# **Способ структурирования знаний с использованием онтологий**

## Понятие онтологии

Введение:

Одной из характерных черт XXI века является лавинообразное увеличение объёмов информации. В таких условиях перед людьми стоит важная задача: найти в этом информационном хаосе способ структурировать, систематизировать накопленные знания для их последующего эффективного использования.

Существует большое количество различных способов (моделей) представления знаний, на слайде показана одна из классификаций.

Онтологии относятся к сетевым моделям, которые являются наиболее мощными по своим описательным возможностям. В основе таких моделей лежит предположение о том, что любые знания можно представить в виде совокупности понятий (объектов) и отношений (связей) между ними.

Определение:

У термина «онтология» существует несколько различных определений, которые зависят, в том числе, от контекста рассмотрения. Зачастую приводятся два определения: из области философии и из области информатики.

* В первом случае под «онтологией» понимается изучение бытия как такового, его фундаментальных категорий и принципов.
* А во втором – формальная спецификация разделяемой концептуальной модели, где:
  + под «концептуальной моделью» подразумевается абстрактная модель предметной области, описывающая систему понятий данной области;
  + «спецификация» подразумевает описание системы понятий в явном виде;
  + «формальная» подразумевает, что концептуальная модель задана на формализованном языке,т.е. онтологии используют способ представления, который может обрабатываться компьютером.
  + под «разделяемой» подразумевается согласованное понимание концептуальной модели определённым сообществом;

В то время как компьютеры могут манипулировать и выполнять логический вывод на знаниях о предметной области с помощью онтологии, понимание того, что обозначают эти символы и правила, остается функцией человека. Поэтому онтология не может существовать без сообщества, которое ее поддерживает.

Формальное задание:

Одним из основных инструментов для математического описания онтологии является алгебраическая логика. Она используется для определения множества понятий и отношений между ними. Алгебра логики также позволяет устанавливать правила, которые определяют, как понятия связаны между собой. Формально онтологию можно представить следующим образом:



где C – конечное множество классов, описывающих понятия некоторой предметной области,

P – конечное множество свойств (атрибутов) этих понятий,

R – конечное множество связей (отношений) между понятиями,

A – множество аксиом, утверждений, построенных из этих понятий, их свойств и связей между ними.

Классификация :

• метаонтология – содержит общие понятия и отношения, не зависящие от предметной области («объект», «свойство», «значение» и т.п.);

• предметная онтология – содержит понятия, описывающие конкретную предметную область и отношения, семантически значимые для данной предметной области;

• онтология задач (и методов) – содержит в качестве понятий типы решаемых задач, а отношения специфицируют декомпозицию задач на подзадачи;

• прикладная онтология (онтология приложения) – описывает концепты, зависящие как от предметной области, так и от задач

• сетевая онтология – используется для описания конечных результатов действий, выполняемых объектами предметной области или задачи.

Задачи:

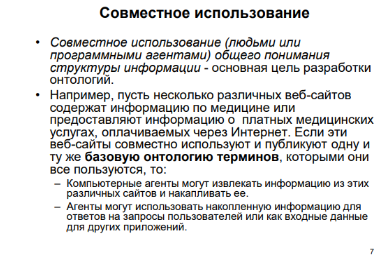
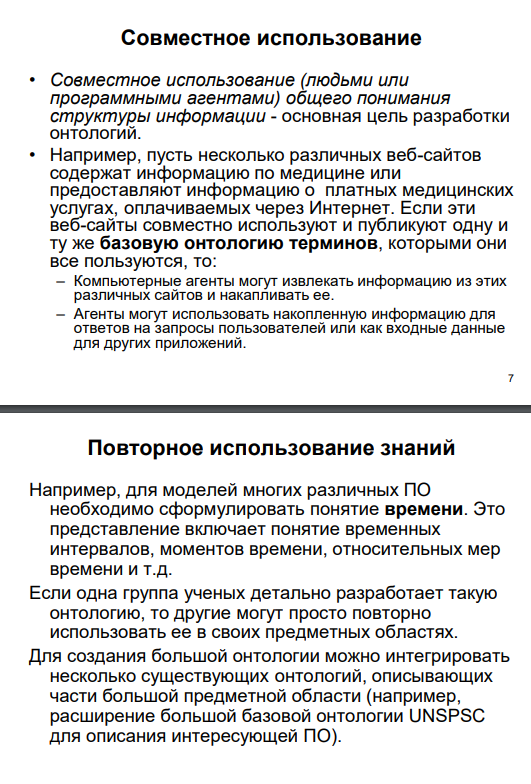
1. Совместное использование людьми или программными агентами общего понимания структуры информации;

