

**ВЫСШАЯ ШКОЛА ЭКОНОМИКИ**  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

*И.В. Ефименко, В.Ф. Хорошевский*

**ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
ЭКОНОМИКИ ПРЕДПРИЯТИЙ  
И ОТРАСЛЕЙ СОВРЕМЕННОЙ РОССИИ**

**Часть 1. Онтологическое моделирование:  
подходы, модели, методы, средства, решения**

Препринт WP7/2011/08 (ч. 1)

Серия WP7

Математические методы анализа решений  
в экономике, бизнесе и политике

Москва  
2011

УДК 004.78:658

ББК 65.39

О59

Редакторы серии WP7

«Математические методы анализа решений в экономике,  
бизнесе и политике»

*Ф.Т. Алескеров, В.В. Подиновский, Б.Г. Миркин*

О 59 **Ефименко, И. В., Хорошевский, В. Ф.** Онтологическое моделирование экономики предприятий и отраслей современной России: Часть 1. Онтологическое моделирование: подходы, модели, методы, средства, решения : препринт WP7/2011/08 (ч. 1) [Текст] / И. В. Ефименко, В. Ф. Хорошевский ; Нац. исслед. ун-т «Высшая школа экономики». – М. : Изд. дом Высшей школы экономики, 2011. – 76 с. – 150 экз.

Представленный аналитический обзор вводит в проблематику онтологического моделирования и онтологического инжиниринга и показывает ретроспективу работ в этих областях. Исследования и разработки обсуждаются в контексте искусственного интеллекта, информационных технологий, семантических технологий и семантического веба. Рассматриваются модели, методы и средства построения онтологий «с нуля», реинжиниринг онтологий, коллективная разработка онтологий, а также вопросы объединения и выравнивания онтологий. Представлены базисные онтологические теории и модели, существующие классификации онтологий, рассмотрены известные онтологии верхнего уровня (CYC, DOLCE, SUMO и онтология Дж. Совы). Кратко обсуждаются новые теории создания онтологических моделей, в частности, теория активностей и ее связь с онтологическими моделями. Работа формирует базу для понимания материалов, представленных в последующих препринтах авторского коллектива.

УДК 004.78:658

ББК 65.39

Работа осуществлена при частичной финансовой поддержке Государственного контракта № 13.G25.31.0033 от 07 сентября 2010 г. между ЗАО «Авикомп Сервисез» и Министерством образования и науки Российской Федерации на реализацию комплексного проекта «Создание высокотехнологичного производства инновационных программно-аппаратных комплексов для эффективного управления предприятиями и отраслями экономики современной России».

*Ефименко И.В.* – НИУ ВШЭ (Москва), [iefimenko@hse.ru](mailto:iefimenko@hse.ru).

*Хорошевский В.Ф.* – Вычислительный центр им. А.А. Дородницына РАН, [khor@ccas.ru](mailto:khor@ccas.ru).

**Препринты Национального исследовательского университета**

«Высшая школа экономики» размещаются по адресу: <http://www.hse.ru/org/hse/wp>

© Ефименко И. В., 2011

© Хорошевский В. Ф., 2011

© Оформление. Издательский дом  
Высшей школы экономики, 2011

## Предисловие редактора серии

В результате конкурса, проведенного Министерством образования и науки РФ летом 2010 г. во исполнение Постановления Правительства РФ № 218 «О мерах государственной поддержки развития кооперации российских высших учебных заведений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства», НИУ ВШЭ стал победителем сразу по двум проектам в области информационных технологий (в консорциумах с двумя российскими IT-компаниями). Одним из проектов, отобранных комиссией Министерства, стал проект НИУ ВШЭ (на тот момент – Государственного университета) и ЗАО «Авикомп Сервисез» по теме «Создание высокотехнологичного производства инновационных программно-аппаратных комплексов для эффективного управления предприятиями и отраслями экономики современной России».

Проект комплексный, наукоемкий и выполняется на стыке трех областей: экономический анализ, семантические технологии и современные математические методы.

Задача комплексного проекта состоит в разработке типовых программно-аппаратных комплексов (ПАК) для анализа информационной среды предприятий. Это позволит выполнять многоаспектную оценку экономических и общественно-политических факторов, характеризующих текущее состояние, зрелость, конкурентоспособность, перспективы развития предприятия или определенного сегмента рынка.

Уникальность проекта заключается в объединении в создаваемых ПАК, во-первых, новых технологий сбора и обобщения разрозненных фактов из внешней информационной среды (семантических технологий, позволяющих извлекать значимые сведения из источников неструктурированной информации), во-вторых, инновационных экономических моделей оценки состояния предприятий и рынков и, в третьих, современных математических методов и алгоритмов.

В соответствии с целями и задачами постановления № 218 Правительства РФ предполагается, что у проекта будут и дополнительные результаты. Так, например, предполагается, что в процессе выполнения комплексного проекта улучшится взаимодействие различных подразделений НИУ ВШЭ при реализации сложных междисциплинарных исследований, в новые исследования и разработки более активно будут вовлекаться сотрудники университета, студенты и аспиранты, а в учеб-

ный процесс будут внедряться новые программы, ориентированные на подготовку профессионалов по перспективным специальностям.

