

Метод оценивания свойств многоуровневых онтологий*

Е.А. Шалфеева¹

¹Институт Автоматики и Процессов Управления ДВО РАН, ул. Радио, д. 5, г. Владивосток, 690041, Россия.

shalf@iacp.dvo.ru

Аннотация. Предложен метод получения объективной информации для оценивания многоуровневых онтологий. Метод основан на измерении структурных свойств онтологий по их графовым моделям. Графы синтаксической структуры онтологии позволяют проводить оценивание межмодульных связей многомодульной онтологии, связей между предложениями онтологии и связей компонентов внутри одного предложения. Представлена модель связи структурных свойств онтологии с ее внешними характеристиками

Ключевые слова: многоуровневая онтология, структурные свойства, метод оценивания свойств онтологии, графовые модели структуры онтологии

1 Введение

Одной из характеристик предметных областей, связанных с научной деятельностью, является наличие в них разных разделов, характеризующихся своей системой понятий, своими знаниями и своим множеством задач [1]. Поэтому при создании интеллектуальных систем, таких как системы, основанные на знаниях, важно хранить онтологию и знания каждого раздела, обеспечивать интеграцию знаний и онтологии разных разделов области.

Средством обеспечения интеграции в этих системах является метаонтология, определяющая систему понятий, используемую при задании онтологии разделов области. Такая метаонтология является онтологией более абстрактного уровня по отношению к онтологиям разделов предметной области. Например, в многоуровневой онтологии химии [1] *метаонтология химии* построена в результате анализа онтологии физической и органической химии, а также некоторых разделов аналитической химии. *Метаонтология химии* (называемая онтологией четвертого уровня) определяет систему понятий, в терминах которой разрабатываются метаонтологии разных разделов химии. Эти метаонтологии разделов используются при создании онтологии второго уровня (онтологии разделов). Предполагается, что теории с еще меньшим уровнем абстракции будут уже представлять сами знания соответствующей предметной области.

Потенциальные преимущества от использования многоуровневых онтологий таковы:

- получение дополнительных уровней представления информации, которые будут повторно использоваться для создания следующих уровней;
- получение более компактного представления текста онтологии за счет введения абстрактных терминов-связей между сущностями и использовании их при определении других терминов.

Методика построения онтологий абстрактных уровней (третьего и/или четвертого) включает создание в них так называемых конструкторов, с помощью которых могут строиться непосредственно термины предметной области. Наличие конструкторов в модели (теории, модуле

* Работа выполнена в рамках проекта ДВО РАН «Развитие систем управления базами знаний с коллективным доступом».

онтологии) более абстрактного уровня обеспечивает возможность создавать с их помощью термины менее абстрактного уровня (в теориях с «меньшим» уровнем абстракции). Конструктор «выглядит как» функция, результатом которой является конструируемая функция («обычная» функция или более простой конструктор). Например, в предметной области «химия» в теории наиболее абстрактного уровня, имеется конструктор *Собственные свойства сущностей*, в теории менее абстрактного уровня с его помощью определяется конструктор *Собственные свойства элементов*, а в теории наименее абстрактного уровня определяется термин *Атомный номер* как одно из собственных свойств элементов.

Введение этих конструкторов позволяет видеть общие виды связей между терминами предметной области и применять их единым образом. С другой стороны, они являются дополнительными абстрактными сущностями, с которыми приходится иметь дело инженеру знаний или эксперту. Это значит, что сложность работы с терминами на некоторых этапах может возрастать. Например, повышается сложность восприятия определений, в которых используются «абстрактные термины-функции», т.е. конструкторы.

Целью исследования является установление зависимости структурных свойств многоуровневой онтологии с внешними свойствами этой онтологии, касающихся людей, работающих с ней, и программных средств, поддерживающих работу с такой онтологией.

Структура изложения материала такова. После постановки задачи представлен подход к измерению структурных свойств многоуровневых онтологий. Далее представлена модель связи структурных свойств онтологии с ее внешними характеристиками в трех разделах, соответствующих разным видам структуры онтологии.

2 Постановка задачи

Исходным материалом для исследования является произвольная *многоуровневая онтология*¹ предметной области. Такая онтология обязательно включает одну или несколько теорий (модулей) уровня непосредственных онтологий некоторой предметной области и хотя бы одну теорию, являющуюся метаонтологией.

Существующая многоуровневая онтология [1] включает несколько теорий для раздела физической химии, несколько – для рентгено-флуоресцентного анализа (они названы в [1] уровнем 2). Имеется метаонтология для этих разделов (уровень 3). Имеется метаонтология (уровня 4), соответствующая структуре всей предметной области (химии).

Таким образом, рассматриваемая в рамках одного раздела химии *многоуровневая онтология* = (Онт4, Онт3, {Онт2₁, ... Онт2_n}). Такая многоуровневая онтология формализована на ЯПЛ [2]. *Метаонтология* определяет систему понятий, в терминах которой разработана онтология третьего уровня для раздела. Она используется при создании онтологий второго уровня: каждая *онтология второго уровня* получается из модели *третьего уровня* с помощью так называемого ее обогащения.

