

# Описание понятий-ролей в лингвистических и онтологических ресурсах \*

© Н.В. Лукашевич

Научно-исследовательский вычислительный центр МГУ им. М.В. Ломоносова  
louk@mail.cir.ru

## Аннотация

В статье рассматриваются онтологические характеристики таких понятий как понятий-роли, показывается их отличие от понятий-типов. Мы покажем, что различие между онтологическими характеристиками понятий-типов и понятий-ролей обязательно нужно учитывать при разработке онтологических и лингвистических ресурсов для автоматической обработки текстов, если предполагается использовать логический вывод по иерархическим отношениям. На примере описаний Тезауруса русского языка RuТез рассматриваются способы описания понятий-ролей в лингвистических онтологиях.

## 1 Введение

Во многих компьютерных ресурсах и приложениях таких как онтологии, системы искусственного интеллекта, объектно-ориентированное программирование, информационно-поисковые тезаурусы и многие другие центральное место занимают отношения между классами и подклассами понятий. Такие отношения обычно рассматриваются как иерархические, транзитивные и обладающие свойством наследования: атрибуты, характеристики вышестоящего понятия наследуются на нижестоящее понятие.

Отношение между классами и подклассами понятий может носить разное название в зависимости от области использования ресурса: таксономическое отношение, родовидовое отношение, IS-а отношение, отношение гипонимии и гиперонимии (в лексических ресурсах). Далее в тексте статьи мы будем ссылаться на это отношение как «родовидовое отношение».

При установлении родовидовых отношений разработчики ресурсов явно или неявно используют ставшие классическими диагностические

высказывания. Например, если понятие X является видом понятия Y, то можно сказать, что «X – это Y», «X, Z и другие Y», «к числу Y относятся X» [3, 12].

Однако с онтологической точки зрения эти высказывания могут соответствовать несколько различным отношениям между сущностями внешнего мира, а именно, отношениям между разделяемыми онтологами типами и ролями («тип-тип», «тип-роль», «роль-тип», «роль-роль»). Например, отношения «тип-тип» (*береза – это дерево*) и отношения «тип-роль» (*яблоко – это пища*) в равной степени могут быть выражены всеми вышеперечисленными тестами. Различие заключается в том, что береза остается деревом в каждый момент своего существования, а яблоко может быть использовано в пищу, может быть использовано для других целей, может вообще никак не использоваться.

Как отмечает Н.Гуарино [9], частично возникновение таких различий связано с тем, что то, что описывается как лексическое отношение между словами, не всегда соответствует онтологическим отношениям между сущностями внешнего мира, в то время как в тех приложениях, в которых используются онтологии, часто речь идет именно об отношениях между сущностями предметной области.

Особенно серьезно проблемы смешения отношений стоят перед разработчиками понятийных ресурсов для автоматической обработки текстов, информационно-поисковых приложений в широких предметных областях. В таких приложениях ресурсы, с одной стороны, должны в значительной мере учитывать существующую понятийную систему языка (группы языков) - это особенность отражается в специальном термине «лингвистические онтологии» [7]. С другой стороны, для сохранения необходимых свойств моделируемых отношений эти отношения должны устанавливаться на основе понятийного, онтологического анализа, а не только с использованием языковых диагностических высказываний.

В то же время нужно подчеркнуть, что онтологии, создаваемые и для другого рода компьютерных приложений, достаточно трудно отделить полностью от естественного языка.

Единицы онтологий часто носят языковые или мнемонические названия, тем самым дополнительно «провоцируя» применение неоднозначных языковых тестов.

Так, Йорик Вилкс в [14] утверждает, что символы в языках представления фундаментально базируются на естественном языке, что язык представления - это средство человеческой коммуникации с присущими ему динамизмом, многозначностью и возможностью расширенного толкования.

В любом случае, на наш взгляд, в «языковую ловушку» может попасть разработчик понятийных ресурсов в самых различных областях и для различных компьютерных приложений. Поэтому важно описать наиболее частые случаи проблемного установления родовидовых отношений, а также возможные способы выявления таких неточностей в момент описания. Кроме того, при использовании транзитивности родовидовых отношений локальная неточность может перерасти в серьезное искажение в процессе многошагового логического вывода.

Данная статья посвящена рассмотрению концепции «роли» в онтологических исследованиях, представлению ролей в онтологиях и лингвистических ресурсах, а также будет рассмотрен опыт описания ролей, и отношений «тип-роль» в тезаурусе русского языка RuTez, предназначенном для автоматической обработки текстов в рамках информационно-поисковых задач [1].