Работы по комплексному проекту ведутся в рабочих группах, отвечающих за тот или иной «участок фронта».

Всего в работах первых двух этапов было задействовано около 150 сотрудников и около 300 студентов НИУ ВШЭ.

Учитывая важность онтологического моделирования для целей комплексного проекта, а также дефицит русскоязычной литературы по этому направлению, я, как редактор серии препринтов WP7, пригласил ведущих специалистов в этой области д.т.н., проф. В.Ф. Хорошевского и к.фил.н. И.В. Ефименко написать три препринта под общим названием «Онтологическое моделирование экономики предприятий и отраслей современной России».

В первом препринте «Онтологическое моделирование: подходы, модели, методы, средства, решения», подготовленном в соавторстве И.В. Ефименко и В.Ф. Хорошевским, представлены основные понятия этого все еще нового для России научного направления, обсуждаются исследования и разработки в области формирования пространств знаний, а также базисные онтологические теории и модели.

Готовы к выпуску еще два препринта – «Мировые исследования и разработки: аналитический обзор» (авторы М.А. Гурьянова, И.В. Ефименко и В.Ф. Хорошевский), где представлены канадская школа разработки онтологических моделей и проект TOVE, моделирование бизнеса в эдинбургском проекте Enterprise Project, а также другие исследования и разработки в области онтологических моделей предприятий и отраслей экономики, и «Российские исследования и разработки в области онтологического инжиниринга и бизнес-онтологий» (авторы И.В. Ефименко и В.Ф. Хорошевский), в котором собрана для заинтересованных читателей рассеянная по разным статьям в разных изданиях научно-техническая информация об отечественных работах в этой области.

Предлагаемые препринты будут полезны не только специалистам, работающим в области онтологического инжиниринга, но и студентам и аспирантам соответствующих специальностей, которые смогут познакомиться с уже полученными здесь научно-техническими результатами, а также с перспективными направлениями исследований и разработок.

*Ф.Т. Алескеров*

# Содержание

<b>1. Введение</b> .....	6
<b>2. Исследования и разработки в области формирования пространств знаний</b> .....	8
2.1. Предварительные замечания.....	8
2.2. От исследований в области ИИ к пространствам знаний.....	9
2.2.1. Инициатива (KA) <sup>2</sup> и инструментарий Ontobroker.....	9
2.2.2. Проект SHOE.....	15
2.3. Информационные технологии и пространства знаний.....	21
2.3.1. Freebase.....	21
2.3.2. Проекты, связанные с использованием микроформатов.....	24
2.4. Семантические технологии для пространств знаний и Semantic Web.....	27
2.4.1. Онтологический инжиниринг и Semantic Web.....	27
2.4.2. Построение онтологий «с нуля».....	32
2.4.3. Реинжиниринг онтологий.....	38
2.4.4. Коллективная разработка онтологий.....	39
2.4.5. Объединение и выравнивание онтологий.....	44
2.5. Предварительные выводы.....	51
<b>3. Базисные онтологические теории и модели</b> .....	53
3.1. Классификации онтологий.....	53
3.2. Онтологическая модель СУС.....	56
3.3. Онтология верхнего уровня DOLCE.....	57
3.4. Онтология верхнего уровня SUMO.....	59
3.5. Онтология Джона Совы.....	61
3.6. Теория активностей и онтологические модели.....	63
3.7. Предварительные выводы.....	64
<b>4. Заключение</b> .....	65
<b>Литература</b> .....	67

# 1. Введение

Онтологическое моделирование представляет в настоящее время весьма популярное и быстро развивающееся научно-техническое направление, а полученные здесь результаты активно используются во многих областях информатики.

Как показывает анализ литературы в области онтологического моделирования, научно-техническими предпосылками этого направления были исследования и разработки в таких областях, как искусственный интеллект (ИИ), информационные технологии (ИТ) и, в последнее время, Semantic Technologies и Semantic Web (ST и SW).

В рамках первой группы исследований и разработок особое место занимают инициатива (КА)<sup>2</sup> и проект SHOE, которые, по сути дела, и заложили научно-технические основы формирования пространств знаний и онтологического моделирования, а затем и онтологического инжиниринга.

Информационные технологии и, в частности, веб-технологии тоже внесли и продолжают вносить свой вклад в онтологическое моделирование и онтологический инжиниринг, формирование Web 2.0 и таким образом, в определенном смысле, готовят почву для Semantic Web.

Последняя из перечисленных выше групп работ направлена на решение проблем формирования пространств знаний и развития концепции Semantic Web, что, в свою очередь, предполагает активное использование результатов, полученных в области Semantic Technologies. В работах классиков этих направлений – Т. Бернерса-Ли (Т. Berners-Lee), В. Бенджаминса (V. Benjamins) и др. отмечается, что SW-эра, в отличие от эпохи Интернет, еще только приближается, и на этом пути существует значительное число научных, технических, технологических и чисто человеческих проблем, основными из которых являются доступность семантического контента, доступность онтологий и средств их разработки, а также эволюция онтологий, масштабируемость, мультязыковость, визуализация и стабильность.

Учитывая вышесказанное, а также дефицит русскоязычных работ по онтологическому моделированию, изложение материала в настоящей работе организовано следующим образом.

