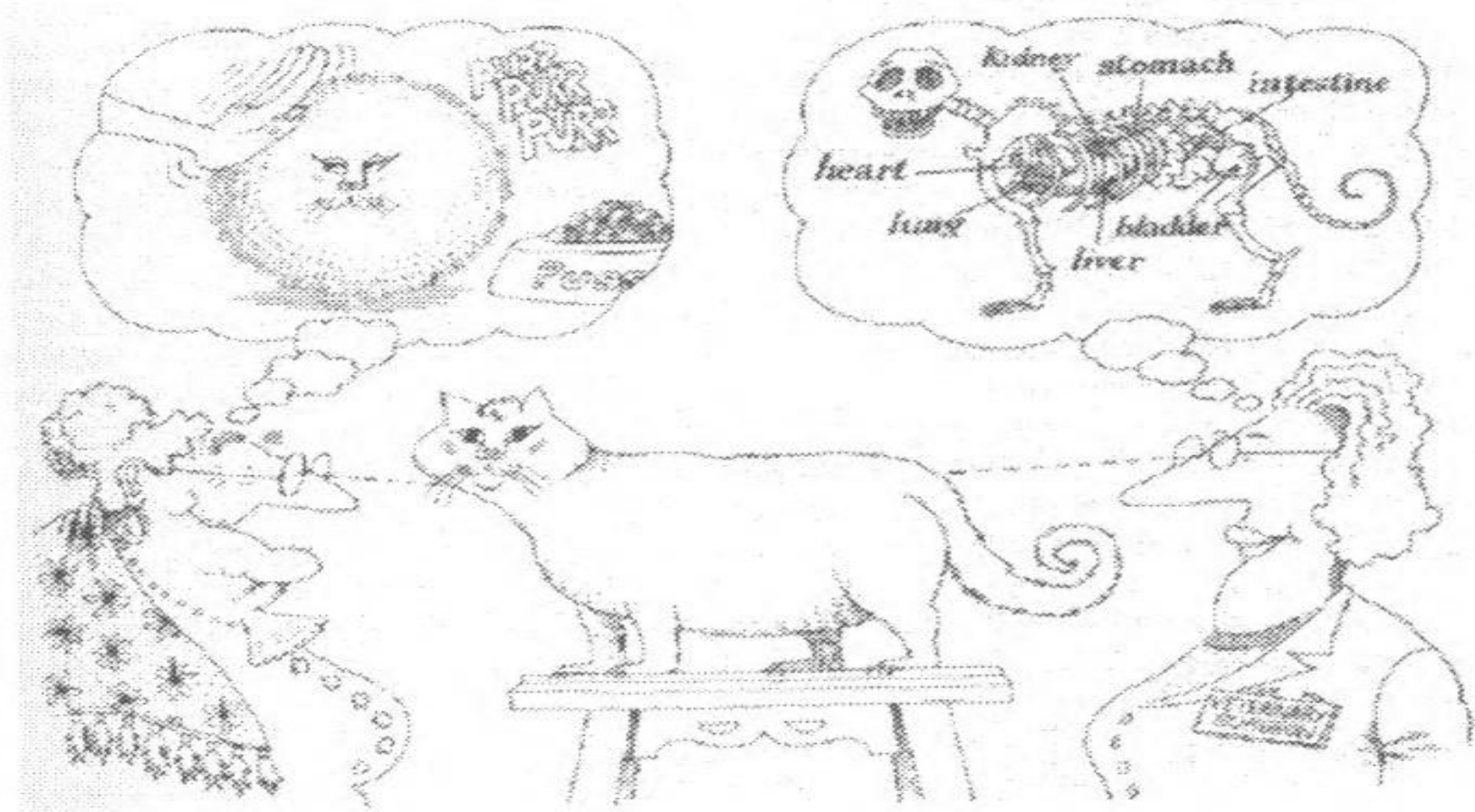
Более того, во многих предметных областях знание, нужное для распространения и интеграции, **содержится в основном в текстах.**

Из-за внутренних свойств человеческого языка непростой задачей является связать знания, содержащиеся в текстах, с онтологиями, даже если бы была построена подробная онтология предметной области.

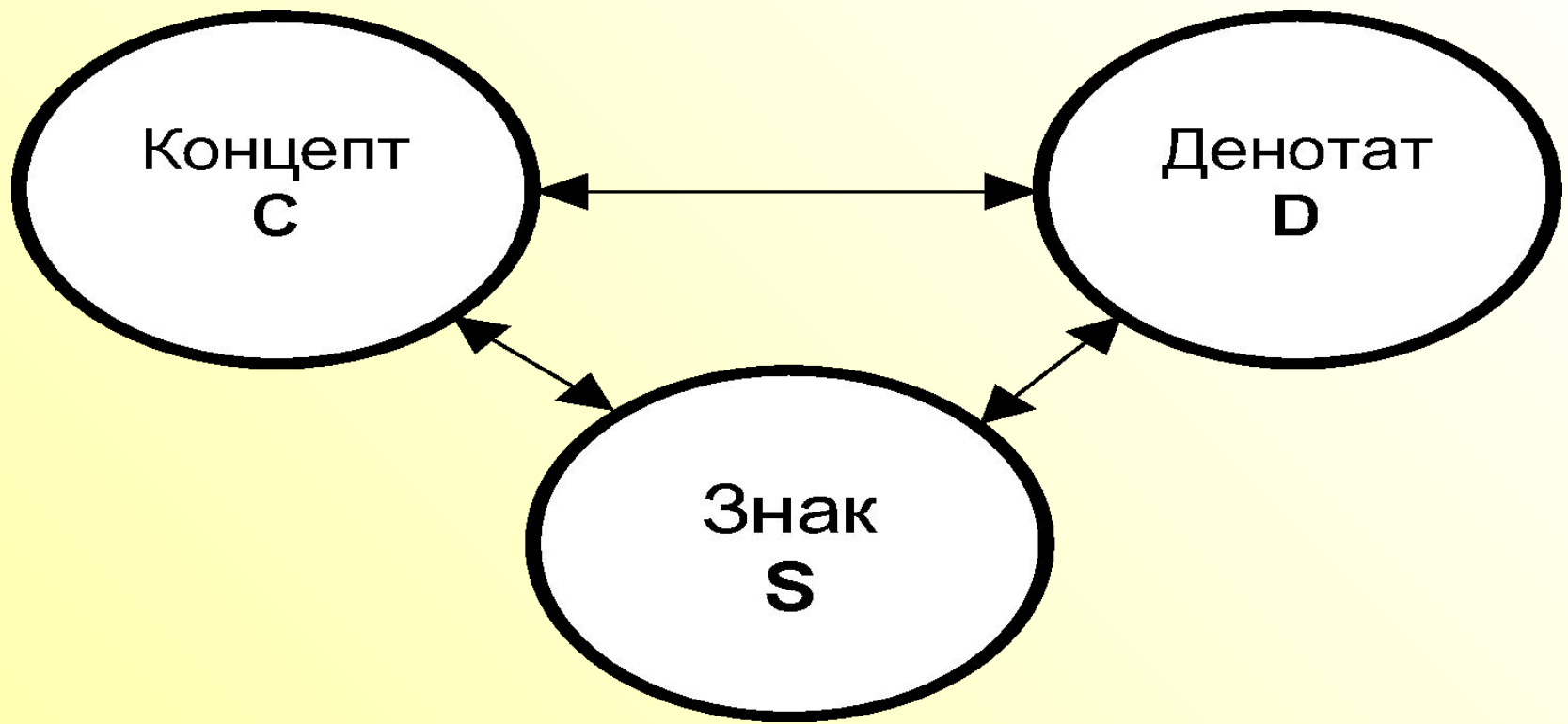
Некоторые исследователи, такие как известный британский лингвист Йорик Вилкс, считают, что

"несмотря на то, что все авторы статей по онтологиям подчеркивают, что понятия являются кирпичиками любой онтологии, мы манипулируем понятиями посредством слов. Во всех онтологиях, которые известны, слова используются для того, чтобы представлять понятия. Следовательно, то множество явлений в мире, которые не вербализованы, не могут быть смоделированы. Мы можем описать это явление как Онтологическую гипотезу Сепира-Уорфа, то есть то, что не описывается словами, не может быть отражено в онтологии".

Различие в абстрактных моделях разных субъектов при рассмотрении одного объекта



Д. В. Кудрявцев «Системы управления знаниями и применение онтологий»
Санкт-Петербург, Изд-во Политехнического университета, 2010

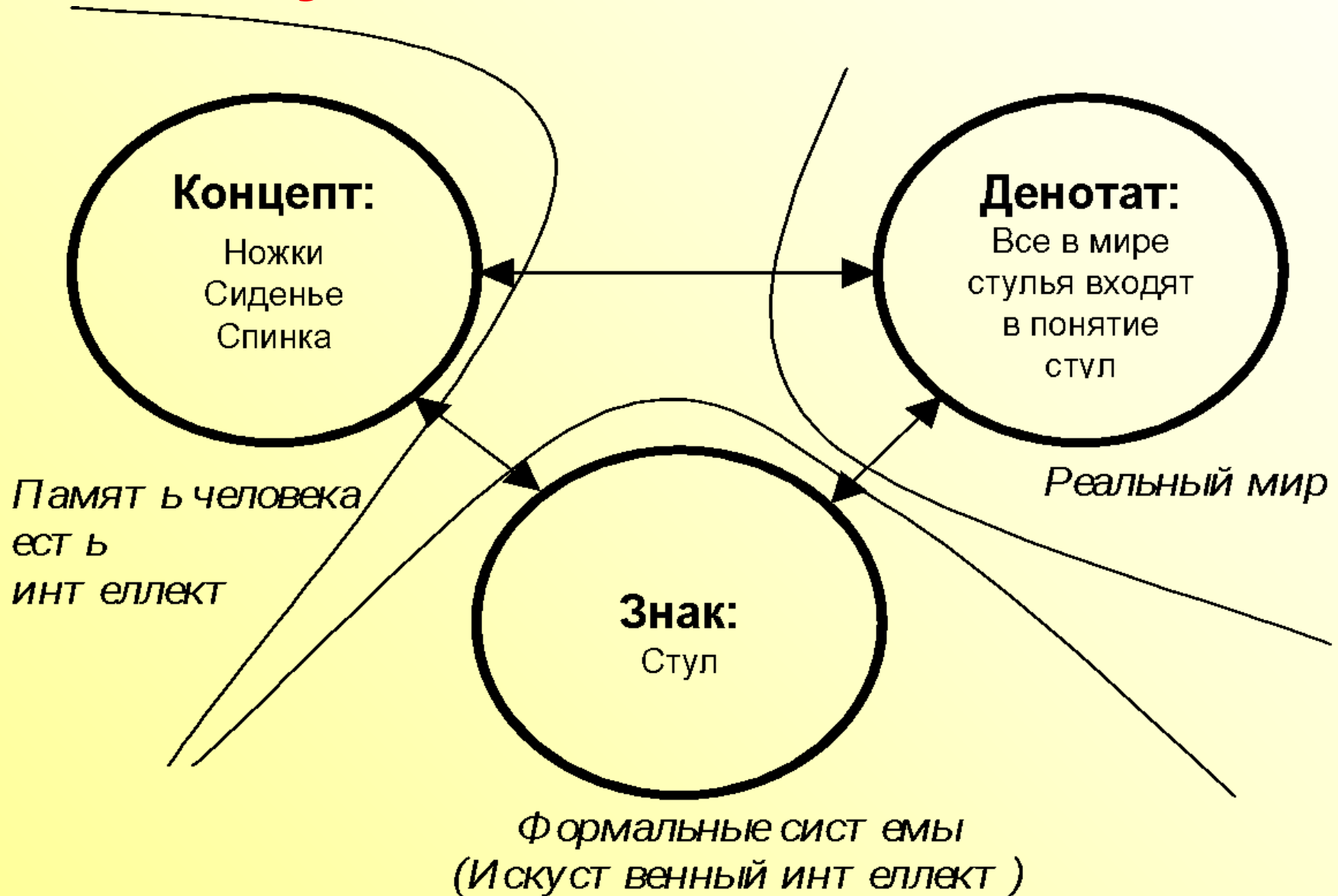


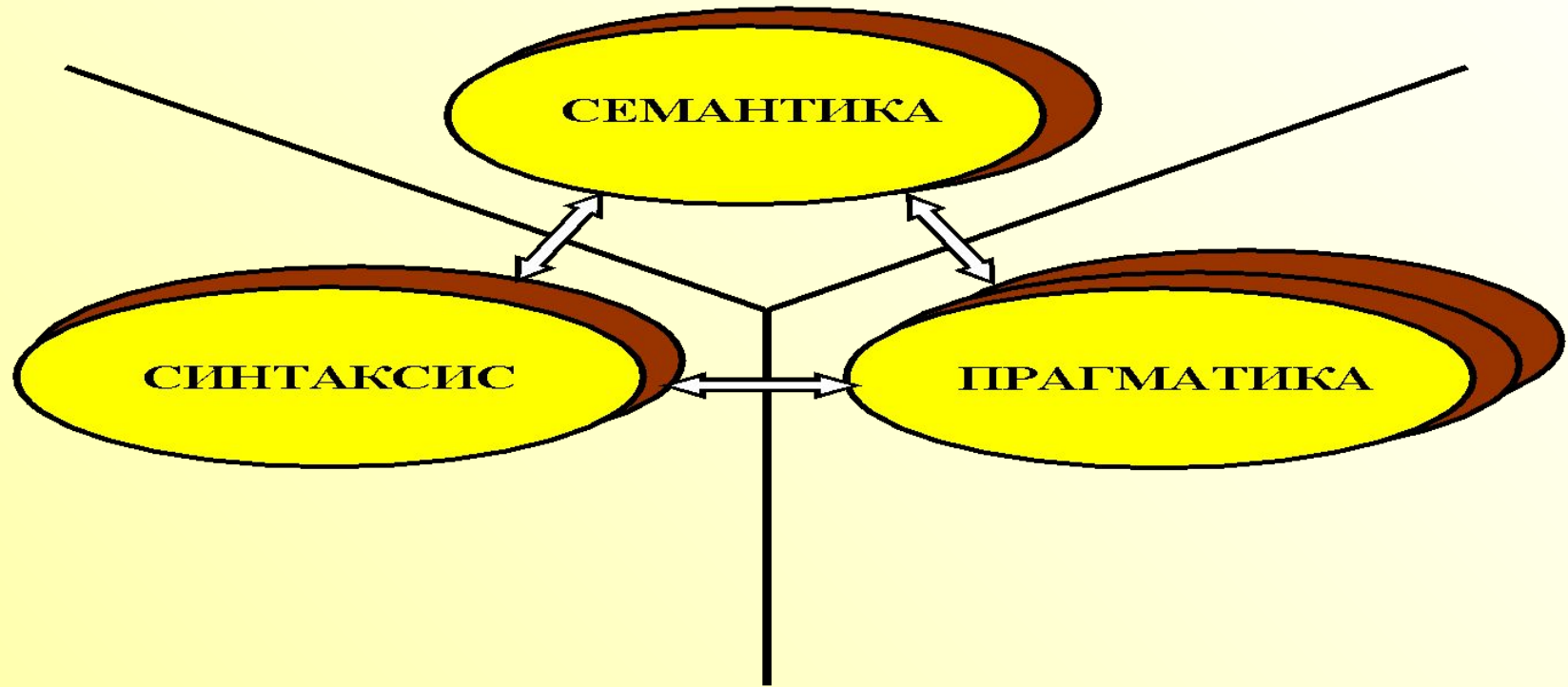
«ЗНАК» – материальный, чувственно воспринимаемый объект, который символически, условно представляет и отсылает к обозначаемому им предмету, явлению, действию или событию.

«КОНЦЕПТ» (*понятие*) - определение обозначаемого посредством *знака* объекта.

«ДЕНОТАТ» – любой предмет, процесс, явление, составляющее содержание языкового (*знакового*) выражения и определяемого посредством *денотата*. Иными словами, - сущность, явление реального или ментального мира, кодируемая языковым знаком.

Представление понятия «стул» в знаковой системе





В семиотике различают следующие разделы:

синтактику, имеющую дело со структурой (синтаксисом) знаковых систем,

семантику, рассматривающую смысл (интерпретацию) знаковых систем (другими словами, соответствие знаковой системы другой знаковой системе),

прагматику, имеющую дело с целенаправленностью знаковых систем.