## 2 Концепция роли в онтологиях

Онтологи обычно различают сущности (то, что есть) и события (то, что случается). Роли в этом разделении занимают «промежуточную» позицию: роли – это то, что есть, но только в контексте того, что случается. В течение многих лет понятие роли активно обсуждается в таких областях как концептуальное моделирование и представление знаний. Наиболее часто роль рассматривается посредством двух дополнительных понятий: игрок и контекст. Например, для роли *студент* игроком является человек, а контекст определяется отношением к высшему учебному заведению [4].

При рассмотрении ролей существенным фактором является то, что роль сообщает сущности некоторые внешние характеристики.

Внутренние характеристики, такие как размер или форма, описывают сущность в изоляции. В противоположность, внешние характеристики описывают сущность относительно других сущностей и событий. Например, свойство «обычно использоваться для забивания гвоздей» является внешней характеристикой молотка, поскольку устанавливает отношение между молотком и гвоздями.

Внешние и внутренние характеристики сущности обладают разными свойствами.

Например, такие внутренние характеристики как возраст могут изменяться в течение времени, но они всегда приложимы к сущности. Напротив, внешние характеристики, такие как зарплата работника, могут оказаться совершенно неприменимыми в некоторых ситуациях, например, когда работник участвует в спортивном соревновании. Более того, в отличие от внутренних характеристик, внешние характеристики могут быть противоречивы, например, как величина зарплаты человека на разных работах.

Steimann [17] выделяет 15 характеристик ролей, из которых наиболее существенными являются следующие пять:

### 2.1 Роли создаются и исчезают динамически.

Из-за того, что роль представляет внешние характеристики сущности, связанные с его участием в некотором событии, роль создается, когда начинается событие. Если сущность прекращает участие в событии, то роль может прекратить существовать и ее свойства более не будут действительными.

### 2.2 Роль может передаваться между сущностями

Например, роль менеджера может переходить от одного человека к другому. Заметим, что многие из характеристик ролей могут передаваться без изменения, тогда как другие могут «перевычисляться» в зависимости от новой сущности, играющей роль. Например, работник получает премию 20 процентов за то, что он менеджер, а заработная плата может пересчитываться в зависимости от конкретного лица.

### 2.3 Сущность может играть разные роли одновременно

Например, одно и то же лицо может быть работодателем и наемным работником.

### 2.4 Сущности разных, не связанных между собой типов, могут выступать в одинаковых ролях.

Например, крекер (для человека) и муха (для лягушки) могут выступать в роли пищи.

### 2.5 Роли могут играть роли

Это отражает ситуацию, что работник может быть, например, руководителем проекта, что, в свою очередь, является ролью работника.

## 3 Критерии распознавания ролей

Сова [15] определяет роль следующим образом: «Подтипы сущности могут быть двух видов: натуральные типы и ролевые типы, которые являются подтипами натуральных типов в конкретных образцах отношений (particular pattern of relationships). Человек, например, является натуральным типом, а учитель – это подтип

человека в ситуации обучения». Сова предлагает простой тест для определения, является ли понятие ролью: *r* – является ролевым типом, если сущность может быть охарактеризована как *r* только при рассмотрении другой сущности, действия или состояния.

В соответствии со взглядом Совы роли ассоциируются с отношениями, но при этом они понятия, а не отношения.

В работе [8] Никола Гуарино отмечает, что тест Совы для различения типов и ролей недостаточен: например, нечто может быть охарактеризовано как автомобиль, только если оно имеет, по крайней мере, колеса и мотор, но автомобиль является типом, а не ролью.

Гуарино заменяет условие, сформулированное Совой, на условие так называемой внешней онтологической зависимости:

*Понятие C1 называется внешне зависимым от понятия C2, если для всех примеров C1 должен существовать пример C2, который не является частью или материалом примера C1 [10].*

Masolo et.al [11] замечают, что точнее условие сформулировать так:

*пример C2 не должен быть внутренним для примера C1, то есть не должен быть частью, или материалом, или качеством (цвет).*

Таким образом, данное условие базируется на родовой зависимости сущности от внешних свойств.

Например, *сын* является внешне зависимым, поскольку существует только в рамках семьи по отношению к своим родителям. С другой стороны, автомобиль не является внешне зависимым, поскольку требует существования мотора, который является частью автомобиля.

Таким образом, данное условие формализует определение ролей, данной Совой.

Вместе с тем, на основе такого определения в класс ролей попадают еще дополнительные сущности такие, как качества: цвет, вес, скорость. Например, если синий – это цвет, то обязательно существует хотя бы один объект, цвет которого синий, при чем этот объект не является частью цвета. Понятие цвета поэтому является зависимым, но не кажется подходящим к понятию роли.

Поэтому Гуарино вводит еще одно условие, которое вместе с условием внешней зависимости дает лучшее определение понятию «роль».

*Понятие C является семантически жестким (rigid), если любой пример понятия C остается примером C в течение всего своего существования.*

Например, щенок перестает быть щенком, все еще оставаясь собакой, поэтому собака и животное

– это жесткие сущности, а щенок не является жестким понятием.

