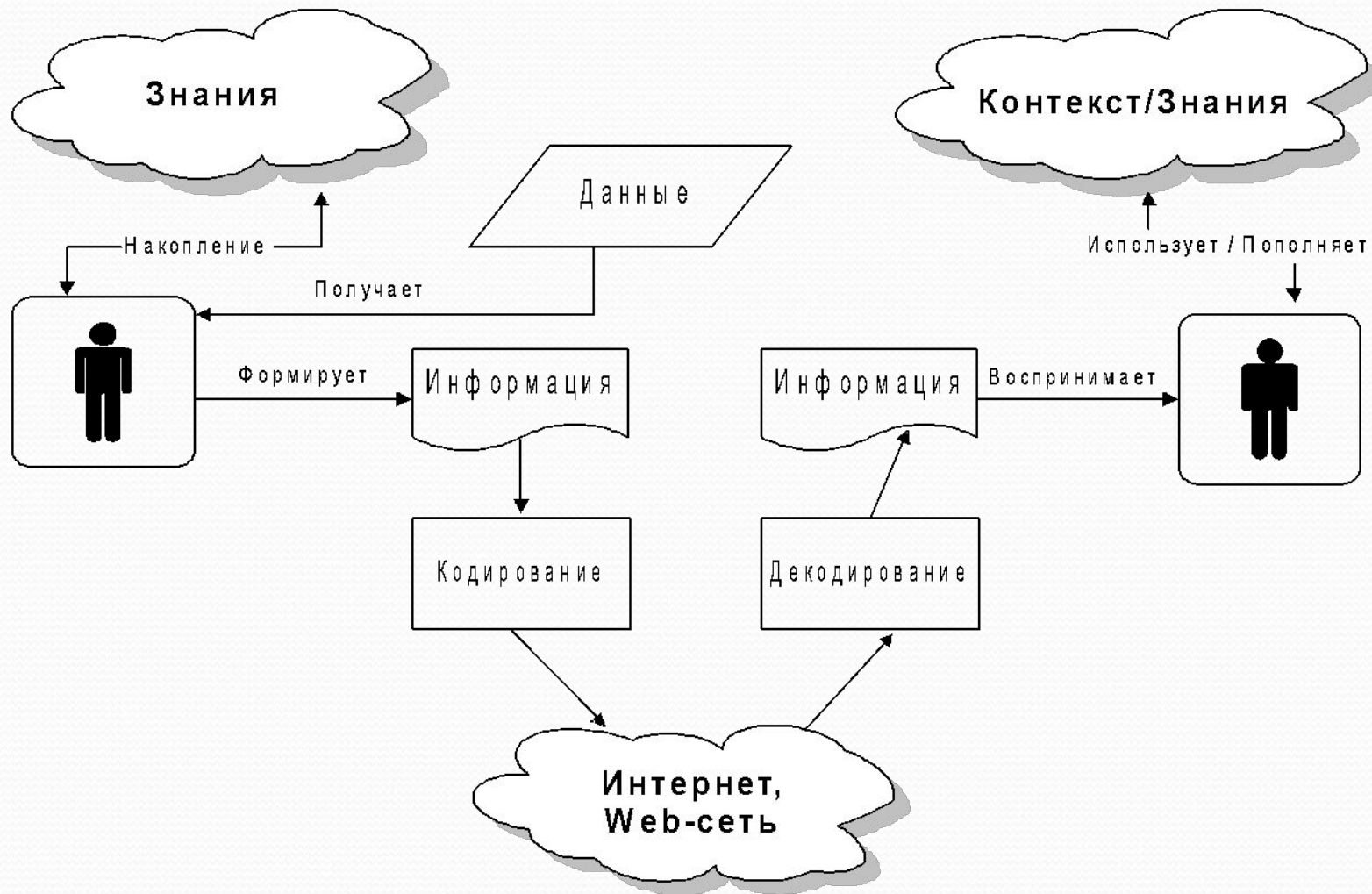


# Онтологические модели представления знаний

- Существует множество обстоятельств, которые затрудняют распространение и обмен знаниями между людьми, столь необходимыми для их систематического пополнения и накопления.
- Прежде всего, это **обстоятельства количественного характера**, связанные с быстрым ростом численности населения, с вовлечением новых поколений в различные сферы деятельности, требующие постоянно растущего уровня знаний, умений и навыков.
- Особым обстоятельством являются **фундаментальные отличия и множественность национальных языков народов, населяющих планету Земля**. По данным ЮНЕСКО на нашей планете существует более 2700 языков, народов и народностей.

- Но дело не только в том, что народы мира думают, говорят и пишут на разных языках. Многие проблемы в обмене и создании знаний связаны с неоднозначным или неадекватным восприятием смысла данных, информации, знаний различными участниками знаниевого процесса. Дело в том, что в цепи передачи знаний (рис) отправитель и получатель знания зачастую пользуются различными представлениями, различной терминологией и понятийным аппаратом. Из-за различий в образовании и в предшествующем опыте они могут руководствоваться различными моделями деятельности и культурой мышления.