Семантический треугольник

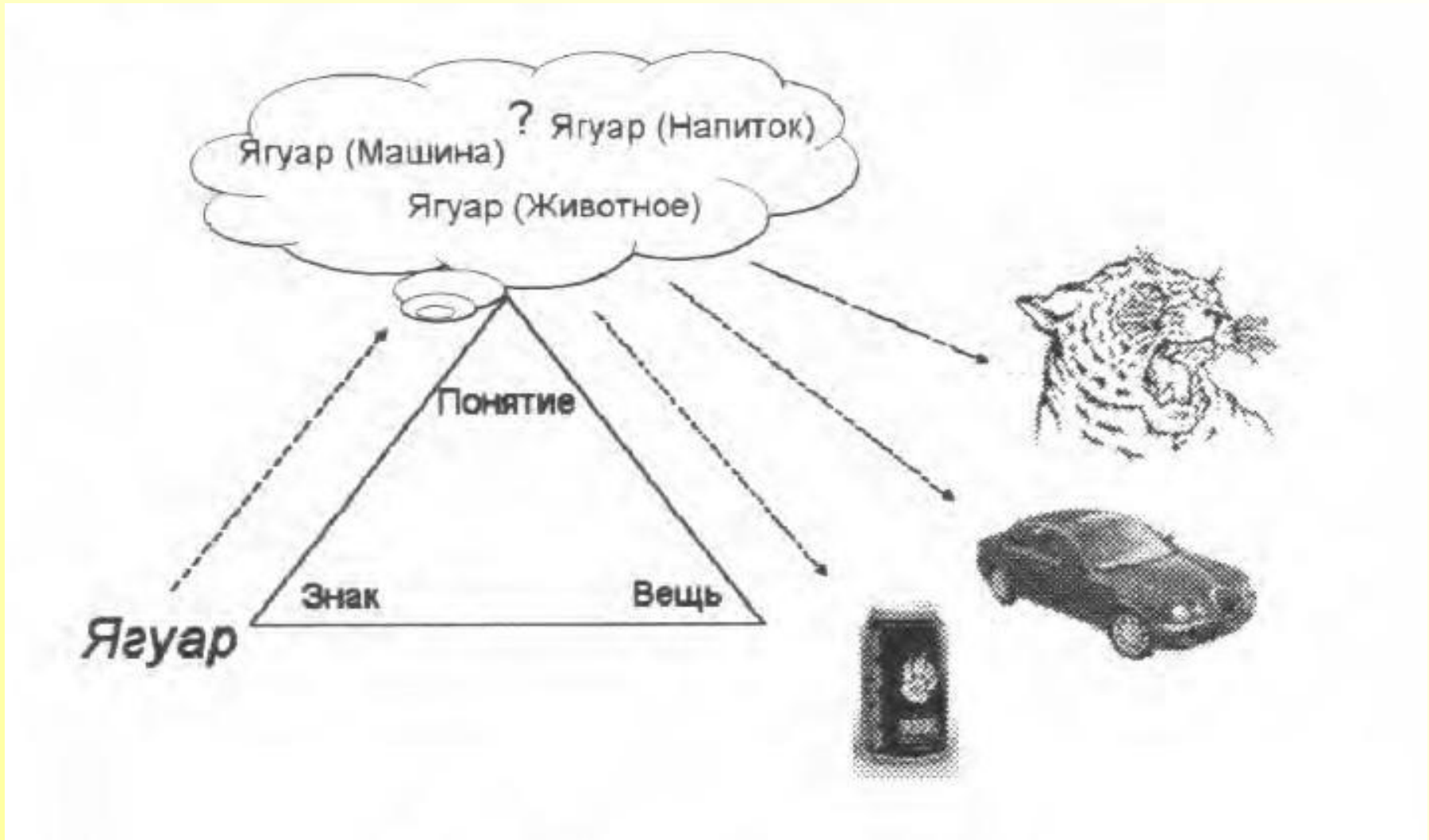


В этом треугольнике часто возникают неоднозначности:

- **Синонимия,**
- **Омонимия,**
- **Полисемия.**

**Д. В. Кудрявцев «Системы управления знаниями и применение онтологий»
Санкт-Петербург, Изд-во Политехнического университета, 2010**

Пример омонимии



Пример распространения синонимии в организации

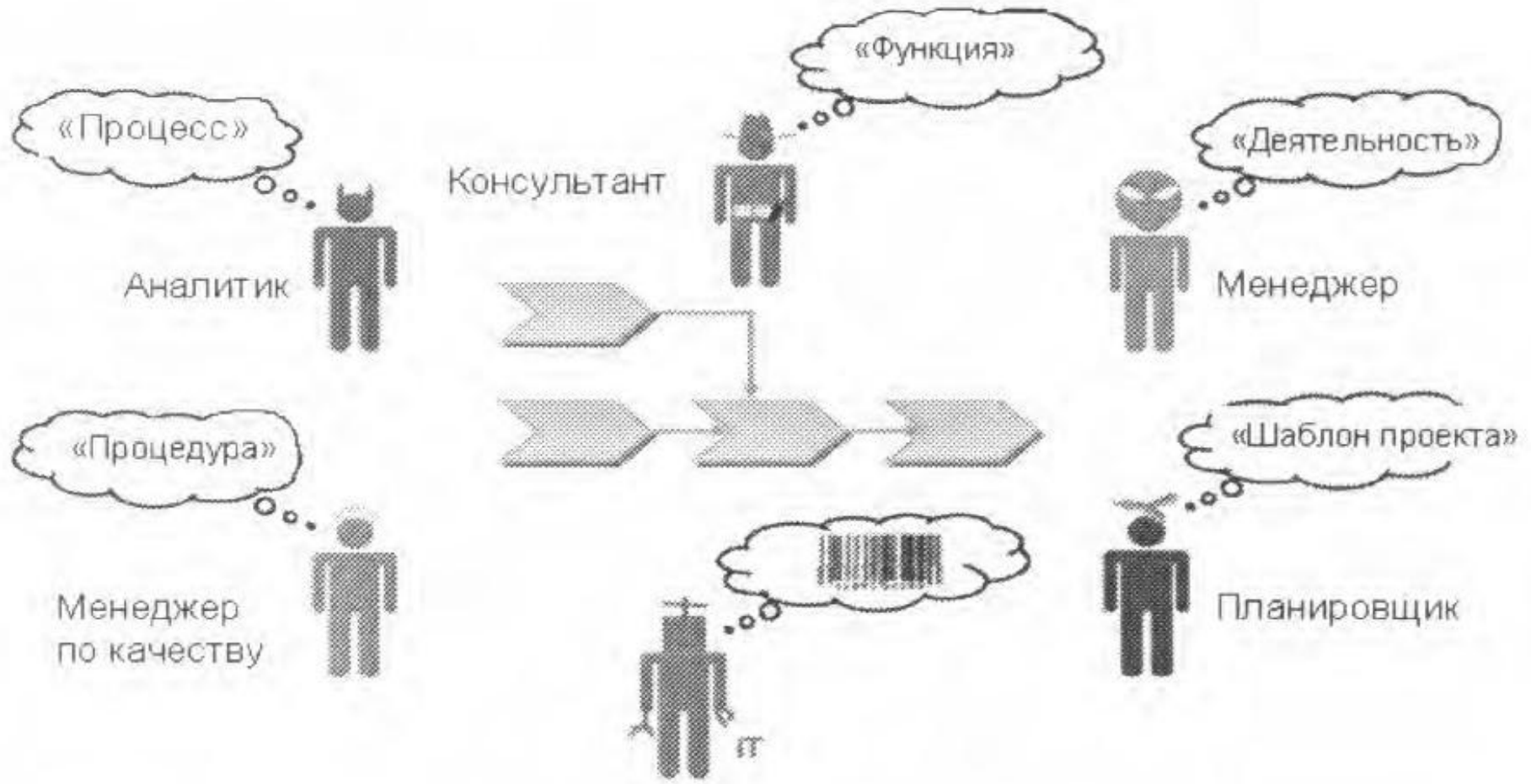


СХЕМА МИРОВ СКОВОРОДЫ

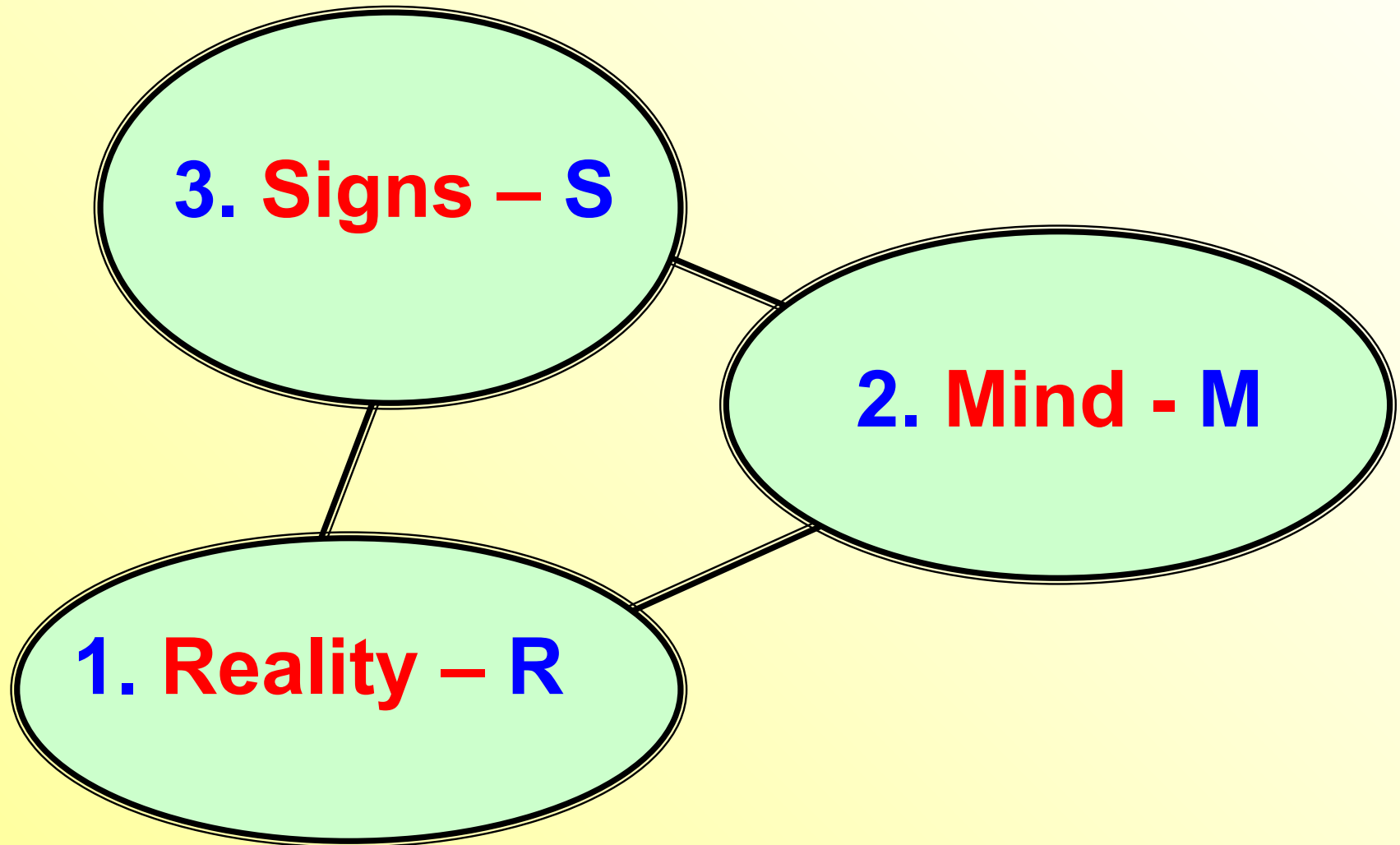
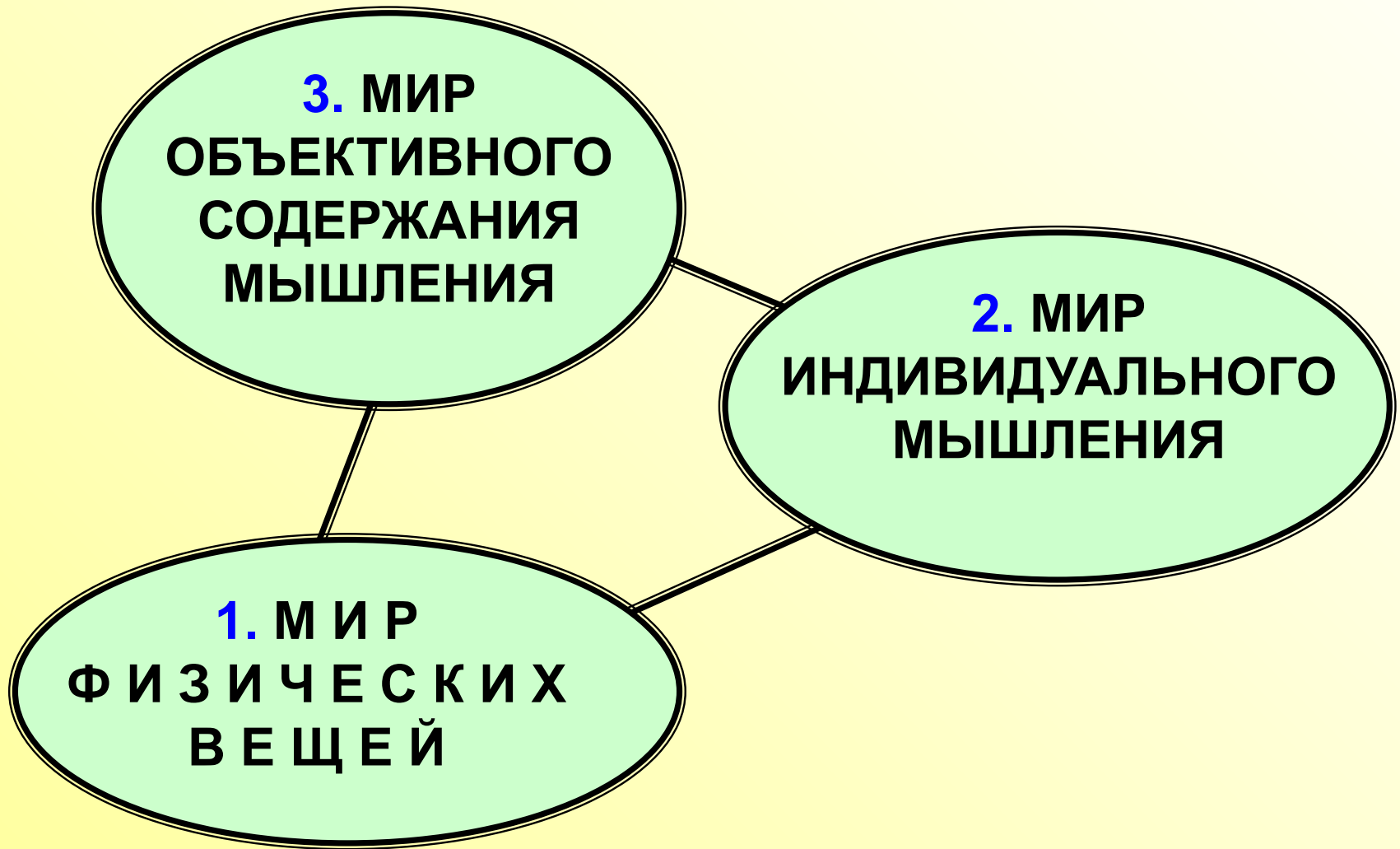


СХЕМА МИРОВ ПОПШЕРА



2. Второе направление, которое обычно обсуждается, — это установление соответствий между **иерархическими лексическими ресурсами типа WordNet** и некоторой онтологией. WordNet-ресурсы описывают лексические отношения между значениями слов, представленные в виде отдельных единиц в иерархической сети — синсетов.

Отношения между лексическими единицами в значительной мере отражают отношения объектов внешнего мира, поэтому такие ресурсы часто рассматриваются как особый вид онтологий — лексические или лингвистические онтологии.

Главной характеристикой лингвистических онтологий является то, что они привязаны к значениям ("are bound to the semantics") языковых выражений (слов, именных групп и т.п.).

Лингвистические онтологии охватывают большинство слов языка и одновременно имеют онтологическую структуру, проявляющуюся в отношениях между понятиями. Поэтому лингвистические онтологии могут рассматриваться как особый вид лексической БД и особый тип онтологии.

Лингвистические онтологии отличаются от формальных онтологий по степени формализации.

Поэтому предполагается, что разработчики такого рода ресурсов разрабатывают иерархию лексических значений ЕЯ, а для более строгого описания знаний о мире необходимо сопоставить такие ресурсы с какими-либо формальными онтологиями.

Так, содержанием одного из проектов является установление отношений между **WordNet** и **EuroWordNet**, с одной стороны, и формальной онтологией **SUMO — Standardized Upper Merged Ontology** — с другой. Проект состоит в том, чтобы установить соответствие между синсетами **WordNet** и понятиями онтологии, при котором каждый синсет **WordNet** либо напрямую сопоставляется с понятием онтологии, либо является гипонимом для некоторого понятия или экземпляром (элементом) понятия онтологии.

Участники другого проекта — **OntoWordNet** — считают, что недостаточно провести формальную склейку ресурса типа **WordNet** и формальной онтологии: необходима значительная реструктуризация исходного лексического ресурса.

Аналогичная работа проводится и в проекте Сус.

3. Третий путь — попытаться

разработать единый ресурс, в котором были бы сбалансированы обе части:

система понятий и

система лексических значений, —

что заключается в разумном разделении этих единиц в создаваемом ресурсе и аккуратном описании их взаимосвязей.

Попытка такого подхода реализуется в

онтологиях **MikroKosmos** и **OntoSem**.