Онтология 4-го уровня может включать такие предложения:

1) *обычные определения терминов-сущностей* и терминов-связей, например, «Сорт Типы сущностей : $\{N \setminus \emptyset\}$ » или «Сорт Типы компонент сущности: (Типы сущностей \rightarrow $\{ \}$ Типы сущностей)»;

2) *ограничения на интерпретацию терминов*, например, «(Тип: Типы сущностей) Тип \notin Типы компонент сущности(Тип)»;

3) *простые конструкторы*, которые выглядят как конструктор-функция, результатом которой является «обычная» функция (термин-связь), имя которой будет задано при ее построении с помощью такого конструктора, например,

«Компоненты сущности \equiv

$(\lambda(\text{Тип1: Типы сущностей}) (\text{Тип2: Типы компонент сущности}(\text{Тип1}))$

$(j(\text{Тип1}) \rightarrow \{ \{ (v: \text{Сущности}) \text{ Тип сущности}(v) = \text{Тип2} \} \setminus \emptyset \})$

или

«Собственные свойства системы \equiv

$(\lambda(\text{Область возможных значений: } \{ \}(\text{Множества значений} \cup \{ \}(\text{Кортежи значений}))$

$([1, \text{Число шагов процесса}] \rightarrow \text{Область возможных значений}))$ »;

¹ Всюду в работе при употреблении термина «онтология» подразумевается «модель онтологии», т.е. формализованное представление онтологии. Здесь, говоря «*многоуровневая онтология предметной области*» автор подразумевает «*многоуровневая модель онтологии предметной области*».

4) *конструкторы с вложенностью* (КсВ), которые выглядят как конструктор-функция, результатом которой является тоже конструктор-функция, а ее результатом является «обычная» функция:

«Собственные свойства сущностей \equiv

$(\lambda(\text{Тип сущности: Типы сущностей})$

$(\lambda(\text{Область возможных значений: } \{ \} (\text{Множества значений} \cup \{ \} \text{Кортежи значений}))$

$(j(\text{Тип сущности}) \rightarrow \text{Область возможных значений}))$ ».

Онтология 3-го уровня может включать:

- 1) *обычные определения терминов-сущностей* и терминов-связей,
- 2) *ограничения на интерпретацию терминов*,
- 3) *простые конструкторы* (ПК),
- 4) *определения терминов-связей через простые конструкторы*, например, «Сорт Участники реакции: Компоненты сущности (Химические реакции, Химические вещества)»;
- 5) *определения простых конструкторов через конструкторы с вложенностью* (ОПКчКсВ), например, «Собственные свойства веществ \equiv Собственные свойства сущностей (Химические вещества)».

В предложениях вида 4 и 5 происходит обогащение онтологии предыдущего уровня, т.е. «значения параметров третьего уровня» подставляются вместо аргументов конструкторов-функций 4-го уровня. Предложения, задающие значения параметров, можно считать предложениями вида 6. Например, для физической химии значение параметра Типы сущностей задается как множество {Химические элементы, Химические вещества, Химические реакции, Табличные значения температуры, Табличные значения давления, Пути реакций, Фазы}.

Однако, таково представление онтологии 3-го уровня «в исходном виде». Если раскрыть *определения терминов-связей через простые конструкторы* и *определения простых конструкторов через конструкторы с вложенностью*, то получим ее представление «в раскрытом виде». В частности, вместо каждого *определения термина-связи через простой конструктор* будет получено *обычное определение термина-связи*, а вместо *определения простого конструктора через конструктор с вложенностью* будет получен *Простой конструктор*. Именно они и будут использованы в онтологиях 2-го уровня.

Онтология 2-го уровня может включать:

- 1) *Обычные определения терминов-сущностей* и терминов-связей (ООТ),
- 2) *Ограничения на интерпретацию терминов* (ОИТ),
- 3) *определения терминов-связей через простые конструкторы* (ОТчПК).

Здесь строятся по конструкторам онтологии третьего уровня некоторые термины-связи (функции над терминами-сущностями). А именно: на основе некоторой абстрактной функции (с абстрактной областью определения и указанной областью значений) вводятся новые термины-функции с конкретизированной областью определения и той же областью значений.

Для многоуровневой онтологии такой структуры (структура которой есть тройка вида (*Онт4*, *Онт3*, {*Онт2₁*, ... *Онт2_n*}), где $\text{Онт4} = \text{ООТ}^2 \cup \text{ОИТ} \cup \text{ПК} \cup \text{КсВ}$, $\text{Онт3} = \text{ООТ} \cup \text{ОИТ} \cup \text{ПК} \cup \text{ОТчПК} \cup \text{ОПКчКсВ}$, $\text{Онт2} = \text{ООТ} \cup \text{ОИТ} \cup \text{ОТчПК}$) важно оценить:

- характеристики *повторной используемости* элементов многоуровневой онтологии,
- характеристики *сложности работы человека с текстом* многоуровневой онтологии,
- характеристики *затрат на создание* многоуровневых редакторов и
- характеристики *сложности работы человека с многоуровневыми редакторами*.