В следующем разделе обсуждаются исследования и разработки в области формирования пространств знаний в контексте трех направле-

ний – от исследований в ИИ, ИТ, ST и SW к пространствам знаний. При этом особое внимание уделяется:

- инициативе (КА)<sup>2</sup>, инструментарию Ontobroker и проекту SHOE;
- подходу компании Metaweb, реализованному в базе данных Freebase, которая позиционируется как «открытая и разделяемая база данных знаний о мире или значительная по объему и кооперативно создаваемая и редактируемая база взаимосвязанных данных»;
- реализации социальных сетей на базе микроформатов, основная цель которых в достижении компактности и использования возможностей XHTML для добавления семантики к страницам Интернет;
- семантическим технологиям для пространств знаний и Semantic Web.

С учетом общих целей и задач настоящей работы в последнем подразделе акцент сделан на исследованиях в области онтологического моделирования и онтологического инжиниринга и, в частности, на обсуждении моделей, методов и средств построения онтологий «с нуля», реинжиниринга онтологий, коллективной разработке онтологий, а также объединении и выравнивании онтологий.

По сути, этот раздел вводит читателя в проблематику онтологического моделирования и онтологического инжиниринга и показывает ретроспективу исследований и разработок в этих областях.

Последний раздел настоящей работы, «Базисные онтологические теории и модели», посвящен более глубокому обсуждению онтологических моделей. Здесь представлены существующие классификации онтологий и на этой базе рассмотрены наиболее известные из онтологий верхнего уровня (Upper Ontology), среди которых особое внимание уделено онтологиям CYC, DOLCE, SUMO и онтологии Дж. Совы. В этой же части обсуждаются новые теории создания онтологических моделей, в частности, теория активностей и ее связь с онтологическими моделями.

Таким образом, данный раздел формирует у читателя базу для понимания материалов, представленных в следующих препринтах, посвященных

- работам канадской школы создания онтологических моделей и проекту TOVE, моделированию бизнеса в эдинбургском проекте Enterprise Project, другим исследованиям и разработкам в области онтологических моделей предприятий и отраслей экономики;

- российским исследованиям и разработкам в области онтологического инжиниринга и бизнес-онтологий, где собрана рассеянная по разным статьям в разных изданиях научно-техническая информация об отечественных работах в этой области.

Авторы надеются, что предлагаемый вниманию читателя препринт, в совокупности с остальными препринтами, достойным образом представит результаты работ по комплексному проекту «Создание высокотехнологического производства инновационных программно-аппаратных комплексов для эффективного управления предприятиями и отраслями экономики современной России», который в настоящее время выполняется специалистами НИУ ВШЭ и ЗАО «Авикомп Сервисез», и привлекут к участию в нем новых талантливых исполнителей.

## **2. Исследования и разработки в области формирования пространств знаний**

### ***2.1. Предварительные замечания***

Анализ литературы по моделям, методам и проектам, связанным с формированием пространств знаний, показывает, что с позиций сегодняшнего дня можно указать следующие основные парадигмы таких исследований и разработок:

- от искусственного интеллекта (ИИ) к пространствам знаний;
- из области информационных технологий (ИТ) к пространствам знаний;
- из области семантических технологий (СТ) к пространствам знаний и Semantic Web (SW).

К первой группе исследований и разработок, безусловно, относятся инициатива (КА)<sup>2</sup> – Knowledge Acquisition Initiative of the Knowledge Acquisition Community [Benjamins, et al., 1999] и проект SHOE – Simple HTML Ontology Extensions [Hefflin, et al., 2000].

Среди множества разработок, «пришедших» к формированию пространств знаний из ИТ, на наш взгляд, наиболее значимыми являются социальные сети.

И, наконец, среди множества проектов третьей группы в данной работе нам представляется целесообразным хотя бы кратко обсудить про-



екты онтологического инжиниринга – направления, в рамках которого в настоящее время интегрируются наиболее наукоемкие и технологически зрелые решения в области формирования пространств знаний.

## ***2.2. От исследований в области ИИ к пространствам знаний***

Исследования и разработки в области Искусственного Интеллекта внесли серьезный вклад не только в формирование проблематики представления и обработки знаний, но и в саму идею пространств знаний, а в последнее десятилетие и в Semantic Web. При этом работы в рамках обсуждаемых ниже исследовательских проектов, по существу, сформировали достаточно интересный и практически значимый взгляд на проблематику пространств знаний и те подходы, модели и методы, которые развивались в дальнейшем с учетом новых интернет-технологий и достижений в области программного обеспечения.

### ***2.2.1. Инициатива (КА)<sup>2</sup> и инструментарий Ontobroker***

Аннотация знаний сообществом приобретения знаний – так расшифровывается аббревиатура (КА)<sup>2</sup>. Целью работ по этому международному проекту [Maedche, et al., 2001], выполнявшемуся в 1996–1999 гг., были в конечном счете интеллектуальный поиск в среде Интернет и автоматическое накопление новых знаний.

В рамках инициативы (КА)<sup>2</sup> были сформулированы три основных направления исследований:

- онтологический инжиниринг;
- аннотация веб-страниц;
- запросы к информации на Web-страницах и вывод ответов на базе знаний.

