

Сопровождение ОНТОЛОГИЙ

Курс: Концептуальное
моделирование предметных
областей





- Создание онтологий
- Согласование онтологий



Развитие методов сопровождения онтологий

- Развитие философских моделей и ручные разработки (1990-е)
 - Разработка экспертами на основе логики, философии и формальной семантики
 - верхнеуровневые онтологии (DOLCE), онтологии широкого охвата (Cyc), тезаурусы (WordNet)
 - неформализуемые подходы к разработке, высокая стоимость, слабая масштабируемость
- Классические методологии разработки онтологий (2000–е)
 - формализация процесса разработки
 - методологии Methontology, Uschold & King, Grüninger & Fox, On-To-Knowledge и другие
 - ориентация на инжиниринг знаний, частичная автоматизация
 - Определение целей → Сбор знаний → Формализация → Реализация (Представление в онтологической модели) → Оценка
- Процессы жизненного цикла онтологий (2010–е)
 - эволюция онтологий, поддержка изменений, повторное использование онтологий
 - управление версиями, согласование и слияние онтологий
 - обнаружение необходимости развития → представление → реализация → проверка согласованности
- Автоматизация разработки онтологий (2020+)
 - извлечение, построение, обновление
 - применение нейросетей и языковых моделей для извлечения и выравнивания онтологий



Спецификации предметных областей

- **Онтологии**
 - Понятия и отношения, характерные для предметной области
 - Формальные ограничения на них
- **Концептуальные схемы**
 - Стандартизованные структуры и ограничения данных предметной области
 - Спецификации поведения



Разработка спецификаций предметной области

- Постановка задач предметной области
- Описание знаний предметной области, используемых при их решении
- Концептуализация предметной области: глоссарий
 - Существенные термины, их определения на естественном языке
- Тезаурус с определениями
 - Глоссарий дополняется лингвистическими семантическими связями общего вида
- Разработка онтологии предметной области на основе определений глоссария
 - Образуются понятия, именованные отношения сущностей, ограничения
- Разработка концептуальной схемы из онтологии предметной области
 - Абстрактные типы для сущностей предметной области, структур, типов данных, наборов, методов
 - Критерии идентификации сущностей предметной области
 - Описание поведения объектов предметной области
- Семантическое аннотирование компонентов концептуальной схемы в терминах онтологии



Формирование глоссариев и тезаурусов

- **Подход к концептуализации предметной области**
 - Постановка представительных задач, решаемых в предметной области
 - Формирование требований для детализации постановок задач в предметной области
 - Выделение терминов в постановках задач
- **Добавление в глоссарий существенных терминов, выделенных из постановок задач**
 - Выделение или формирование определений терминов экспертом предметной области, либо определений из текстов
 - Определения терминов должны описывать свойства сущностей, существенные для данной предметной области, и их ограничения в предметной области
 - Существенные термины из определений также добавляются в глоссарий и определяются до некоторого насыщения характерными терминами
- **Выявление семантических отношений и свойств терминов для формирования тезауруса**
 - Синонимия, гипер/гипонимия, часть/целое, зависимость, ассоциативные отношения
 - Свойства: сущности, виды, роли, состояния, процессы, измеримые характеристики



Пример

- Существенные термины предметной области двойных звёзд
 - *Astronomical Object, Star, Stellar System, Component, Pair, Observational Type, Visual Pair, Physical Visual Pair*
- Определения терминов
 - *Visual pair is a pair of stellar objects observed photometrically as two visually distinct components. Observational type of such pairs is visual. Observation of relative positions of visual binary stars components includes angular separations and position angles. Among visual pairs, optical and physical ones are distinguished. A physical pair can be recognized by common parallaxes of components or orbital movement of its components, or if the pair is also observed as one of more close type of pairs, for instance spectroscopic, eclipsing one*
- Термины из определений
 - *Coordinate, Coordinate System, Equatorial Coordinate System, Right Ascension, Declination, Precession, Parallax, Proper Motion; Photometric System, Passband, Magnitude; Orbital Movement, Period, Position Angle, Angular Separation, Eccentricity*
- Семантические связи терминов
 - *Visual Pair “is a” Pair, Component “is part of a” Pair*