Информация об этих характеристиках важна для аргументации применимости многоуровневых онтологий к моделированию различных сложно-структурированных областей знания.

Характеристики *повторной используемости* элементов многоуровневой онтологии могут быть получены на основе исследования структуры связей онтологий друг с другом. Идеальным случаем представляется такой, когда малое количество абстрактных функций используются большим количеством разных теорий следующего уровня (и в них происходит к ним обращение по многу раз).

Характеристики *сложности работы человека с текстом* многоуровневой онтологии и характеристики затрат на создание многоуровневых редакторов могут быть, в частности, получены на основе исследования структуры предложений - определений конструкторов и определений терминов, в которых используются конструкторы.

² Здесь и далее аббревиатура обозначает множество предложений одного из вышеуказанных видов с например, *ООТ* - *обычные определения терминов (сущностей и связей)*.

Идеальным случаем представляется такой, когда малое количество не очень сложных конструкторов используются при определении целого множества терминов предметной области. Это способствует снижению *трудоемкости разработки многоуровневого редактора*, включающей проведение его испытаний, так и трудоемкость поиска дефектов онтологии (таких, например, как несогласованности определений с конструкторами) при ее создании и усовершенствовании.

Характеристики *сложности работы человека с многоуровневыми редакторами* могут быть получены на основании измерения показателей реальной работы с такими редакторами и описания документации по их использованию людьми.

3. Использование графовых моделей

Одним из методов исследования свойств онтологий является применение единого подхода к оцениванию структурных свойств онтологий [3]. Предлагаемые в его рамках графовые модели онтологий разбиты на группы *графов синтаксических связей*, *графов стандартных связей* и *графов концептуальных связей*. Каждая из групп имеет свои границы применимости. Графы синтаксических связей могут быть созданы на трех уровнях структуры онтологии: на уровне межмодульных связей многомодульной онтологии, на уровне связей между предложениями онтологии (или модуля) и на уровне связей компонентов внутри одного предложения.

Граф использования модулей онтологии обычно полезен при оценивании разных аспектов повторного использования онтологий и адекватности и целесообразности разбиения по уровням, когда важны степень зависимости модулей друг от друга, согласованность и избыточность для всей совокупности определений.

Определение и способ построения такого графа приведены в [3, 4]. Пример фрагмента такого графа, построенного для многоуровневой онтологии химии, представлен на рисунке 1. Стрелка направлена от модуля в сторону другого модуля, конструкторы которого используются. Иногда удобно анализировать связи между уровнями вместо связей между отдельными модулями, относящимися к таким уровням.

Граф связей определяемых понятий [3, 4] обычно используется для контроля достаточности и избыточности множества определяемых терминов, а также сложности связей, отражающих особенности предметной области или особенности ее представления. Эти особенности связаны с трудозатратами при моделировании, сопровождении. При исследовании многоуровневых онтологий такой граф позволяет оценивать разные аспекты повторного использования терминов и конструкторов и целесообразности их введения. Рекомендуется для этих исследований строить *граф связей определяемых понятий*, дополненный информацией о терминах-конструкторах.

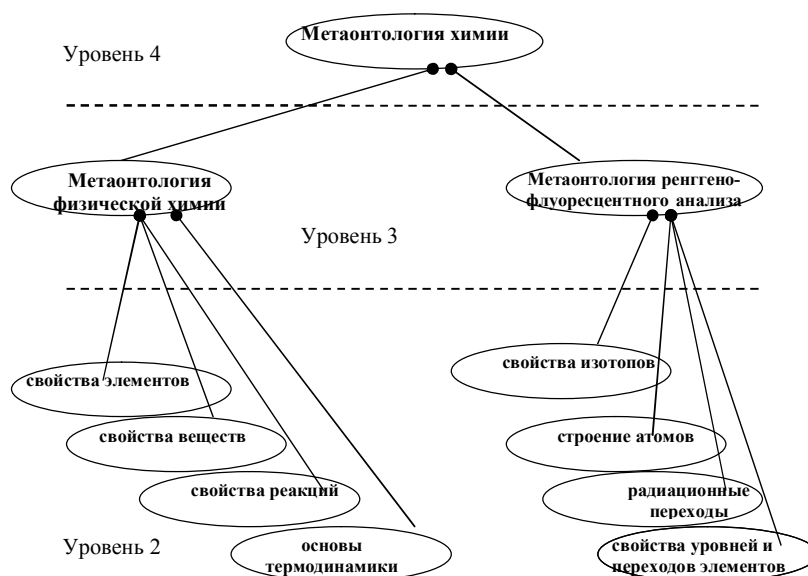


Рис. 1. Фрагмент графа использования модулей для многоуровневой онтологии химии.

