

МЕТОДЫ И МОДЕЛИ РАСШИРЕНИЯ СЕМАНТИКИ ОТНОШЕНИЙ В ОНТОЛОГИИ С ПОМОЩЬЮ ОПИСАНИЯ МАТЕМАТИЧЕСКИХ ЗАВИСИМОСТЕЙ МЕЖДУ ПАРАМЕТРАМИ ОБЪЕКТОВ

Диссертация на соискание ученой степени
кандидата технических наук

Научная специальность:

05.13.23 – системы и средства искусственного интеллекта

Белоиваненко Максим Викторович

Научный руководитель:

Терзиян Ваган Яковлевич,
д.т.н., профессор кафедры
искусственного интеллекта ХНУРЭ

Актуальность темы исследования

- Семантическое описание числовых характеристик объектов в онтологии не позволяет устанавливать связь между ними. Знание о функциональной зависимости между свойствами объектов нельзя описать существующими средствами Semantic Web.
- Описание математических зависимостей дает возможность решать проблему проверки непротиворечивости знаний в онтологии. Каждое из числовых значений является результатом выполнения математической функции над известными в базе знаний другими значениями свойств связанных объектов.
- Обеспечивается возможность получения новых знаний путем выполнения вычислений согласно сохраненным в онтологии знаниям об имеющихся зависимостях между характеристиками объектов.
- Выделяются новые средства для классификации и установления отношений для объектов онтологии согласно описанию функционально связанных характеристик.
- Рассматривается задача семантического аннотирования математической зависимости как отношения в онтологии, а также входящих в нее параметров, что ранее не было реализовано при описании математических выражений.

Цель исследования

Разработка метода и моделей описания математического знания, представленного в форме функциональной зависимости, усовершенствование онтологических баз знаний путем расширения семантики связей между объектами предметной области

Объект исследования

Создание формального описания знания об отношениях, которые представлены в виде математических зависимостей, между объектами предметной области в онтологических базах знаний

Предмет исследования

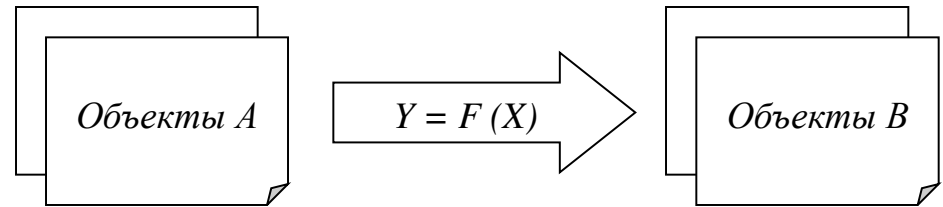
Методы и модели расширения семантики отношений между объектами онтологии предметной области на основе интеграции в нее семантических аннотаций математических зависимостей

Задачи исследования

- анализ методов формального описания математических выражений и способов установления связей между объектами в онтологии;
- разработка модели семантической аннотации математической зависимости как внешнего по отношению к онтологии информационного объекта;
- разработка формата семантической аннотации математической зависимости с помощью средств языка *RDFS*;
- разработка метода расширения семантики отношений между объектами в онтологии посредством добавления знания о соответствующих математических зависимостях;
- разработка модели интеграции семантической аннотации математической зависимости в онтологию предметной области и ее формализация средствами языков *OWL* и *SWRL*;
- создание программных средств на основе разработанных метода и моделей, применение их в системах онтологического моделирования для осуществления логического вывода с учетом включенных в онтологию знаний о математических зависимостях.

Анализ математической функции как способа формализации отношения между объектами реального мира

- Связь некоторых объектов реального мира называется *отношением*. Устойчивая закономерность в таком отношении представляет зависимость
- *Функциональная зависимость* – связь между величинами, описывающими объекты, при которой изменение одних величин вызывает изменение других



- Математическая функция – формальное представление функциональной связи между числовыми параметрами объектов
- Функция F описывает некоторый «закон», который управляет соответствием элементов X одного множества некоторым элементам Y другого множества
- Конкретная функция имеет смысл только для связанных объектов A и B , то есть обладает контекстом своего применения
- Математическая формула, представляющая функцию, является записью отношения между абстрактными объектами

Анализ существующих средств аннотирования информационных ресурсов для записи математических выражений

- *RDFS* – определяет словарь терминов и отношения между этими терминами;
- *OWL* – семантика языка сосредоточена на описании логических отношений;
Языки RDFS, OWL не содержат механизмов для описания математических отношений между объектами
- *TeX, LaTeX* – разметка без возможности указывать семантические характеристики объектов, которые являются элементами документов;
- *MathML* – отсутствует возможность описания области применения формулы при наличии определения семантических характеристик ее частей;
- *OpenMath* – наличие возможности описать семантику объектов, входящих в математическое выражение, без возможности задать контекст;
- *Mizar Language* – язык, предназначенный для записи и автоматического доказательства формально записанных математических выражений.

Анализ существующих решений онтологического описания математических зависимостей

- *OWL2* – предоставляет механизм описания диапазонов значений для свойств онтологии и задает связь между ними системой линейных уравнений; дает возможность записать лишь небольшой класс математических зависимостей;

`owl:DataRange owl:DataComparison`

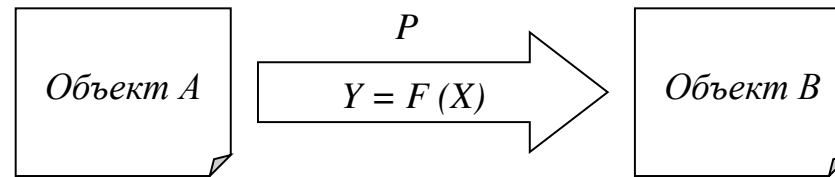
- *SWRL* – строгое установление (непосредственная запись) выражения для вычисления указанных числовых свойств онтологии или частей *RDF*-графа;

`Implies: Antecedent(...) → Consequent(...)`

- *SPIN* – запись сложных семантических правил на основе *SPARQL*-запроса с возможностью подключения внешних сервисов и произвольных функций, в том числе и для реализации вычислений, в онтологической базе знаний;
- *OMDoc* – семантический язык разметки, который позволяет описывать смысл математических текстов и предоставляет возможности для интеграции математических систем.