*Таким образом, понятие C называется ролью, если оно является внешне зависимым и не является семантически жестким [10, 11].*

В соответствии с этим определением качества не могут быть ролями, поскольку они являются семантически жесткими: если цвет прекратит быть цветом, то он станет чем-то еще, потеряет свою идентичность.

Таким образом, оба условия вносят вклад в определение роли. Первое условие описывает, что сущность-роль должна рассматриваться в рамках чего-либо объемлющего, что в определении Совы называлось *pattern of relationships*, второе условие помогает формализовать условие конкретности, особенности, упоминаемое в определении Совы [15, 16].

Введенные понятия онтологической зависимости и семантической жесткости помогают формализовать понятие натурального типа.

*Понятие C называется натуральным типом, если оно существенно независимо и семантически жестко.*

Таким образом, собака – это пример натурального типа. Цвет не является ни ролью, ни натуральным типом: цвет – семантически жесткий, но является онтологически зависимым. Щенок или хромая собака также не являются ни натуральным типом, ни ролью, поскольку они являются независимыми от внешних сущностей и являются семантически жесткими.

По типу контекста, определяющего роль, можно выделить три вида ролей: реляционные роли (способы участия в отношении), процессуальные роли (способы участия в событии), и социальные роли (способы участия в некотором сообществе) [13].

Роли не ограничиваются временем участия в конкретном событии: музыкант остается музыкантом, даже когда спит. Такая особенность характерна для так называемых социальных ролей.

В отличие от реляционных и процессуальных ролей социальные роли онтологически зависят от социальных институтов: организаций, нормативных систем, предприятий.

В работе [13] реляционные роли рассматриваются как части произвольного отношения *r*, а процессуальные роли как части процесса *p*. Утверждается, что и для социальной роли всегда можно найти социальное сообщество, частью которого является эта социальная роль. Такими социальным сообществом может быть юридически оформленная организация (профессор в ВУЗе), группа лиц (предводитель банды), социальная система (президент государства).

#### 4 Представление ролей в компьютерных ресурсах

Обычно рассматривается три основных подхода к представлению ролей [4]. Первый подход (1) рассматривает роль только как метку, приписанную к участнику ситуации. Например, роль нанимателя отмечает роль агента в событии найма на работу. Такой подход сходен с описанием семантических ролей в семантической модели управления глагола.

Этот подход прост, но он не представляет роли отлично от сущностей (комбинируя внешние и внутренние характеристики понятий в единое представление сущности). Как следствие, становится невозможным описать собственные свойства ролей. Например, пятое свойство ролей из раздела 2 говорит о том, что роли могут играть роли, то есть у ролей есть своя собственная классификация, которую при данном способе представления невозможно отразить.

Второй подход представляет роли и сущности отдельными понятийными единицами, однако комбинирует эти два типа понятий в рамках одной иерархии.

В одной иерархии роли и типы могут быть объединены двумя способами.

Во-первых (2.1), роли могут описываться как вышестоящие понятия для типов, которые могут их занимать.

Например, поскольку работодателем может быть человек или организация, то понятие *РАБОТОДАТЕЛЬ* (Рис.1) представляется как вышестоящее, родовое понятие, а понятия *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ* представляются как нижестоящие, видовые понятия. Против такого представления выступают многие онтологи [9, 16, 17]. Действительно, такое представление неточно описывает свойства сущностей, поскольку не каждый человек является работодателем.

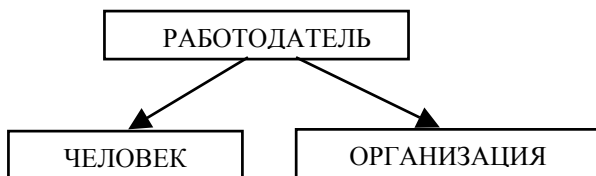


Рис.1

Во-вторых (2.2), роли можно описывать как нижестоящие понятия для сущностей, которые могут их занимать.

Тогда понятие *РАБОТОДАТЕЛЬ* может быть представлено, например, как подтип понятия *ОРГАНИЗАЦИЯ* (Рис.2). Однако если нужно отразить знание, что работодателем может быть и человек, то ситуация несколько усложняется. Если теперь представить понятие *РАБОТОДАТЕЛЬ* как подтип понятий *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ*, то получится, что работодатель одновременно и человек, и организация.

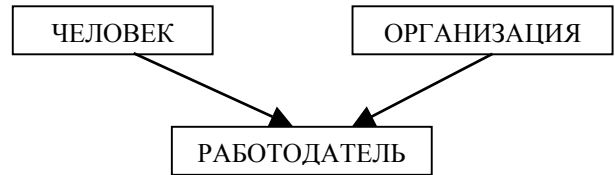


Рис.2

Чтобы описать, что работодатель может быть человеком или организацией, может быть введено дополнительное понятие, например, с названием *СУБЪЕКТ ДЕЯТЕЛЬНОСТИ* (Рис.3), подтипами которого являются понятия *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ*.