Онтологический инжиниринг – одно из основных направлений, в рамках которого предполагалось, что сообщество (КА)<sup>2</sup> создаст свою собственную и достаточно общую систему онтологий на основе использования средств Ontolingua [Ontolingua, 2010]. К моменту окончания проекта (1999 г.) были разработаны восемь онтологий, которые рассматривались как разделы общей онтологии – организации (organization ontology), проекта (project ontology), личности (person ontology), направления исследований (research-topic ontology), публикаций (publication ontology), событий (event ontology), исследовательских продуктов (research-product ontology) и исследовательских групп (research-group ontology).

При этом разработка примеров онтологий осуществлялась и управлялась участниками проекта – так называемыми провайдер-агентами (*provider agents*), а размещались эти онтологии на их веб-страницах. Такие страницы аннотировались с использованием нового типа HTML-тэгов (*ONTO*), информация в пределах которых обрабатывалась специальной компонентой, работающей на основе онтологий, – системой *Ontocrawler*. В рамках этой компоненты, в зависимости от «богатства» используемой онтологии, должна была выводиться новая информация, релевантная запросам, но не присутствующая явно на веб-страницах в сети Интернет. Сама система *Ontocrawler* разрабатывалась в рамках отдельного проекта инициативы (КА)<sup>2</sup> – проекта *Ontobroker* [Decker, et al., 1999], который, собственно, и интересен с точки зрения обсуждения средств представления и обработки онтологических знаний.

В *Ontobroker* было три основных подсистемы: интерфейс формулирования запросов (*query interface*), машина вывода ответов (*inference engine*) и собственно машина доступа к интернет-ресурсам – *Webcrawler*, используемый для накопления требуемых знаний из этой среды.

Для спецификации онтологий в рамках обсуждаемого проекта был разработан специальный язык представления знаний, подмножество которого использовалось и для формулировки запросов, а язык аннотирования – для «обогащения» веб-документов онтологической информацией.

Формализм *Ontobroker* ориентирован на фреймовое представление знаний, а базисом представления являются так называемые фреймовые логики (*Frame-Logic*) [Kifer, et al., 1995].

Базисными конструкциями в данном случае являются:

- Подклассы –  $C1 :: C2$  (класс  $C1$  является подклассом  $C2$ ).
- Экземпляры –  $O : C$  ( $O$  является экземпляром класса  $C$ ).
- Декларация атрибутов –  $C1[A=>>C2]$  (для экземпляра класса  $C1$  определен атрибут  $A$ , значением которого должен быть экземпляр  $C2$ ).
- Значения атрибутов –  $O[A->>V]$  (экземпляр  $O$  имеет атрибут  $A$  со значением  $V$ ).
- Системное отношение «Часть – целое» – *Part-of* ( $O1 < : O2$ ) специфицирует, что объект  $O1$  является частью объекта  $O2$ .
- Другие отношения – *Relations*, предикаты вида  $p(a1, \dots, a2)$ , которые могут использоваться, как и в обычных логических формализ-

мах, но с тем расширением, что в качестве аргументов здесь могут выступать не только термы, но и выражения.

Из базисных конструкций строятся более сложные – факты (facts), правила (rules), «двойные» правила (double rules) и запросы (queries), причем факты, по сути дела, являются элементарными выражениями.

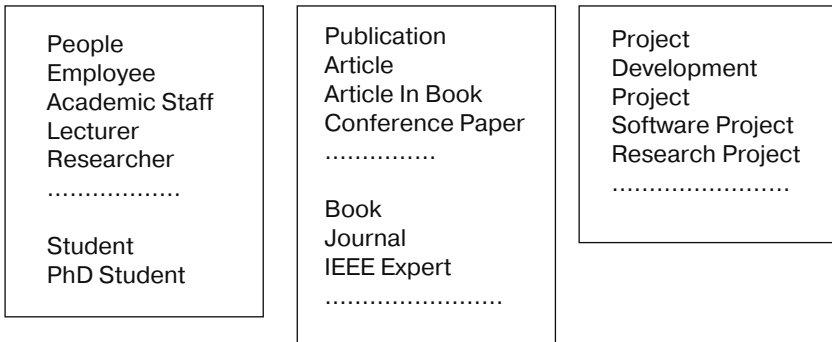
Правила, как и обычно, имеют левую и правую части, причем левая часть (здесь она называется «головой») является конъюнкцией элементарных выражений, а правая («тело») – сложная формула, термы которой – суть элементарные выражения, связанные обычными предикатными символами типа implies:  $\rightarrow$ , implied by:  $\leftarrow$ , equivalent:  $\leftrightarrow$ , AND, OR и NOT. Отличие между обычными и «двойными» правилами в симметричности последних. Важное достоинство формализма – возможность использования переменных в «голове» правил (с квантором FORALL) или в его «теле» (с кванторами FORALL и EXISTS). Фрагмент онтологии в формализме Ontobroker, адаптированный из работы [Fensel et al. 1999], приведен ниже, а фрагменты иерархии ее концептов и диаграммы бинарных отношений, выделяемых в рамках инициативы (KA)<sup>2</sup>, – на Рис. 1 и 2.