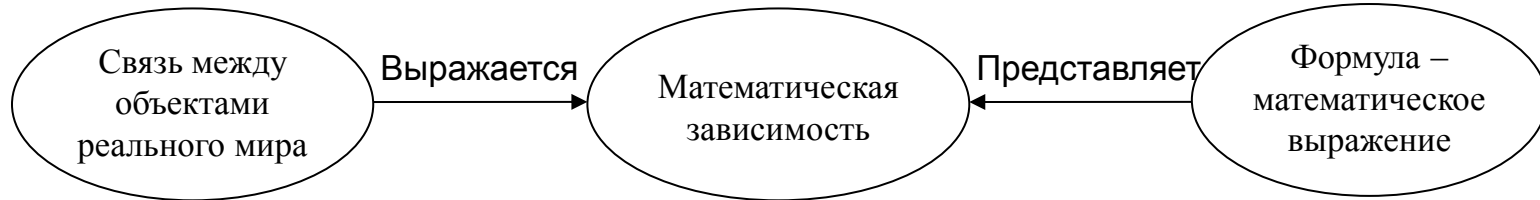
Определение назначения семантической аннотации математической зависимости в онтологических базах знаний

- Семантическая связь между объектами онтологии A и B существует на основании зависимости. Если рассматривается функциональная зависимость F , то она может быть представлена некоторым выражением, которое задает смысл данной связи
- Семантическая аннотация позволяет объяснить, на основании каких объектов A и B , их параметров X и Y , каким именно путем стал истинным факт связи P в онтологии



- Назначение семантической аннотации:
 - *описание аспектов*, которые характеризуют механизм, связывающий объекты онтологии и устанавливающий природу отношений;
 - *предоставление возможности* производить вычисления для решения различных задач в контексте предметной области;
 - *установление происхождения* фактов и значений, которые были получены в результате логического вывода с использованием математических вычислений.

Выбор семантических характеристик математических зависимостей

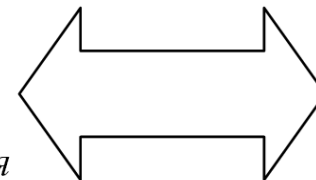


- *Математическая зависимость* определена набором условий, описывающих состояние среды, при котором наблюдается зависимость
- *Математическая зависимость* представлена механизмом, который связывает формальные признаки объектов, т.е. выражает формулу зависимости

Интеграция математической зависимости в онтологию с использованием ее семантической аннотации требует установления соответствия между параметрами среды и переменными формулы

Среда

- объекты
- условия
- отношения
- параметры



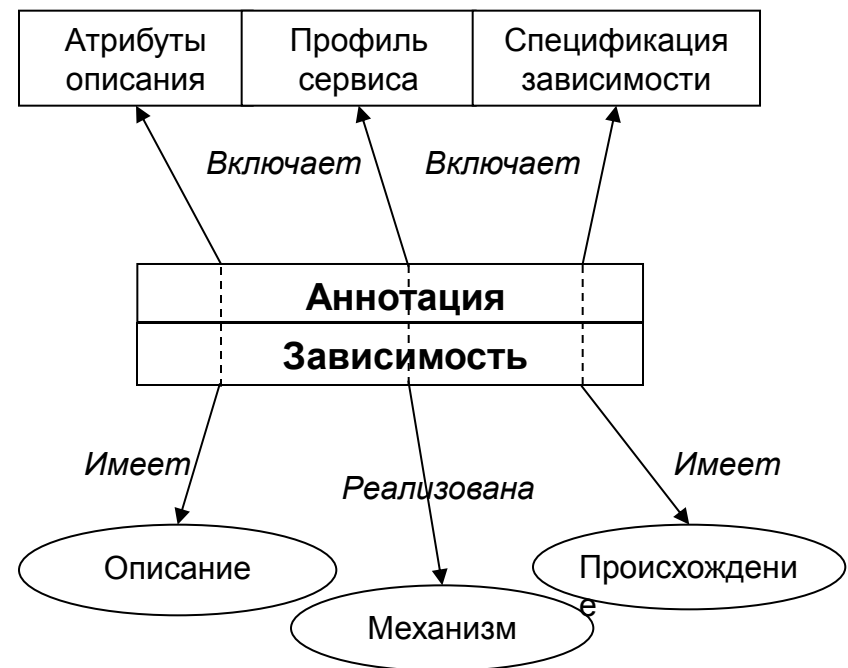
Формула

- математическая запись
- переменные

Разработка модели семантической аннотации математической зависимости (1)

Создание модели **семантической аннотации** математической зависимости предполагает, что:

1. Существует предметная область знаний Ω , которая формально описана онтологией O , и где наблюдается некоторая функциональная зависимость φ между объектами.
2. Зависимость формально представлена математическим выражением ψ , численная обработка которого реализована внешним сервисом σ .
3. Заданная зависимость φ может быть описана в виде информационного объекта f , семантическая запись которого производится в терминах модели онтологии O .