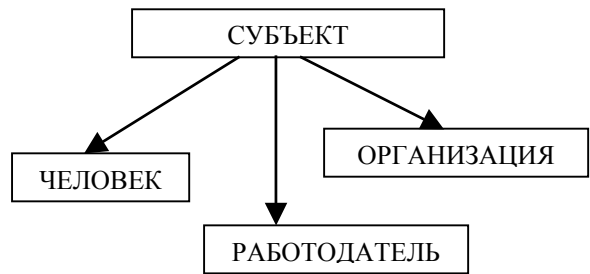


Рис.3

Далее устанавливается отношение между понятием *РАБОТОДАТЕЛЬ* и понятием *СУБЪЕКТ ДЕЯТЕЛЬНОСТИ*. Таким образом, понятие *РАБОТОДАТЕЛЬ* оказывается на одном уровне иерархии с понятиями *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ*, что, с одной стороны, не описывает первоначального утверждения, что работодатель является либо человеком, либо организацией, а с другой стороны все-таки может использоваться как модель для представления ролевых понятий.

В работе [6] авторы для уточнения возможностей совмещения представления ролей и типов в одних и тех же иерархиях разделяют роли на два подтипа: *материальные роли* и *формальные роли*.

Как указывают авторы, формальные роли не несут идентичности, то есть не относятся ни к какому конкретному типу, могут быть применены к любому типу. В качестве формальных ролей можно привести пример таких ролей как *часть* и *целое*, *инструмент* и т.п.

В качестве материальных ролей рассматриваются такие роли как *студент* (должен относиться к типу человек) или *еда* (является физической сущностью).

Авторы считают, что формальные роли должны представляться только в таксономиях ролей, материальные роли могут быть представлены как видовые понятия для классов и ролей, а сами могут

подчинять как видовые понятия только материальные роли.

Наконец, в третьем подходе (3) предполагается, что иерархия ролей должна быть представлена отдельно от иерархии типов. Иерархия ролей подчиняется самому верхнему узлу иерархии. В таком представлении понятия ролей описываются независимо от типов, но каждый пример роли существует только как пример типа, то есть пример роли не может существовать независимо от примера типа, которые может занимать эту роль.

## 5 Роли в WordNet

В системах автоматической обработки естественно-языковых текстов хорошо известен компьютерный тезаурус WordNet, который представляет лексическую систему английского языка в виде иерархической сети синсетов – совокупностей синонимов [12]. Основным отношением в этой сети является родовидовое отношение, называемое отношением гипонимии и гиперонимии.

Указывая на смешение типов и ролей в Wordnet, Никола Гуарино [9] привел следующие примеры описания из WordNet:

*Человек – это живое существо и каузальный агент.*

*Яблоко – это фрукт и еда.*

Человек всегда живое существо, но он (она) начинает играть роль каузального агента только в некоторых ситуациях. Та же проблема возникает для яблока, которое всегда плод растения, и в некоторых ситуациях может быть пищей. Проблема в том, что человек и яблоко – это типы сущностей, в то время как каузальный агент и пища – это роли.

Одна из авторов WordNet Кристиан Фелбаум отвечая на эту критику Николы Гуарино, заявляет [5], что в таких ресурсах как WordNet неоднородные классификации имеют право на существование, поскольку такие ресурсы рассматриваются в настоящее время, прежде всего, как инструменты для компьютерной обработки текстов, а не только как совершенные онтологии, которые должны соответствовать строгим онтологическим принципам.

Во-первых, если уничтожить «неправильные» отношения, то теряется важная информация. В некоторых случаях семантическая информация об отношениях между словами, не отвечающая строгим принципам, может быть более полезна, чем более обоснованное семантическое отношение.

Во-вторых, считает Кристиан Фелбаум, для лексических ресурсов, которые используются для компьютерной обработки текстов, полезно иметь подробную сеть отношений. Это важно для разрешения многозначности, разрешения референции, методов выявления лексической связности текстов.

## 6 Роли в информационно-поисковых тезаурусах

При разработке информационно-поисковых тезаурусов между единицами тезаурусов (дескрипторами) обычно устанавливаются два типа связи: иерархические и ассоциативные. Иерархические отношения обычно рассматриваются как несимметричные и транзитивные.

Многие руководства и стандарты [18, 19] подчеркивают, что иерархические отношения в информационно-поисковых тезаурусах должны устанавливаться в тех случаях, когда отношения истинны независимо от контекста, - только в таких случаях дескрипторы информационно-поискового тезауруса могут быть организованы в иерархии. Эта рекомендация связана с тем, что обычно в информационном поиске очень трудно четко определить контекст употребления термина и понять, применимо ли в данном контексте то или иное отношение.