Формализация онтологии на основе тезауруса

- Онтологические спецификации создаются из терминов в соответствии со свойствами и связями терминов
 - Понятия генерируются для терминов со свойствами сущностей, видов, ролей, иногда процессов
 - Понятия связываются с онтологиями верхнего уровня в соответствии со свойствами терминов
 - Отношения гиперонимии/гипонимии определяют иерархию понятий
 - Бинарные или более арные именованные отношения понятий генерируются на основе семантических отношений с терминами-характеристиками, действиями (сгенерированными из глаголов)
 - Модули онтологии объединяют понятия, полученные из терминов различных ветвей деревьев требований (подзадач)
- Вербальные определения анализируются для дальнейшей формализации онтологии
 - Лингвистический анализ определений глоссария по набору правил
 - Большинство правил не зависят от предметной области
 - Формализация ограничений средствами языковых моделей
 - Генерация ограничений зависит от выразительной мощности и разрешённых конструкций языка онтологии



Пример определения ПОНЯТИЙ ОНТОЛОГИИ

```
Class(Pair partial CompoundObject
restriction(isPairOf allValuesFrom(StellarSystem))
restriction(hasPrimaryComponent
    allValuesFrom(Component) maxCardinality (1))
restriction(hasSecondaryComponent
    allValuesFrom(Component) maxCardinality (1))
restriction(hasObservationType
    allValuesFrom(ObservationalType))
restriction(hasOrbit allValuesFrom(Orbit) maxCardinality (1))
restriction(hasComponent
    allValuesFrom(Component) maxCardinality(2)))
```

```
Class(VisualPair partial Pair
restriction(hasRelationalPosition
    allValuesFrom(RelationalPosition))
restriction(hasObservationalType hasValue(Visual)))
```



Методы создания онтологий: Сус

- Кодирование знаний из источников
 - с помощью программных средств разбора текстов, используя знания, уже сохраненные в базе знаний
- Выявление примитивов представления знаний
 - Понятия, атрибуты, значения атрибутов
- Выявление понятий верхнего уровня
 - Вещь (thing)
 - Неосвязаемое (intangible)
 - Коллекция (collection)
- Использование этих представлений для остальных понятий



Методы создания онтологий: Ушолд-Кинг

- Методика использована при создании онтологии Enterprize
- Концептуализация
 - Выделение важных понятий
 - middle-out подход: дополнение общими (обобщение) и частными понятиями (специализация)
 - Подробное текстовое определение понятий
 - Выявление из определений существенных терминов и их определение
- Кодирование:
 - Разработка логического эквивалента определений
- Интеграция с существующими онтологиями