2. Возможность повторного использования знаний в предметной области;

3. Создания явных допущений в ПрО;

4. Отделение знаний в предметной области от оперативных знаний;

5. Анализ знаний в предметной области.



Поиск:

Полноценные онтологии имеют больше возможностей для описания отношений, чем простые таксономические иерархии. Связи между понятиями значительно увеличивают количество способов осуществлять навигацию и поиск по предметной области, выполнять анализ, классификацию и визуализацию знаний.

В настоящее время проблема поиска в больших массивах знаний становиться не менее актуальной, чем их структуризация. Эта проблема усугубляется еще и тем, что существующие поисковые механизмы осуществляют поиск информации без учета семантики слов, входящих в запрос, а также контекста, в котором они используются. Благодаря использованию вывода на онтологии удается строить образ достаточно релевантный запрашиваемому. Этот образ может использоваться для формирования более эффективных запросов для поисковой системы.

БД и БЗ:

Традиционным способом хранения и использования данных являются реляционные базы данных, в которых данные хранятся в виде связанных таблиц. Однако не всегда табличные данные эффективны в использовании, поэтому все чаще используются альтернативные формы.

Классические базы данных используют «гипотезу закрытого мира», когда все, чего нет в базе данных – не существует. В противоположность этому онтологии используют «гипотезу открытого мира», то есть если чего-то нет в базе знаний, то это не обязательно не существует, а просто не описано.

миша любит яблоки

маша любит груши

миша любит груши?

онт. неизвестно

бд. нет

Роль в обработке естественного языка:

онтологии могут применяться для широкого круга задач обработки естественного языка: аннотирование текстов с помощью онтологий, извлечение знаний, распознавание именованных сущностей (Named Entity Recognition или NER), связывание именованных сущностей Named Entity Linking, автоматический вывод новых знаний, ризонинг (reasoning).

## Компоненты онтологии

## Процесс создания онтологии

В настоящее время известен только один стандарт, регламентирующий процесс

разработки онтологий и связанных с этим исследований: IDEF5.

Несмотря на это существует множество предложений по методикам разработки

онтологий, но во всех них выделяют следующие основные задачи:

1. Анализ целей создания и области применения создаваемой онтологии.

2. Построение онтологии:

2.1 Сбор и фиксация знаний о предметной области

2.2 Кодирование

2.1 Сбор и фиксация знаний, в свою очередь состоит из

● Определение основных понятий и их взаимоотношений в выбранной ПО

● Создание точных непротиворечивых определений для каждого основного понятия и отношения

● Определение терминов, которые связаны с основными понятиями и отношениями.

● Согласование перечисленных компонентов онтологии

2.2 Кодирование:

● разбиение совокупности основных терминов, используемых в онтологии, на классы

● выбор или разработку специального языка для представления знаний.

● формирование концептуализации в рамках выбранного языка представления знаний

диаграммы:

В стандарте IDEF5 предусмотрены четыре основных вида схем для представления

онтологической информации в наглядной графической форме

1. Диаграммы классификации

Служат средством логической систематизации знаний, накопленных при изучении

системы. Существуют два типа таких диаграмм:

● диаграмма строгой классификации (Description Subsumption – DS) и

● диаграмма естественной или видовой классификации (Natural Kind Classification

– NKC)

2. Композиционные схемы

Composition Schematics — служат для графического представления состава

классов онтологии. В частности, с помощью них можно наглядно отобразить

состав объектов, относящихся к тому или иному классу.