Граф связей определяемых понятий (и конструкторов) – графовая модель $\langle V, D \rangle$, где вершины $V = \{v_i\}$, v_i – вершина, соответствующая отдельному предложению,

определяющему термину онтологии; она характеризуется *типом*, принимающим одно из значений {сущность, связь, конструктор простой, конструктор с вложенностью} и меткой – именем термина, частью которого может быть префикс, содержащий имя модуля онтологии, внутри которого он определен;

дуги $D = \{d-j\}$, $d-j$ – связь *использования* от вершины-определяемого-термина к вершине, соответствующей используемому в определении термину.

Пример такого графа представлен на рис.2, здесь имеются вершины–сущности (*типы сущностей* и *типы компонентов сущности*), вершины-связи (*тип сущности*), вершины-конструкторы простые (*сущности процесса*, *компоненты сущности*, *выделенные сущности процесса*), вершины - конструкторы с вложенностью.

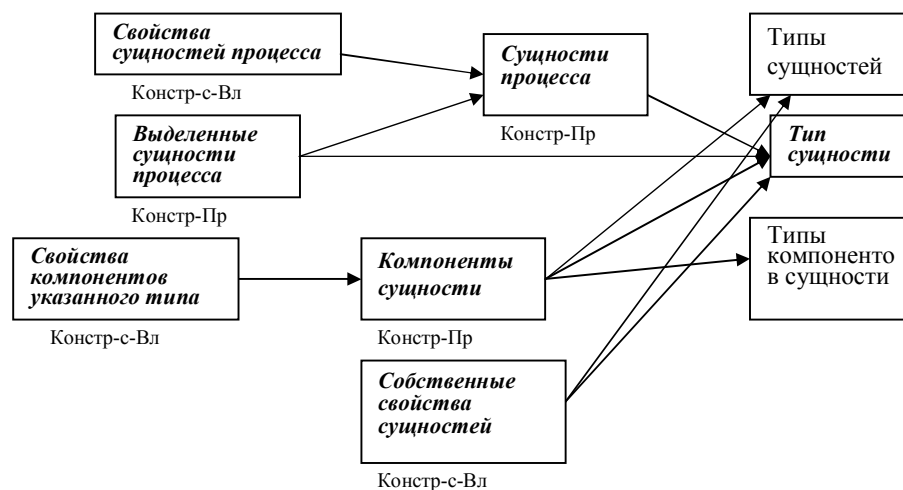


Рис.2. Фрагмент графа связей определяемых понятий для многоуровневой онтологии химии.

Граф структуры предложения наиболее полезен при оценивании трудоемкости разработки компонентов интеллектуальных систем, которые отвечают за содержание знаний предметной области, их создание, хранение, использование или демонстрацию. В частности, сложность структуры предложения связана с трудоемкостью разработки многоуровневого редактора, со сложностью работы в нем эксперта.

Для многоуровневых онтологий требуется расширить определение графовой модели с учетом наличия таких типов предложений, как *простой конструктор*, *конструктор с вложенностью*, *определение простого конструктора через конструктор с вложенностью* и *определение термина-связи через простой конструктор*.

Граф структуры определения термина онтологии – графовая модель $\langle B, D \rangle$,

где вершины $B = \{v-i \mid v-i \text{ принадлежит множеству } \{v-Abстр, v-Ком, v-Огр, v-Опр, v-Перем, v-РезAbстрФ, v-РезФун, v-ТипСуцн}\}$, при этом:

корневая вершина **v-Опр** соответствует названию определяемого термина,

вершины, которые могут быть непосредственно связаны с корневой, соответствуют:

v-Ком - комментарию,

v-ТипСуцн - описанию типа определяемого сущности (для определения термина-сущности),

v-Перем - аргументу определяемой связи или конструктора,

v-Рез - выражению, отвечающему за результат определяемой функции (если есть),

v-Abстр - конструируемой (или используемой) абстрактной функции или абстрактному простому конструктору (если есть),

остальные вершины соответствуют:

v-ТипСуцн - описанию типа переменной (для задания типа аргументу),

v-РезФун - описанию переменной как результата некоторой функции (для задания типа переменной),

в-РезАбстрФ - описанию переменной как результата некоторой абстрактной функции, задаваемой здесь конструктором,

в-Перем - аргументу абстрактной функции или абстрактного простого конструктора,

в-Абстр - конструируемой абстрактной функции (если есть),

в-Огр - некоторому выражению-ограничению на связь значений переменных и результатов функций (если есть);

а дуги $D = \{d-j\}$, $d-j$ – связи, метки которых зависят от типов связываемых вершин: «включает комментарий», «имеет тип», «строится через», «имеет аргумент (номер аргумента)», «задает результат», «является результатом-Ф», «является результатом-К», «включает операнд», «включает ограничение».

Пример такого графа представлен на рис.3, корневая вершина соответствует конструктору онтологии 4-го уровня, с ней связаны вершина **в-Ком**, две вершины **в-Перем** («Тип сущности» и «ОВЗ») и вершина **в-Абстр**. Кроме того, есть три вершины **в-ТипСущн**.