Так, в [19] указывается, что для мышей можно указать, что они грызуны, поскольку это внутренняя характеристика мышей. В то же время неправильно указывать, что мыши – вредители, поскольку имеются лабораторные мыши и домашние мыши, которые не являются вредителями.

Американский стандарт на информационно-поисковые тезаурусы [20] предлагает при описании родо-видовых отношений использовать тест «все-некоторые». Например, *все мыши являются грызунами, но некоторые мыши являются вредителями*. «Вредитель» – это роль и, таким образом, не рекомендуется устанавливать родовое отношение ВЫШЕ от типа к роли некоторой сущности.

## 7 Описание ролей в Тезаурусе русского языка РуТез

### 7.1 Особенности тезауруса РуТез как ресурса для автоматической обработки текстов

Тезаурус русского языка РуТез – представляет собой иерархическую сеть 50 тысяч понятий, ассоциированных с более чем 125 тысячами слов и выражений [1].

С самого начала тезаурус РуТез создавался как ресурс для автоматической обработки больших текстовых коллекций. С 1996 года Тезаурус работает в разнообразных приложениях автоматической обработки больших текстовых коллекций таких как автоматическое концептуальное индексирование, визуализация и уточнение результатов поиска, автоматическая рубрикация, автоматическое аннотирование. Тезаурус тестируется и пополняется на основе анализа результатов обработки в конкретных приложениях.

В отличие от традиционных информационно-поисковых тезаурусов, которые предназначены для ручного индексирования экспертами-индексаторами и разработку которых регулируют международные и национальные стандарты, тезаурус РуТез предназначен работы в автоматических режимах обработки текстов, что приводит к необходимости описания в тезаурусе:

- значительно большего объема информации: большего количества текстовых входов понятий (языковых выражений, с помощью которых понятие может быть выражено в тексте), конкретных понятий (например, конкретных видов рыб, фруктов, произведений искусства и т.п.), и, как следствие, большему количеству отношений между понятиями;
- представлению разных значений многозначных языковых единиц;
- использованию правил логического вывода, основанных на транзитивности отношений и наследовании свойств [2].

## 7.2 Отношения в тезаурусе РуТез

Функции отношений в ресурсе, предназначенном для использования в автоматических режимах обработки текстов, весьма разнообразны. Отношения могут использоваться при расширении запроса пользователя, при выводе рубрики для текста, для разрешения многозначности терминов, для установления близких по смыслу терминов в тексте и т.п.

Для обеспечения такой функциональности отношения тезауруса должны обладать рядом свойств.

Во-первых, набор отношений должен быть небольшим и быть применимым в любых предметных областях.

Во-вторых, отношения должны иметь набор простых свойств (таких как транзитивность и наследование), регулирующих использование этих отношений в логическом выводе.

В-третьих, отношения между понятиями должны быть «надежными» [2], то есть их применение не должно зависеть от того контекста, в котором упомянуты эти понятия. Это условие установления отношений в тезаурусе РуТез полностью согласуется с общепринятыми стандартами на разработку информационно-поисковых тезаурусов.

Именно обеспечение надежности отношений как основы надежного логического вывода делает важным аккуратное описание в тезаурусе ролевых понятий, поскольку в силу перечисленных в разделах 2 и 3 динамических свойств ролей некоторые отношения с ними могут быть ошибочными или неприменимыми в отдельных контекстах, из-за чего может пострадать логический вывод. Поскольку цепочки логического вывода могут быть достаточно длинными, одно «вкравшееся» неточное отношение может привести к серьезным проблемам.

В Тезаурусе имеется четыре основных типа отношения.

Первый тип отношений – таксономическое отношение НИЖЕ-ВЫШЕ, обладает свойством транзитивности и наследования.

Два следующих типа отношений определяются на основе используемого Н. Гуарино отношения онтологической зависимости [6, 10].

Второй тип отношений – отношение ЧАСТЬ-ЦЕЛОЕ. Используется не только для описания физических частей, но и для других внутренних сущностей понятия таких как свойства или роли для ситуаций. Важным условием при установлении этого отношения является то, что понятия-части должны быть жестко связаны со своим целым, то есть каждый пример понятия-части должен все свое существование являться частью для понятия-целого, а не относиться к чему-либо другому. В терминах Н. Гуарино понятия-части должны онтологически зависеть от понятия-целого.

При таком условии удастся ввести свойство транзитивности отношения ЧАСТЬ-ЦЕЛОЕ, что очень важно для автоматического вывода в процессе автоматической обработки текстов.