Коллекции электронных документов и задачи их автоматической обработки

- Миллионы текстов в электронной форме
- Множество разнообразных насущных задач по автоматической обработке электронных документов
- **Но:** для решения этих задач используются пословные статистические методы (“bag of words” models)
- Information retrieval community:
текст – это набор features, закономерности которых хорошо учитываются статистическими методами

Онтологии. Концептуальное индексирование

- Ресурс для автоматического индексирования.
- Индекс: не слова, а понятия.
 - Многозначные слова разведены к разным понятиям
 - Синонимы приводят к одному понятию
 - Отношения могут использоваться для расширения или уточнения запроса

Онтологии. Semantic Web (2001)

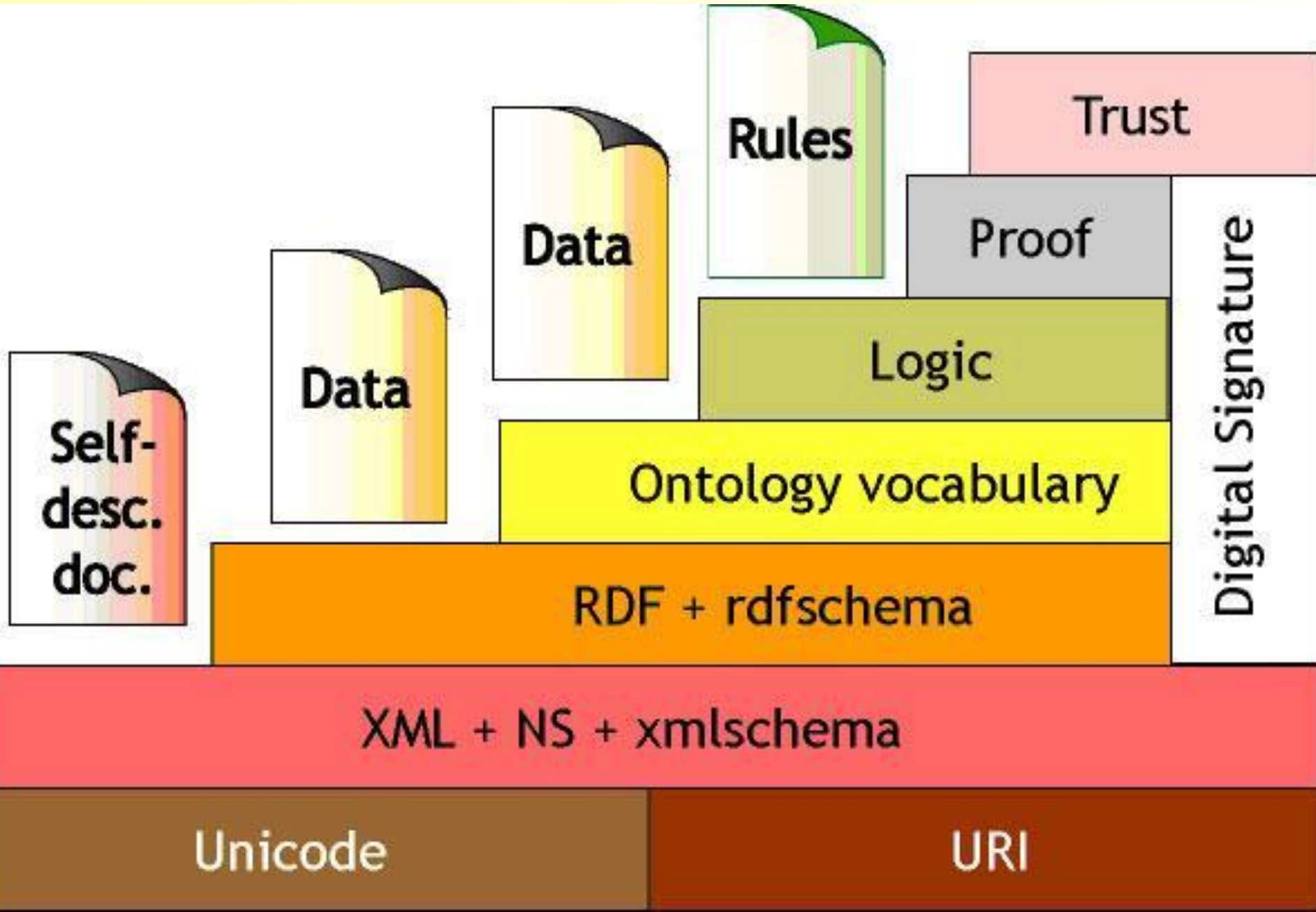
Тим Бернес-Ли, Джеймс Хендлер, Ора Лассила

- Страницам сайта приписана некоторое формально описание, которое помогают автоматическим процессам в сети взаимодействовать
- **RDF** (Resource Description Framework)
- Web Ontology Language (**OWL**)
- Единицы описаний – из Онтологий
- «Сеть наполнится семантикой»

ЯЗЫКИ ДЛЯ ОПИСАНИЯ ОНТОЛОГИЙ

- Ontobroker
- CycL
- Description Logics
- RDF/RDFS
- OWL

Диаграмма "слоеного пирога" Semantic Web



WordNet версии 2.1
охватывает приблизительно
155 тысяч различных лексем и
словосочетаний,
организованных **в 117 тысяч**
понятий, или совокупностей
синонимов (synset); общее
число пар "лексема-значение"
насчитывает **200 тысяч**.

Словарь состоит из 4 сетей для основных знаменательных частей речи: **существительных, глаголов, прилагательных и наречий.**

Базовой словарной единицей в WordNet является не отдельное слово, а так называемый синонимический ряд («**синсеты**»), объединяющий слова со схожим значением и по сути своей являющимися узлами сети.

Синсеты в WordNet связаны между собой различными семантическими отношениями:

- **гипероним** (breakfast → meal) (завтрак → пища);
- **гипоним** (meal → lunch) (пища → обед);
- **has-member** (faculty → professor) (факультет → профессор);
- **member-of** (pilot → crew) (пилот → экипаж);
- **мероним**: has-part (table → leg) (стол → ножка);
- **антоним** (leader → follower) (лидер → последователь).

PyTез

- **Онтологии по естественным наукам и технологиям**
- **Технология построения онтологии сочетает традиции разработки трех видов ресурсов**
 - **Информационных-поисковых тезаурусов**
 - **Ресурсов типа wordnet,**
 - **Формальных онтологий**
- **Сочетание этих трех методологий позволяет создавать большие онтологии для широких предметных областей**
- **В 2008 году: Свободное распространение Онтологии по естественным наукам и технологиям для некоммерческого применения.**

С 1994 года в **АНО ЦИИ** (Автономная некоммерческая организация Центр информационных исследований) началась разработка **Общественно-политического информационно-поискового тезауруса (русскоязычного)** как ресурса для автоматического индексирования.

Сравнение Тезауруса Исследовательской службы Конгресса США (LIV) с
Общественно-политическим тезаурусом

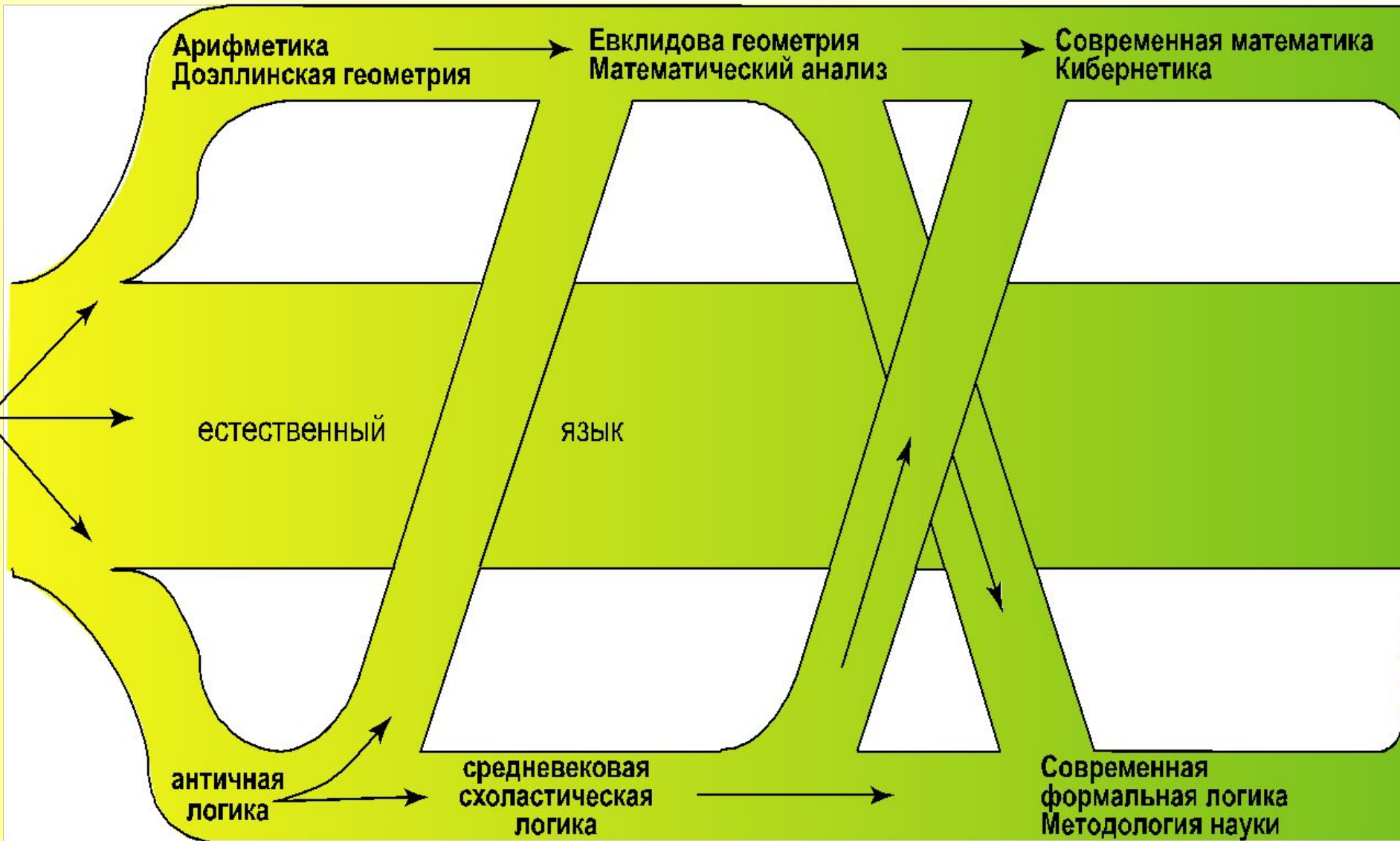
Характеристика	Общественно-политический Тезаурус	LIV
Число понятий	33 тысяч	6,8 т
Число терминов	87 тысяч	9,8 т
Термины, описанные как многозначные	4,5 тысяч	Нет
Общее количество описанных отношений между понятиями	130 тысяч	15 т
Количество отношений, полученных	700 тысяч	Не опр

МЕРОНОМИЯ

- **Сущности, такие как группы, классы и коллекции, состоят в отношении меронимии со своими элементами:**
- **Примеры групп:** племя, команда, комитет, семья, оркестр, суд, отряд и др.
- **Примеры классов:** пролетариат, аристократия, буржуазия.
- **Примеры коллекций:** куча, лес, библиотека (как коллекция книг).

4. Об эволюции структур данных и знаний

Историческое развитие языково-мыслительных и математико-формализованных средств познания



Для того чтобы уверенно оперировать понятиями "информация", "данные", "знание", необходимо не только понимать суть этих понятий, но и прочувствовать отличия между ними.

Однако, одной интуитивной интерпретации этих понятий здесь недостаточно.

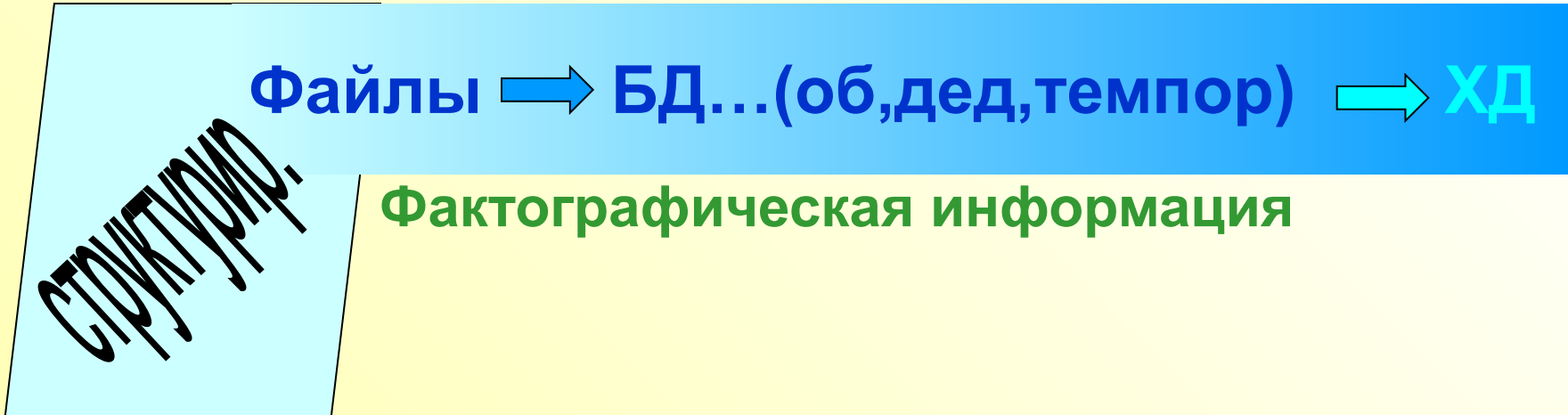
Сложность понимания отличий вышеупомянутых понятий - в их кажущейся синонимичности.

Вспомним, что понятие Data Mining переводится на русский язык при помощи этих же трех понятий: **как добыча данных, извлечение информации, раскопка знаний.**

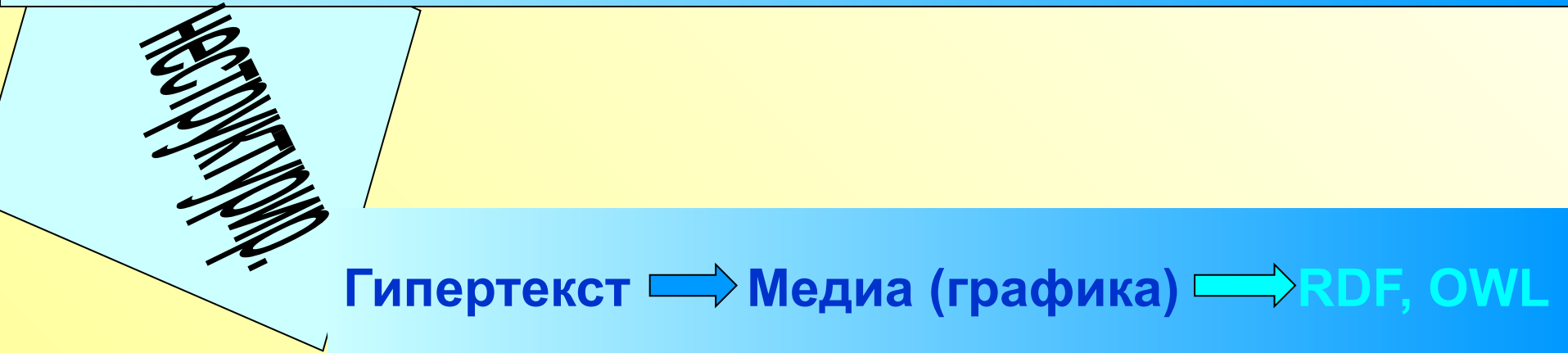
Для начала сделаем попытку разобраться в этих терминах на простых примерах.

1. Студент, который сдает экзамен, нуждается в данных.
2. Студент, который сдает экзамен, нуждается в информации.
3. Студент, который сдает экзамен, нуждается в знаниях.

Эволюция структур и представлений данных и знаний

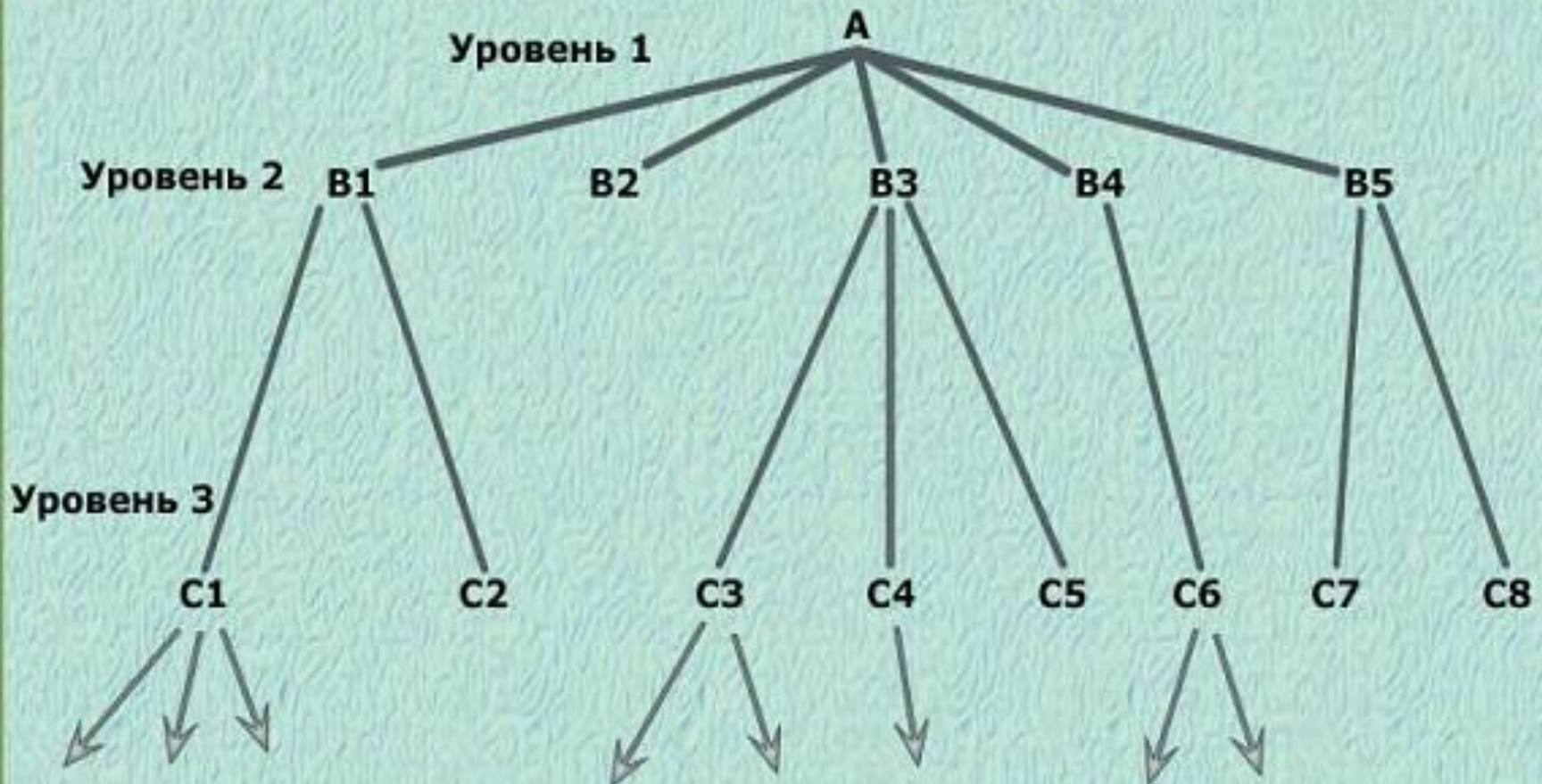


ЗНАНИЯ → лог... фреймы... продукции... сем. сети



Виды моделей данных БД

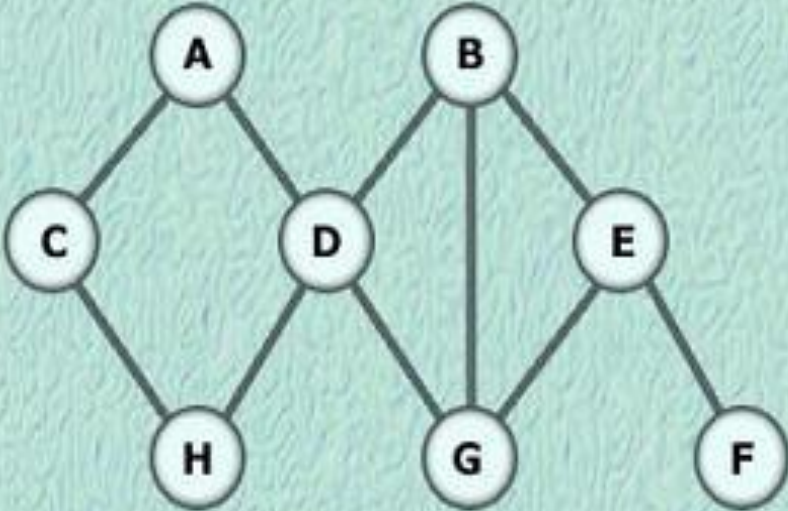
Иерархическая модель



Виды моделей данных БД

Сетевая модель

Взаимосвязь данных



Пример реализации

Студент (№ зач.книжки, Фамилия, группа)

87695
Иванов
113

87697
Петров
116

87699
Сидоров
120

Работа (шифр, рук-ль, область)

757
Сергеев Н.П.
Информатика

1001
Петухова А.Е.
Экономика

344
Кофа Л.Д.
История

1037
Яковлева С.Т.
Право

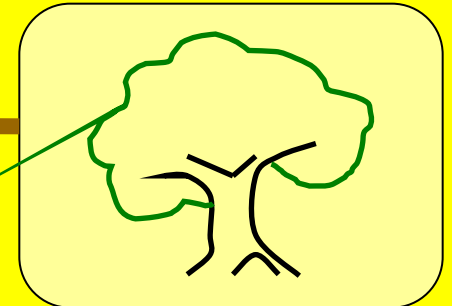


РЕПРЕЗЕНТАЦИЯ ПОНЯТИЙ С ТОЧКИ ЗРЕНИЯ ПСИХОЛОГОВ

МНОЖЕСТВЕННАЯ

Дуб, каштан,
береза, ель,
сосна, ...