Разработка модели семантической аннотации: параметры математической зависимости (2)

$$f = \{ P, C, X, A \}$$

- P – множество свойств заданной онтологии O , которые соответствуют всем переменным выражения ψ математической зависимости φ ;
- C – набор классов заданной онтологии O , который включает:
 - C_L – множество классов, функциональные отношения между которыми определяются математической зависимостью φ ;
 - C_D – множество классов, соответствующих области применения переменных ψ ;
 - C_R – множество классов, задающие допустимые значения переменных ψ ;
- X – множество неравенств, задающих диапазоны значений параметров, при которых выражение ψ имеет смысл в заданной предметной области;
- A – аннотация математической зависимости, представленная множеством упорядоченных пар типа $\langle n, v \rangle$, где
 - n – идентификатор терма аннотации;
 - v – значение терма аннотации.

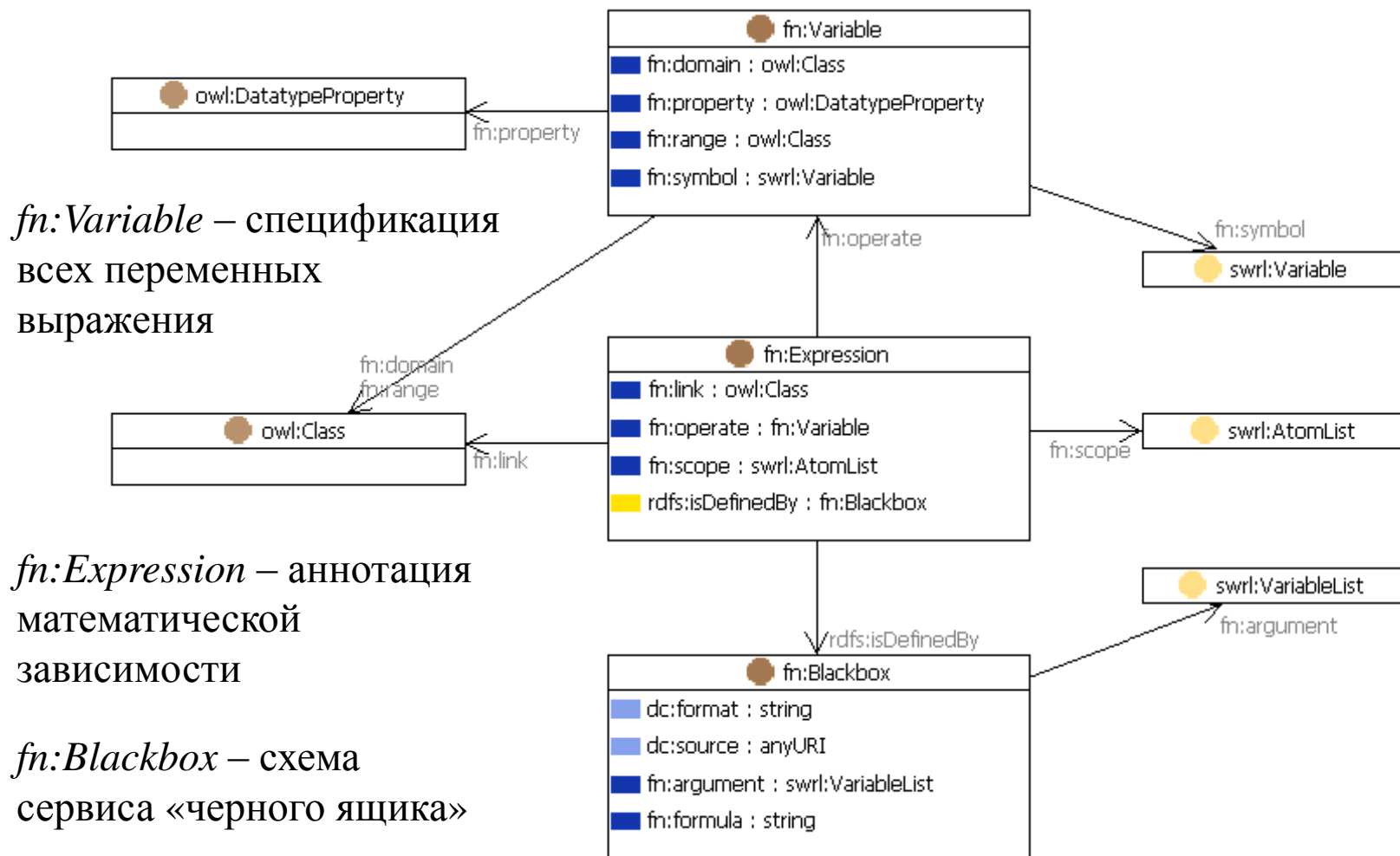
Разработка модели семантической аннотации: профиль математической зависимости (3)