3. Схемы взаимосвязей

Relation Schematics – позволяют разработчикам визуализировать и изучать связи

между различными классами объектов системы. В некоторых случаях эти схемы

используются для представления зависимостей между взаимосвязями классов

4. Диаграмма состояния объекта

Object State Schematics – позволяет описать процесс изменения состояния

объекта. С объектом могут произойти два типа изменений: он может поменять либо

своё состояние, либо класс.

Правила:

выделим некоторые фундаментальные правила разработки. Эти правила могут

показаться довольно категоричными. Тем не менее, во многих случаях они

могут помочь принять проектные решения.

1) Не существует единственного правильного способа моделирования

предметной области – всегда существуют жизнеспособные альтернативы.

Лучшее решение почти всегда зависит от предполагаемого приложения и

ожидаемых расширений.

2) Разработка онтологии – это обязательно итеративный процесс.

3) Понятия в онтологии должны быть близки к объектам (физическим или

логическим) и отношениям в интересующей вас предметной области.

Скорее всего, это существительные (объекты) или глаголы (отношения) в

предложениях, которые описывают вашу предметную область.

Этапы:

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

1. Изучение и систематизация начальных условий. Этот этап устанавливает основные цели и контекст разработки онтологии, а также распределяет роли членов проекта.

2. Сбор и накопление данных для построения онтологии.

3. Анализ и группировка собранных данных для облегчения согласования терминологии.

4. Начальное развитие онтологии. На этом этапе формируется предварительная онтология на основе систематизированных данных.

5. Уточнение и утверждение онтологии (заключительный этап).

шаг1

перейдем к самому алгоритму построения онтологии.

шаг 1. Определение области и масштаба онтологии

Начало разработки онтологии предлагается с определения ее области и

масштаба. То есть, с ответа на несколько основных вопросов:

1. Какую область будет охватывать онтология?

2. Для чего мы собираемся использовать онтологию?

3. На какие типы вопросов должна давать ответы информация в онтологии?

4. Кто будет использовать и поддерживать онтологию?

Ответы на эти вопросы могут измениться во время процесса

проектирования онтологии, но в любой заданный момент времени они

помогают ограничить масштаб модели

шаг2

Шаг 2. Рассмотрение вариантов повторного использования

существующих онтологий

Оно может быть необходимым, если проектируемой системе нужно взаимодействовать с другими приложениями, которые уже вошли в отдельные онтологии или

контролируемые словари.

Многие онтологии доступны в электронном виде и могут быть

импортированы в используемую среду проектирования онтологии.

Формат онтологии часто не имеет значения, т.к. многие системы

представления знаний могут импортировать и экспортировать онтологии.

Даже если система представления знаний не может работать напрямую с

отдельным форматом, задача перевода онтологии из одного

формата в другой обычно не является сложной.

шаг 3

Шаг 3. Перечисление важных терминов в онтологии

Полезно составить список всех терминов, о которых мы хотели бы сказать

что-либо или которые хотели бы объяснить пользователю, ответив на представленные вопросы .

В начале важно получить полный список терминов, не беспокоясь о

пересечении понятий, которые они представляют, об отношениях между

терминами, о возможных свойствах понятий или о том, чем являются

понятия – классами или слотами.

шаг4

Шаг 4. Определение классов и иерархии классов

Существует несколько возможных подходов для разработки иерархии

классов.

● Процесс нисходящей разработки начинается с определения самых

общих понятий предметной области с последующей конкретизацией

понятий.

● Процесс восходящей разработки начинается с определения самых

конкретных классов, листьев иерархии, с последующей группировкой

этих классов в более общие понятия.

● Процесс комбинированной разработки – это сочетание нисходящего и

восходящего подходов: Сначала определяются более заметные понятия,

а затем они соответствующим образом обобщаются и ограничиваются.