#### Определения объектов и отношений с их атрибутами:

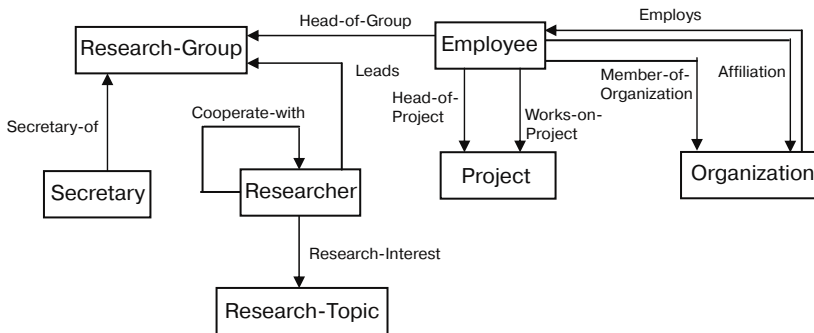
```
Person [firstName =>> STRING;  
lastName =>> STRING;  
eMail =>> STRING; ...  
publication =>> Publication].  
Employee [affiliation =>> Organization; ...].  
Researcher [researchInterest =>> ResearchTopic;  
memberOf =>> ResearchGroup;  
cooperatesWith =>> Researcher].  
Publication [ author =>> Person;  
          title =>> STRING;  
          year =>> NUMBER;  
          abstract =>> STRING].
```

#### Правила:

```
FORALL Person1, Person2  
    Person1:Researcher [cooperatesWith ->> Person2] <-  
Person2:Researcher [cooperatesWith ->> Person1].  
FORALL Person1, Publication1  
    Publication1:Publication [author ->> Person1] <->  
    Person1:Person [publication ->> Publication1].
```



**Рис. 1.** Фрагмент таксономии объектов, выделяемых в рамках инициативы (KA)<sup>2</sup>



**Рис. 2.** Фрагмент диаграммы бинарных отношений, выделяемых в рамках инициативы (KA)<sup>2</sup>

Первое из приведенных выше правил фиксирует симметричность отношения cooperatesWith, а второе утверждает, что если конкретная личность (экземпляр класса Person) имеет публикацию, то последняя имеет автора, который тоже является экземпляром класса Person, и обратно.

Машина вывода Ontobroker состоит из двух основных компонент: транслятора с расширенного языка представления в ограниченный и собственно вычислителя выражений ограниченного языка, который является обычным языком логического программирования.

Формализм запросов ориентирован на фреймовое представление онтологий, в рамках которого, как и обычно, определены понятия примеров, классов, атрибутов и значений.

Для примера, запрос вида

**FORALL R** <- R:Researcher

предполагает поиск всех объектов, являющихся экземплярами класса Researcher. Если предположить, что идентификатором объекта служит URL домашней страницы специалиста, в качестве результата по этому запросу будет выдан список соответствующих интернет-ссылок. Если необходимо найти всех специалистов по фамилии Иванов и при этом выдать в качестве результата их имена и электронные адреса, то приведенный выше запрос модифицируется следующим образом:

**FORALL Obj, FN, EM** <- Obj: Researcher[firstName->>FN; lastName->>"Иванов"; email->>EM].

В качестве ответа при этом могут быть получены значения переменных:

Obj = <http://www.anywhere.ru/~ivanov/>

FN = Иван

EM = <mailto:ivanov@anywhere.ru>

Имеются в языке Ontobroker и средства вывода значений свойств. Так, некоторые из атрибутов объекта могут задавать отношения, свойства которых известны машине вывода. Для примера, в запросе вида **FORALL Obj, CP** <- Obj: Researcher[lastName->>"Иванов"; cooperatesWith->>CP]

атрибут cooperatesWith является отношением, обладающим свойством симметричности. Это означает, что даже если у объекта, описывающего специалиста по фамилии Иванов, свойства cooperatesWith нет, Ontobroker выведет его, если в онтологии представлен объект, описывающий другого специалиста, который имеет такое свойство со значением Иванов.

В данном языке представления знаний присутствуют и другие правила вывода значений атрибутов, эксплицитно не представленных в Web-документах. При этом язык запросов Ontobroker может использоваться и для формирования репозиториев с информацией, удовлетворяющей заданным ограничениям. И более того, с помощью запросов можно получать и метаинформацию: запрос вида

**FORALL Att, T** <- Researcher[Att=>>T]

вернет в качестве результата имена всех атрибутов класса *Researcher* и связанных с ним классов.

В самой системе *Ontobroker* поддерживаются два типа интерфейсов при формировании запросов – текстовый (для экспертов) и графический (для пользователей). Первый из них предполагает, что запросы формулируются непосредственно во входном языке описания онтологий. Понятно, что при этом знание синтаксиса языка *Ontobroker* и знакомство с онтологией, для которой запрос формулируется, должны у эксперта присутствовать.

Проблема знания синтаксиса решается в данном случае, как, впрочем, и в большинстве других инструментальных средств, за счет диалогов, управляемых системой (*system-driven dialogue*). Пользователю выдается соответствующая панель, где могут быть определены (путем выбора из меню) компоненты запроса и связки между ними. Такой подход обеспечивает синтаксическую корректность и однозначность интерпретации запроса.