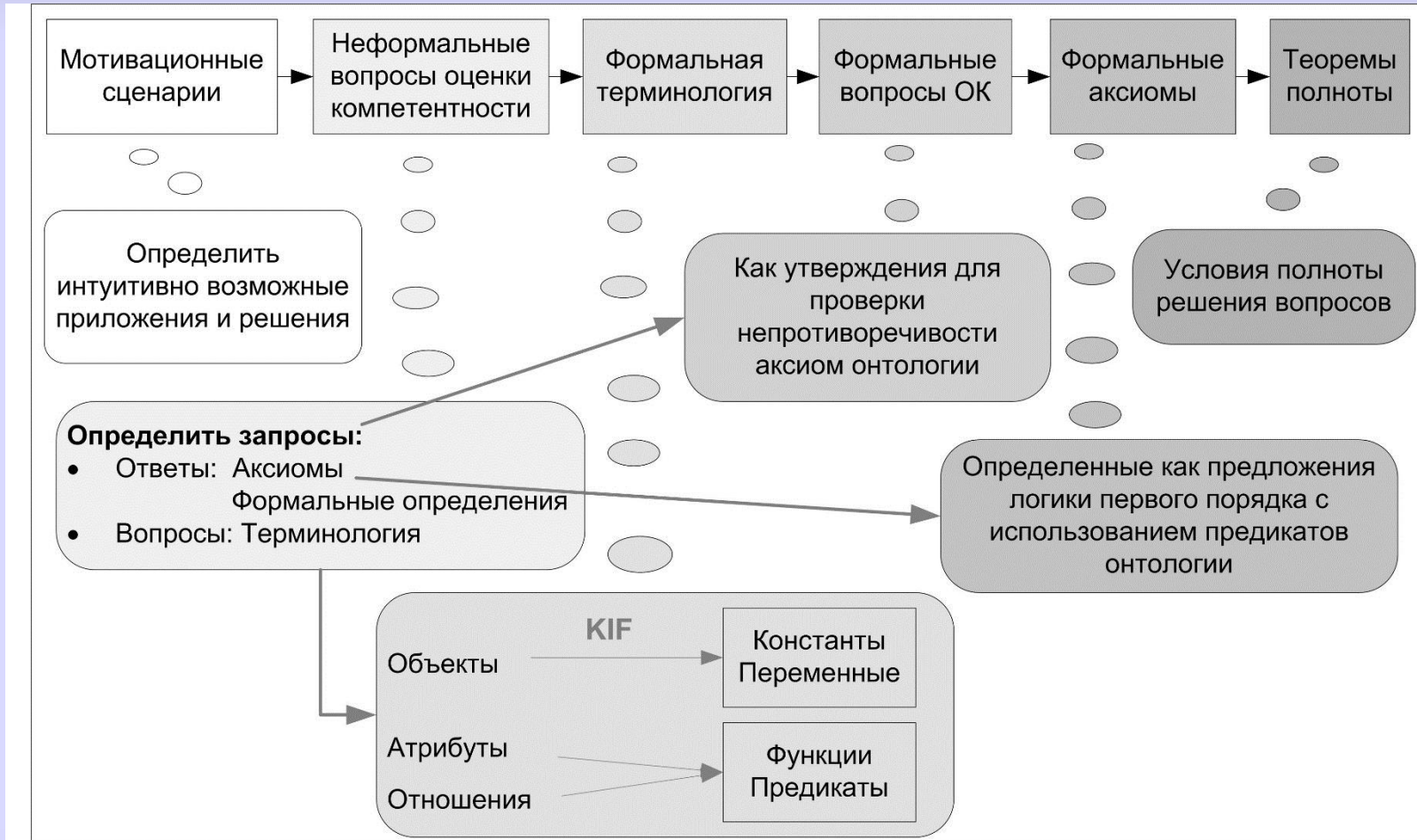


Методы создания онтологий: Грюнингер-Фокс

- (TOVE Ontology)
 - мотивационный сценарий и вопросы для оценки компетентности (онтология должна отвечать на вопросы предметной области)
 - терминология – на основе вопросов
 - аксиомы – на основе ответов
 - теоремы – условия полноты решений для вопросов