Ни один из этих трех методов не лучше других по своей сути. Выбор подхода

в большой степени зависит от личного взгляда на предметную область.

шаг 5

Шаг 5. Определение свойств классов – слотов

Классы сами по себе не предоставляют достаточно информации для ответа

на вопросы проверки компетентности из Шага 1.

После определения некоторого количества классов требуется описать

внутреннюю структуру понятий.

Классы выбираются из списка терминов, который был создан на Шаге3.

Большинство оставшихся терминов, вероятно, будут свойствами этих

классов.

Для каждого свойства из списка требуется определить, какой класс оно

описывает. Эти свойства станут слотами, привязанными к классам.

Вообще, в онтологии слотами могут стать несколько типов свойств

объектов:

- «внутренние» свойства;

- «внешние» свойства;

- части, если объект имеет структуру; они могут быть как физическими, так и

абстрактными «частями»;

- отношения с другими индивидными концептами; это отношения между

отдельными членами класса и другими элементами

шаг 6

Шаг 6. Определение фацетов (ограничений) слотов

Слоты могут иметь различные фацеты, которые описывают тип данных,

разрешенные значения, число значений (мощность) и другие свойства

значений, которые может принимать слот.

Несколько общих фацетов:

● Мощность слота: определяет, сколько значений может иметь слот

● Тип значения слота: какие типы значений можно ввести в слот

● Строка

● Число

● Булевы слоты

● Нумерованные слоты (список конкретных значений)

● Слоты-экземпляры (отношения между индивидными концептами)

●

● Домен слота и диапазон значений слота: Разрешенные классы для

слотов типа Экземпляр часто называют диапазоном значений слота.

Некоторые системы позволяют ограничить диапазон значений слота,

если слот привязан к определенному классу. Классы, к которым слот

привязан, или классы, свойство которых слот описывает, называются

доменом слота.

шаг 7

Шаг 7. Создание экземпляров

Последний шаг – это создание отдельных экземпляров классов в иерархии.

Для определения отдельного экземпляра класса требуется

1) выбрать класс,

2) создать отдельный экземпляр этого класса и

3) ввести значения слотов.

языки

В связи с использованием онтологий в различных приложениях возникла необходимость создания стандартизированных способов их представления.

Онтология может быть разработана с использованием различных языков и инструментов, таких как: RDF (Resource Description Framework; язык представления информации о ресурсах сети, а также информации о ресурсах, на которые можно ссылаться) и RDFS (семантическое расширение RDF); OWL (Web Ontology Language; словарь, расширяющий набор терминов, определенных RDFS; имеет 3 степени детализации, легко масштабируется и согласовывается с современными сетевыми стандартами, построен на форматах RDF и RDFS); SPARQL (язык запросов к RDF-хранилищам) и др.

Соответственно, существуют различные среды разработки онтологий: Protégé (поддерживает форматы RDF, OWL, XML Schema), Web-Protégé и др.

## Примеры применения онтологий

## Сравнение с другими методами структурирования знаний

Достоинства:

К достоинствам онтологий относится их высокая эффективность при интеграции разнородных источников информации и знаний, которые могут быть размечены с помощью терминов онтологии, независимо от их форматов и представлений. Онтологии позволяют эффективно устранить проблемы обнаружения взаимосвязей и структурированности форматов, возникающие при разнородности источников.

Применение онтологий при структуризации знаний позволяет сделать модель предметной области более наглядной и простой для понимания. Фрагменты знаний, объединенные в онтологии, становятся более пригодными для эффективного повторного применения. Технология поиска с применением онтологий позволяет учитывать семантику и контекст запроса, повышая релевантность искомых знаний.

В явном виде онтологии используются как источники данных для многих компьютерных приложений (для информационного поиска, анализа текстов, извлечения знаний и в других информационных технологиях), позволяя более эффективно обрабатывать сложную и разнообразную информацию. Этот способ представления знаний позволяет приложениям распознавать те семантические отличия, которые являются само собой разумеющимися для людей, но не известны компьютеру.

## Заключение