$$s = \{ M, g, t, d \}$$

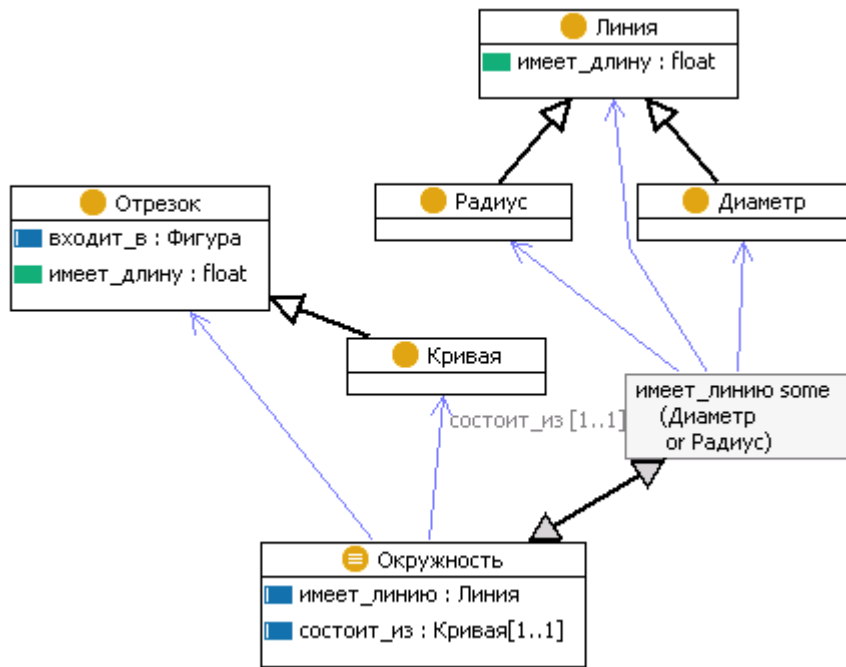
- M – перечень сопоставления свойств p_i из множества P с переменными q_i математического выражения ψ , который представлен в виде набора упорядоченных пар $\langle p, q \rangle$, где
 - p – идентификатор свойства онтологии;
 - q – символьная запись переменной выражения;
- g – формальная (математическая) запись математического выражения ψ ;
- t – формат сервиса σ , заданный некоторым протоколом доступа к нему;
- d – адрес или идентификатор сервиса σ .

Сервис – программа, реализующая обработку соответствующего математического тождества. Сервису на вход поступают значения всех переменных выражения, выходом является заключение – верно ли тождество для данных значений или нет.

Реализация модели семантической аннотации средствами Semantic Web



Пример семантической аннотации математической зависимости для представления взаимосвязанных характеристик объектов



➤ Онтология предметной области «Окружность»

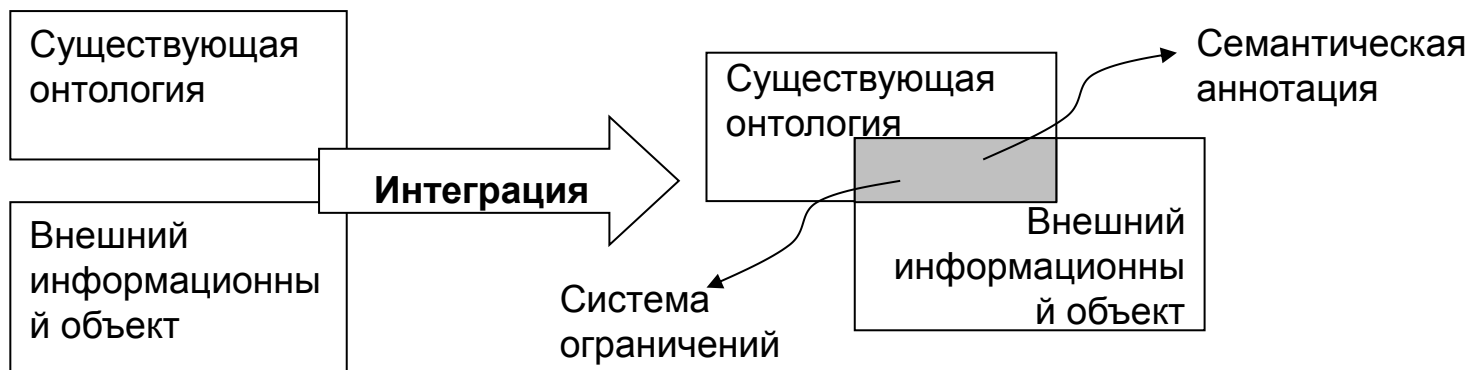
The screenshot shows a semantic editor interface with two main panels:

- Left Panel (fn:operate)**:
 - `п_длина_отрезка`:
 - `rdftype`: `fn:Variable`
 - `fn:domain`: `Отрезок`
 - `fn:property`: `имеет_длину`
 - `fn:range`: `xsd:positiveDouble`
 - `fn:symbol`: `|`
 - `fn:link`: `Кривая`, `Радиус`
 - `fn:scope`: `r > 0, l > r`
 - `rdftype`: `fn:Expression`
- Right Panel (fn:operate)**:
 - `п_характ_радиус`:
 - `rdftype`: `fn:Variable`
 - `fn:domain`: `Радиус`
 - `fn:property`: `имеет_длину`
 - `fn:range`: `xsd:positiveDouble`
 - `fn:symbol`: `r`
 - `rdfis:isDefinedBy`:
 - `bb_длина_окружности`:
 - `rdfs:label`: `Формула длины окружности`
 - `rdftype`: `fn:Blackbox`
 - `fn:argument`: `[l, r]`
 - `fn:formula`: `| == 2 * [Pi] * r`
 - `dc:format`: `Mathematica`

➤ Математическое выражение «Длина окружности»

Определение принципов включения нового информационного объекта в онтологию предметной области

- Интеграция семантической аннотации внешнего информационного объекта и заданной онтологии не является объединением двух онтологических моделей. Это расширение существующей модели онтологии новым знанием.
- Способ включения в онтологию внешнего информационного объекта определяется исходя из описания самого объекта и задач, для которых он будет использован.
- Система ограничений служит способом для сопоставления понятий онтологии информационного объекта с существующей онтологией.
- Включение нескольких внешних объектов, связанных между собой, требует некоторого метаописания для представления правил интеграции данных объектов.



Разработка метода расширения семантики отношений в онтологии предметной области (1)

Описание метода расширения семантики отношений в онтологии путем интеграции аннотации математической зависимости включает компоненты:

Ω – знания об объектах и отношениях между ними в предметной области. Числовые параметры объектов связаны посредством математического выражения ψ .