# Движение знаний между отправителем и получателем



- Вследствие этих объективных обстоятельств весьма желательно, чтобы информация и знания были структурированы и описаны таким образом, чтобы получатель (пользователь) был способен понять и текст, и контекст (смысл) сообщения. В идеале, сообщение (знаниевая сущность) должна структурироваться таким образом, чтобы компьютер, а не только образованный человек был способен «понять» его. Под словом «понять» здесь имеется в виду, что компьютер будет способен обработать документ (знаниевую сущность) посредством использования известных ему правил с помощью некоторого логического языка, а также будет способен вывести новые факты и знания из данного документа.

# Онтологии предметных областей

- Онтологии предметной области описывают явные знания, которые имеются в компании (или в отдельных ее частях). Описанием знаний уже достаточно давно занимается такая дисциплина, как «Искусственный интеллект» (ИИ), а также такие ее разделы, как «Представление знаний» и «Инженерия знаний». Учитывая, что ИИ занимается работой со знаниями с 50-х годов, в данной дисциплине накоплен достаточно большой опыт в области представления (моделирования) знаний. В вопросе описания знаний дисциплина «Управления знаниями» имеет общие интересы с ИИ. У них общий объект исследования – знания, но цели его исследования в этих дисциплинах разные.
- Цель ИИ заключается в создании моделей и методов работы со знаниями, которые позволят их использовать без участия (или почти без участия) человека. Например, в [Люггер Д.Ф., 2003] дается следующее определение: «ИИ можно определить как область компьютерной науки, занимающейся автоматизацией разумного поведения».

- **Онтология** (от древнегреч. онтос – сущее, логос – учение, понятие) – термин, определяющий учение о сущем, бытии, в отличие от гносеологии – учение о познании. В философском смысле, а этот термин заимствован из философии, онтология есть определенная система категорий, являющихся следствием определенных взглядов на мир.
- Термин «онтология» был использован рядом исследовательских сообществ по ИИ вначале в области инженерии знаний, в обработке естественных языков, а затем в представлении знаний. В конце 1990-х годов понятие онтологии также стало широко использоваться в таких областях, как интеллектуальная интеграция информации, поиск информации в Интернет и управление знаниями [Jos de Bruijn, Fensel D., [Staab](#). В конце 1990-х годов понятие онтологии также стало широко использоваться в таких областях, как интеллектуальная интеграция информации, поиск информации в Интернет и управление знаниями [Jos de Bruijn, Fensel D., Staab S., [Studer R.](#), 2004]. Позже онтологии стали рассматриваться в качестве ключевого элемента в проекте Семантической Сети – нового этапа развития сети WWW (Word Wide Web). Если существующая Web-сеть – это огромное множество документов, которые связаны перекрестными ссылками, то создаваемая Семантическая Сеть должна добавить к существующей сети множество онтологий и метаописаний знаний, содержащихся в документах Web-сети (включая стандарты и программные инструменты) [Berners-Lee T., Hendler J., Lassila O., 2001].

- Определение онтологии, которым руководствуются многие исследователи в данной области, было дано в [Gruber T.A., 1995].
- «Онтология – это формальное, явное, точное определение (спецификация) совместно используемой концептуализации». *Концептуализацией* именуется абстрактное упрощенное представление мира, которое формируется для некоторых целей. Онтология является *точным определением (спецификацией)* потому, что она представляет концептуализацию в конкретной форме. Она является *явной*, потому что все используемые в ней ограничения явно определены. Слово *формальная* означает, что онтология должна пониматься машиной. Слово *совместно используемая* указывает на то, что онтология содержит согласованные знания.

- Рабочим и более приближенным к управлению знаниями можно признать определение, приведенное в [Гаврилова Т.А., Хорошевский Ф. В., 2001]: «Онтология – это базы знаний специального типа, которые могут «читаться» и пониматься, отчуждаться от их разработчика и /или физически разделяться их пользователями».
- Онтология состоит из терминов, организованных в таксономию, их определений, атрибутов, а также связанных с ними аксиом и правил вывода.



- **Онтология, таким образом, соединяет человеческое и компьютерное понимание символов. Эти символы, также называемые терминами (точными определениями понятий), могут интерпретироваться как людьми, так и машинами. Термин понятен для человека, так как это слово, написанное на естественном языке. Понятны человеку и связи между терминами типа «суперпонятие – подпонятие» (род – вид), обычно обозначаемые как *is-a* (являться). Эта связь обозначает тот факт, что одно понятие (субпонятие) является более общим, чем другое (подпонятие). В качестве примера возьмем такое понятие, как компьютер, которое является менее общим, чем понятие машина (автомобиль, трактор, танк и т.д.).**

В соответствии с этим пространство знаний (интеллектуальное пространство) организации предлагается [Тузовский А.Ф., Ямпольский В.З. Интеллектуальное пространство, 2004] описывать следующим образом:

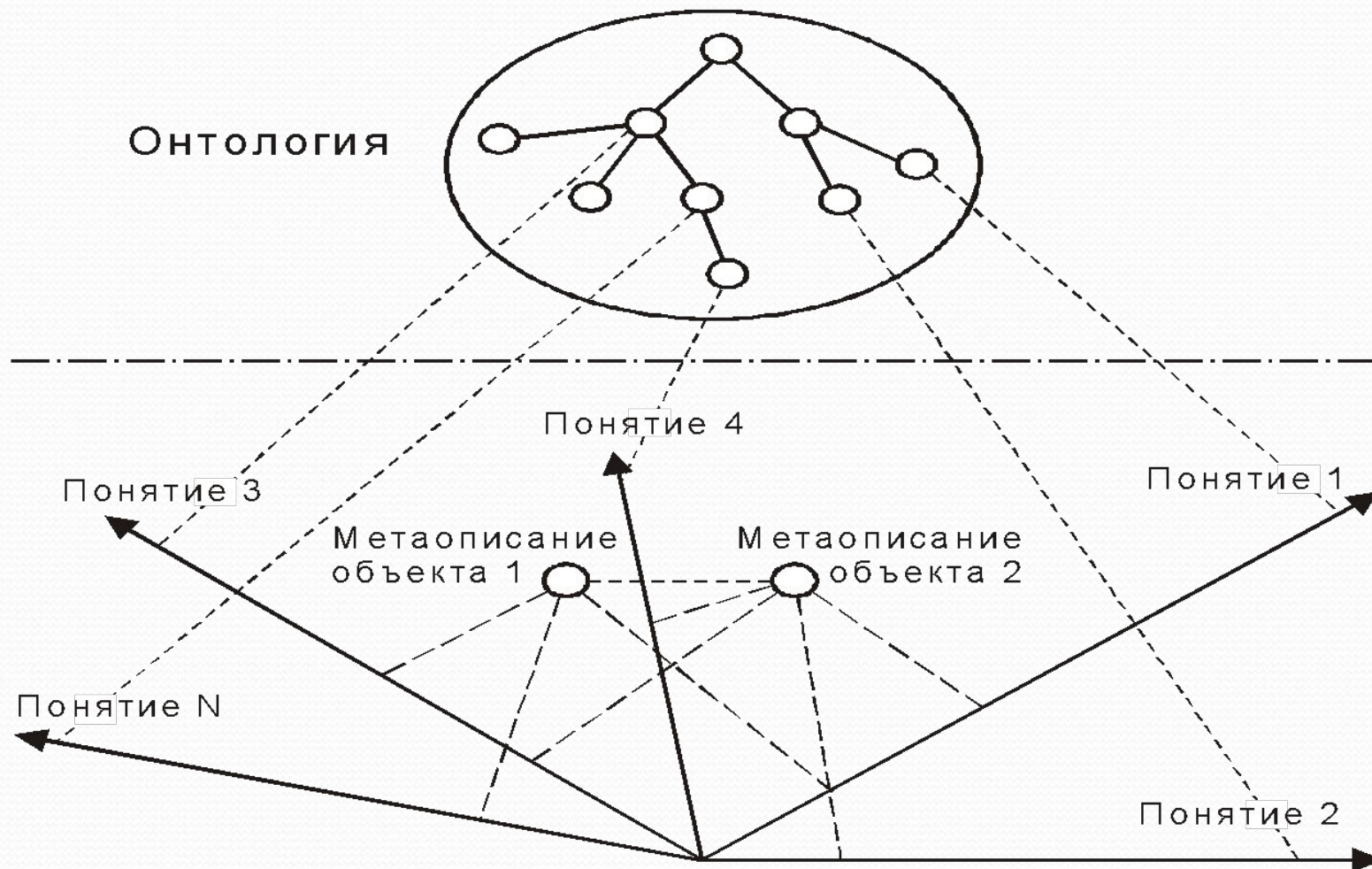
- В качестве системы координат использовать *онтологию предметной области*.
- Описания объектов, содержащие знания, задавать в виде их *метаописаний*, составленных из основных понятий онтологии.
- В качестве меры близости объектов (метрики) использовать *семантическую близость их метаописаний*.

# Модель интеллектуального пространства

Модель многомерного пространства является признанной в разных областях науки абстракцией, которая используется для работы с различными и не только математическими описаниями объектов. Известно, что описание любого пространства включает такие элементы, как:

- выбор системы координат;
- задание способа описания положения объектов в выбранной системе координат;
- задание метрики (способа вычисления) близости объектов в данном пространстве.

# Схема интеллектуального пространства



В настоящее время существуют и развиваются разные методы представления и описания знаний, например, такие, как: **продукционные модели, семантические сети, фреймы, онтологии.**

- **Продукционная модель** или модель, основанная на правилах, позволяет представлять знания в виде предположения типа «**if – then**»: если (условие), то (действие). Под «условием» понимается некоторое предложение – образец, по которому осуществляется поиск в базе знаний, а под «действием» – выполняемые в результате успешного поиска действия.
- Продукционные модели чаще всего применяются в промышленных экспертных системах при фиксации совокупности правил поведения персонала в некотором множестве ситуаций.

**Семантическая сеть** означает «смысловая» сеть, а, собственно, семантика – это наука, устанавливающая отношения между символами и объектами, которые они обозначают. Иначе говоря, семантика – это наука, определяющая смысл знаков [Люггер Д.Ф., 2003].