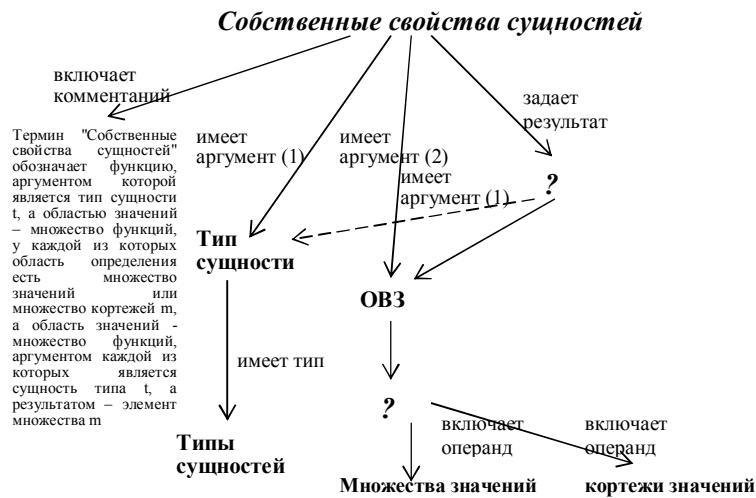


Рис.3. Граф структуры определения термина-конструктора «Собственные свойства сущностей»

4. Связь межмодульных свойств онтологии с ее внешними характеристиками

Граф использования модулей, построенный для многоуровневой онтологии, дает возможность оценить структуру разбиения всей онтологии на составные части и структуру их взаимного использования. Структурные свойства совокупности модулей (онтологии) таковы.

Цикличные использования между модулями - число циклов в *графе использования модулей*.

Цикличные использования между уровнями - число циклов в *графе использования уровней* (граф, объединяющий в одну вершину все модули одного уровня).

Наличие связей между несоседними уровнями - наличие дуги между двумя вершинами, номера которых различаются более чем на единицу, в *графе использования уровней* (см. выше).

Использованность отдельной теории абстрактного уровня - число дуг, входящих в вершину, соответствующую выбранному модулю выбранного уровня в *графе использования модулей*.

Внешние характеристики совокупности модулей связаны со структурными свойствами следующим образом.

Соответствие разбиения онтологии структуре предметной области (осмысленность разбиения по уровням и по модулям) связано со всеми перечисленными структурными свойствами. Оно влияет на эффективность работы эксперта с такой онтологией. *Наличие циклических связей между онтологиями* - признак неудачного разбиения большой онтологии на составные части. Циклы между уровнями – явный дефект. *Возможность связей между несоседними уровнями* в общем случае предполагается. С одной стороны *связей между несоседними уровнями* свидетельствует о востребованности отдельного модуля на разных уровнях, с другой, если один модуль востребован

только на несоседнем уровне, может свидетельствовать о неудачности разбиения. *Использованность отдельной теории абстрактного уровня* со значением единица – показатель ее низкой повторной используемости и свидетельствует о необязательном выделении в отдельный модуль, т.к. этот модуль скорее всего не будет использован и в новых создаваемых онтологиях.

Таким образом, измерение простых структурных свойств по *графу использования модулей* позволяет сделать предварительные выводы об адекватности и целесообразности разбиения по уровням, что особенно важно для оценивания разрабатываемой онтологии. Более подробную оценку можно получить на основе измерения совместных свойств *графа использования модулей и графа связей определяемых понятий* (см. в следующем разделе совместное рассмотрение связей модулей и связей терминов).

Пример. Оценивается *многоуровневая онтология рентгено-флуоресцентного анализа*³. Для нее характерно отсутствие циклических связей между теориями, использование терминов несоседних уровней: имеется 8 терминов модулей уровня 2, которые определены непосредственно на конструкторах уровня 4. Такие показатели свидетельствуют об адекватности и целесообразности разбиения по уровням.

5. Связь свойств структуры связи терминов онтологии с ее внешними характеристиками

Анализ *графа связей определяемых понятий*, построенного для многоуровневой онтологии, дает возможность оценить достаточность и избыточность определяемых терминов и сложность их связей.

Структурные свойства совокупности терминов онтологии таковы: *глубина зависимости определяемых понятий, число неиспользуемых понятий, наличие циклов между определениями, наличие вершин с одинаковым именем* (или *однозначность терминов*), *число неопределяемых понятий* (или *использование вершин не определенных в совокупности модулей*). Эти свойства (их определения приведены в [4]) применимы к любым онтологиям, особенно полезны для масштабных и сложных онтологий, например, многоуровневых.

Специфичные для многоуровневых онтологий структурные свойства таковы.