Еще один тип отношения, называемого несимметричной ассоциацией АСЦ2-АСЦ1, связывает два понятия, которые не могут быть связаны выше рассмотренными отношениями, и когда одно из понятий не существует без существования другого. Например, понятие САММИТ требует существования понятия ГЛАВА ГОСУДАРСТВА.

Таким образом, отношение несимметричной ассоциации связывает два понятия, если одно из них онтологически зависит от другого и при этом не является его частью.

Последний тип отношений – симметричная ассоциация связывает, например, понятия очень близкие по смыслу, но которые не были соединены в одно понятие.

## 7.3 Принципы описания ролевых понятий в Тезаурусе русского языка РуТез

При разработке ресурса для автоматической обработки текстов приходится строить онтологические описания на основе знаний, полученных из текстов.

А тексты часто указывают на связь между ролью и типом.

Например, следующий фрагмент (<http://www.giord.ru/0705211117391.php>):

*наиболее используемыми консервантами являются: поваренная соль, этиловый спирт, уксусная, сернистая, сорбиновая, бензойная кислоты и некоторые их соли*

может показаться хорошим источником информации для того, чтобы описать виды консервантов.

Определение электролита:

*Электролит - проводник второго рода; вещество, обладающее ионной проводимостью. Электролитами являются: расплавы солей, оксидов или гидроксидов; растворы солей, кислот или оснований в полярных растворителях; а также + твердые электролиты.*

может показаться основанием, например, для установления отношения, что соль является видом электролита.

Однако в таких случаях нужно помнить, что консервант и электролит являются ролями веществ - вещество становится консервантом или электролитом только, если попадает в некоторые условия. А поваренная соль и соль как химическое соединение являются типами веществ.

Устанавливая родовидовую связь от типа ко роли, мы сообщаем системе некорректное знание, состоящее в том, что любое вещество, относящееся к классу солей, в любой момент времени своего существования в любой ситуации, является электролитом, что далеко не так.

Возникает вопрос, можно ли отразить полученную из вышеприведенных фрагментов информацию, выразив ее набором «надежных» отношений (см.п.7.2). В тезаурусе РуТез мы обычно пытаемся применить несколько способов.

Во первых, если мы предполагаем, что в нашей предметной области большинство примеров того или иного типа будут использованы в некоторой роли, то все-таки устанавливается родовидовое отношение от типа как вида к роли как роду, которое снабжается пометкой В – что означает «возможно по умолчанию».

Так, например, мы можем установить такое отношение между понятием *СОРБИНОВАЯ КИСЛОТА* и *КОНСЕРВАНТ*, если посчитаем, что это основное применение сорбиновой кислоты в нашей предметной области и вероятность встретить в текстах обсуждение сорбиновой кислоты в других применениях (например, в органическом синтезе) в нашей области не слишком велико:

*СОРБИНОВАЯ КИСЛОТА*  
*ВЫШЕ\_В КОНСЕРВАНТ*

Однако не рекомендуется устанавливать такое отношение между понятиями *ПОВАРЕННАЯ СОЛЬ* и *КОНСЕРВАНТ*, поскольку основное применение поваренной соли совсем другое. Даже если бы мы установили такое отношение (ввели бы еще пометку для неосновных ролей), то нужно учитывать, что для автоматической системы обработки текстов невозможно качественно учитывать контекст употребления поваренной соли в тексте, чтобы разобраться можно использовать это отношение или нет.

Таким образом, в некоторых случаях мы все-таки размещаем понятия-роли выше по иерархии, чем понятия-типы, однако отмечаем такое отношение специальной пометкой. Мы применяем это отношение только для описания знания о предметной области, которое верно по умолчанию, то есть, с одной стороны, оно может пригодиться при обработке текстов, с другой стороны, относительно редко может привести к ошибке вывода. Для каждого типа может быть описано максимум одно такое отношение, а описания многих понятий-типов не включают такие отношения, поскольку могут выступать в самых разных ролях.

Именно с использованием отношения *ВЫШЕ\_В* может быть отражено критикуемое Николой Гуарино отношение *яблоко – пища*, если будет известно, что в рабочей предметной области использование в пищу – это основная роль яблок.

В перечисленных в разделе 4 способах представления иерархических отношений между типами и ролей такое решение соответствует способу 2.1, однако корректируется дополнительной пометкой

На примере описания понятия *ЭЛЕКТРОЛИТ* может быть продемонстрирована еще одна возможность описания отношений между ролями и типами в тезаурусе РуТез.

Мы можем попытаться ввести дополнительное понятие для ситуации соли в роли электролита. Если это важно для данной сферы, то это наше желание обычно поддерживается и языком предметной области – для такого понятия существует одно или более употребительных языковых выражений. И в нашем случае существует и активно употребляется такое словосочетание как *солевой электролит*.