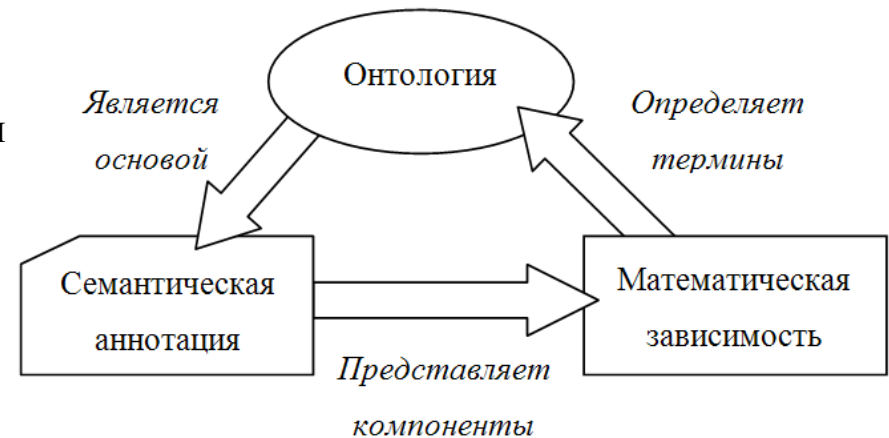
O – базовая онтология, представляющая знания Ω , необходимые для описания математической зависимости φ . В терминах данной модели должно быть создано семантическое описание f .

Θ – знание о некоторой предметной области, для объектов которой характерна рассматриваемая математическая зависимость φ из области знаний Ω .

Q – целевая онтология, представляющая знание Θ , в которой предполагается существование математического отношения, ранее описанного в контексте онтологии O .

Q' – расширенная онтология Q , включающая знания Ω о природе математической зависимости φ и ее семантическую аннотацию f .

Φ – модель интеграции математических отношений онтологии O в онтологию Q с помощью семантической аннотации f . Задаёт связь соответствующих элементов онтологий O и Q .



Разработка метода расширения семантики отношений: процесс включения нового знания (2)

Метод расширения семантики отношений в онтологии предметной области Q предполагает аналитическое описание **двух связанных моделей**:

- семантическая аннотация f математической зависимости
- модель интеграции Φ данной аннотации в целевую онтологию



Результатом процесса интеграции является модифицированная онтология Q' , которая содержит знания о включенной в нее математической зависимости и способе ее вычисления. Описание таких знаний опирается на систему понятий онтологии O , записанных в определении самого математического выражения.

Разработка модели интеграции семантической аннотации математической зависимости в онтологию предметной области (1)

$$\Phi = \{ F, R, E \}$$

- F – множество семантических аннотаций f , которые представляют математические зависимости φ и выражают все математические отношения, характерные заданной связи или сущности онтологии Q ;
- R – набор продукционных правил, определяющих связь между экземплярами онтологии Q и соответствующими элементами онтологии O , вида $H \rightarrow B$, где
 - H – логическая формула, которая определяет экземпляры онтологии Q , связанные при помощи математической зависимости φ_i ;
 - B – логическая формула, которая сопоставляет свойства экземпляров онтологии Q , которые были выбраны в H , с параметрами математической зависимости φ_i ;
- E – спецификация ограничений, при помощи которых класс или отношение в онтологии Q определяется согласно набору математических зависимостей φ , представленных множеством семантических аннотаций F .

Разработка модели интеграции семантической аннотации математической зависимости: правила логической интерпретации (2)

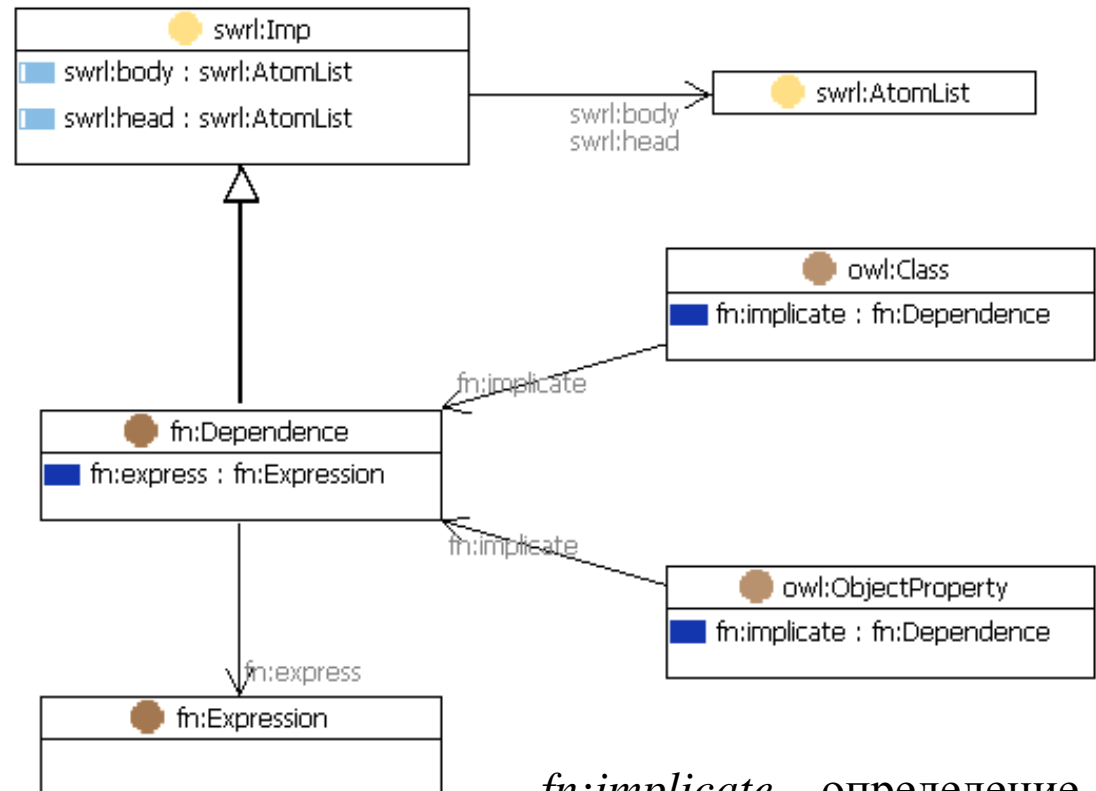
1. При нахождении во множестве R такого правила, которое сопоставляет все переменные некоторой семантической аннотации математической зависимости $f_i \in F$ со свойствами классов заданной онтологии Q , следует считать, что соответствующая зависимость φ_i **определяет спецификацию ограничений E** .
2. Если **все значения** числовых свойств соответствующих экземпляров определены через правила R в заданной онтологии Q и математическое выражение ψ_i аннотации $f_i \in F$ при этом является верным тождеством, то все спецификации ограничений E являются истинными.
3. Если одно или несколько значений свойств экземпляров классов, выбранных согласно правилам R для соответствующей аннотации $f_i \in F$, **не заданы** в онтологии Q , то при истинности ограничений E недостающие значения могут быть вычислены как неизвестные уравнения, выведенного из математического выражения ψ_i аннотации f_i .

