

Онтологии и онтологические системы

Понятие онтологии

Возникновение онтологий и их стремительное развитие связано с проявлением в нашей реальности следующих новых факторов:

- колоссальный рост объемов информации, предъявляемых для обработки (анализа, использования) специалистам самых различных областей деятельности;
- чрезвычайная зашумленность этих потоков (повторы, противоречивость, разноуровневость, и т.п.);
- острая необходимость в использовании одних и тех же знаний разными специалистами в разных целях;
- всеобщая интернетизация нашей жизни и острая необходимость в структуризации информации для её представления пользователям и более эффективного поиска;
- необходимость сокращения времени на поиск нужной информации и повышения качества информационных услуг в Интернете.

Наиболее распространенными стали следующие определения.

Онтология – это точная спецификация концептуализации.

Онтология – это формальная точная спецификация совместно используемой концептуализации.

Онтологии – это базы знаний специального типа, которые могут *читаться* и *пониматься*, *отчуждаться* от разработчика и/или физически *разделяться* их пользователями.

Формальная модель онтологии

Тогда, формальная модель онтологии (O) может быть определена как упорядоченная тройка вида:

$$O = \langle X, R, F \rangle, \text{ где}$$

X – конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология; R – конечное множество отношений между концептами (понятиями, терминами) заданной предметной области; F – конечное множество функций интерпретации (аксиоматизации), заданных на концептах и/или отношениях онтологии.

Естественным ограничением, накладываемым на множество X , является его конечность и не пустота. Иначе дело обстоит с компонентами F и R в определении онтологии. Очевидно, что F и R должны быть конечными множествами.

Рассмотрим граничные случаи, связанные с пустотой и не пустотой F и R .

1. Пусть $R = \emptyset$ и $F = \emptyset$. Тогда онтология $O = V = \langle X, \{ \}, \{ \} \rangle$, что соответствует простому словарю (V). Очевидно, что это случай вырожденной онтологии. Она может быть полезна для спецификации, пополнения и поддержки словарей предметной области. Онтологии-словари имеют ограниченное применение, поскольку, когда используемые термины принадлежат очень узкому (например, техническому) словарю, их смыслы уже заранее хорошо согласованы в пределах определенного (например, научного) сообщества. Например, онтологию этого типа образуют индексы машин поиска информации в Интернете.

2. $R = \emptyset$, $F \neq \emptyset$, $O = V = \langle X, \{ \}, F \rangle$. Тогда каждому элементу множества терминов из X ставится в соответствие функция (аксиома) интерпретации (одна или несколько) $f \in F$. Здесь могут быть варианты.

Допустим, что $X = X_1 \cup X_2$, причем $X_1 \cap X_2 = \emptyset$, где X_1 – множество интерпретируемых терминов, X_2 – множество не интерпретируемых терминов. При этом $\exists (x \in X_1, y^1, y^2, \dots, y^k \in X_2)$ такие, что $x = f(y^1, y^2, \dots, y^k)$, $f \in F$.

Введение в рассмотрение функции k аргументов обеспечивает более полную интерпретацию $x \in X$. Вид отображения $f \in F$ определяет выразительную мощность и практическую полезность этого вида онтологии.

Так, если предположить, что функция интерпретации задается оператором присваивания значений ($X_1 := X_2$, где X_1 – имя интерпретации X_2), то онтология трансформируется в пассивный словарь $O = V^P = \langle X_1 \cup X_2, \{ \}, \{ := \} \rangle$, так как все определения терминов из X_1 берутся из уже существующего и фиксированного множества X_2 . Практическая ценность данного словаря выше, чем простого, но он явно недостаточен, например, для представления знаний в задачах обработки информации в Интернете в силу динамического характера этой среды. Смысл таких терминов «вычисляется» каждый раз при их интерпретации. Ценность словаря для задач обработки информации, например в Интернет среде, выше, чем у предыдущей модели, но еще недостаточна, так как интерпретируемые элементы из X_1 никак не связаны между собой и, следовательно, играют лишь роль ключей входа в онтологию.

3. $R \neq \emptyset$, $F = \emptyset$, $O = V = \langle X, R, \{ \} \rangle$, т.е. множество F – пусто, множество отношений на концептах онтологии – не пусто и включает одно из отношений типа иерархии: $R = IS_A$ (*быть элементом класса*) ИЛИ POW (*part of while – часть - целое*) ИЛИ AKO (*a kind of – вид - род*) ИЛИ $contained_in$ (содержится в) ИЛИ see_also (смотрите также) и др.