Таким образом, мы можем ввести понятие *СОЛЕВОЙ ЭЛЕКТРОЛИТ* и установить следующие отношения:

*СОЛЕВОЙ ЭЛЕКТРОЛИТ*  
*ВЫШЕ СОЛИ*  
*ВЫШЕ ЭЛЕКТРОЛИТЫ*

Тем самым мы корректно отражаем знание, полученное нами из прочитанного определения. В перечисленных в разделе 4 способах представления иерархических отношений между типами и ролей такое решение соответствует способу 3, однако иерархии типов и ролей пересекаются не только на примерах понятий, которые относятся к обеим иерархиям, но и в специально введенных понятиях.

Если рассмотреть такое решение для отражения отношений между понятиями *РАБОТОДАТЕЛЬ*, *ЧЕЛОВЕК И ОРГАНИЗАЦИЯ*, то нужно ввести два дополнительных понятия, например, *РАБОТОДАТЕЛЬ-ФИЗИЧЕСКОЕ ЛИЦО* и *РАБОТОДАТЕЛЬ-ЮРИДИЧЕСКОЕ ЛИЦО*.

Тогда можно сделать следующие описания (Рис.4):

*РАБОТОДАТЕЛЬ-ФИЗИЧЕСКОЕ ЛИЦО*  
ВЫШЕ *РАБОТОДАТЕЛЬ*  
ВЫШЕ *ЧЕЛОВЕК*

*РАБОТОДАТЕЛЬ-ЮРИДИЧЕСКОЕ ЛИЦО*  
ВЫШЕ *РАБОТОДАТЕЛЬ*  
ВЫШЕ *ОРГАНИЗАЦИЯ*

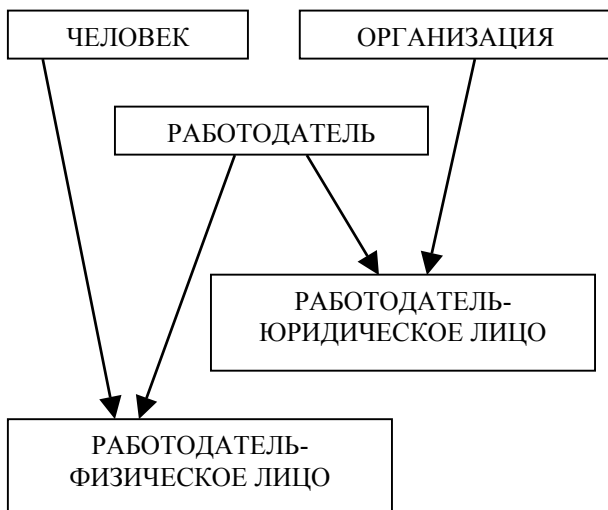


Рис.4

Поскольку введение дополнительных понятий может серьезно усложнять описание понятий в ресурсе, такой способ используется лишь в тех случаях, когда такие дополнительные понятия действительно используются в предметной области, как в случае понятия *СОЛЕВОЙ ЭЛЕКТРОЛИТ*. Интересно отметить, что введенные дополнительные понятия *РАБОТОДАТЕЛЬ-ФИЗИЧЕСКОЕ ЛИЦО* и *РАБОТОДАТЕЛЬ-ЮРИДИЧЕСКОЕ ЛИЦО* также имеют реальное основание в правовой области, поскольку отношения разных типов работодателей с работниками по-разному регулируются законодательством.

Понятия-типы *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ* могут равным образом выступать во многих ролях, поэтому в тезауусе РуТез действительно введено такое понятие как *СУБЪЕКТ ДЕЯТЕЛЬНОСТИ*, как обсуждается в методе 2.1 раздела 4, и в качестве нижестоящих к нему понятий размещены многие ролевые понятия, в которых могут выступать и примеры понятия *ЧЕЛОВЕК*, и примеры понятия *ОРГАНИЗАЦИЯ*.

Таким образом, в реальном ресурсе, создаваемом для работы в приложениях в широкой предметной области, приходится применять несколько разных подходов к описанию иерархий типов и ролей, обсуждаемых в литературе как альтернативные.

## Заключение

Термин «онтология» как средство моделирования предметной области в информационных системах является в настоящее время очень популярным. При этом наблюдается большое число работ, посвященных созданию той или иной программной оболочки для ведения онтологии, возможностям выгрузки онтологии в разных форматах языков RDF, OWL и т.п.

В данной статье мы хотели подчеркнуть важность рассмотрения содержательных проблем разработки онтологий, когда трудности могут встретиться при описании самых широко известных отношений - родовидовых отношений

В статье была рассмотрена одна из наиболее распространенных проблем моделирования родовидовых отношений – проблема смешения типов и ролей в одних и тех же иерархиях.