«Взвешенная» глубина зависимости определяемого понятия – максимальная сумма «стоимостей» всех дуг на путях, исходящих из рассматриваемой вершины, при этом стоимость дуги равна 2, если дуга входит в вершину-конструктор-простой, равна 3, если дуга входит в вершину-конструктор с вложенностью, и равна 1 в остальных случаях. *Примечание.* Возможно более чувствительное определение способа вычисления такого показателя, если использовать вместо 1, 2 и 3 значение *сложности определения соответствующего термина* (см. ниже).

Число конструкторов в определении – число вершин с *типом* «конструктор простой» или «конструктор с вложенностью» на всех путях, исходящих из рассматриваемой вершины в *графе связей определяемых понятий*.

Другой набор измеряемых показателей связан с анализом *графа связей определяемых понятий* совместно с *графом использования модулей*. Такое комбинирование позволяет учитывать значения префиксов вершин (какому модулю или даже уровню принадлежит термин).

Неиспользованные термины абстрактных уровней - наличие в *графе связей определяемых понятий* вершины, полустепень захода которой равна нулю, а уровень абстракции содержащего ее модуля больше двух.

Использованность конструктора - полустепень захода рассматриваемой вершины с типом конструктор-простой или конструктор с вложенностью.

Использованность конструкторов некоторого уровня – число дуг, входящих в вершины, относящиеся к модулям указанного уровня в *графе связей определяемых понятий*.

Использование конструкторов на некотором уровне – число вершин с типом конструктор-простой или конструктор с вложенностью, в которые входят дуги из вершин рассматриваемого уровня.

Использованность типов данных – средняя полустепень захода вершин с типом «сущность» и нулевой полустепенью исхода в *графе связей определяемых понятий*.

Внешние характеристики многоуровневой онтологии связаны со свойствами структуры связей терминов следующим образом.

Затраты на проверку правильности определения каждого термина онтологии тем больше, чем больше значение свойств *«взвешенная» глубина зависимости определяемого понятия* и *число конструкторов в определении*. Это также отражается на затратах по реализации ввода либо

³ Разрабатывается в лаборатории интеллектуальных систем ИАПУ ДВО РАН

вычисления значений такого термина (с учетом реализации всех необходимых проверок допустимости значений термина).

Затраты на создание редактора некоторого уровня зависят еще и от свойства *использование конструкторов на некотором уровне*. Чем больше этот показатель, тем больше влияние двух вышеуказанных свойств, измеренных для каждого определяемого термина рассматриваемого уровня. Повторная используемость реализуемых в редакторе раскрывающихся списков для выбора правильных областей значений описываемых свойств зависит от свойства *использованность типов данных*. Высокие значения этого свойства предпочтительнее.

О потенциальной *неполноте онтологии*, либо о неоправданных затратах на введение абстракций, не пригодившихся в рамках онтологии, можно судить по свойству *неиспользованные термины абстрактных уровней*. Хорошим показателем повторной используемости является *использованность отдельных конструкторов*, большая двух. И чем больше терминов, для которых этот показатель равен 1, тем менее оправданы затраты на введение этих абстракций.

Другая внешняя характеристика, связанная с вышеописанными структурными свойствами – удобство работы пользователя-эксперта с многоуровневым редактором.

Чем больше значения свойства *Использованность конструктора теории абстрактного уровня* для каждого абстрактного термина, тем легче пользователю-эксперту заполнять однотипные связи в онтологии 2-го уровня. Чем больше значение свойства *использованность конструкторов 4-го уровня*, тем более упрощена задача эксперта по заполнению однотипных связей в онтологии 3-го и 2-го уровней. Это особенно заметно для связей, явно задаваемых экспертом (например, для связей *Типа сущности с типами ее компонентов*).

Пример. При разработке многоуровневой онтологии *рентгено-флуоресцентного анализа* свойства *наличие вершин с одинаковым именем, число неиспользуемых понятий и использование вершин не определенных в совокупности модулей* позволили своевременно обнаружить дефекты (к сожалению, введение в онтологию разных уровней абстракции не способствуют уменьшению дефектов).

Имеются четыре термина-конструктора онтологии уровня 3, которые использованы однократно на уровне 2. Предварительный вывод: единообразие (введение этих четырех абстракций) не оправдано с точки зрения усилий эксперта на прочитывание онтологии.

Есть три конструктора уровня 4, которые определены в терминах других конструкторов этого же уровня. Это повышает сложность проверки правильности, внесения изменений и понимания смысла определяемого термина. Свидетельство тому – определение одного из них:

свойства компонентов указанного типа $\equiv (\lambda(\text{Тип1: Типы сущностей}) (\text{Тип2: Типы компонентов сущности (Тип1)}) (\lambda(\text{Область возможных значений: } \{\}\{\text{Множества значений} \cup \{\}\text{Кортежи значений}) (\text{Сущность типа 1} \rightarrow j(\text{Тип1}), \text{Сущность типа 2} \rightarrow \text{Компоненты сущности (Тип1, Тип2) (Сущность типа 1)}) \rightarrow \text{Область возможных значений}))$.