Реализация модели интеграции семантической аннотации средствами Semantic Web

swrl:Imp – правила сопоставления объектов онтологии и переменных выражения

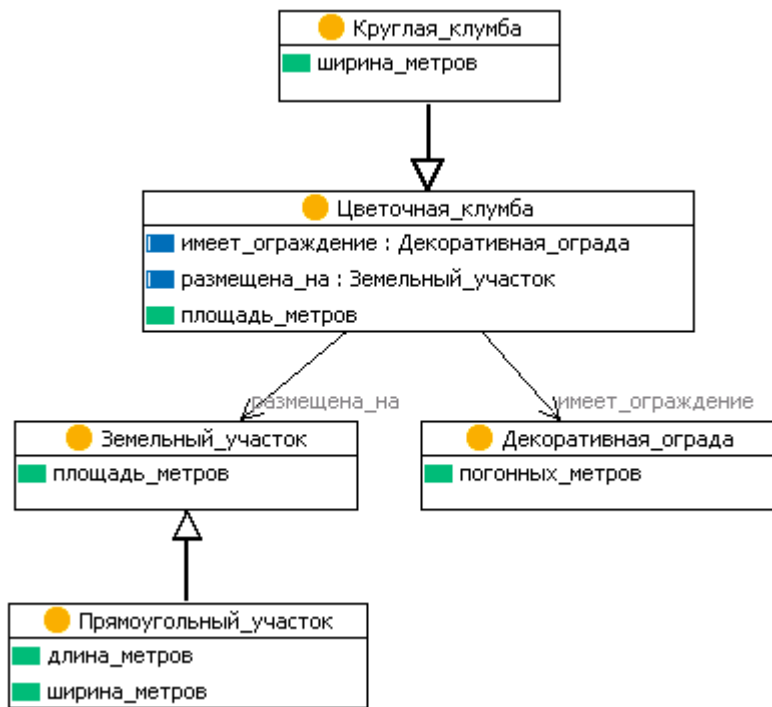
fn:Dependence – описание «природы» отношения или класса