Метод Грюнингера-Фокса



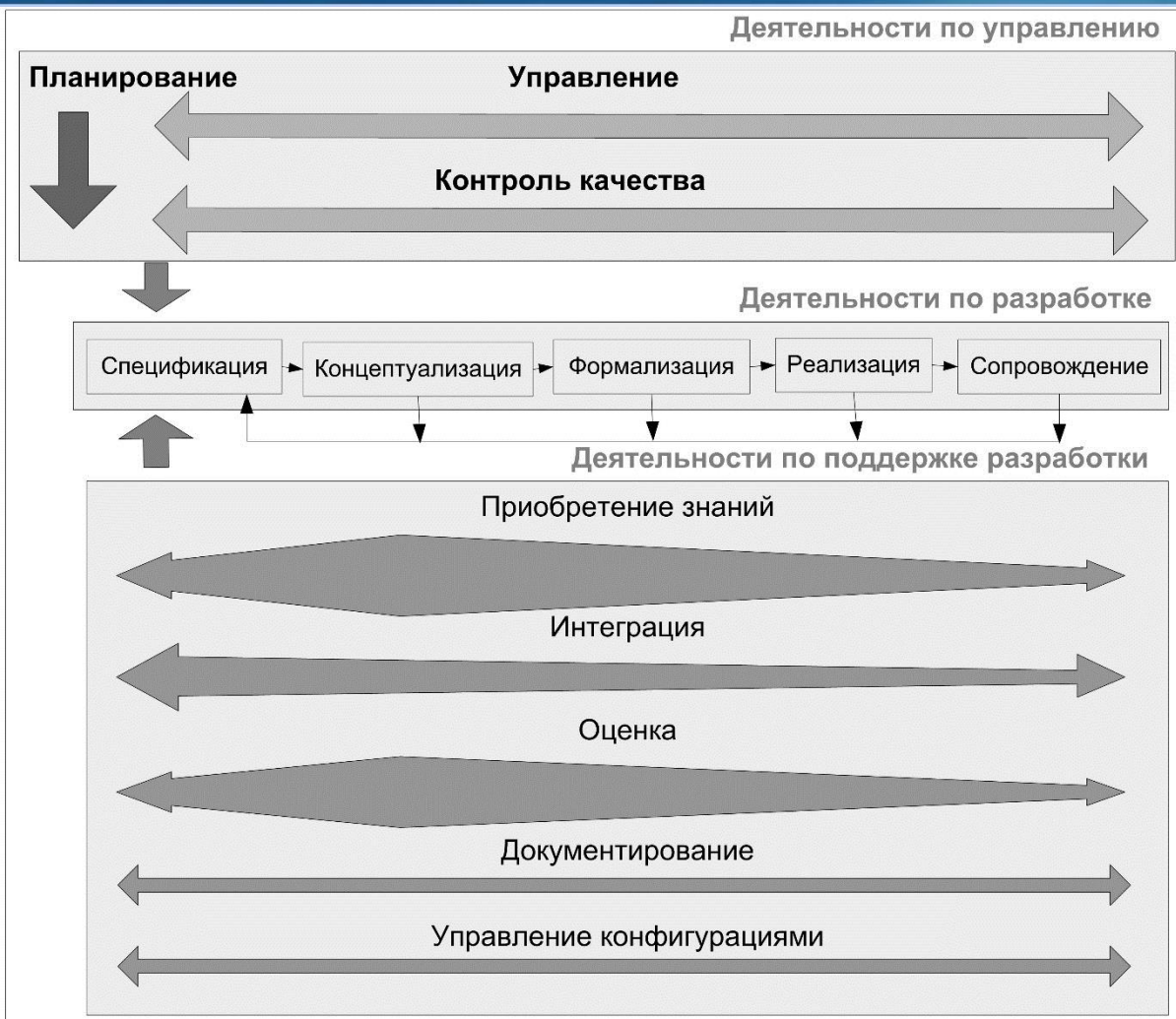


Методы создания онтологий: METHODOLOGY

- METHODOLOGY (KA²)
- Жизненный цикл разработки
 - Спецификация
 - Концептуализация
 - Формализация
 - Реализация
 - Сопровождение



Жизненный цикл разработки онтологий





TOVE

- Вопросы к компетенциям авторства, полномочий, обязанностей
 - Какие роли выполняют агенты?
 - К каким подразделениям организации относится тот или иной агент?
 - С кем агент должен взаимодействовать?
 - Какой информацией пользуется агент в процессе коммуникации?
 - Кому агент должен отчитываться?