ПРОТОТИПА



ПРИЗНАКОВ

Ствол, корни,
крона, ветки,
листья (хвоя),
растение, ...

Д Е Р Е В О

Реляционная модель

Экс-нал Ин-нал	Ствол	Ветки	Корни	Листья	...
Дуб					
Каштан					
Береза					
Ель					
Сосна					
...	...				

$R \subseteq A_1 \times A_2 \times \dots \times A_n$

Отображение признаков понятия при построении классификатора



ГИПЕРКУБЫ И ПОЛИКУБЫ



Иерархические пути суммирования



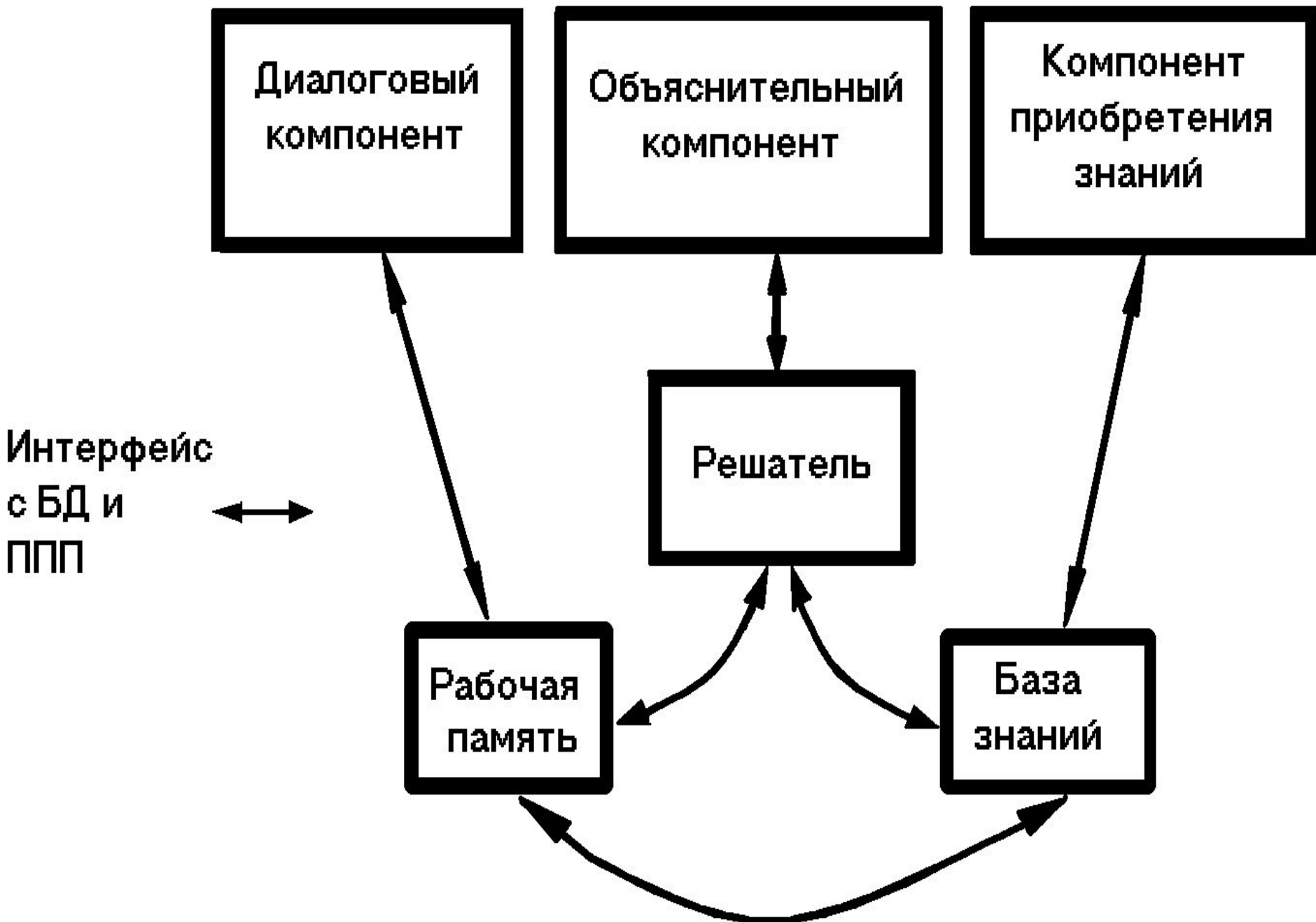


Рис.1. Структура статической ЭС.

5. Применение ОНТОЛОГИЙ

**По материалам книги Д. В. Кудрявцева
«Системы управления знаниями и применение
онтологий»**

**Санкт-Петербург, Изд-во Политехнического
университета, 2010**

- 1. Поиск**
- 2. Общий словарь**
- 3. Индексирование**
- 4. Схема данных**
- 5. Посредник для интеграции**
- 6. Семантический анализ**
- 7. Извлечение информации**
- 8. Моделирование знаний**
- 9. Систематизация знаний**

(1) Общий словарь

Онтология предоставляет общий

словарь для описания информационных ресурсов, а также общения сотрудников.

Общий словарь снижает

неоднозначность соответствия между знаком, понятием и денотатом. Данный

пример использования был рассмотрен

выше, и он является основой для

обоснования целесообразности

использования онтологии в УЗ.

Использование онтологии в качестве общего словаря и указателя



(2) Поиск

При использовании онтологии поисковая машина будет выдавать только такие ресурсы, где упоминается в точности искомое понятие, а не произвольные страницы, в тексте которых встретилось данное многозначное ключевое слово.

Сегодняшние поисковые системы зачастую выдают бесчисленное множество совершенно не относящихся к запросу «хитов», обрекая пользователя на длительный ручной отбор материала. Например, если вы ввели для поиска слово «ягуар», то компьютеру совершенно непонятно, имеете ли вы в виду автомобиль, животное или напиток. Вся проблема в том, что для компьютера слово «ягуар» не имеет четкого смысла, или другими словами, семантического содержания.

Онтология позволяет указать контекст для поиска. Указывая в поисковой строке слово «ягуар» необходимо дополнительно указать, идет ли речь об автомобиле, животном или напитке. На техническом языке это может называться сужением области поиска. Онтологии дают возможность производить запросы на основе понятий, а не на основе совпадения строк. Например, если пользователь задаст вопрос: «Какие транспортные средства производятся в России?», то он получит из базы ответ, в который попадут автомобили (= подкласс транспортных средств), производимые во Всеволожске (= город, который находится в России).

Система на основе онтологии может предлагать пользователю возможные интерпретации его поискового запроса на основе анализа взаимосвязей понятий, а также распределять результаты поиска по различным интерпретациям запроса пользователя.

Например, для запроса «организационное проектирование» система может уточнить интересуется ли пользователем вопросом проектирования организаций (область менеджмента) или проектированием любых объектов, но в рамках организации.

(3) Указатель

В данном применении онтологии используются в качестве структурированного указателя на ресурсы. Онтология предоставляет возможности для навигации, по аналогии с папками на домашнем компьютере или каталогом ресурсов.

Часто пользователь не может точно сформулировать запрос, например, из-за новизны предметной области для него или нечеткости информационной потребности.

В таком случае использование указателя (каталога) может быть более эффективной стратегией, нежели формирование запросов.

Семантические описания также могут направлять пользователя в процессе поиска информации, предлагать детализацию, более общий взгляд, смежные области.

В случае малого количества результатов поиска, иерархия понятий из онтологии может быть использована для расширения запроса добавлением более общих и более частных понятий, например, добавлением понятий «транспортное средство» и «внедорожник» к понятию «автомобиль».

Использование онтологии в качестве общего словаря и указателя



(4) Схема данных

Онтология может использоваться для описания и спецификации схемы базы данных.

Конкретными примерами в данной группе является схема метаданных единого научного информационного пространства Российской академии наук, а также масштабные базы данных по исследованиям.

Использование онтологии для описания схемы БД и структуры БЗ



(5) Посредник для интеграции знаний

Онтологии позволяют интегрировать (обмениваться) знания, находящиеся в различных источниках, - документы, базы данных и знаний, сотрудники. Потребителями интегрированной информации могут быть как люди, так и программы (или программные агенты). Возможны два варианта интеграции:

- (а) потребители работают **с одной общей онтологией**, с помощью которой описываются все информационные ресурсы,
- (в) потребители имеют свои **собственные частные онтологии**, между элементами которых установлены соответствия.

Использование в роли посредника при интеграции знаний

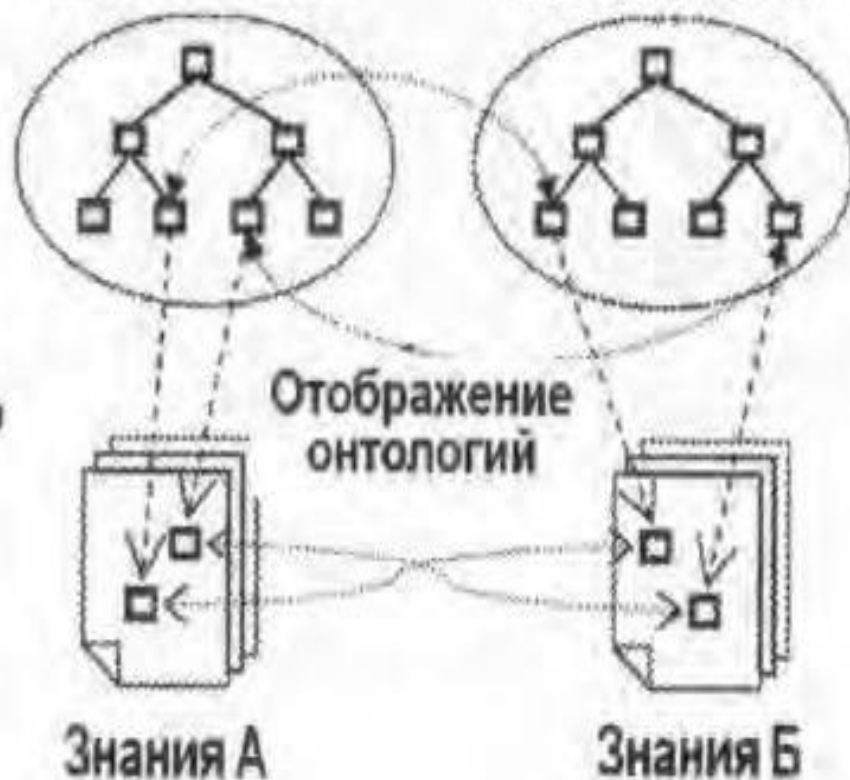
Референтная онтология



(a) Интеграция знаний с помощью референтной онтологии

Онтология А

Онтология Б



(b) Интеграция знаний с помощью нескольких онтологий

В традиционных ИС потребность в знаниях нужно удовлетворять путем раздельного поиска по различным источникам



В целом, доступ к информации может быть организован за счет ее совместного использования потребителями или посредством обмена информацией между ними. Так как сейчас все большее внимание уделяется распределенным информационным системам моделирования организаций, которые обмениваются знаниями друг с другом, совместно используют знания, реализуют возможности повторного использования знаний, то возникает вопрос о представлении знаний в формальной форме.

Несмотря на то, что такие системы имеют стилистические и организационные различия в представлении знаний, необходимо, чтобы они сохраняли декларативное содержание этих знаний и, таким образом, могли формулировать запросы к другим информационным системам, обмениваться с ними данными и знаниями.

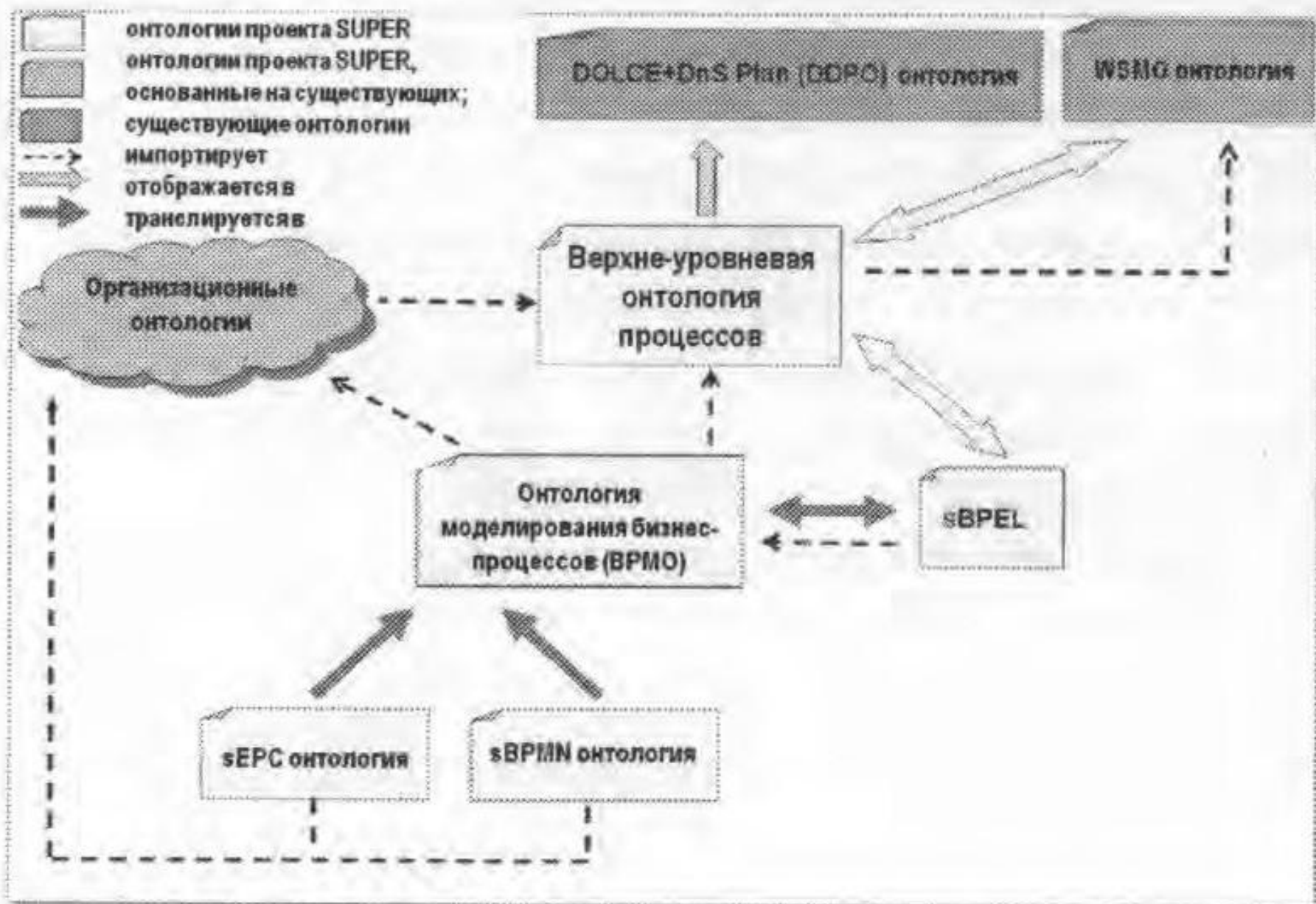
В данном случае возникает проблема необходимости организации взаимодействия с учетом нескольких созданных онтологий. В результате, онтология становится *средством предоставления возможности взаимодействия между компьютерными системами независимо от индивидуальных технологий* систем, архитектуры информации и областей применения.

В качестве реального примера использования онтологии для интеграции знаний можно рассмотреть интеграцию знаний о бизнес-процессах, представленных в разных нотациях.