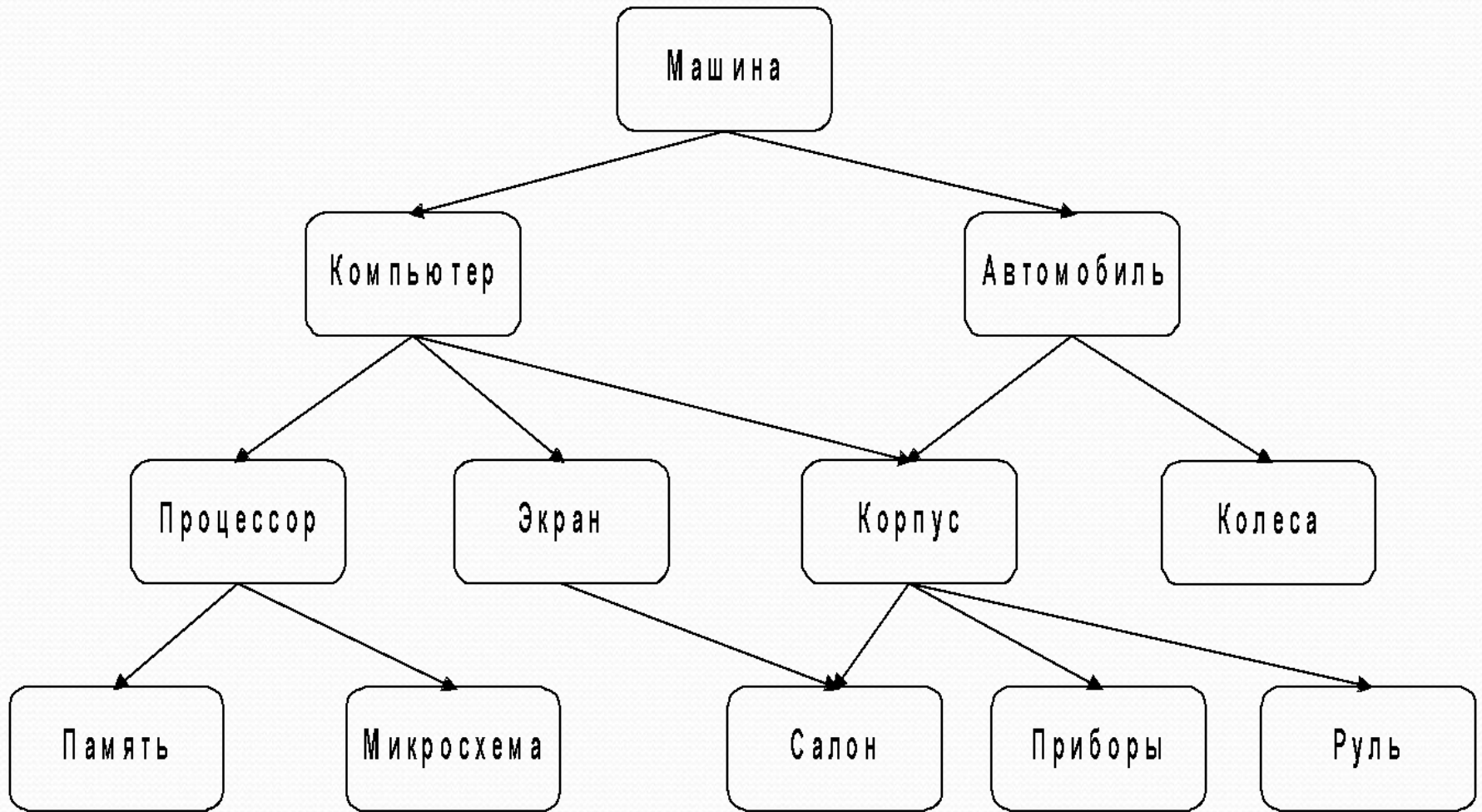
По своей структуре семантическая сеть – это ориентированный граф, вершины которого – понятия, а дуги – отношения между ними. Характерной особенностью семантических сетей является наличие трех типов отношений:

- отношение класс – элемент класса;
- отношение свойство – значение свойства;
- отношение фрагмент – элемент класса.

## Чаще всего в семантических сетях используются следующие отношения:

- связи типа часть – целое (например, элемент – класс);
- функциональные связи (определяются глаголами типа «производит», «влияет»...);
- количественные отношения (больше, меньше, равно);
- пространственные отношения (далеко от, близко от...);
- временные отношения (раньше, позже, в течение ...);
- атрибутивные связи (иметь свойство, иметь значение);
- логические связи (И, ИЛИ, НЕ);
- лингвистические связи и др.

# Пример *is-a* иерархии (таксономии)





# Формальная модель онтологии

В общем виде формальная модель онтологии может быть описана следующим кортежем [Meadche A., Zacharias V., 2002]:

$$O = \{L, C, F, G, H, R, A\}, \text{ где}$$

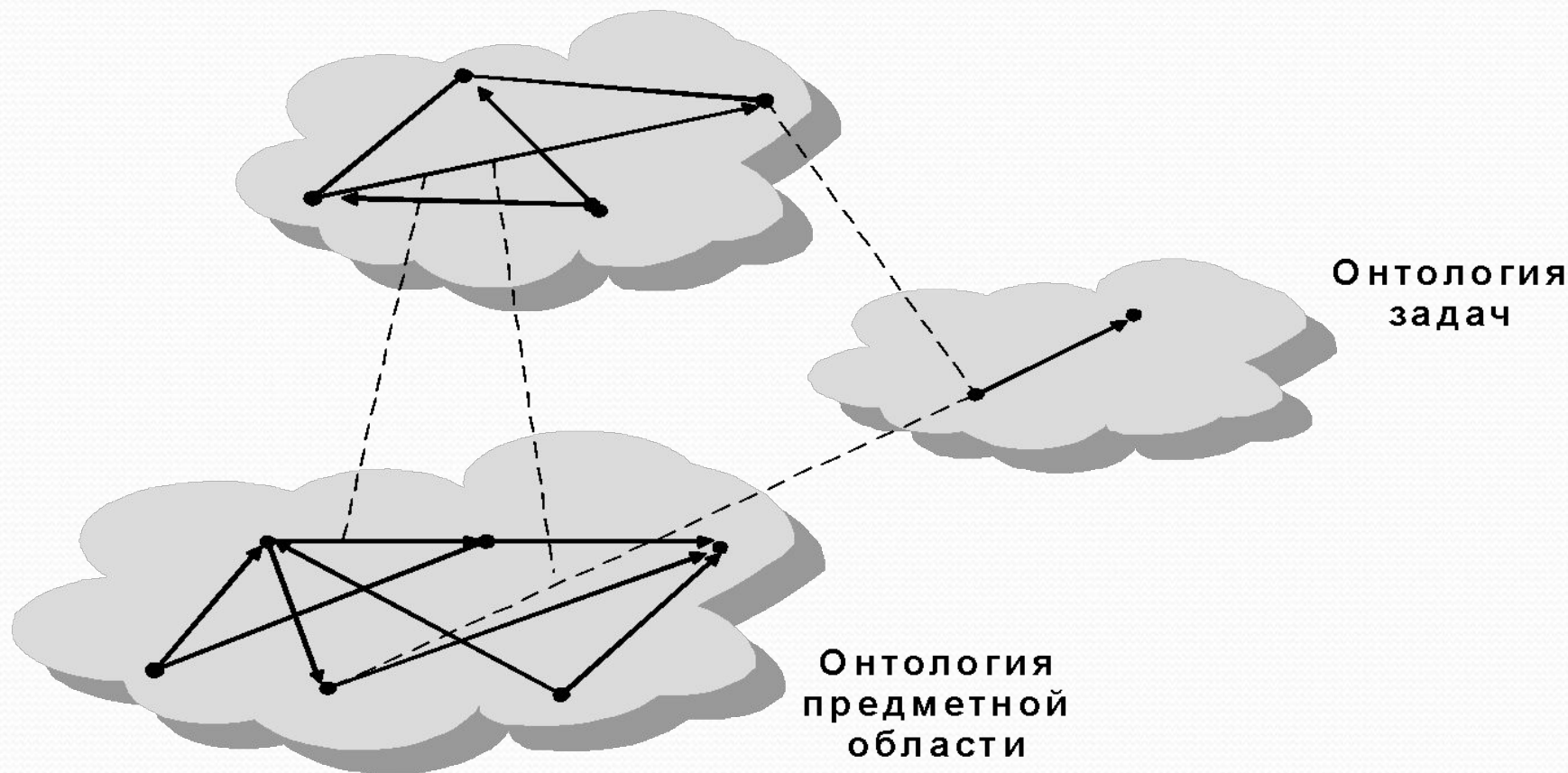
- $L = LC \cup LR$  – словарь онтологии, содержащий набор лексических единиц (знаков) для понятий  $LC$  и набор знаков для отношений  $LR$ ;
- $C$  – набор понятий онтологии, причем для каждого понятия  $c \in C$  в онтологии существует по крайней мере одно утверждение;
- $F$  и  $G$  – функции ссылок такие, что  $F: FLC \rightarrow {}_2C$  и  $G: FLR \rightarrow {}_2R$ . То есть  $F$  и  $G$  связывают наборы лексических единиц  $\{L_j\} \subset L$  с наборами понятий и отношений, на которые они соответственно ссылаются в данной онтологии. При этом одна лексическая единица может ссылаться на несколько понятий или отношений и одно понятие или отношение может ссылаться на несколько лексических единиц. Инверсиями функций ссылок являются  $F^{-1}$  и  $G^{-1}$ ;
- $H$  – фиксирует таксономический характер отношений (связей), при котором понятия онтологии связаны нереклексивными, ациклическими, транзитивными отношениями  $H \subset C \times C$ . Выражение  $H(C_1, C_2)$  означает, что понятие  $C_1$  является подпонятием  $C_2$ ;
- $R$  – обозначает бинарный характер отношений между понятиями онтологии, фиксирующие пары области применения (domain)/области значений (range), то есть пары  $(D R)$  с  $D, R \in C$ ;
- $A$  – набор аксиом онтологии.

По уровню универсальности выделяют три типа онтологий:

- **Онтологии верхнего уровня**, или **метаонтологии**, описывающие общие понятия, независимо от задач конкретного домена. Примером такой онтологии может служить WordNet [Fellbaum С., 1998]. Масштаб WordNet весьма обширный – весь английский язык с описанием каждого термина, его синонимов и гипер/гипо (более/менее) общих терминов и отношений между ними. В то же время уровень детальности в WordNet очень низкий, имеются лишь описания на естественном языке терминов, которые не могут быть поняты машиной, и зафиксированы только самые простые отношения между ними.
- **Онтологии предметных областей** описывают относительно общие понятия для общих задач. В какой-то мере она относится к онтологиям верхнего уровня, так как ее можно использовать во множестве предприятий различных предметных областей.
- **Онтологии приложений** описывают понятия, зависящие как от домена, так и от решаемой задачи.