Легковесные подходы

- От логической теории к онтологии, которая решает задачу
- SAMOD (Simplified Agile Methodology for Ontology Development), UPONLite (Unified Process for ONtology)
 - Описание на естественном языке (story)
 - Выделение сущностей и связей
 - Строится прототип онтологии
 - оценка соответствия и итеративная разработка
 - нет глубокого анализа предметной области, онтологии под задачу
- Обратная разработка RapidOWL
 - полуавтоматическая генерация OWL из структурированных таблиц (Excel/CSV) с помощью шаблонов
 - разработка неспециалистом
 - D2RQ (SQL → RDF), Ontop (виртуальная онтология поверх баз данных)
- Ontology Design Patterns (ODP)
 - микро-онтологии для типовых задач (участие в событии, часть-целое, изменение параметров)
 - образец как модуль онтологии
- Коллаборативные подходы
 - краудсорсинг онтологий
 - открытые платформы сбора понятий, экземпляров, отношений
 - выделение понятий в открытых энциклопедиях (Wikidata)
 - механизмы проверки: консенсус, рейтинг правок, модерация
- Автоматическое извлечение онтологий (Ontology Learning)
 - Из текстов Text2Onto, OntoLearn, FRED (NLP)
 - Из веба: краулинг страниц со структурированными данными (schema.org, микроформаты) и извлечение понятий
 - LLM-извлечение: «выдели сущности и отношения из текста»



Отображение онтологий и схем

- Отображение онтологий чаще всего не различают с сопоставлением схем
- Лингвистические и структурные подходы к связыванию понятий
 - Сходство по вербальным определениям
 - Эвристические формулы и правила (PROMPT)
 - Сравнение графовых структур, включая правила на графах, алгоритмы на взвешенных графах (SODA)
 - обучающиеся подходы, основанные на статистике (RiMOM)
 - формальный вывод и доказательство (S-Match)
 - и другие



Термины деятельности по сопровождению онтологий

- **Отображение (Mapping)**
 - Однонаправленный процесс, выражение терминологии одной онтологии через другую
- **Интеграция (Integration)**
 - Встраивание одной онтологии в другую
- **Сопоставление (Matching)**
 - вычисление или выявление связей или соответствий между понятиями онтологий (с применением лексических, структурных и других методов), может быть первым этапом выравнивания онтологий
- **Выравнивание (Alignment)**
 - установление соответствий между двумя или более онтологиями
- **Гармонизация (Harmonization)**
 - связывания онтологий на уровне концептуализации в процессе обсуждения
- **Слияние (Merging)**
 - создание новой онтологии из двух или более исходных, которые остаются неизменными
- **Согласование (Reconciliation)**
 - приведение в соответствие двух или более онтологий, обычно требующий изменений одной или более сторон
- **Сравнение (Comparing)**
 - нахождение различий между двумя или более онтологиями



Графы знаний

- **Создание графов знаний**
 - Исходные данные: неструктурированный текст (статьи, базы знаний, веб-страницы)
 - Извлечение сущностей и отношений,
 - Вложения (эмбеддинги) предложений или слов позволяют кластеризовать схожие понятия и разрешать неоднозначности
 - на основе векторов строится граф знаний: узлы — сущности, рёбра — связи
 - масштабируемые графы, созданный без полного ручного моделирования
- **Обратное направление: вложения, порождённые графами знаний**
 - Существующие графы знаний: связанные сущности
 - обучение вложений.
 - Каждый узел или отношение отображается в вектор
 - трансляционные модели (TransE, RotatE: голова + отношение \approx хвост), случайные блуждания (node2vec) и графовые нейросети (GNN)
- **Вложения сохраняют топологию и семантические связи графа**



- Однако предыдущие методы не учитывают особенности схем как представлений данных и онтологий как системы понятий
 - Структура типа в схеме может содержать элементы сразу нескольких сущностей реального мира, данные в схеме в виде, удобном для решения задачи
 - Понятия с разной структурой могут соответствовать совпадающим экстенционалам сущностей реального мира
 - Необходимы подходы, различающие семантику понятий и интерпретации в соответствии с ней