Этот случай соответствует специальному подклассу онтологий, называемых простыми таксономиями:

$$O = T^O = \langle X, \{IS_A\}, \{ \} \rangle, O = T^O = \langle X, \{POW\}, \{ \} \rangle, O = T^O = \langle X, \{AKO\}, \{ \} \rangle \text{ и др.}$$

Под *таксономической структурой* понимается иерархическая система понятий, связанных между собой одним из отношений IS_A , POW , AKO и т.п. Поскольку перечисленные отношения имеют заранее фиксированную семантику, это позволяет организовывать структуру понятий онтологии в виде дерева. Такой подход имеет свои преимущества и недостатки, но в общем случае является адекватным и удобным для представления иерархий понятий.

4. $R \neq \emptyset$, $F \neq \emptyset$ и $O = V = \langle X, R^I, F \rangle$, где R^I – это какое-либо из отношений иерархии (IS_A , POW , AKO , и т.п.). В этом случае строится иерархия терминов, в которой каждому из них ставится определенная процедура или аксиома интерпретация. Это, например, онтологии, в которой термины раскрываются по матрешечному принципу.

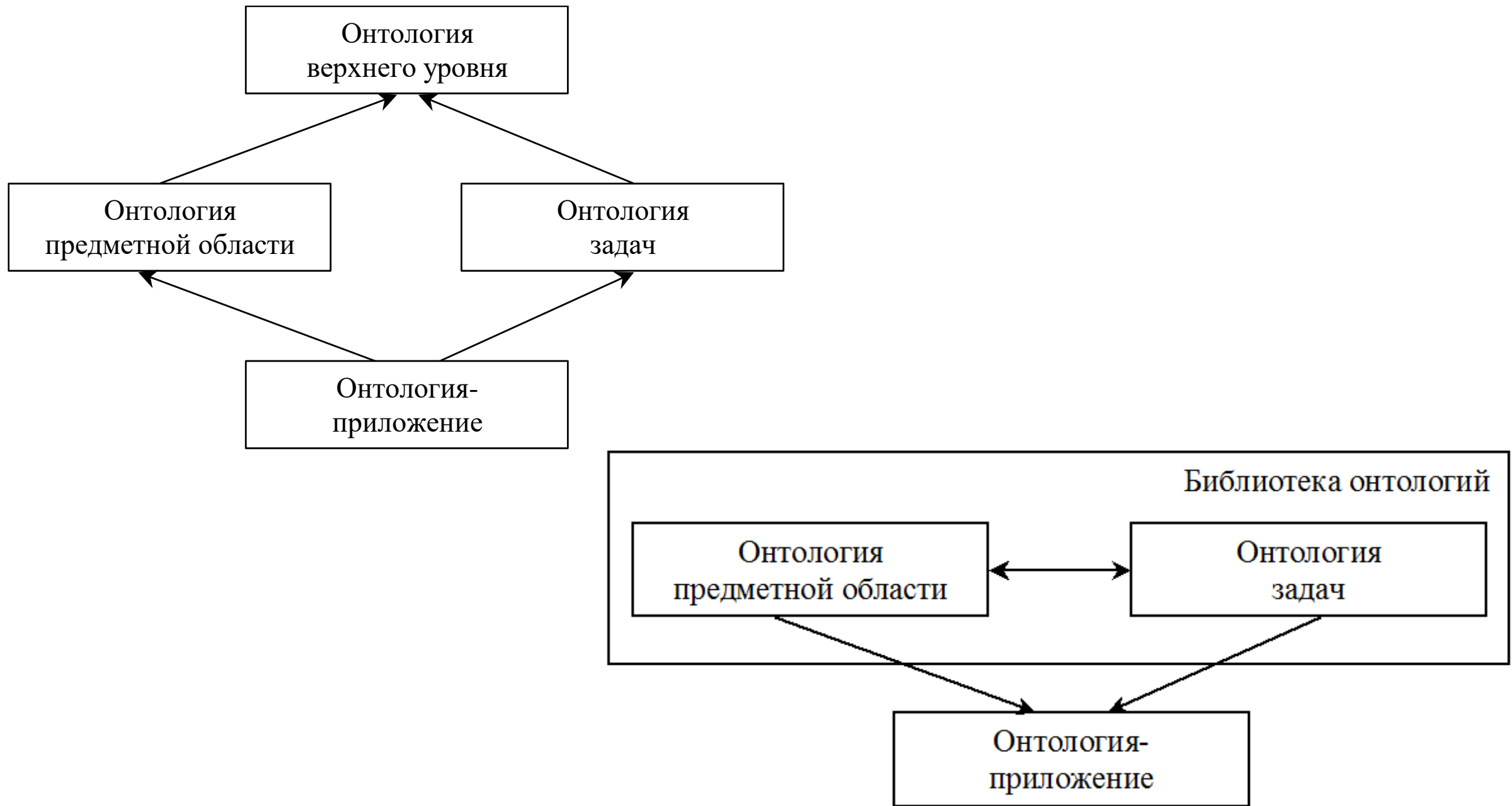
2.4. $R \neq \emptyset$, $F \neq \emptyset$, $O = V = \langle X, R^I + R^C, F \rangle$ и $R = R^I + R^C$, где R^I – это какое-либо из отношений иерархии, R^C – специфические отношения универсального толкования типа: временные, пространственные каузальные, сравнения и т.п. Таким образом, в данном случае речь идет об иерархических структурах, на которых вводимые сущности имеют соответствующие толкования и на них определены отношения, в том числе, часто с нечетко фиксируемой семантикой. Это приводит к значительным трудностям по обработке таких структур, в связи с необходимостью учитывать свойства вводимых отношений: рефлексивность (нерефлексивность, антирефлексивность) симметричность (несимметричность, антисимметричность), транзитивность (нетранзитивность, антитранзитивность). Именно по этой причине это самые сложные онтологии и их практически очень мало, да и те, которые есть, являются очень упрощенными. Онтологии данного типа соответствуют модели расширяемой онтологии, достаточно мощной для спецификации процессов формирования пространств знаний.