Пример одного из путей в графе (в скобках указана стоимость): *Энергия связи электронов на энергетическом уровне элемента* \rightarrow *Свойства уровней элемента* (2) \rightarrow *Свойства компонентов указанного типа* (3) \rightarrow *Компоненты сущности* (2) \rightarrow *Типы компонентов сущности* (1). Длина пути равна 4 (4 дуги), а вычисляемое значение «взвешенной» глубины – 8 (на фоне желаемого значения не более 5).

Высокие значения свойств «взвешенная» глубина зависимости определяемого понятия и число конструкторов в определении увеличивают трудоемкость работы эксперта и инженера знаний по проверке онтологии и внесению изменений приведен ниже.

Многоуровневый редактор призван облегчить пользователю определение терминов 2-го уровня. Одна из реализуемых цепочек связи терминов разных уровней (2-го, 3-го и 4-го), в рассматриваемой онтологии - *Энергия связи электронов на энергетическом уровне элемента* (уровень 2) \rightarrow *Свойства уровней элемента* (уровень 3) \rightarrow *Свойства компонентов указанного типа* (уровень 4). Самым сложным термином является конструктор со вложенностью, т.е. термин 4-го уровня (определение выше). Остальные определения намного проще, например, «*сорт Энергия связи электронов на энергетическом уровне элемента: Свойства уровней элемента (R(0, ∞))*».

Но при проведении проверки корректности онтологии проводятся достаточно трудоемкие шаги перехода от конструктора к конструктору (в частности, для одного этого термина - около получаса).

Сначала по четвертому уровню уточняется третий, на место аргументов конструктора подставляются указанные значения:

Свойства уровней элемента \equiv *Свойства компонентов указанного типа (Химические элементы, Энергетические уровни) = Свойства компонентов указанного типа* $(\lambda(\text{Химические элементы}) (\text{Энергетические уровни: Типы компонентов сущности (Химические элементы}))$

(λ (Область возможных значений: $\{\}$ (Множества значений \cup $\{\}$ Кортежи значений))
 (Сущность типа 1 \rightarrow j(Химические элементы),
 Сущность типа 2 \rightarrow Типы компонент сущности (Химические элементы)) \rightarrow Область возможных значений))

При этом (в соответствии с описанием второго аргумента в конструкторе «Тип2: Типы компонент сущности (Тип1)») делается проверка принадлежности подставляемого значения результату функции Типы компонент сущности от аргумента «Химические элементы». И получается такой конструктор:

свойства уровней элемента (Область возможных значений) ==
абстрактная функция
 (λ (Область возможных значений: $\{\}$ (Множества значений \cup $\{\}$ Кортежи значений)) \rightarrow
 (Химические элементы, Энергетические уровни) \rightarrow Область возможных значений))

Далее по этому построенному конструктору (по этой полученной абстрактной функции без имени) уточняется требуемый термин второго уровня, на место аргументов конструктора подставляются указанные значения:

Сорт Энергия связи электронов на энергетическом уровне элемента: Свойства уровней элемента ($R(0, \infty)$) = *абстрактная функция*
 (λ ($R(0, \infty)$: $\{\}$ (Множества значений \cup $\{\}$ Кортежи значений)) \rightarrow
 (Химические элементы, Энергетические уровни) $\rightarrow R(0, \infty)$))

Также осуществляется проверка принадлежности аргумента указанному типу значений в построенной абстракции: $R(0, \infty)$ принадлежит множеству « $\{\}$ (Множества значений \cup $\{\}$ Кортежи значений)».

В результате (при раскрытии термина *Энергия связи электронов на энергетическом уровне элемента* через предназначенные конструкторы) получается достаточно легко воспринимаемое предложение:

Сорт Энергия связи электронов на энергетическом уровне элемента: (Химические элементы, Энергетические уровни) \rightarrow ($R(0, \infty)$).

Его простота - одно из преимуществ, предоставляемых многоуровневой онтологией. Этого термина нет в тексте онтологии, но его построили с помощью онтологии.

Предполагается, что для каждого такого термина, раскрываемого через 2 уровня (2 дуги в подцепочке) и характеризующего сложностью, близкой к 8, а не к 5, будет затрачено ручной работы по проверке корректности онтологии при ее создании и сопровождении не менее получаса.

Более точные методы подсчета трудозатрат при создании редактора могут быть установлены экспериментальным путем и с учетом более «точных» характеристик сложности определений используемых конструкторов (при вычислении «взвешенной глубины» вместо сложности «3» для *Свойства компонентов указанного типа* может рассматриваться «21» (сложность как число дуг в графе структуры предложения - см. ниже) или «8» (длина максимального пути в этом графе).

Пример удобства работы с многоуровневым редактором таков. На 3-ем уровне многоуровневого редактора по химии пользователь задает один раз все множество «Типы сущностей» - {Химические элементы, Оболочки, Энергетические уровни, Радиационные переходы орбитальных электронов, Радиоактивные изотопы, Энергии излучения} и один раз - все возможные значения функции «Типы компонент сущности»: аргументу «Химические элементы» - соответствует множество {Энергетические уровни, Радиационные переходы орбитальных электронов}, а аргументу Оболочки - {Энергетические уровни}. Других соответствий нет. Благодаря этому на следующих шагах в редакторах 3-го и 2-го уровней выбор пар аргументов пользователем-экспертом ограничивается только этими значениями.