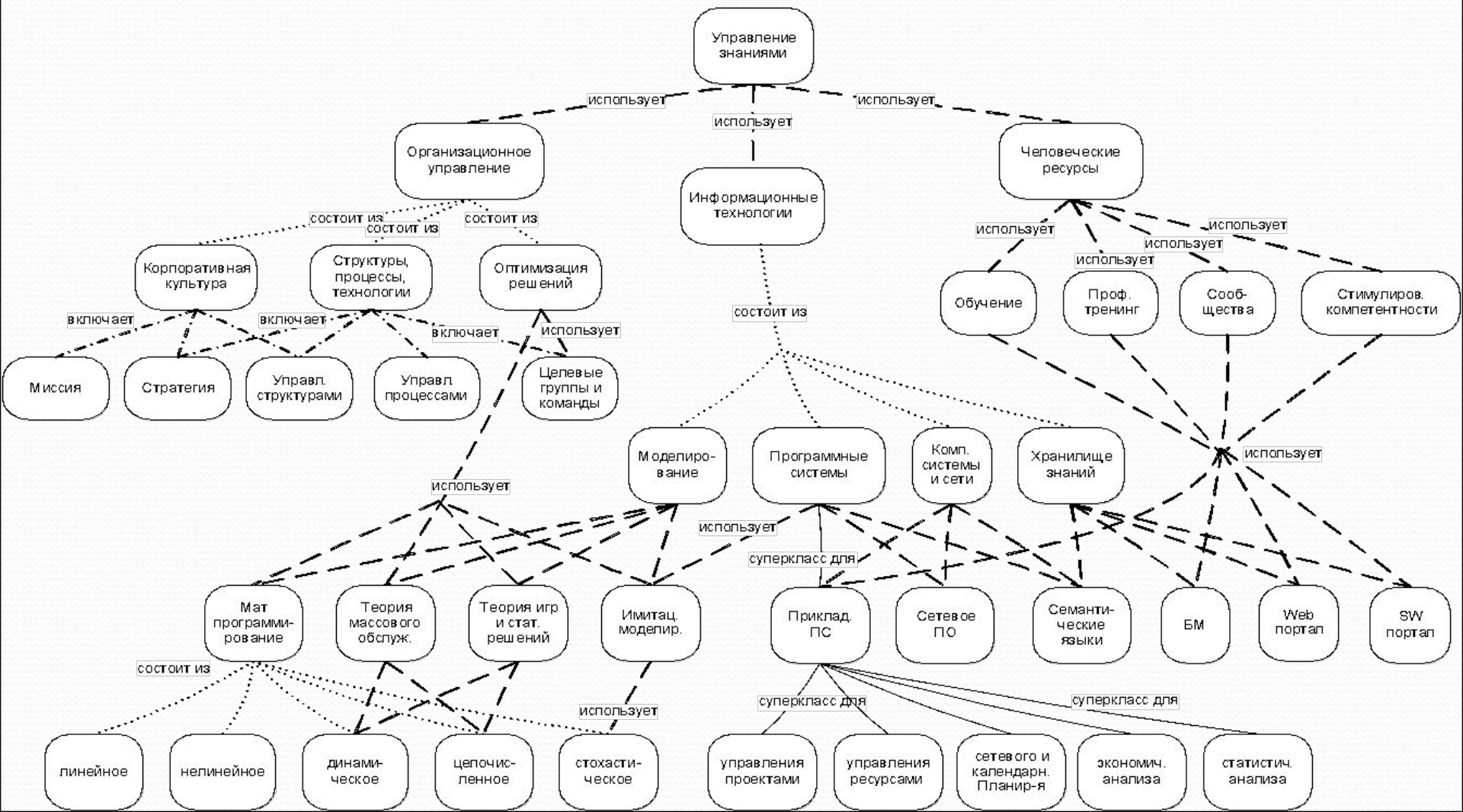
# Взаимосвязь между различными онтологиями формальной модели онтологической системы

Метаонтология



- В качестве примера онтологии предметных областей приведем краткое описание онтологии предприятия «The Enterprise Ontology» (далее ЕО), разработанную Эденбургским университетом совместно с такими партнерами как IBM, Lloyd's Register и др. [Uschold M., King M., 1998; TOVE Ontology Project].
- Целью создания ЕО было обеспечение предприятию возможности успешно справляться с быстро изменяющейся внешней средой. Основным средством для достижения этой цели признано совершенствование бизнес-планирования на основе моделирования, улучшения коммуникаций и интеграции информационных и бизнес-процессов.