fn:Expression – семантическая аннотация математического выражения

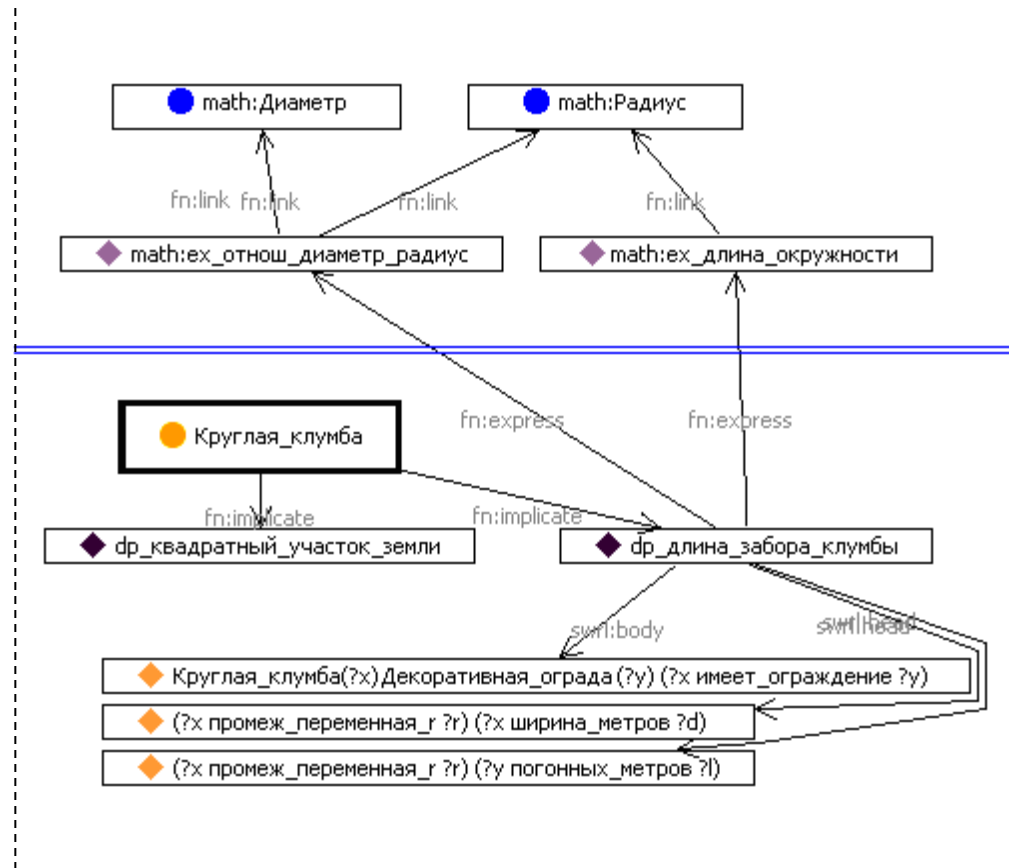


fn:implicate – определение класса или отношения указанной зависимостью

Пример интеграции онтологий путем установления связи между характеристиками объектов из разных предметных областей



➤ Онтология предметной области «Цветочная клумба»

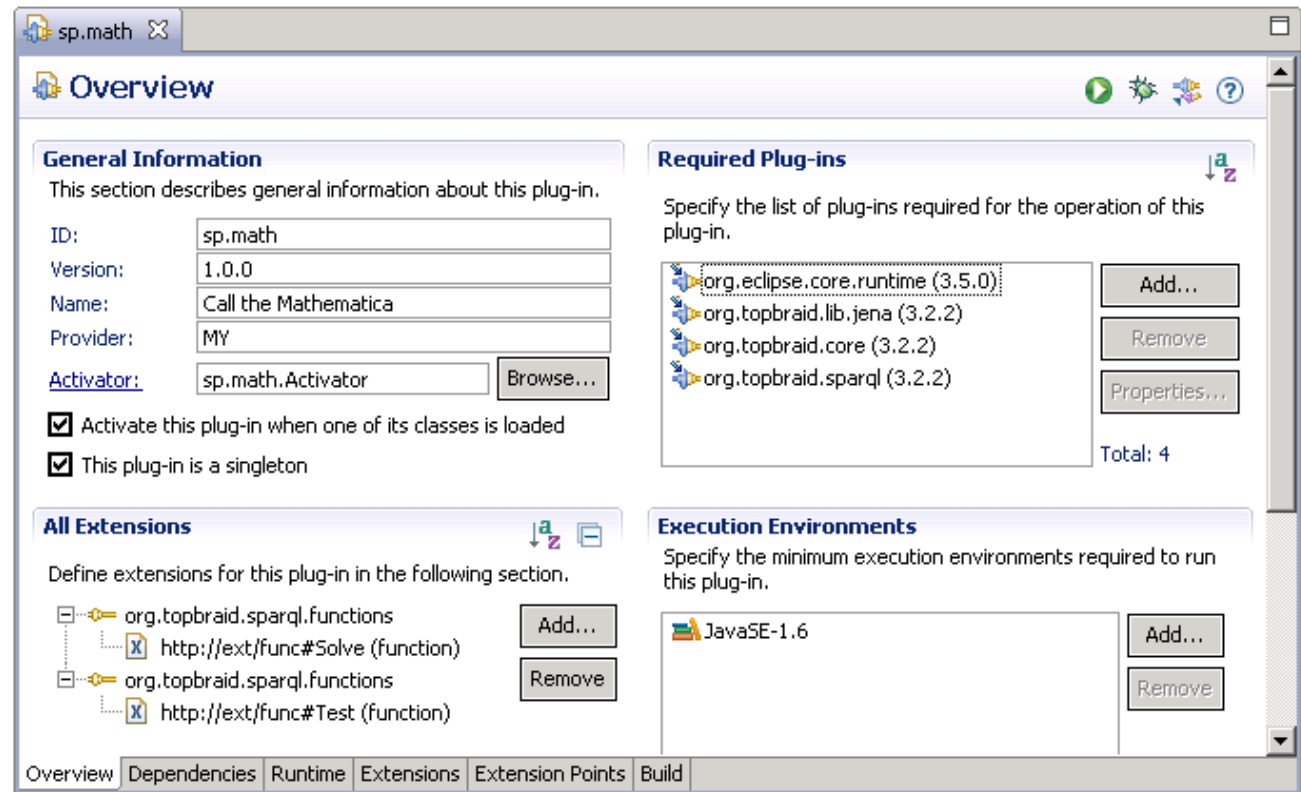


➤ Включение отношений из онтологии «Окружность»

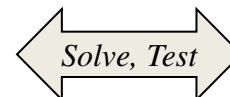
Расширение среды моделирования онтологий TopBraid Composer возможностью обработки математических зависимостей

Использование встроенной машины логического вывода **TopSPIN** для обработки математических зависимостей, включенных в онтологию по разработанному методу:

1. **Открыть** OWL-онтологию с интегрированными аннотациями зависимостей.
2. **Выполнить** SPARQL-запрос *DepsQuery* и сохранить результат в файл *DepsList*.
3. **Запустить** скрипт *ConvMath* с загрузкой данных из файла *DepsList* и записью выдачи в *DepsImpl*.
4. **Импортировать** в открытую модель файл SPIN-правил *DepsImpl*.
5. **Загрузить** Eclipse-модуль *ExtMath* для интерпретации функций *Solve* и *Test*, используемых в правилах.