Виды онтологий

С точки зрения концептуализации онтологии делятся на:

- онтологии представления, определяющие концептуализацию, которая лежит в основе формализма представления знаний. Понятия и отношения, определенные в других видах онтологий, считаются конкретизацией понятий онтологии данного вида. Предполагается, что они не зависят от конкретной ПрО. Онтологии данного вида могут использоваться для поддержки нескольких альтернативных теорий, относящихся к одной и той же предметной области;
- общие (родовые) онтологии, затрагивающие общие, фундаментальные аспекты концептуализации такие, как «часть», «причина», «участие», «представление»;
- промежуточные онтологии, содержащие общие понятия и отношения, характерные для конкретной предметной области. В идеальном случае, они используются в качестве интерфейса между онтологиями предметных областей и общими онтологиями, но могут выступать как онтологии верхнего уровня для описания знаний конкретной ПрО;

- онтологии верхнего уровня, являющиеся конкретным назначением понятий общих и промежуточных онтологий. Это уникальный модуль, находящийся над онтологиями предметной области, или являющийся самостоятельной, независимой от предметной области теорией;
- онтологии предметных областей, содержащие понятия определенной области знаний или входящих в нее областей;
- онтологии задач, описывающие определенные задачи или деятельность, используя словарь, введенный в общих, промежуточных онтологиях и онтологиях предметных областей;
- онтологии-приложения, являющиеся специализацией онтологий предметных областей и онтологий задач, и опирающиеся на определения, характерные для конкретного приложения.



↔ - правила отображения

Операции над онтологиями

операции по редактированию;

В данную группу операций входят операции, связанные с разработкой и редактированием онтологий на этапе их построения.

Создание (creating) онтологии состоит из четырех основных этапов:

- *составление спецификации*, определяющей допустимую область применения описываемых онтологией знаний,
- *определение задач*, в решении которых используются рассматриваемые знания;
- *разработка концептуализации*, то есть определение объектов предметной области, составляющих описываемую область знаний, и отношений между ними;
- *формализация представляемых знаний*.

Возможны два способа построения новой онтологии:

- (1) создание новой онтологии с нуля;
- (2) создание новой онтологии на основе имеющихся.

алгебра онтологий;

1. В онтологиях предметных областей осуществляется *поиск терминов*, представляющих интерес для конкретной прикладной задачи.

2. Для найденных терминов на основании правил *строятся отображения для извлечения и соединения контекстов*. Контекстом считается часть более общей онтологии, являющаяся внутренне согласованной, достоверной самой по себе. В рамках каждой предметной области *формируются* контексты на основании отображений для извлечения. Выполняется *реструктуризация* для представления данных во внутреннем формате с внесением при необходимости семантических исправлений и уточнений. Сформированные контексты образуют срезы онтологий, значимые для рассматриваемой прикладной задачи, сами по себе также являющиеся онтологиями и называемые *соединяемыми онтологиями*.