Сложнее преодолеть разрыв в знаниях эксперта и пользователя, особенно новичка, об используемой онтологии, поскольку для правильной формулировки запроса необходимо знать, по крайней мере, какие концепты в онтологии присутствуют и какие атрибуты имеются у концептов. Поэтому все системы представления онтологических знаний предоставляют своим пользователям средства визуализации онтологий и навигации по онтологии. В *Ontobroker* для визуализации онтологий используется подход, основанный на идеях гиперболической геометрии (*Hyperbolic Geometry*) [Lamping, et al., 1995], которые в данном случае реализуются следующим образом: текущий класс, интересующий пользователя, представляется «большим шаром», а классы, с ним непосредственно связанные, представлены «маленькими шарами» и располагаются по границе круга, «очерчивающего» соответствующий слой. Используя данный интерфейс, и эксперт, и пользователь могут легко и эффективно включать в свой запрос нужные концепты и их атрибуты, а система *Ontobroker* переводит их в нужное и синтаксически правильное представление автоматически.

Реализован интерфейс онтологий *Ontobroker* как Java-апплет, что обеспечивает работу с Web-браузерами на любых платформах, где поддерживается Java-технология.

Поскольку в те годы, когда выполнялся проект инициативы (КА)<sup>2</sup>, веб-информация чаще всего представлялась на языке HTML, в рамках

проекта Ontobroker было разработано простое его расширение для аннотации веб-страниц. Основная идея этого расширения состояла в следующем: в язык HTML были добавлены несколько релевантных для решения поставленных задач тэгов, использование которых позволяло Ontobroker интерпретировать аннотированные фрагменты HTML-текста как факты языка представления онтологических знаний. При этом веб-страницы оставались приемлемыми для стандартных браузеров типа MS Explorer.

В язык были введены три эпистемологически различных примитива:

- идентификация объекта, который может быть определен как экземпляр определенного класса, с помощью URL;
- установка значения атрибута объекта;
- определение отношений между объектами.

Все примитивы синтаксически расширяли тэг `<a ...>` языка HTML. Так, например, если специалист Иванов захочет определить себя как объект обсуждавшейся выше онтологии, он может на своей домашней странице ввести конструкцию вида:

```
<a onto="" `http://www.anywhere.ru/~ivanov/` : Researcher`` > </a>
```

что позволяет для объекта Иванов класса Researcher ввести атрибут email и его значение с помощью следующей конструкции:

```
<a onto="" `http://www.anywhere.ru/~ivanov/` [email=""mailto:ivanov@anywhere.ru`] `` > </a>
```

Аналогичным образом вводятся и отношения:

```
<a onto=""REL(Obj1, Obj2, Obj3, ..., Objn)`` > ... </a>
```

Имеются в языке и средства, которые обеспечивают уменьшение сложности аннотирования: например, возможности именованя «длинных» конструкций и последующего использования этих имен.

Оценивая проект инициативы (КА)<sup>2</sup> в целом, можно констатировать, что очень многие идеи, разработанные и реализованные в этом проекте, сейчас «переоткрываются» снова.

### 2.2.2. Проект SHOE

Проект SHOE (Simple HTML Ontology Extensions) выполнялся в 1997–1999 гг. и был ориентирован на решение проблемы добавления к веб-страницам семантической информации и соотнесения ее с онтологиями соответствующих предметных областей. Предполагалось, что, используя эту информацию, поисковые системы смогут обеспечивать

более релевантные ответы на запросы, чем это возможно на базе использования стандартных машин поиска, функционирующих в сети Интернет.

Для поддержки процессов аннотирования в рамках проекта SHOE был разработан специальный набор инструментальных средств (*suite of tools*), основой которых был язык интернет-совместимого представления знаний, давший название всему проекту.

Основными направлениями исследований в рамках данного проекта были:

- разработка множества повторно используемых онтологий (*reusable ontologies*) для концептов, которые наиболее частотны для веб-ресурсов;
- создание аннотаторов знаний (*Knowledge Annotator*), которые бы упростили процессы проектирования онтологий.

Предполагалось также, что в составе SHOE-инструментария будет «несложная» обработка естественного языка (*lightweight natural language processing techniques*), которая обеспечит представление пользователей аннотаций документов.

По своей идее проект SHOE близок к инициативе (KA)<sup>2</sup>, но концепция языка представления знаний здесь базировалась не на фреймовых логиках, а лежала в русле расширения HTML специальными тэгами. При этом основное отличие языка SHOE было в том, что здесь предлагалось «полномасштабное» расширение HTML. Для этого в рамках SHOE в HTML-стандарт были введены следующие новые тэги для спецификации онтологий: ONTOLOGY, USE-ONTOLOGY, DEF-CATEGORY, DEF-RELATION, DEF-ARG, DEF-RENAME, DEF-CONSTANT, DEF-TYPE, DEF-INFERENCE, INF-IF, INF-THEN, COMPARISON, CATEGORY, RELATION, ARG и др.

Для аннотирования HTML-документов в формализме SHOE, кроме того, были введены новые тэги: INSTANCE, метатэг <META HTTP-EQUIV =...> и др.

Для определенности в рамках спецификации языка SHOE предполагалось, что онтология представляется в виде *is\_a* иерархии классов/категорий, множества атомарных отношений между категориями и множества правил вывода в форме простых клауз Хорна [Клоксин др., 1987]. В качестве термов языка служили термы HTML и, дополнительно к этому, понятия Category (Class), Data (с типами STRING, NUMBER,



DATE, TRUTH), Element, Instance, Instance Key, Name, Ontology, Relation (Relationship), Rule и некоторые другие.