MATHEMATICA



Применение разработанного метода и моделей в среде онтологического моделирования

Программная реализация метода расширения семантики отношений в онтологии с помощью описания математических зависимостей позволяет вычислять недостающие свойства экземпляров, выполнять классификацию объектов и обеспечивать проверку непротиворечивости данных в процессе редактирования онтологической модели

The screenshot shows the 'Resource Form' for the instance 'прямоугольник_1'. The 'Name' field contains 'прямоугольник_1'. Under 'Other Properties', the following are visible:

- base:имеет_линию**: diagonal_1
- base:имеет_площадь**: 100
- base:состоит_из**:
 - сторона_2 (rdf:type: base:Сторона, base:имеет_длину: 10)
 - сторона_3 (rdf:type: base:Сторона, base:имеет_длину: 10.0)
- rdf:type**: base:Квадрат, base:Прямоугольник

At the bottom, there are tabs for 'Form', 'Graph', and 'Source Code'.

The screenshot shows the 'Resource Form' for the instance 'окружность'. The 'Name' field contains 'окружность'. Under 'Other Properties', the following are visible:

- base:имеет_линию**: диаметр_1
- base:имеет_длину**: 10
- base:состоит_из**:
 - кривая_1 (rdf:type: base:Кривая, base:имеет_длину: 50). A tooltip for this property shows the formula: $l == 2 * \sqrt{[Pi]} * r$ (SPIN constraint at base:Окружность).
- base:имеет_длину**: 10
- base:имеет_площадь**: 100

At the bottom, there are tabs for 'Form', 'Graph', and 'Source Code'.

Онтологический портал менеджмента образовательных и научных ресурсов Украины

Субъект информационной системы		Способ внутреннего представления субъекта
Тип	Назначение	
Модель предметной области	Формирование интерфейса, автоматическое распределение прав доступа	Онтология предметной области, технологии Semantic Web, язык OWL
Информационный ресурс	Оценка качества, учет и систематизация ресурсов предметной области	Классы и свойства онтологической модели, информационные ресурсы как экземпляры онтологии
Математическая формула	Вычисление обобщенных характеристик ресурсов предметной области	Дерево операций, записанное при помощи языка RDF
Пользователь	Отображения информации о ресурсах предметной области согласно поставленной задаче	Реквизиты авторизации, семантическое описание ролей пользователей в Web-системе
Документ	Подтверждение значений характеристик информационных ресурсов предметной области	Файл произвольного формата, истинность которого подтверждается цифровой подписью

Информационно-аналитическая система хранения статистической информации



Внедрение результатов диссертационной работы

- Информационно-аналитическая система хранения статистической информации ООО «ИА “Интерфакс-Украина”»
- Учебный процесс кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники по дисциплинам «Интеллектуальные технологии в Internet» и «Системы распределенного искусственного интеллекта»
- Госбюджетные темы:
 - №195 «Розробка теоретичних засад, методів та моделей інтелектуальної обробки інформації та менеджменту знань у системах розподіленого штучного інтелекту» (№ ГР 0106U003286)
 - №219 «Розробка Web-орієнтованої системи для підтримки процедур акредитації та ліцензування вищих навчальних закладів України» (№ ГР 0108U010139)
 - №233 «Розробка системи підтримки семантичних запитів до онтологічної бази акредитації і ліцензування» (№ ГР 0109U001647)

Выводы (1)

- Разработана новая формальная модель семантической аннотации математической зависимости на основе анализа характеристик функциональной связи между объектами предметной области, заданной с помощью математического выражения. Применение предложенной модели в онтологических базах знаний расширяет возможности описания предметной области.
- Впервые предложена формальная модель интеграции нового знания, заданного в виде математического выражения, в онтологию предметной области. Модель является связующим звеном между целевой онтологией, в которой производится расширение семантики отношений, и онтологией описания данного математического знания.
- Усовершенствован метод расширения семантики отношений между объектами онтологической базы знаний, основанный на методе согласования онтологий, который уточняет описание отношений с помощью системы специальных ограничений. Метод позволяет интегрировать онтологии с разными контекстами применения и описывать объекты одной онтологии в терминах другой.

Выводы (2)

- Опираясь на стандарты Semantic Web, впервые разработаны структуры описания семантической аннотации математической зависимости и модели ее интеграции в онтологию предметной области с помощью языков SWRL, RDFS и OWL и словаря Dublin Core.
- Экспериментальная реализация разработанного расширения языка OWL была использована для усовершенствования среды онтологического моделирования TopBraid Composer. Добавлены возможности вывода новых знаний и проверки непротиворечивости существующих за счет обработки математических зависимостей, описанных в онтологии, с помощью пакета Mathematica.
- Предложенные метод и модели расширения семантики отношений в онтологии позволяют создавать интеллектуальные информационные системы, использующие базы знаний для хранения и обработки числовых характеристик объектов, функциональные зависимости между которыми заданы математическими формулами.
- Предложенная модель семантической аннотации дает новые возможности онтологического описания предметных областей, в которых математические взаимосвязи имеют важную роль.

Выводы (3)

- Разработанные модели позволяют описывать с помощью онтологий и автоматически обрабатывать знания о предметных областях, в которых взаимосвязи определены сложными законами, реализованными в отдельных программных пакетах или модулях.
- Разработанные принципы описания семантики функциональной зависимости как информационного объекта применимы в онтологическом инжиниринге предметной области, при котором нет необходимости анализа математической структуры такой зависимости.
- Распределенный метод создания Web-онтологий и предложенная модель профиля сервиса, доступного в Интернет, позволяет создавать сложные распределенные системы путем интеграции вычислительных мощностей, описанных формальной онтологической моделью.
- Применение разработанных моделей в открытом информационном пространстве с опорой на публичные стандарты Semantic Web ускоряет семантическое описание мира с помощью математических законов. Существующие модели для выражения математических связей, применяемые в корпоративных системах, являются закрытыми.