Мы рассмотрели подходы к описанию отношений «роль-тип» в онтологических и лингвистических ресурсах. Было показано, что различие между онтологическими характеристиками понятий-типов и понятий-ролей обязательно нужно учитывать при разработке лингвистических ресурсов для автоматической обработки текстов, если предполагается использовать логический вывод по иерархическим отношениям.

При этом на примере тезаууса РуТез видно, что в практической деятельности по разработке лингвистического ресурса приходится использовать несколько способов описания иерархических отношений между типами и ролями в зависимости от ряда дополнительных факторов.

## Литература

- [1] Лукашевич Н.В., Добров Б.В., Тезауус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002. М.: Наука, 2002. Т.2. С.338-346.
- [2] Лукашевич Н.В., Добров Б.В., Отношения в онтологиях для решения задач информационного поиска в больших разнородных текстовых коллекциях // Девятая национальная конференция по искусственному интеллекту КИИ-2004. М.: Физматлит, 2004. Т.2. С.544-551.
- [3] Cruse D. 1986 Lexical Semantics. Cambridge. University Press. 1986.
- [4] Fan J., Barker K., Porter B, Clark P. Representing Roles and Purpose // Proceedings of 1st Int Conference on Knowledge Capture (K-Cap'01). 2001. p. 38-43.
- [5] Fellbaum Ch. Parallel Hierarchies in the Verb Lexicon // Proceedings of 'The Ontologies and Lexical Knowledge bases' workshop (OntoLex 2002). Las Palmas, Spain, 2002.



- [6] Gangemi A., Guarino N., Masolo C., Oltramari A., Understanding Top-Level Ontological Distinctions // Proceedings of IJCAI 2001 workshop on Ontologies and Information Sharing. , 2001.
- [7] Gomez-Perez A., Fernandez-Lopez M., Corcho O. OntoWeb. Technical Roadmap. D.1.1.2. - IST project IST-2000-29243, 2001. ([http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/OntoWeb\\_Del\\_1-1-2.pdf](http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/OntoWeb_Del_1-1-2.pdf))
- [8] Guarino N. Concepts, attributes and arbitrary relations // Data Knowledge Engineering.1992. 8. pp. 249-261.
- [9] Guarino N., Some Ontological Principles for Designing Upper Level Lexical Resources. // Proceedings of First International Conference on Language Resources and Evaluation. Granada, Spain, 1998.
- [10] Guarino N., Welty C. Ontological Analysis of Taxonomic Relationships // Proceedings of ER-2000. The international conference of Conceptual Modeling. Springer Verlag, 2000.
- [11] Masolo C., Vieu L., Bottazzi E. Catenacci C., Ferrario R., Gangemi A., Guarino N. Social roles and their descriptions // Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning. AAAI Press, 2004.
- [12] Miller G. Nouns in WordNet // WordNet – An Electronic Lexical Database / Ed. By Fellbaum, C. The MIT Press, 1998. p.23-47.
- [13] Loebe F. Abstract vs. Social Roles: A Refined Top-level Ontological Analysis // Proceedings of the 2005 AAAI Fall Symposium 'Roles, an Interdisciplinary Perspective: Ontologies, Languages, and Multiagent Systems / Guido Boella, James Odell, Leendert van der Torre and Harko Verhagen (ed.). AAAI Press, 2005. FS-05-08. p.93-100.
- [14] Nirenburg S., Wilks Y., What's in a symbol: Ontology, representation, and language // Journal of Experimental and Theoretical Artificial Intelligence, 2001, 13(1):9-23.
- [15] Sowa J.F. Using a Lexicon of Canonical Graphs in a semantic interpreter // Relational models of lexicon / M.Evens. Cambridge University press, 1988. p.113-137.
- [16] Sowa J. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA, 2000.
- [17] Steimann F. The representation of roles in object-oriented and conceptual modelling // Data and Knowledge engineering. 2000. 35, 1. p. 83-106.
- [18] Z39.19 – Guidelines for the Construction, Format and Management of Monolingual Thesauri. – NISO, 1993.
- [19] Will L. Thesaurus consultancy // The thesaurus: review, renaissance and revision / Sandra K. Roe and Alan R. Thomas, editors. New York ; London : Haworth, 2004.

## **Description of Role Concepts in Linguistic and Ontological Resources**

Natalia V. Loukachevitch

In the paper we consider ontological characteristics of such concepts as roles and show their distinctions from type concepts. We argue that the difference between types and roles is necessary to account in ontological and linguistic resources intended for automatic text processing if the automatic inference procedure is planned.

We show that information about concepts and entities obtained from text definitions often leads to incorrect descriptions of type-role relations, and it is necessary to make special efforts to describe this information appropriate for the logical inference.

We discuss possible means for description of the type-role relations used in the Thesaurus of Russian Language RuThes.

---

\* Данная работа частично поддержана грантом РФФИ № 05-07-90391.