# Таксономия понятия «Управление знаниями» и используемых теорий, методов, процессов



# Метаописания

- **Метаописания (описания об описаниях) – это особо структурированная информация, характеризующая содержание документов, информационных ресурсов и баз знаний, профилей компетенций специалистов и т.п., которая может быть полезна как пользователям, так и самой системе управления знаниями. Метаописания отражают различные свойства и характеристики объекта, такие, как статус, формат, семантика и др.**
- **Разделение описания объекта на информацию и метаописание – процесс не однозначный и зависит от целей описания. То, что может быть метаописаниями для одних целей, может являться частью содержания (информации) для других.**

- **Процесс создания метаописаний иногда именуют аннотированием. Аннотирование может происходить как с участием человека, так и без него, с помощью специальных программно реализованных алгоритмов. Результатом аннотирования является набор метаописаний, который может помещаться в хранилище метаописаний.**

**В метаописаниях выделяют три типа:**

- **Системные (служебные) метаданные.**
- **Структурные метаданные.**
- **Семантические метаописания.**

- ***Системные метаданные*** предназначены для функционирования информационных систем и систем управления знаниями. Они включают имена файлов и баз, даты их создания, тип и формат, размер файла и вид носителя и т.п.
- ***Структурные метаданные*** содержат, как правило, справочную информацию об объектах. Это могут быть наименование, статус, структурная принадлежность, профиль и т.п. То есть описания, используемые при идентификации и категоризации объектов в тех или иных целях.
- ***Семантические метаописания*** – особый вид описаний, включающий концептуальное (аннотированное) изложение содержания и смысла информации об объекте.



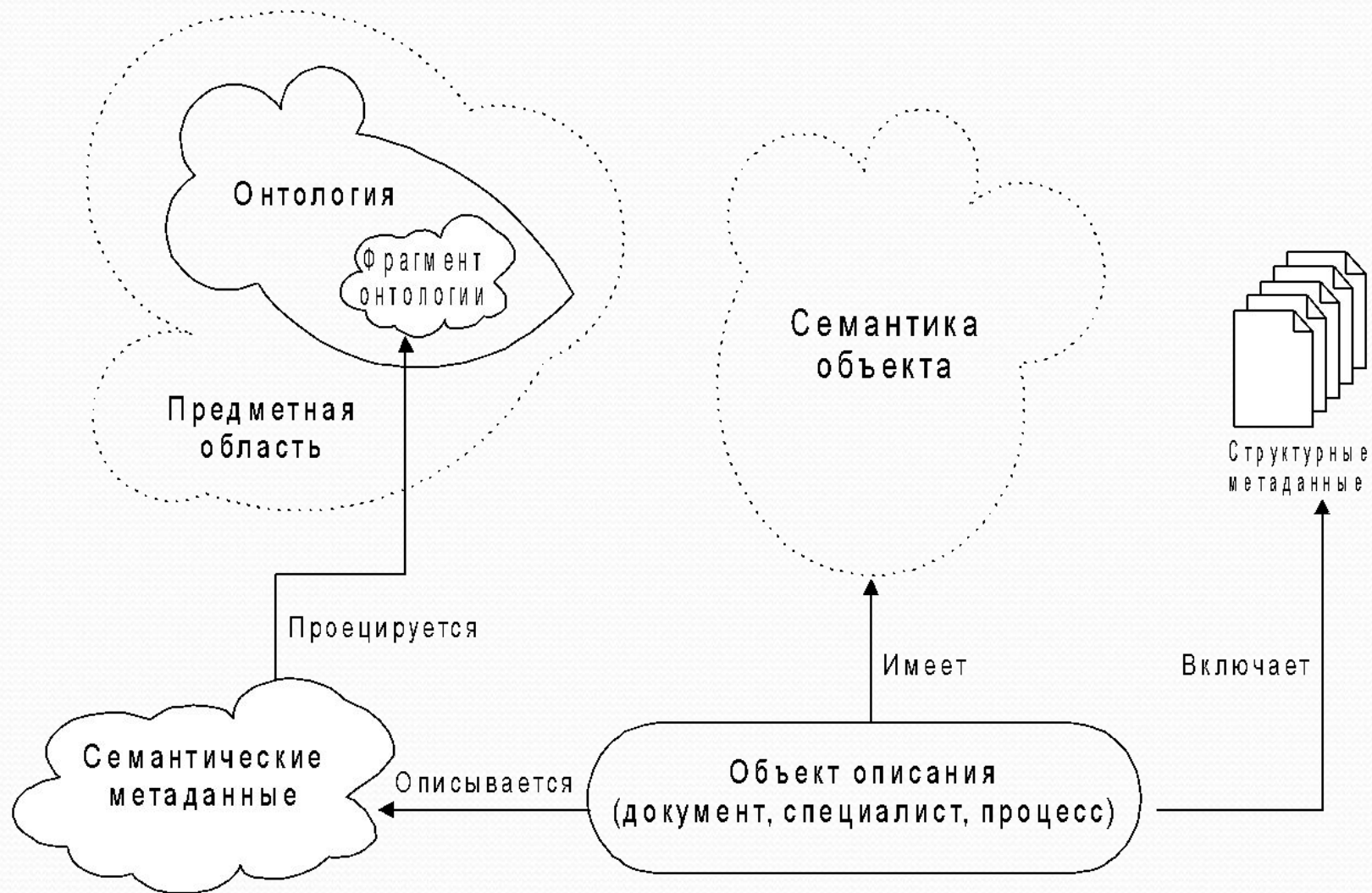
# Иерархия метаописаний



- **Добавление метаданных к электронным ресурсам системы создает возможность более точного определения местоположения информации об объектах, улучшает механизм фильтрации и отбора знаний, упрощает и ускоряет процессы доступа к необходимым программам, серверам, ресурсам дисковой памяти и т.п.**