Декларации онтологий задавались внутри тела HTML-документа и не могли перекрываться с другими тэгами HTML. В одном документе могли быть определены несколько онтологий, но такие определения тоже не могли перекрываться или быть вложенными.

Общая схема определения онтологии в проекте SHOE – следующая [Heflin, et al., 1998]:

```
<ONTOLOGY ID="идентификатор-онтологии"  
  VERSION="версия"  
  [BACKWARD-COMPATIBLE-WITH="список-версий"]  
  [DESCRIPTION="текст"]  
  [DECLARATORS="список-деклар.-экземпляров"]>  
    собственно-декларация-онтологии  
</ONTOLOGY>
```

Данная онтология может расширять другую онтологию:

```
<USE-ONTOLOGY ID="идентификатор-онтологии"  
  VERSION="версия" PREFIX="префикс" [URL="URL"]>
```

Внутри определения онтологии могут специфицироваться новые категории:

```
<DEF-CATEGORY NAME="имя-категории"  
  [ISA="список-родительских-категорий"]  
  [DESCRIPTION="текст"] [SHORT="текст"]>
```

Аналогичный подход применяется и для определения отношений:

```
<DEF-RELATION NAME="имя-отношения"  
  DESCRIPTION="текст" [SHORT="текст"]> список-аргументов  
</DEF-RELATION>
```

Определение правил вывода:

```
<DEF-INFERENCE [DESCRIPTION="текст"]>  
  <INF-IF> тело </INF-IF>  
  <INF-THEN> голова </INF-THEN>  
</DEF-INFERENCE>
```

Ниже, для примера, обсуждается фрагмент определения онтологии в формализме SHOE, коррелирующий с уже обсуждавшимся фрагментом определения аналогичной онтологии в формализме Ontobroker.

Пусть нас интересуют исследователи, имеющие в сети Интернет свои домашние страницы. Для работы с такими страницами можно вос-

пользоваться уже существующей в рамках SHOE онтологией общих понятий, (organization-ontology) по адресу <http://www.ont.org/orgont.html>. Однако предположим, для определенности, что существующую онтологию необходимо расширить понятиями Person и Organization. Тогда спецификация фрагмента новой онтологии (HomePageOntology) может быть представлена в формализме SHOE следующим образом:

```
<ONTOLOGY ID="HomePageOntology" VERSION="1.0">
  <ONTOLOGY-EXTENDS «organization-ontology"
  VERSION="2.1" PREFIX="org"
  URL="http://www.ont.org/orgont.html">
  <ONTDEF CATEGORY="Person" ISA="org.Thing">
  <ONTDEF RELATION="lastName" ARGS="Person STRING">
  <ONTDEF RELATION="firstName" ARGS="Person STRING">
  <ONTDEF RELATION="employee"
  ARGS="org.Organization Person">
  .....
</ONTOLOGY>
```

Аннотация HTML-документов в SHOE осуществляется также с использованием тэгов. В частности, для этого служат тэги USE-ONTOLOGY, INSTANCE, CATEGORY, RELATION. Последние три тэга имеют следующие форматы:

```
<INSTANCE KEY="значение-ключа"
  [DELEGATE-TO="список-примеров"]> ... </INSTANCE>
<CATEGORY NAME="префикс.категория" [FOR="ключ"]>
<RELATION NAME=" префикс.отношение">
  список-аргументов </RELATION>
```

Для поиска и обработки домашних страниц с помощью специфицированной выше онтологии необходимо, чтобы авторы веб-публикаций сами (или на основе инструментария SHOE) проаннотировали свои документы.

Так, например, фрагмент аннотации персональной страницы исследователя Иванова в формализме SHOE выглядит следующим образом:

```
<BODY>
<META HTTP-EQUIV="Instance"
  CONTENT="http://www.anywhere.ru/~ivanov">
<USE-ONTOLOGY «HomePageOntology" VERSION="1.0" PREFIX="our"
```

```
URL="http://www.ont.org/HomePageOntology.html">
<CATEGORY "our.Person">
<RELATION "our.firstName" TO="Ivan">
<RELATION "our.lastName" TO="Ivanov">
.....
<RELATION "our.employee" FROM="http://www.ccas.ru">
.....
```

</BODY>

Анализ приведенного HTML-текста показывает, что даже в таком простом случае задача аннотации веб-документа достаточно сложна. Ситуация становится еще более сложной при аннотировании реальных HTML-документов. Во-первых, уже выбор объектов текста, подлежащих аннотированию, не тривиален, особенно если веб-документ представляет объекты реального мира. Во-вторых, гиперссылки часто фиксируют лишь наличие определенных отношений между объектами, но не их семантику. И наконец, можно, конечно, аннотировать каждую именную группу HTML-страницы, но для реальных документов это слишком трудоемкая задача, которая к тому же чревата большим количеством ошибок.

Поэтому в рамках проекта SHOE для автоматизации процессов аннотирования веб-документов была разработана специальная система Knowledge Annotator [SHOE, 1999]. Одна из экранных форм, иллюстрирующих архитектуру этой компоненты, представлена на Рис. 3.