3. В данных (соединяемых) онтологиях на основании отображений для соединения *определяются общие* для них контексты. В пределах этих контекстов соединяемые онтологии являются *согласованными*.

4. Контексты последовательно *объединяются* в результирующую онтологию, описывающую собранные из разных областей знания применительно к рассматриваемой задаче.

операции по интеграции онтологий;

Выборка (slicing, extraction) – предназначена для формирования внутренне согласованной части исходной онтологии или ее среза, который будет включен в новую онтологию.

Отсечение (pruning) – служит для удаления из максимального среза классов, которые являются несущественными для новой онтологии.

Установка соответствий (alignment) – операция, задающая отношения между онтологиями посредством установки связей между эквивалентными понятиями в онтологиях. Такой вид соединения сохраняет структуру исходных онтологий, не требует переименования терминов, изменений в аксиомах, определениях, доказательствах и т.п., т.е. является наиболее «слабой» формой соединения.

Частичная согласованность – операция, устанавливающая между онтологиями соответствия, поддерживающие эквивалентный логический вывод и вычисления, выполняемые для всех эквивалентных понятий и отношений.

Унификация – операция, результатом которой является новая универсальная онтология. Унификация определена как частичная согласованность онтологий, которая может быть

операции по агрегированию и декомпозиции;

Модуляризация (modularization) или *декомпозиция (decomposition)* – операция по разбиению онтологии на модули, которые могут служить основой для разработки новых онтологий. Возможны подходы к разбиению онтологий на модули, основанные на:

- области применения: упорядочивание базы знаний, описываемой онтологией, по темам, заданным инженером по знаниям;
- решаемой задаче: упорядочивание базы знаний по правилам логического вывода, применяемых для ее решения, а также по используемым в логическом выводе объектам;
- микро-теориях: микро-теория состоит из набора высказываний и ограничена областью применения (для каких объектов, при каких условиях и т.п. данная теория может быть использована);
- контекстах: выявление разобщенных контекстов и помещение их в самостоятельные модули. Для выявления разобщенных контекстов используются, как правило, специальные методы анализа контекстов.

Составление (assembling, composing) – операция конструирования новой онтологии из модулей.

операции по преобразованию;

Переформулирование (reformulation) – операция по преобразованию представления входной теории. Например, данная операция может быть использована для преобразования функциональных предикатов в отношения.

Построение таксономии (taxonomy design) – операция по формированию иерархической структуры, основанной на отношении *IS_A*, в которой свойства более общих понятий наследуются подчиненными им понятиями.

Перевод (translation) – операция по переводу онтологий, разработанных в одном формализме и реализованных с помощью одних языков, в другие формализмы и языки. Реализация операции связана с разработкой средств, способных осуществлять взаимные переводы между несколькими форматами.

операции по сравнению, проверке и оценке

Сравнение (matching) – операция, устанавливающая степень соответствия онтологий. Данная операция включает с себя операции по *сравнению концептуальных структур* и операции, *определяющие степень переводимости одной онтологии в другую*. Сравнение концептуальных структур заключается в *определении степени совместимости понятий*, используемых в разных онтологиях. Концептуальная структура определяется областью применения описываемых знаний. Для сравнения выбираются несколько онтологий, описывающих одну и ту же прикладную проблему. Степень совместимости оценивается в соответствии с классификацией: *согласие / соответствие / конфликт / противоречие*. В соответствии с указанной классификацией концептуальные структуры могут находиться в следующих отношениях:

- согласующиеся (consensus): разные онтологии имеют одинаковую структуру и терминологию;
- соответствующие (correspondence): в разных онтологиях одинаковые структурные элементы имеют разные имена (несовпадения в терминологии);
- конфликтующие (conflict): в разных онтологиях разные структурные элементы названы одинаково;
- противоречивые (contrast): в разных онтологиях нет совпадений ни в структурах, ни в терминологии.

Проверка (validation) – операция, проверяющая корректность и непротиворечивость (внутреннюю согласованность) онтологии. В ходе операции производится семантический и синтаксический анализ онтологии.

Оценка (evaluating) – операция, в результате которой выносится заключение о соответствии онтологии ее назначению. Операция выполняется экспертами перед *выпуском* онтологии. Онтология оценивается по ее соответствию: требованиям, спецификации, реальному миру, степени компетентности.

Инструменты инженерии онтологий

Ontolingua.

OntoEdit.

OilEd.

Protégé.

Web-DESO.