- **Трудно переоценить роль метаописаний в библиотечном деле, в учебном процессе всех видов образовательных систем, включая электронные системы дистанционного обучения в части оперативности и полноты доступа к знаниям.**
- **Семантические метаописания являются частью онтологии, при построении которой максимально учитывается семантика объектов некоторой предметной области.**
- **При этом семантические метаописания отражают не всю семантику объекта, так же как онтология не может охватить всю описываемую ею предметную область (рис.).**

# Отражение смысла объекта в семантических метаописаниях



- Под метаданными объекта  $O_i$  будет пониматься следующее выражение:

$$MD = C_i \cup I_i,$$

где  $C_i$  – множество понятий онтологии  $O$ , имеющих отношение к объекту  $i$ ,

содержащихся в информации об объекте (документы, базы данных и знаний и т.п.) и в интересе пользователя. С каждым понятием связан свой весовой коэффициент  $K_i$ ;

$I_i$  – множество экземпляров понятий онтологии  $O$  с экземплярами отношений между ними.

# Измерение близости объектов в интеллектуальном пространстве

- Формализованное представление онтологий, а также метаописаний объектов создает возможность для измерения близости (подобия) объектов в интеллектуальном пространстве.

- Например, подобие между метаданными  $\text{Sim}(\text{MD}_i, \text{MD}_j)$  может быть определено через подобие входящих в них экземпляров понятий:

где  $\text{Sim}(\text{MD}_i, \text{MD}_j)$  – величина близости метаописания объекта  $i$  и объекта  $j$ ;  $\text{sim}(I_i, I_j)$  – величина близости экземпляров понятий  $I_i$  и  $I_j$ , входящих в сравниваемые метаописания.

Можно выделить следующие составляющие измерения подобия двух экземпляров понятий:

- 1) **таксономическое** (по близости в иерархии онтологии,  $\text{TS}(I_i, I_j)$ );
- 2) **реляционное** (по сходству отношений экземпляров,  $\text{RS}(I_i, I_j)$ );
- 3) **атрибутивное** (по близости значений атрибутов,  $\text{AS}(I_i, I_j)$ ).

# Таксономическое подобие (близость)

- Таксономическое подобие между экземплярами  $I_i$  и  $I_j$ , такими, что  $C_i(I_i)$  и  $C_j(I_j)$ , вычисляется с учетом положения соответствующих им понятий  $C_i$  и  $C_j$  в таксономии НС. Для вычисления семантического расстояния в иерархии понятий используется множество UC (upwards cotory), которое содержит все вышележащие по иерархии НС понятия и само исследуемое понятие:

$$UC(C_i, H^C) = \{C_j \in C \mid H^C(C_i, C_j) \vee C_j = C_i\}$$

- Используются семантические характеристики НС: рассмотрение ограничивается суперпонятиями заданного понятия  $C_i$  и рефлексивным взаимоотношением  $C_i$  к самому себе. Основываясь на определении UC, можно следующим образом определить таксономическое подобие:

$$TS(I_i, I_j) = \frac{|UC(C_i, H^C) \boxtimes UC(C_j, H^C)|}{|UC(C_i, H^C) \boxtimes UC(C_j, H^C)|}$$

# Атрибутивное подобие (Attribute Similarity)

Атрибутивное подобие основывается на подобии значений атрибутов для определения подобия между экземплярами. Так как атрибуты очень сходны с отношениями (например, в RDF атрибуты являются отношениями с интервалами (область допустимых значений), которые содержат литералы), то большая часть того, что было сказано об отношениях, также применимо и здесь.

Для вычисления атрибутивного подобия вначале определим набор сравниваемых атрибутов для двух экземпляров:

$$PA_i(I_i) := \{A : A \in A\},$$

$$PA(I_i, I_j) := PA_i(I_i) \cap PA_i(I_j),$$

а также значения их атрибутов:

$$As(A, I_i) := \{L_x : L_x \in L \wedge A(I_i, L_x)\}.$$



Подобие для одного атрибута равно

$$\begin{aligned}
 OA(I_1, I_2, A) = & \begin{cases} \frac{\sum_{a \in A_s(A, I_1)} \max\{\text{sim}(a, b, A) \mid b \in A_s(A, I_2)\}}{|A_s(A, I_1)|} & \text{если } |A_s(A, I_1)| \geq |A_s(A, I_2)|, \\ \frac{\sum_{a \in A_s(A, I_2)} \max\{\text{sim}(a, b, A) \mid b \in A_s(A, I_1)\}}{|A_s(A, I_1)|} & \text{если } |A_s(A, I_1)| < |A_s(A, I_2)|. \end{cases}
 \end{aligned}$$

И, наконец, подобие атрибутов вычисляется следующей формулой:

$$AS(I_i, I_j) = \frac{\sum_{a \in P_A(I_i, I_j)} OA(I_i, I_j, a)}{|P_A(I_i, I_j)|}$$