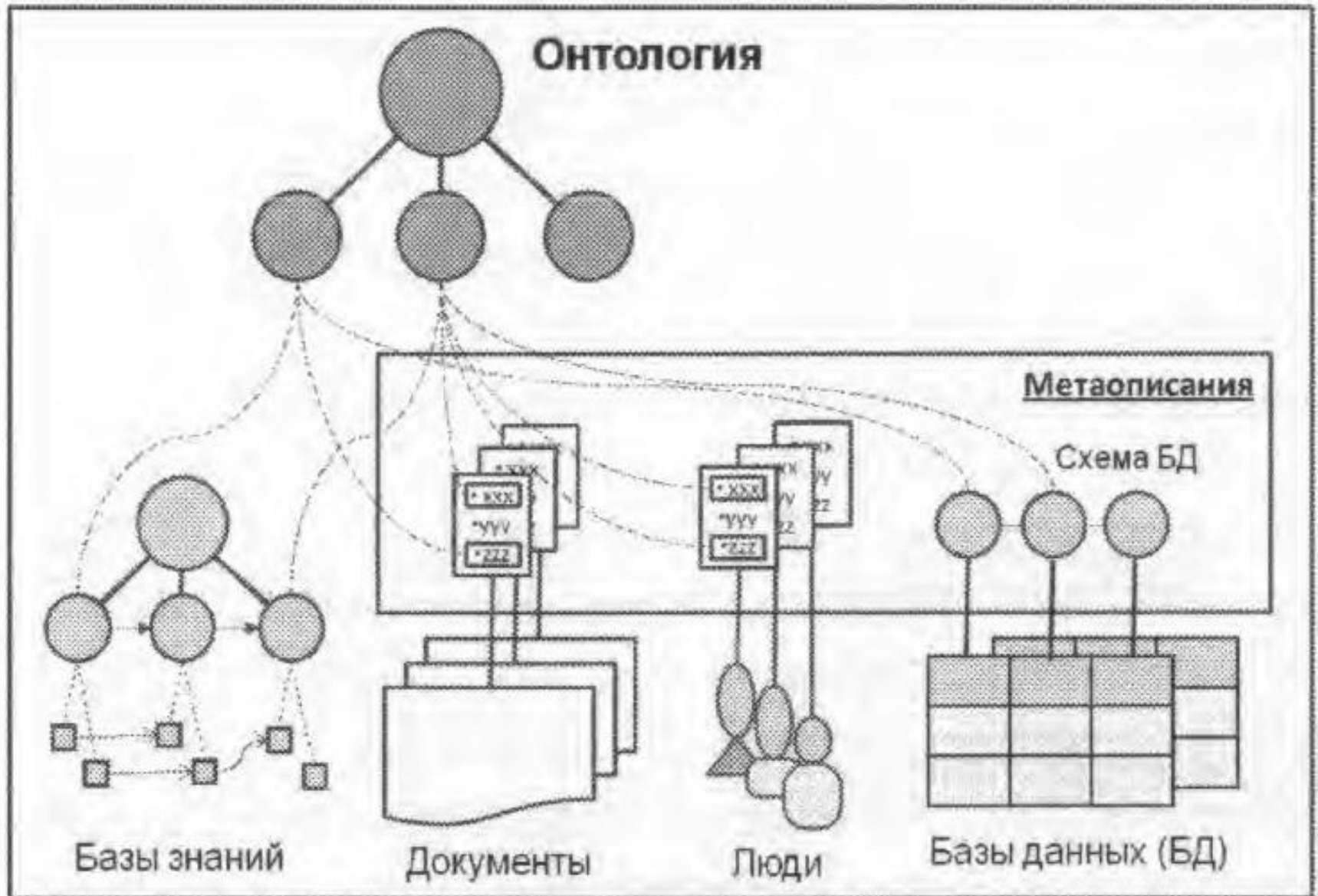
В рамках проекта SUPER были разработаны онтологии, соответствующие стандартным нотациям (EPC, BPMN, BPEL), обобщенная онтология моделирования бизнес-процессов, верхнеуровневая онтология процессов, а также схемы соответствия между ними.

Данные онтологии и схемы, заложенные в программную среду, позволяют преобразовывать модели процессов из одной нотации в другую, а также осуществлять интегрированный поиск по всем моделям.

Интеграция онтологий моделирования бизнес-процессов



Онтология позволяет интегрировать разнородные источники информации



(6) Семантический анализ

К знаниям, представленным с помощью онтологии, могут быть применены автоматизированный логический вывод и семантическая обработка, которые позволяют проводить семантический анализ и создавать новые знания. Одним из распространенных примеров вывода является автоматическая классификация объектов с помощью инструментов логического вывода.

Например, зная, что ингредиентами какого-либо блюда являются только *овощи и крупа*, система может автоматически отнести данное блюдо к *классу вегетарианских*.

Интеграция онтологии с правилами позволяет проводить автоматическую классификацию, проверку соблюдения политик и может использоваться для поддержки принятия решений. Некоторые приложения визуализируют информацию с помощью онтология для поддержки работ по анализу.

Использование информации программными агентами при решении рутинных умственных задач, которое было предложено в рамках концепции семантического веба, также относится к семантическому анализу на основе онтологии.

Семантический веб привносит смысловую структуру в содержание веб-страниц. Это не какая-то отдельная сеть, а расширение уже существующей, такое, что в ней информация снабжена точно определенным смыслом, позволяющим человеку и машине успешно взаимодействовать.

Сценарий работы таких программных средств представлен в основополагающей статье идеологов семантического веба — Тима Бернерса-Ли и Джеймса Хендлера, опубликованной в 2001 году в журнале Зиеппйс Атепсап:

«Пит поднял трубку, звонила его сестра Люси из кабинета врача: «Моей маме нужно попасть на прием к врачу, а потом ей требуется пройти несколько сеансов физиотерапии. Примерно два раза в неделю. Я сейчас поручу своему агенту записать нас на прием к врачу». Пит сразу же согласился подвезти ее маму на своей машине.

В кабинете врача Люси дала указания своему электронному агенту семантического веба через свой портативный веб-браузер. Агент сразу же получил *информацию о назначенном маме* лечении от агента ее врача, просмотрел несколько перечней *медицинских учреждений*, оказывающих подобные услуги, проверил, какие из них *подходят* по ее страховому полису и находятся *в окрестности 20 миль от ее дома* и при этом имеют *рейтинг -«отличный»* либо *«очень хороший»* согласно достоверным рейтинговым источникам.

Далее агент попытался сопоставить имеющиеся **часы приема** врачей (предоставленные агентами соответствующих медучреждений на их веб-сайтах) с расписаниями Пита и Люси. (Слова, выделенные курсивом, указывают на термины, семантика или смысл которых была предоставлена агентам посредством семантического веба.)

Уже через несколько минут агент представил им план. Питу он не понравился - университетская клиника расположена на другой стороне города, и ему пришлось бы возвращаться обратно как раз в час пик.

Он попросил своего агента повторить поиск с более строгими предпочтениями относительно **места** и **времени**. Тем временем агент Люси, **полностью доверяя** агенту Пита в рамках данной задачи, автоматически помогал, предоставляя тому права доступа и ссылки на уже добытую им информацию. Практически мгновенно был предложен новый план: найденная клиника была уже гораздо ближе, и часы приема более ранние».

(7) Извлечение информации

Данное применение онтологии позволяет выявлять экземпляры класса и отношения между ними в тексте на естественном языке, а также записывать выявленные факты в семантическое хранилище (базу знаний).

На рис. указано, что в результате семантической разметки фразы *«XYZ планирует инвестировать 150 млн долларов в строительство завода в Исландии»* некоторые ее элементы позволяют наполнить семантическое хранилище.

В свою очередь, информация в нем удобна для дальнейшей обработки и анализа.

Семантическая разметка текста на ЕЯ

XYZ планирует инвестировать 150 млн долларов в строительство завода в Исландии

Семантическое хранилище

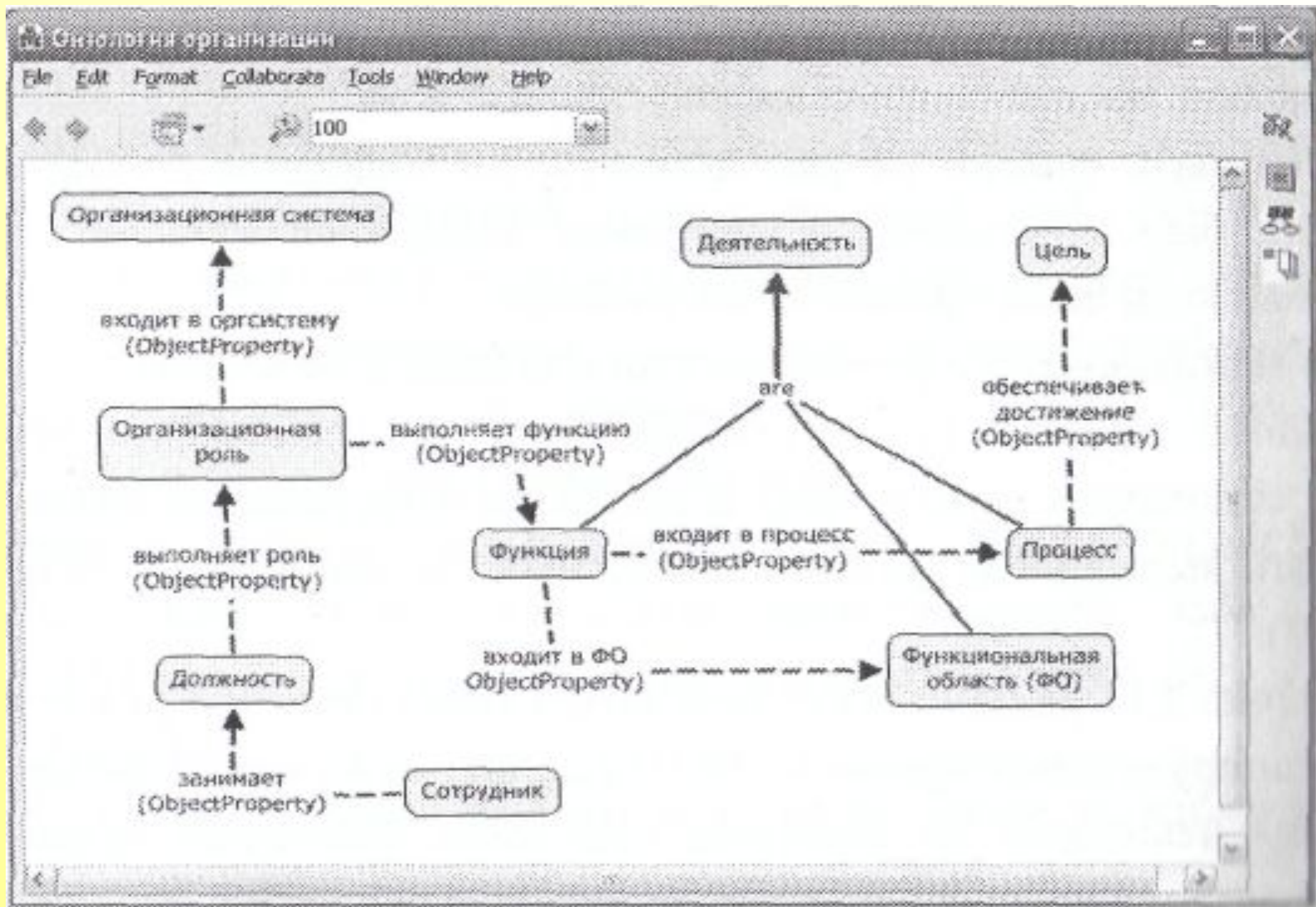


(8) Спецификация структуры моделей знаний

Онтология используется для спецификации структуры моделей знаний, то есть баз знаний.

На одном рис. приведена онтология, которая задает структуру для онтологической модели (модели знаний), представленной на другом рис.

Онтология, задающая структуру БЗ



БЗ на основе онтологий

The screenshot displays a software interface for strategic planning, titled "Правила Стратегические цели ФСУ Процесс". The interface is divided into several sections:

- Left Panel (Functional Systems and Processes):** A tree view showing a hierarchy of functional systems. The selected item is "1.1.5.2.1 Управление оргструктурой и системой процессов".
- Right Panel (Strategic Goals):** A tree view showing strategic goals. The selected goal is "Соответствие системы процессов корпоративной стратегии".
- Bottom Panel (Relationships):** A table with columns "№", "Вид", and "Характеристика". The table shows a relationship between the selected functional system and the selected strategic goal, with the relationship type "обеспечивает достижение".

Table 1: Functional Systems and Processes

Code	Description
1.1.2	Корпоративные отношения и контроль
1.1.3	Отношения с общественностью
1.1.4	Финансово-экономическое управление
1.1.5	Стратегические результаты
1.1.5.1	Разработка политик и процедур в области "организации"
1.1.5.2	Управление формированием и совершенствованием бизнес-модели Группы
1.1.5.2.1	Управление оргструктурой и системой процессов
1.1.5.2.2	

Table 2: Strategic Goals

Code	Description
331.1.3.1.1	Оптимальность выбора функционала и его распределения по различным системам
331.1.3.1.2	Соответствие системы процессов корпоративной стратегии
331.1.3.1.3	Оптимальность распределения полномочий и ответственности

Table 3: Relationships

№	Вид	Характеристика
	Влияет на	
	достигается с помощью	
	зависит от	
	обеспечивает достижение	

(9) Систематизация знаний

Структурировать знания об окружающем пространстве –

- для отдела персонала это может быть **каталог компетенций сотрудников,**
- для службы техобслуживания и ремонта — **справочник оборудования и инструментов,**
- для маркетологов - **сегменты рынка,**
- для бизнес-аналитика - **справочник бизнес-процессов.**

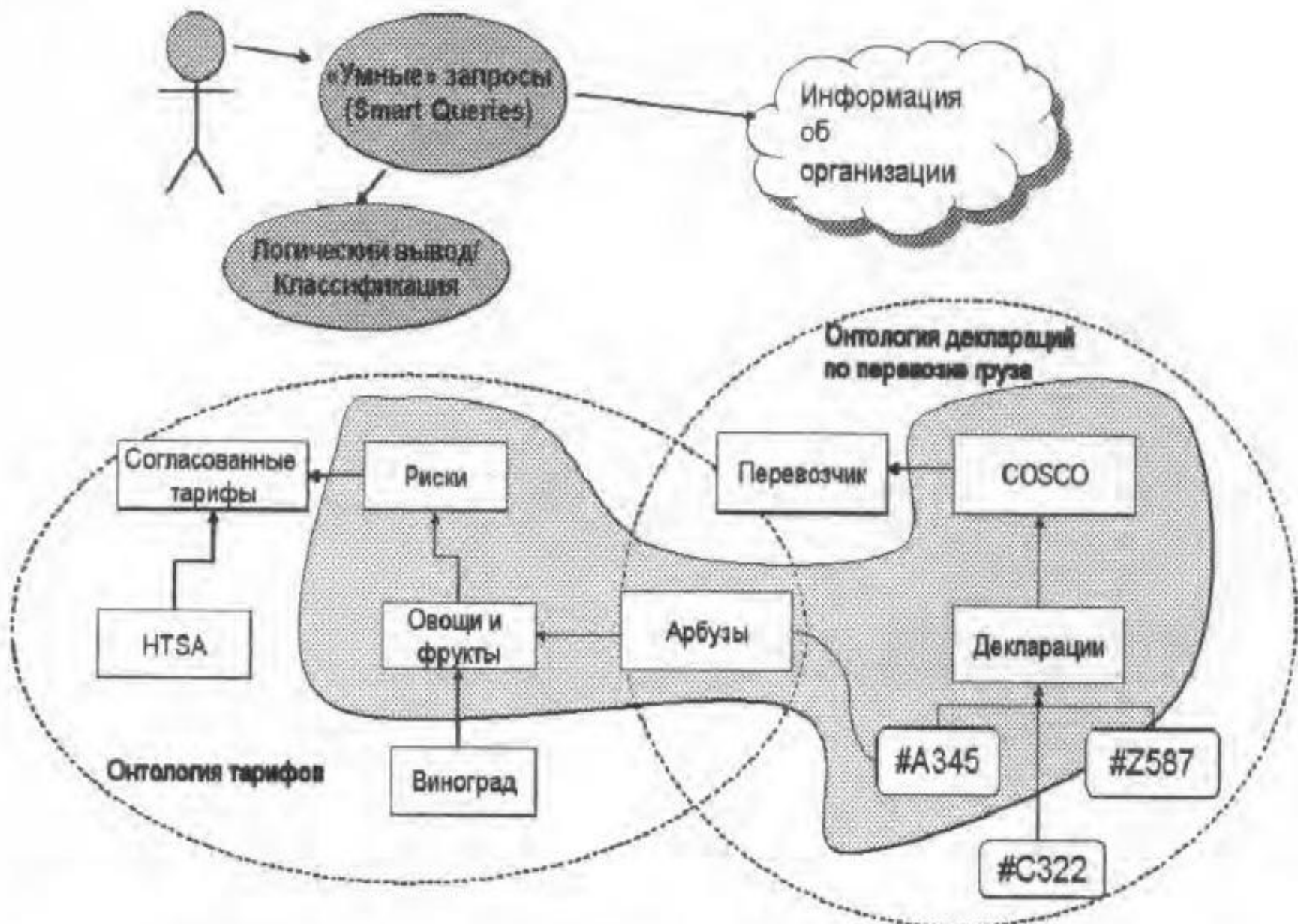
Ориентиром в этой области применения онтологии на предприятии является комплексное структурированное описание архитектуры предприятия.

На практике, приложение часто сочетают указанные выше примеры использования. Например, выполнения простых автоматических рассуждений и интеграция информации могут повышать качество результатов поиска информации. Онтологии дают возможность получения не заданных явно знаний из информационных хранилищ путем логического вывода - поиск «скрытой информации».