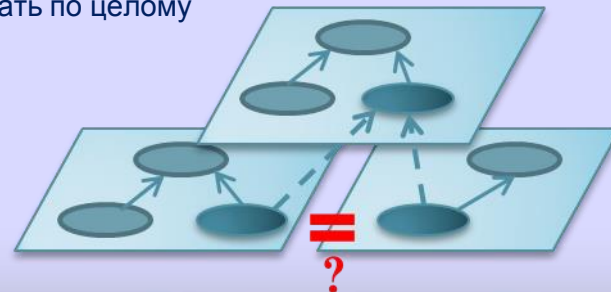
Принципы отображения ОНТОЛОГИЙ

- Установление слабых соответствий между понятиями и отношениями
 - Анализ вербальных определений и структурных определений с применением различных методов
- Унификация онтологических моделей
 - Установление отображений конструкций онтологических языков
- Множества сущностей, соответствующих понятиям
 - Восстановление отношений специализации/генерализации между понятиями (отношения подмножества между интерпретациями/экстенционалами понятий)
 - Проверка включения классов в моделях сводимых к логикам описаний
 - Каноническая модель должна быть ограничена расширениями, соответствующими модели, основанной на конкретной логике (например, OWL DL)
- Другие подходы к формальной верификации спецификаций понятий
 - Связывание на основе метаонтологий
 - непротиворечивость свойств понятий
 - Интерактивное доказательство уточнения спецификаций
 - Для любых спецификаций, выраженных средствами языка СИНТЕЗ или сводимых к логике первого порядка



Метаонтологии

- **Фундаментальные метасвойства понятий (Н. Гуарино, OntoClean)**
 - существенность – неотъемлемость свойства сущности
 - строгость – принадлежность существенного свойства сущности в любом воображаемом контексте
 - идентификация – является ли утверждение об объекте идентифицирующим его свойством
 - собственная идентификация – несёт ли идентификацию само утверждение (а не наследует)
 - неизменность – может ли свойство меняться во времени
 - зависимость – может ли сущность существовать без других
 - постоянство – как долго сущность остаётся таковой
 - объединение – существование экземпляров понятия как целых сущностей в отношении часть/целое и другие
- **Правила, ограничивающие взаимные значения метасвойств отображаемых понятий ($q \supseteq p$)**
 - если q строгое для любых сущностей, то p также строгое для любых сущностей
 - если q несёт критерий идентификации сущностей, то и p также
 - если q несёт критерий объединения, то и p также
 - если q не несёт объединение, то и p также
 - всякая сущность должна быть значением наиболее общего свойства, несущего его идентификацию
- **Следствия**
 - Родовые понятия (сущности), видовые понятия (категории), ролевые понятия
 - Ролевые могут быть подпонятиями родовых, видовые могут быть подпонятиями ролевых или родовых
 - Если часть является неотъемлемой, целое может быть идентифицировано по части
 - Если целое инвариантно, то части можно идентифицировать по целому





Экстенциональный подход к отображению онтологий

- Отображение на основе денотационной семантики
 - «по образцу» (by example)
- Выделение экземпляров для проверки принадлежности отображаемых понятиям
 - объекты, соответствующие сущностям реального мира
 - подпонятия (подмножества экземпляров, определённые по дополнительному признаку)
 - примеры моделей реального мира
 - хорошо классифицированные с помощью онтологий данные
- Формальность подхода
 - Существование хотя бы одного примера модели, в которой сущности не принадлежат одновременно эквивалентным понятиям отображаемых онтологий опровергает корректность отображения



Методы слияния ОНТОЛОГИЙ

- **ONIONS (Гангеми)**
 - Анализ концептуально релевантных данных
 - Аксиоматизация
 - Слияние или интеграция онтологий
- **PROMPT (Ной)**
 - общие суперклассы для классов,
 - слияние слотов объединённых классов
 - классов слияние областей значений для объединённых слотов
- **FCA-Merge**
 - перепостроение решётки понятий на основе документов, релевантных каждой онтологии
- **Ontology Alignment Evaluation Initiative**
 - Вложения
 - Графы знаний
 - Выявление связей
- **Schema matching**
 - "como++" constraint matching unsigned target attributes instance based
 - zero shot station
 - conschema
 - smat - эмбединги ...



Применение онтологий

- Аннотирование информационных сущностей предметной области
- Классификация сущностей
- Идентификация сущностей
- Семантический поиск
- Семантическая интеграция данных
- Создание концептуальных схем для представления и манипулирования данными