6. Связь свойств структуры определений терминов онтологии с ее внешними характеристиками

Анализ *графов структуры предложения* обеспечивает показатели трудоемкости разработки редакторов интеллектуальных систем, сложности работы эксперта в редакторах, показатели проектирования баз данных для хранения знаний.

Структурным свойством, определяемым по таким графам, является *структурная сложность определения*, особенно важная для *определения конструкторов* многоуровневых онтологий. Желательным вариантом ее вычисления является «тройка» простых измерений *в графе структуры определения*: (число всех дуг; число всех вершин, кроме в-Ком; длина максимального пути). Другие показатели таковы:

число особых вершин – число вершин с типами *в-Абстр* и *в-РезАбстрФ* в графе структуры определения;

число проверяемых соотношений - число дуг «является результатом-Ф», «является результатом-К», «включает ограничение» в графе структуры определения;

сложность конструируемых типов - число дуг «включает операнд», выходящих из вершин типа *в-ТипСуцн* в графе структуры определения.

Внешние характеристики многоуровневой онтологии связаны со свойствами структуры предложений следующим образом.

Затраты на проверку корректности онтологии при ее создании и сопровождении тем больше, чем больше значение любой из этих «сложностей» для терминов-конструкторов. Также больше будет *трудоемкость разработки* окон редактирования терминов на основе этих терминов-конструкторов (ввиду необходимости контроля соответствия всех элементов определения и реализации подфункций проверки выполнения указанной связи сущностей). Практика показала, что превышение значений показателей 12, 10, 5 (соответственно) приводит к заметному возрастанию усилий на работу с текстом определения.

Трудоемкость работы программиста по созданию редактора, в том числе по обеспечению в редакторе 3-го уровня выбора допустимого типа значений из соответствующего множества зависит от значений *числа проверяемых соотношений* и *сложность конструируемых типов*.

Проект базы данных для хранения знаний о связях между сущностями в предметной области зависит от структуры определений терминов-функций 2-го уровня. На *число таблиц* влияет *число графов структуры определения*, в которых из *вершины в-Опр* выходят дуги «имеет аргумент (номер аргумента)», а на *число столбцов* в каждой такой таблице - *число дуг* «имеет аргумент (номер аргумента)», выходящих из *вершины в-Опр* в графе структуры определения.

Пример. В многоуровневой онтологии *рентгено-флуоресцентного анализа* сложность определения термина (из рис. 3) в виде трех показателей – 10, 8, 3. Один из них показывает на заметную сложность определения. Однако благодаря этому сложному определению становится возможным упрощение других определений, например термина уровня 3 *собственные свойства оболочек* (сложность характеризуется значениями 2, 3, 2).

7. Заключение

Структурные свойства онтологии отражают не столько особенности предметной области, сколько особенности ее представления инженером знаний. Эти особенности влияют на трудозатраты инженера знаний по созданию, проверке и сопровождению онтологий и моделей знаний. Они влияют и на трудозатраты программистов, реализующих программные системы, создаваемые на основе онтологий.

Многоуровневая онтология призвана наиболее полно отразить терминологию сложно структурированной предметной области, выявить общее и частное в содержании такой сложно структурированной области и обеспечить возможность повторного использования описания терминов и связей при инженерии знаний и при разработке интеллектуальных систем.

Характеристики создаваемой многоуровневой онтологии и показатели трудозатрат по созданию программных систем, создаваемых на основе такой онтологии, сильно связаны с простыми структурными свойствами этих онтологий. Измерение структурных свойств с помощью графового подхода дает возможность получения своевременных оценок. Настоящая статья является демонстрацией такого подхода. Исследование проведено по отношению к одной предметной области. Продолжение исследований к другим предметным областям, в которых целесообразно разработка многоуровневых онтологий и многоуровневых редакторов знаний, позволит создать модель связи структурных свойств многоуровневой онтологии с ее внешними свойствами.

Литература

- [1]. Артемьева И.Л., Рештаненко Н.В. Интеллектуальная система, основанная на многоуровневой онтологии химии // *Программные продукты и системы*, 2008, № 1. С. 84-87.
- [2]. Клещев А.С., Артемьева И.Л. Необогатенные системы логических соотношений // *НТИ. Сер.* 2, 2000, № 7 - 8: № 7 с.18-28, № 8 с. 8-18.
- [3]. Клещев А.С., Шалфеева Е.А. Определение структурных свойств онтологий // *Изв. РАН. Теория и системы управления*, 2008. №2. С. 69–78
- [4]. Шалфеева Е.А. Каталог структурных свойств онтологий. Свойства синтаксической структуры // Владивосток: ИАПУ ДВО РАН. 2007. С. 28.