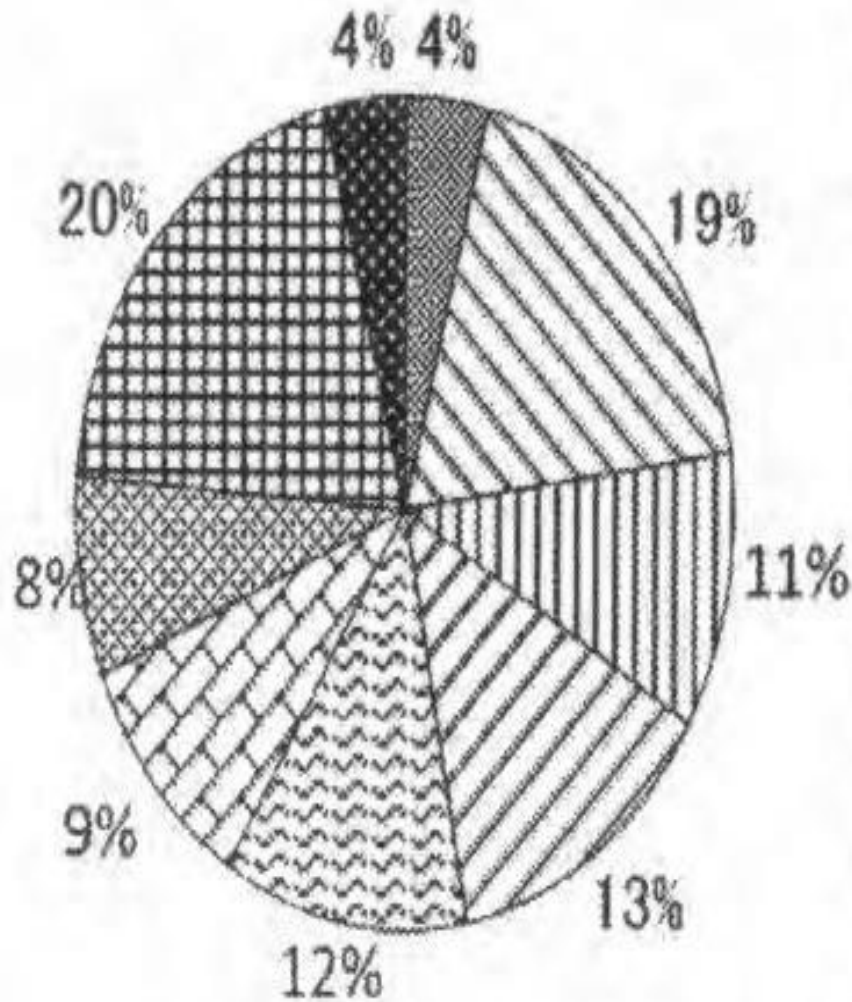
Например, пользователь системы может задать вопрос: *Какие поставки продукции находятся сейчас в состоянии риска?* В ответ на такой вопрос система в одной онтологии тарифов определит, что с учетом текущих условий (например, географических или погодных) существуют риски связанные с перевозкой овощей и фруктов.

А в другой базе или онтологии деклараций по перевозке груза определит, что в декларации № А345 указаны арбузы, которые являются подклассом «Овощей и фруктов». В результате, система сможет дать конкретный ответ на поставленный вопрос: *Поставка СО5СО #А345*

Интеллектуальный поиск на основе логического вывода и интеграции информации



Распределение типов использования онтологии в семантических приложениях



- (1) Общий словарь
- ▧ (2) Поиск
- ▨ (3) Индексирование
- ▩ (4) Схема данных
- (5) Посредник для интеграции
- (6) Семантический анализ
- ▬ (7) Извлечение информации
- ▭ (8) Моделирование знаний
- (9) Систематизация знаний

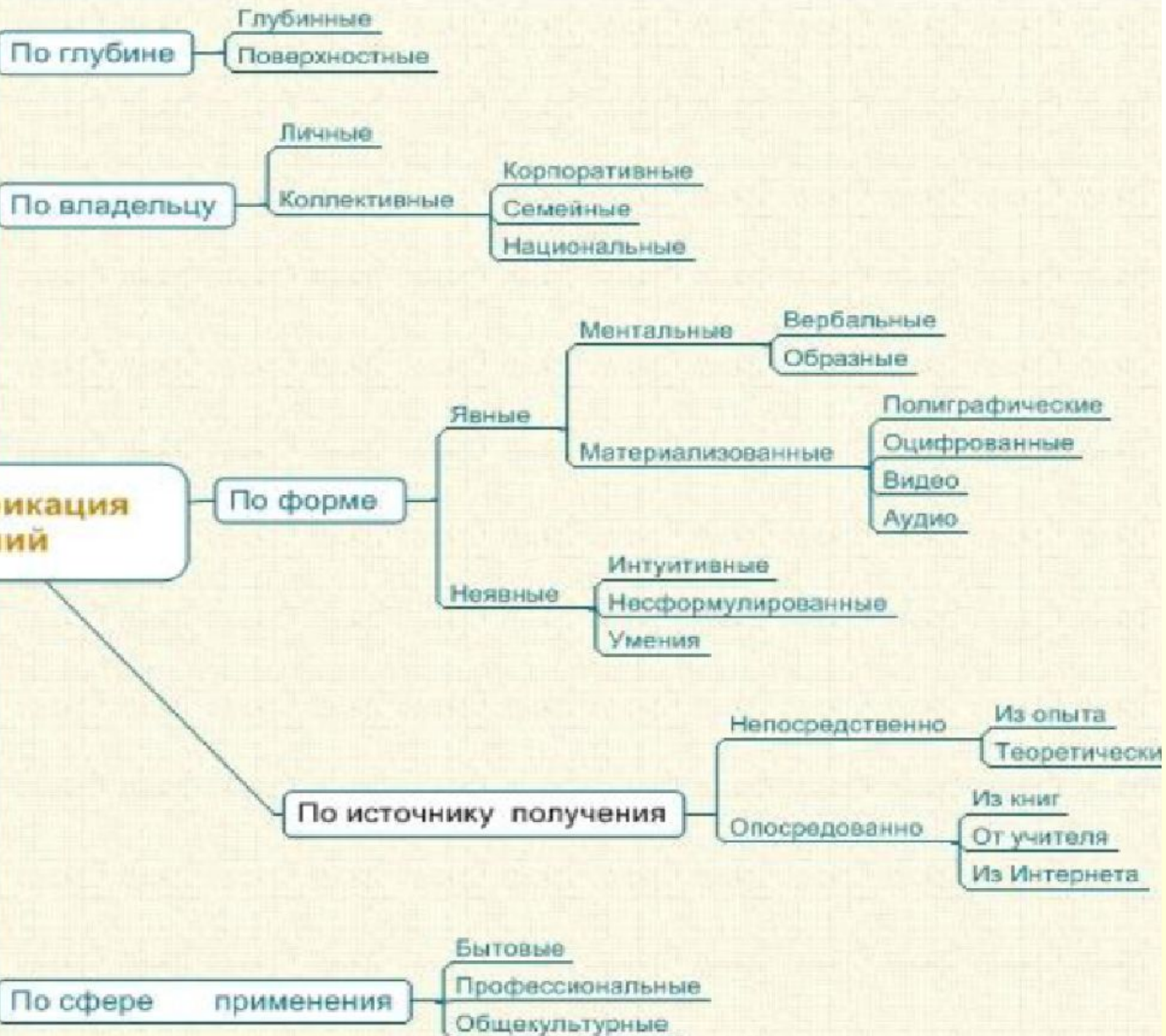
6. ПРИМЕРЫ

Пример таксономии понятия «деятельность»



Д. В. Кудрявцев «Системы управления знаниями и применение онтологий»
Санкт-Петербург, Изд-во Политехнического университета, 2010

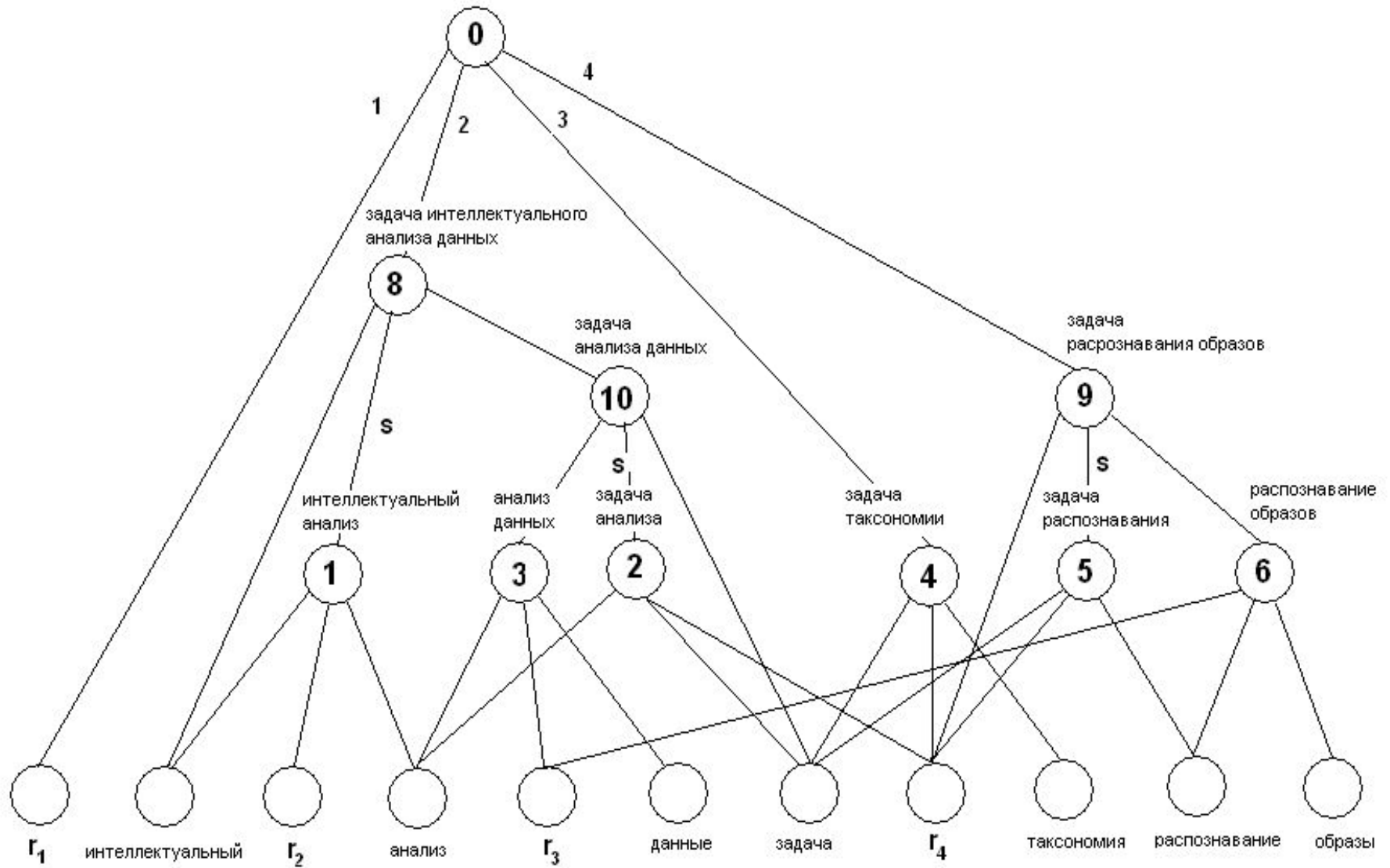
Классификация знаний



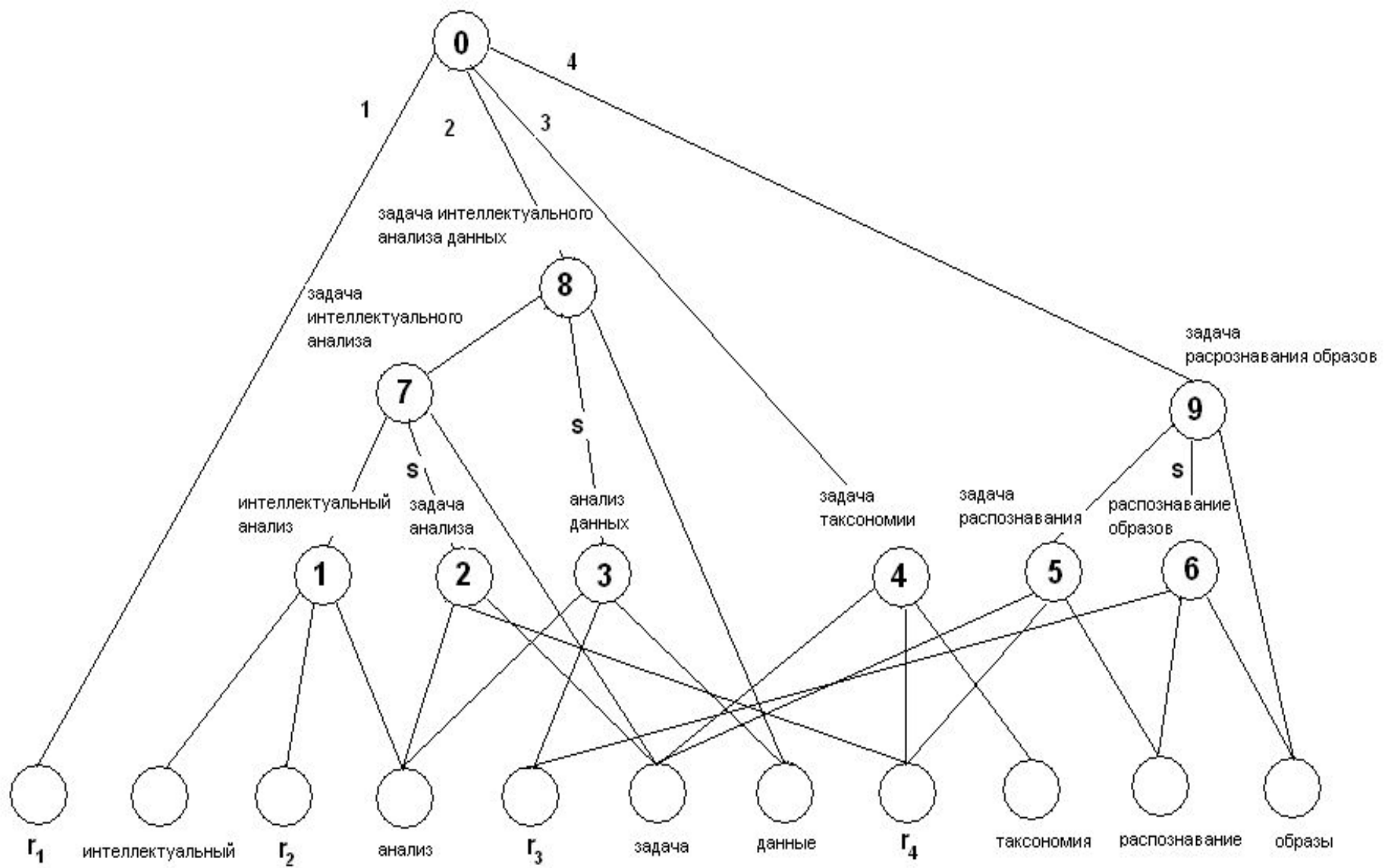
Онтологический инжиниринг



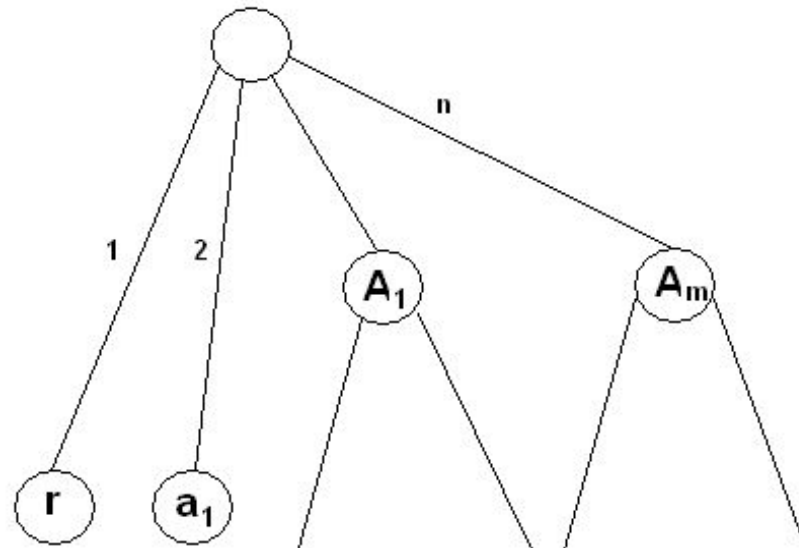
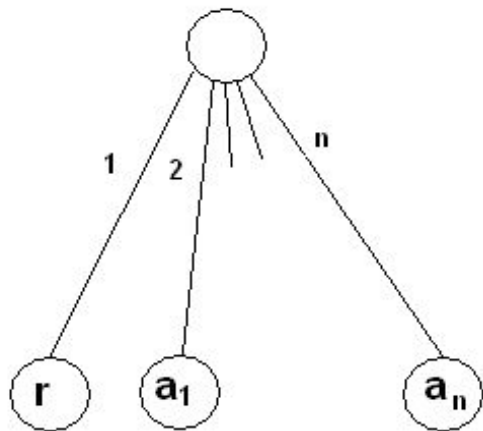
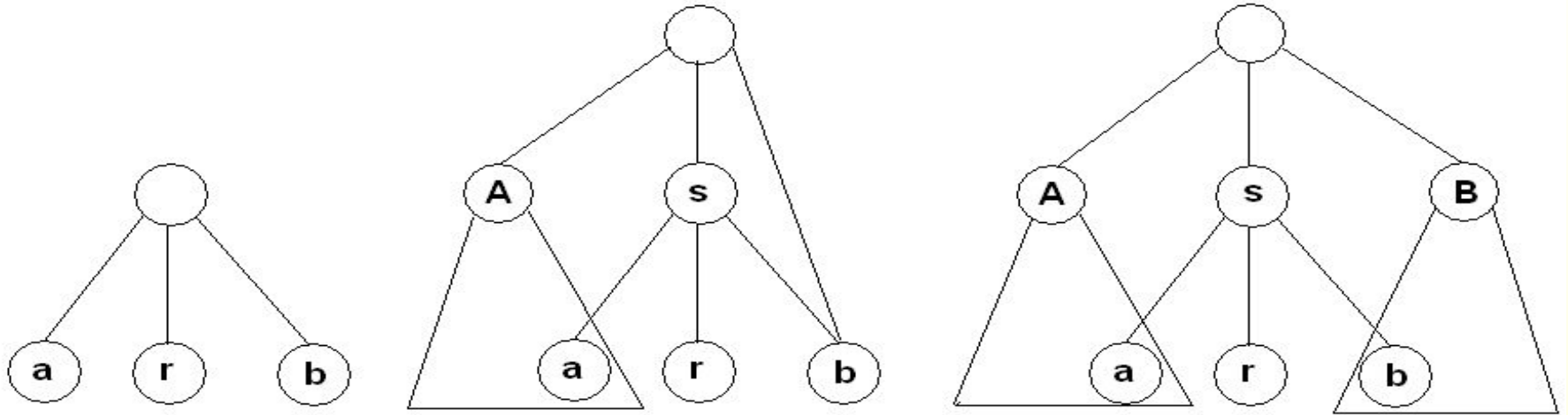
«Среди задач интеллектуального анализа данных имеются задача таксономии и задача распознавания образов»



Семантическая сеть (Q-сеть) ЗАГОРУЙКО Н.Г., которая объединяет в себе основные достоинства пирамидальных сетей В.П. Гладуна и семантических сетей Кузнецова И.П.



Q - сети



Логотипом журнала “Онтология проектирования” является образ, воплощающий в себе связь естественного и искусственного.

В качестве природного - взят важнейший элемент биологического объекта - крыло птицы, а искусственным объектом – выступает самолет. **Онтология этого процесса, процесса моделирования будущего, возвращает нас к сущностям, к существующему, к уже созданному бытию...**

Гибрид самолета и крыла птицы в данном изображении напоминают горящий факел. Созидательный огонь знаний, как луч позволит осветить путь к истине, к тайнам мироздания...



Онтология проектирования как научное направление включает в себя:

- исследование понятийного аппарата и разработки на его основе тезауруса,
- анализ критериев и моделей проектируемого объекта, методов и сценариев проектирования,
- сбор и обработку информации об объекте как системы и составляющих его элементах.

Проектирование как процесс и область знаний об артефакте является предметом онтологического анализа ученых-исследователей и узких специалистов.

Онтология проектирования, ее понятийный аппарат, ее базовые принципы - инвариантны к предметной области, в то время как само проектирование, как деятельность, всегда предметно, всегда объектно-ориентировано.

Развиваясь как научная дисциплина, онтология проектирования вбирает и обобщает накопленный опыт из разных предметных областей.

Онтология проектирования - это структурированное знание об объектах и методах проектирования, это совокупность семантических справочников, баз данных и процедур, алгоритмов оптимизации и учета проектной неопределенности.

Онтология проектирования - это также формализованное описание знаний субъектов проектирования о процессе проектирования новых или модернизаций уже известных артефактов, включая знания о самом объекте проектирования и близких к нему по свойствам артефактов, а также тезаурус предметной области.

Цель журнала - в рамках научной дискуссии попытаться приблизиться к замыслу Создателя, понять и объяснить суть процесса сотворения, развития, самоорганизации, моделирования и проектирования будущего через познание процесса создания артефактов, эволюции и отбора.

ТЕМАТИКА ЖУРНАЛА

“Во что ты веришь? В то, что все вещи должны быть наново взвешены”. Фридрих Ницше “Веселая наука”

Философские аспекты онтологии проектирования:

- истоки онтологического подхода в естествознании и инженерных науках;
- философия техники;
- онтологическая относительность и инженерные метаонтологии;
- парадигмы и принципы проектирования;
- пространство семантики и семиотики, психологии и биологии, акмеологии и социологии, феноменологии в проектной деятельности;
- суть, смысл и цель проектирования;
- роль гносеологии, когнитологии в онтологии проектирования;
- субъект и объект проектирования, природа их взаимодействия и развития.

Инжиниринг онтологий:

- когнитивные принципы формирования онтологий;
- языки формализации онтологии;
- инструментальные интегрированные среды инжиниринга онтологий;
- конструирование онтологий;
- нахождение семантических связей, подобия между онтологиями;
- композиция онтологий;
- динамика онтологий и денотативных баз данных;
- платформы, форматы, стандарты для инжиниринга онтологий;
- экспертные системы, оболочки и компоненты.

Прикладные онтологии проектирования:

- принципы использования онтологий в проектировании;
- проектирование, управляемое онтологией;
- проектирование, использующее онтологий;
- параллельное проектирование;
- онтологий предметных областей проектирования;
- интеллектуальные системы проектирования, использующие онтологий;
- технические справочники на основе онтологий;
- онтологий проектирования и обучающие системы;
- тезаурусы и базы знаний в проектировании.

**8. Формальный
концептуальный анализ
или
Анализ формальных
понятий**

Формальный концептуальный анализ (ФКА) — логико-алгебраический метод анализа данных, предложенный в 1981 году Рудольфом Вилле (1982).

В методе отражено философское понимание понятия как единицы мышления, определяемой своим объемом и содержанием.

ФКА предназначен для исследования объектов, которые задаются имеющимися у них свойствами.

Для установления связи между объектами и их свойствами служит формальный контекст.

Простой формальный контекст это тройка множество объектов, множество свойств, связь между объектами и свойствами.

ФОРМАЛЬНЫМ КОНТЕКСТОМ

называется тройка вида

$$K=(G, M, I),$$

где G и M – множества, а

I – отношение на множестве $G \times M$.

G представляет множество объектов,

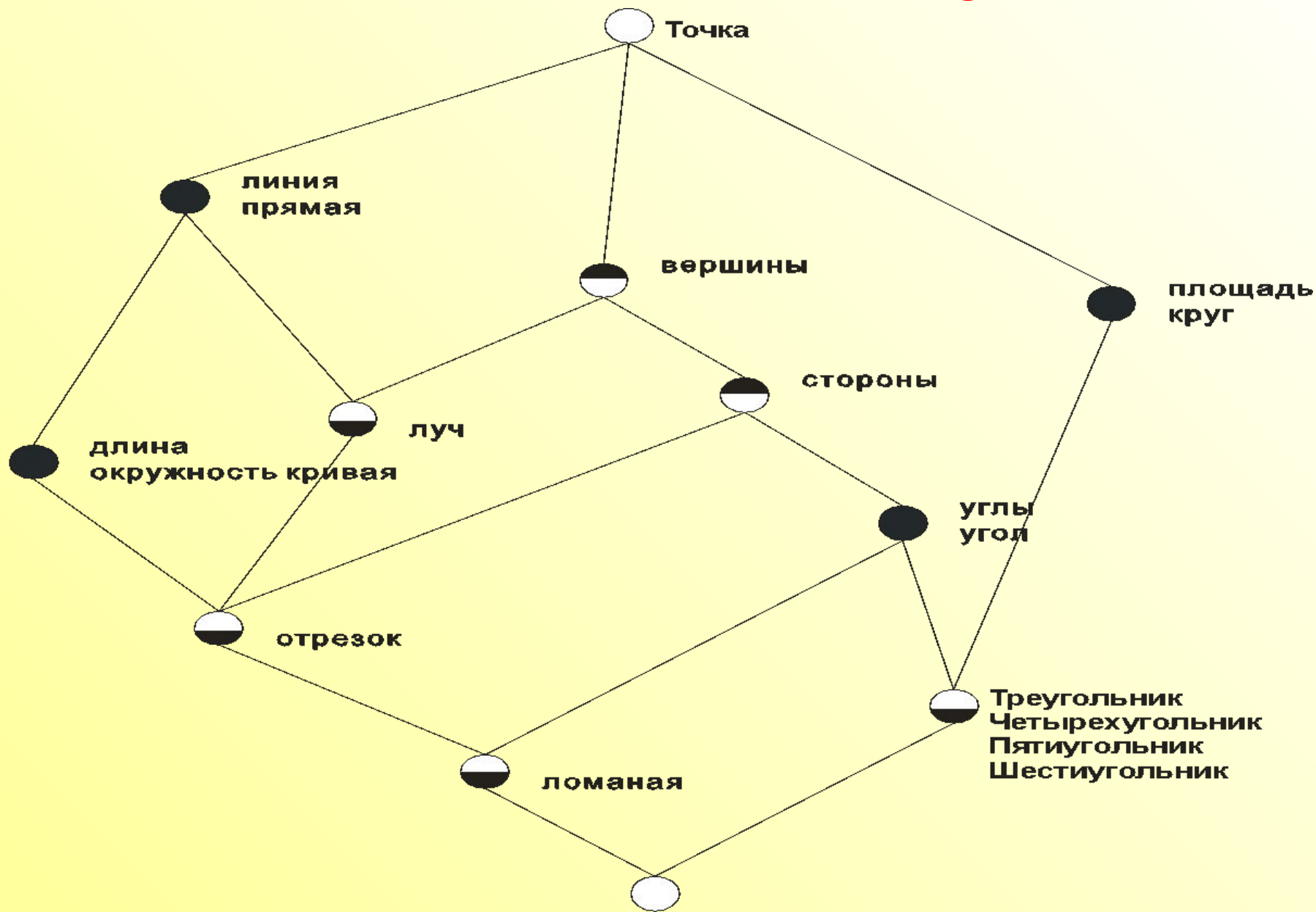
M – множество свойств, а

gIm означает, что объект g обладает

свойством m .

	Имеет вершины	Имеет длину	Есть линия	Имеет площадь	Имеет стороны	Имеет углы
Точка						
Прямая			х			
Луч	х		х			
Отрезок	х	х	х		х	
Угол	х				х	х
Круг				х		
Окружность		х	х			
Кривая		х	х			
Ломаная	х	х	х		х	х
Треугольник	х			х	х	х
Четырехугольник	х			х	х	х
Пятиугольник	х			х	х	х
Шестиугольник	х			х	х	х

Линейная диаграмма контекста «Геометрические фигуры»



В ФКА *понятия* принято называть *формальными концептами*. Множество объектов представляет объем (экстенсионал) понятия (концепта), а множество всех свойств, которыми они обладают, – его содержание (интенсионал).

Математически формальное понятие определяется с помощью соответствий Галуа.

Соответствие Галуа [Биркгоф 1984] – это пара отображений

$$s: G \rightarrow M,$$

$$t: M \rightarrow G,$$

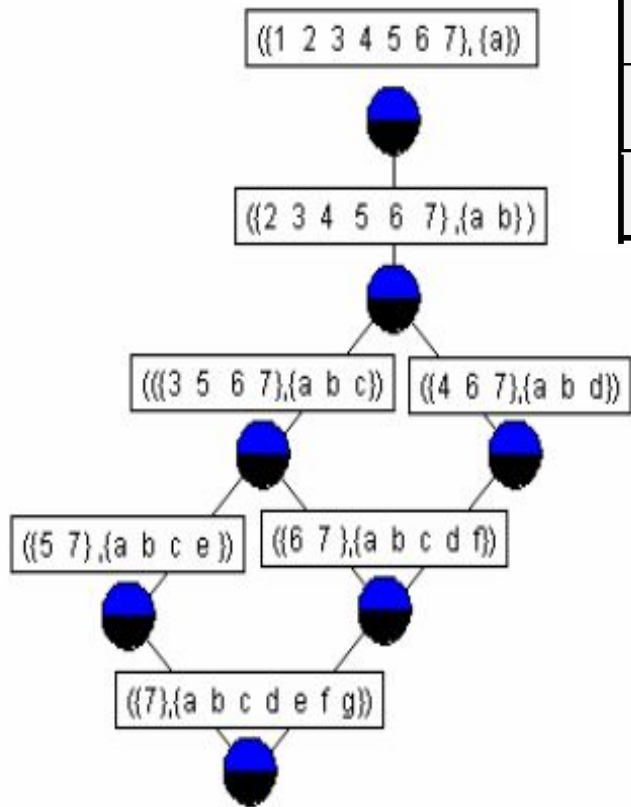
такое, что если $A \subseteq G$ и $B \subseteq M$, то

$$s(A) := \{m \in M \mid glm \text{ для всех } g \in A\},$$

$$t(B) := \{g \in G \mid glm \text{ для всех } m \in B\}.$$

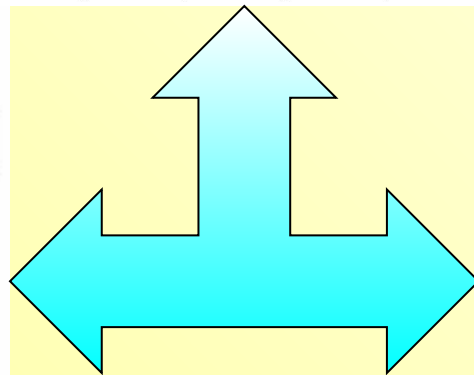
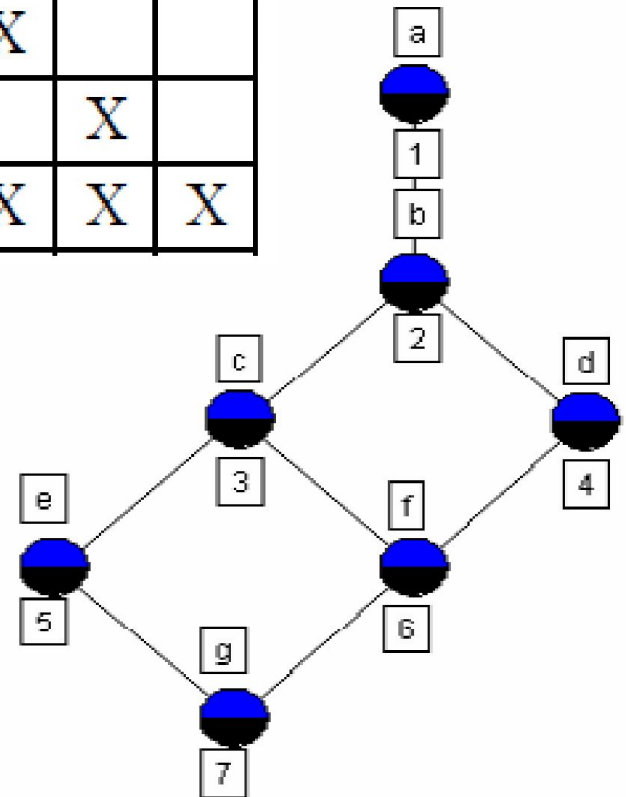
Формальный контекст

Диаграмма Хассе соответствующей концептуальной решетки



	a	b	c	d	e	f	g
1	X						
2	X	X					
3	X	X	X				
4	X	X		X			
5	X	X	X		X		
6	X	X	X	X		X	
7	X	X	X	X	X	X	X

Линейная диаграмма концептуальной решетки



Естественное шкалирование выборки

G \ M	w	y	g	b	f	\bar{f}	s	\bar{s}	r	\bar{r}	фрукт
яблоко		×				×	×		×		+
грейпфрут		×				×		×	×		+
киви			×			×		×		×	+
слива				×		×	×			×	+
кубик			×		×		×			×	-
яйцо	×				×		×			×	-
теннисный мяч	×					×		×	×		-

Сокращения:

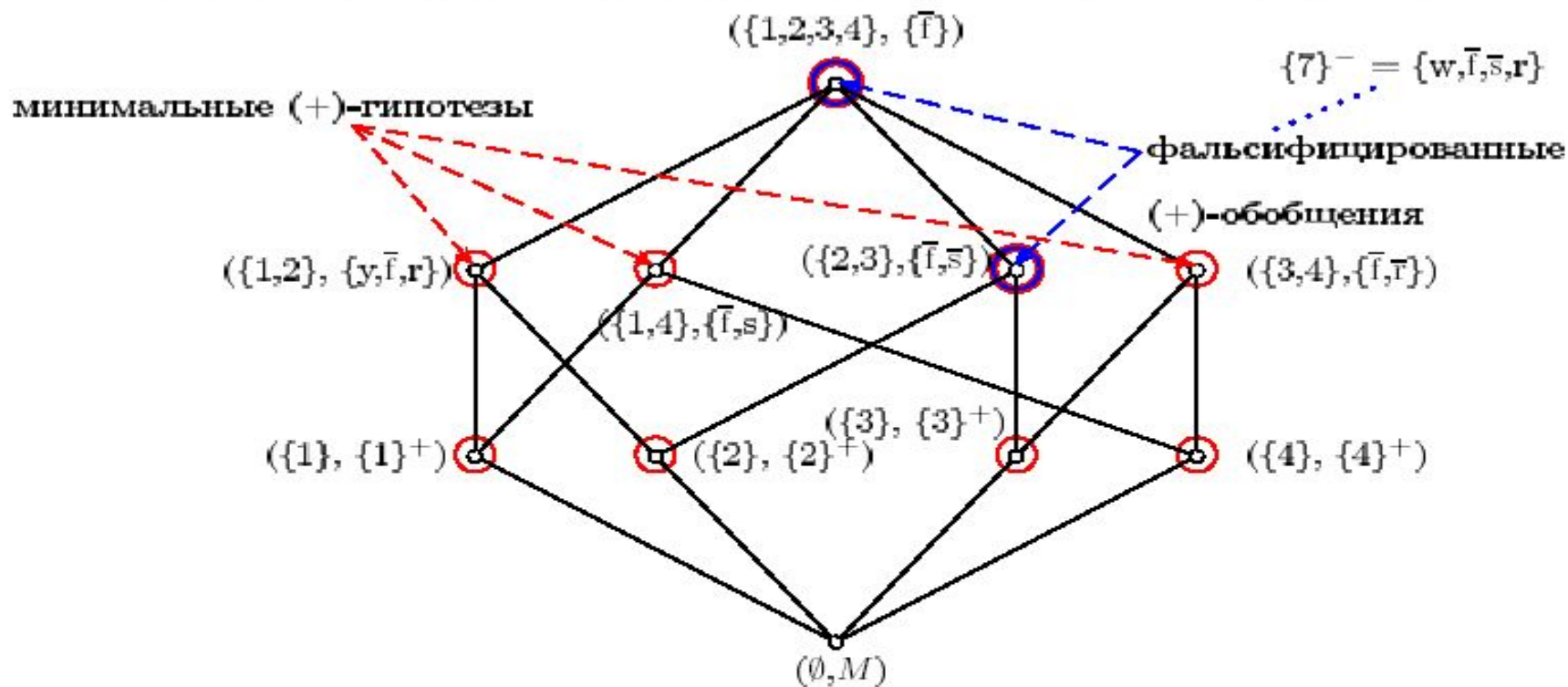
“g” - зеленый, “y” - желтый, “w” - белый, “r” - твердый, “ \bar{r} ” - нетвердый,

“s” - гладкий, “ \bar{s} ” - негладкий, “r” - круглый,

“ \bar{r} ” - некруглый.

Кузнецов Сергей Олегович «Приобретение знаний и построение онтологий с помощью решеток формальных понятий» КИИ-2008, Дубна

Решетка понятий положительного контекста



$G \setminus M$	w	y	g	b	f	f	v	s	r	r	Фрукт
яблоко		x				x	x		x		+
грейпфрут		x				x		x	x		+
киви			x			x		x		x	+
слива				x		x	x			x	+
кубик			x		x		x			x	-
яйцо	x				x		x			x	-
теннисный мяч	x					x		x	x		-

Кластерная понятийная система

Понятие понятийного кластера N_i определяется тройкой:

$$N_i = \langle d_i, F(d_i), V(d_i) \rangle,$$

где d_i — *имя* — это знак, обозначающий понятие N_i ,

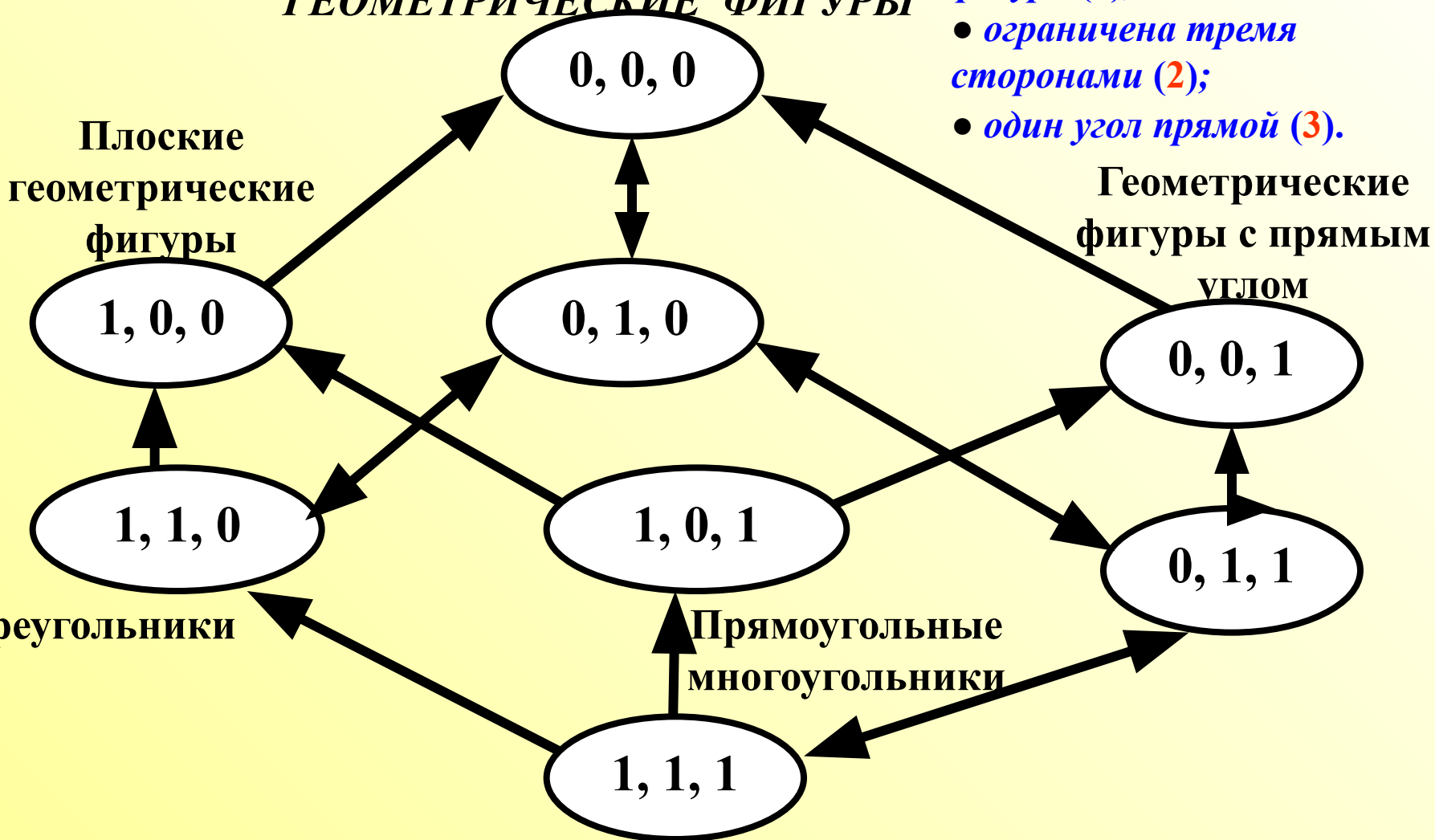
$F(d_i)$ — *содержание* — множество признаков этого понятия,

$V(d_i)$ — *объем понятия* — множество объектов реального мира, удовлетворяющих содержанию понятия, т. е. имеющих все признаки, включенные в содержание понятия.

Понятийный кластер начального понятия *прямоугольный треугольник*

- *плоская геометрическая фигура (1);*
- *ограничена тремя сторонами (2);*
- *один угол прямой (3).*

ГЕОМЕТРИЧЕСКИЕ ФИГУРЫ



ПРЯМОУГОЛЬНЫЕ ТРЕУГОЛЬНИКИ

9. МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В БЗ

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ



Основные типы отношений

- Родо-видовые (АКО A-kind-of, "класс - подкласс", "элемент-множество", «роза- это цветок», «ЭТО», «НАПРИМЕР»),
- связи типа «часть – целое», «имеет частью» и т.п.,
- функциональные связи (определяемые обычно глаголами: "производит", "влияет" ...),
- атрибутивные связи (иметь свойство, иметь значение),
- причинно-следственные (если-то)
- количественные (больше, меньше, равно...),
- пространственные (далеко от , близко от, за, над ...),
- временные (раньше, позже, в течение...),
- логические связи (и, или, не),
- лингвистические связи (синонимия, антонимия) и др.

Фреймы

Термин *фрейм* (от английского *frame*, что означает "каркас" или "рамка") был предложен Марвином Минским в 70-е годы как структура знаний для восприятия пространственных сцен.

Фрейм – это абстрактный образ для представления некоего стереотипа восприятия.

Различают *фреймы-образцы* или *прототипы* и *фреймы-экземпляры* или *примеры*.

Определение и классификация знаний



Логические языки представления знаний

Логические ЯПЗ опираются на представление всех знаний в виде синтаксически правильных формул какой-либо формальной логической системы. Напомним, что формальная логическая система задается четверкой компонент вида

$$F = (T, P, A, W),$$

где **T** — множество базовых элементов, из которых строятся все высказывания в **F**,

P — множество синтаксических правил, определяющие среди всех возможных выражений из базовых элементов те, которые синтаксически правильны,

A — множество аксиом **F**, образующее подмножество в множестве синтаксически правильных формул, которым априорно приписывается статус истинности,

W — правила вывода, позволяющие получать из аксиом новые синтаксически правильные формулы, которым можно приписывать статус истинности.

1. Множество T есть множество базовых элементов различной природы, например слов из некоторого *ограниченного* словаря, деталей детского конструктора, входящих в состав некоторого набора и т.п. Важно, что для множества T существует некоторый способ определения принадлежности или непринадлежности произвольного элемента к этому множеству. Процедура такой проверки может быть любой, но за конечное число шагов она должна давать положительный или отрицательный ответ на вопрос, является ли x элементом множества T . Обозначим эту процедуру $P_1(T)$.

2. Множество P есть множество синтаксических правил. С их помощью из элементов T образуют *синтаксически правильные совокупности*. Например, из слов ограниченного словаря строятся синтаксически правильные фразы, из деталей детского конструктора с помощью гаек и болтов собираются новые конструкции.

Декларируется существование процедуры $P_2(P)$, с помощью которой за конечное число шагов можно получить ответ на вопрос, является ли построенная совокупность *синтаксически правильной*.

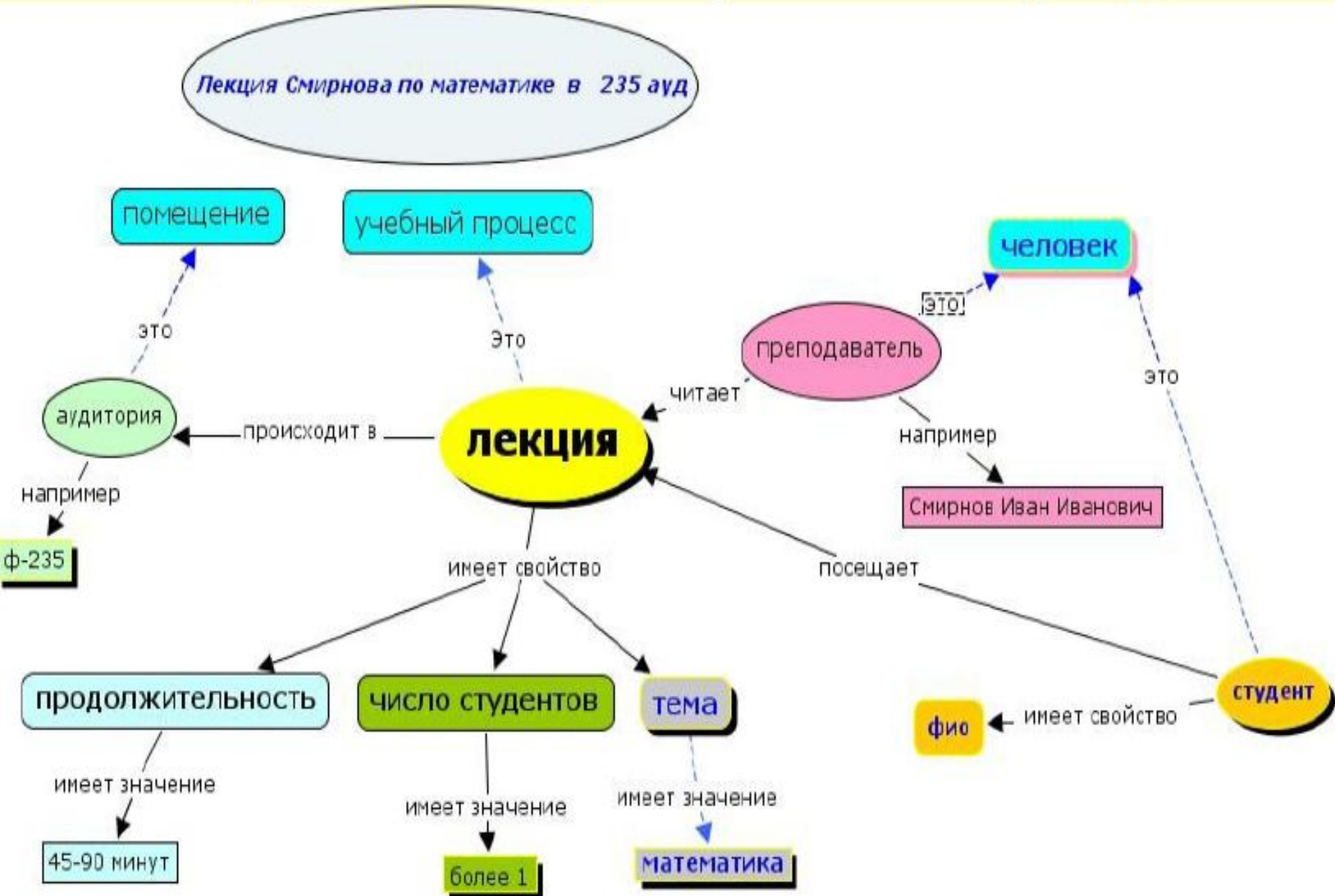
3. В множестве синтаксически правильных совокупностей выделяется некоторое подмножество **A**.

Элементы **A** называются аксиомами. Как и для других составляющих формальной системы, должна существовать процедура $P_3(A)$, с помощью которой для любой синтаксически правильной совокупности можно получить ответ на вопрос о принадлежности ее к множеству **A**.

4. Множество **W** есть множество правил вывода. Применяя их к элементам **A**, можно получать новые синтаксически правильные совокупности, к которым снова можно применять правила из **W**. Так формируется множество выводимых в данной формальной системе совокупностей. Если имеется процедура $P_4(W)$, с помощью которой можно определить для любой синтаксически правильной совокупности, является ли она выводимой, то соответствующая формальная система называется разрешимой.

Это показывает, что именно правило вывода является наиболее сложной составляющей формальной системы.

Пример концептуального графа



УРОВНИ ПОНИМАНИЯ

Результаты исследований по пониманию ЕЯ к середине 80-х годов XX века составили серьезный базис для разработки и реализации ЕЯ-систем. Вместе с тем, даже наличие развитых моделей всех уровней, рассмотренных выше не позволяет, как показали работы Д.А. Поспелова, обеспечить понимание ЕЯ в достаточно полном объеме. Вот почему, Дмитрием Александровичем акцент исследований был сдвинут из сферы лингвистики в области, психолингвистики и когнитивной психологии. Согласно вышеуказанной работе в рамках функционирования ЕЯ-систем можно выделить 5 основных уровней понимания и 2 уровня метапонимания.

1. Первый уровень характеризуется тем, что любые ответы на вопросы система формирует только **на основе прямого содержания** введенного в нее текста (при этом в лингвистическом процессоре осуществляются морфологический, синтаксический и семантический анализы текста и вопросов, интерпретация результатов которых приводит к формированию их внутреннего представления, с чем работает блок вывода).

2. Второй уровень отличается от первого тем, что здесь, в систему **включаются средства вывода на логиках текста** (временных, пространственных, каузальных и др.), что дает возможность порождать информацию, отсутствующую в тексте явно.

3. Третий уровень характеризуется тем, что к средствам первых двух уровней добавляются **правила пополнения текста знаниями системы о среде** (как правило, в виде сценариев или процедур иного типа).

4. Следующие два уровня понимания даже в наиболее продвинутых системах того времени были реализованы лишь частично. При этом на **четвертом уровне предполагается пополнение текста невербальной** (например, зрительной) информацией, позволяющей соотнести текст с реальной ситуацией его порождения.

5. А на пятом уровне кроме текста и его расширения четвертого уровня используются информация **о конкретном субъекте, являющемся источником текста, и знания о коммуникации** (модель диалога, цели общения, нормы коммуникации, оценки коммуникантов и т.п.), которые опирались в то время на теорию речевых актов.

7 и 8. Последние два уровня в момент разработки модели понимания были исследованы слабо.

На первом метауровне происходит изменение содержимого БЗ системы, что приводит к существенным сложностям, связанным с немонотонностью вывода на знаниях, а на втором метауровне идет порождение метафорического знания на основе вывода по аналогии и ассоциации.

Для метауровней важно и то, что здесь возникают новые процедуры манипулирования знаниями и процесс этот носит принципиально открытый характер.

7. Разработка ОНТОЛОГИЙ 101

Разработка онтологий 101: руководство по созданию Вашей первой онтологии

В большинстве американских колледжей

**вступительный курс любого предмета имеет номер «101»:
«Химия 101», «Биология 101» и т.д.**

**Следующие два более углубленных курса по химии
назывались бы «Химия 102» и «Химия 103»
соответственно. В США номер «101» означает
«Введение». Т.е., название работы нужно понимать как
«Введение в разработку онтологий: Руководство по
созданию Вашей первой онтологии».**

Вопросы для проверки компетентности

Один из способов определить масштаб онтологии – это набросать [список вопросов](#), на которые должна ответить база знаний, основанная на онтологии, т.е. **ВОПРОСЫ ДЛЯ ПРОВЕРКИ КОМПЕТЕНТНОСТИ.**

Эти вопросы будут служить лакмусовой бумажкой:

- Содержит ли онтология достаточно информации для ответа на эти типы вопросов?
- Требуется ли для ответов особый уровень детализации или представление определенной области?

Эти вопросы для проверки компетентности являются всего лишь формальными и не должны быть исчерпывающими.

В области вина и еды возможны следующие

вопросы для проверки компетентности:

- 1. Какие характеристики вина мне следует учитывать при выборе вина?**
- 2. Вино Bordeaux красное или белое?**
- 3. Хорошо ли сочетается Cabernet Sauvignon с морскими продуктами?**
- 4. Какое вино лучше всего подойдет к жареному мясу?**
- 5. Какие характеристики вина влияют на его сочетаемость с блюдом?**
- 6. Влияет ли с год производства вина на его букет или крепость?**
- 7. Какие урожаи Napa Zinfandel были хорошими?**

Судя по этому списку вопросов, онтология будет включать информацию

- о различных характеристиках вина и типах вин,**
- годах производства вин (хороших и плохих),**
- классификациях еды, которые нужно учесть при выборе подходящего вина,**
- рекомендуемых сочетаниях вина и**

Шаг 1. Определение области и масштаба онтологии

Мы предлагаем начать разработку онтологии с определения ее области и масштаба.

То есть, ответим на несколько основных вопросов:

- Какую область будет охватывать онтология?
- Для чего мы собираемся использовать онтологию?
- На какие типы вопросов должна давать ответы информация в онтологии?
- Кто будет использовать и поддерживать онтологию?

**СПАСИБО ЗА
ВНИМАНИЕ**

**ТВОРЧЕСКИХ
УСПЕХОВ!**