Основными информационными блоками системы Knowledge Annotator были экземпляры (instances), онтологии (ontologies) и утверждения (claims). Пользователь мог добавлять, редактировать и/или удалять любой из элементов этих блоков. При создании новых объектов пользователю выдавались соответствующие подсказки, например, в виде списка доступных онтологий, описанных в них категорий, отношений и т.п.

Для визуализации знаний, содержащихся в обрабатываемом документе, в Knowledge Annotator использовались различные методы, начиная с аннотированного HTML-текста и заканчивая описаниями утверждений на естественном (английском) языке. Кроме того, система осуществляла проверку корректности действий пользователя и транслировала его выборы в синтаксически правильные конструкции SHOE.

В процессе выполнения проекта были разработаны различные языки запросов к документам, проаннотированным на основе формализмов

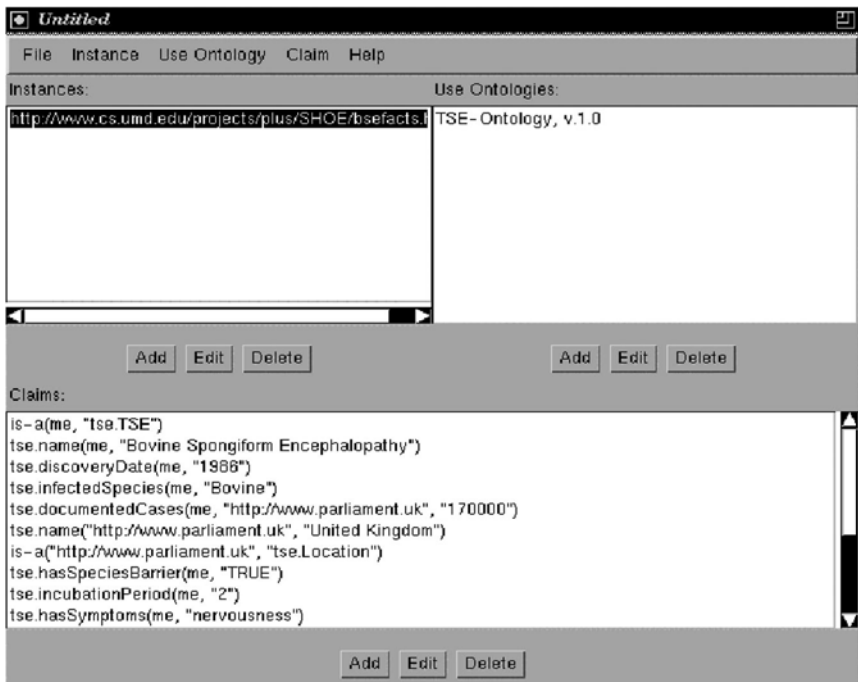


Рис. 3. Экранная форма системы Knowledge Annotator

SHOE. Так, в университете Мэрилэнд (University of Maryland at College Park) был разработан робот Exposй, который обрабатывал SHOE-документы и добавлял их в свою базу знаний, используя систему представления знаний PARKA [Evetт, et al., 1994].

PARKA-запрос для поиска домашних страниц специфицировался в таком случае следующим образом:

```
(query! '(and
  (!instanceOf ?X #!Person) (!instanceOf ?Y #!Person)
  (!instanceOf ?Z #!Organization)
  (!lastName ?X "Ivanov") (!lastName ?Y "Ivanova")
  (!employee ?Z ?X) (!employee ?Z ?Y)
  (!involvedIn ?Z "РФФИ-проекты")))
```

По существу, это достаточно простой SQL-запрос [Клайн, 2006], расширенный за счет использования понятий онтологии, переменных и

ограниченных по мощности образцов. Заметим, что в настоящее время идеи, заложенные в основу языка запросов PARKA, нашли развитие в языке SPARQL [SPARQL, 2008].

Оценивая формализм представления онтологических знаний SHOE и поддержку процессов аннотирования веб-ресурсов в этом проекте в целом, можно отметить, что это достаточно мощная система методов и средств, которая сложнее для пользователя, чем Ontobroker. В своей реализации проект SHOE «тяготел» к существующим в то время информационным технологиям обработки интернет-страниц больше, чем реализация в рамках инициативы (KA)<sup>2</sup>, и в этом смысле был более прагматичным. Вместе с тем мощность средств представления знаний, а следовательно, и возможности формирования пространств знаний здесь были ниже, чем у инициативы (KA)<sup>2</sup>.

Тем не менее можно констатировать, что именно рассмотренные выше два проекта, по сути дела, и заложили научно-технические основы формирования пространств знаний.

### ***2.3. Информационные технологии и пространства знаний***

Информационные технологии и, в частности, веб-технологии уже внесли и продолжают вносить существенный вклад в формирование Социального Веба и таким образом, в определенном смысле, готовят почву для Semantic Web.

В настоящее время существует много проектов уровня Web 2.0, но особую роль в этой области сыграл, на наш взгляд, подход компании Metaweb, реализованный в базе данных Freebase.

#### ***2.3.1. Freebase***

Одним из самых успешных проектов в области Web 2.0 уже несколько лет остается база данных Freebase, реализованная компанией Metaweb Technologies [Freebase, 2010]. Freebase позиционируется как «открытая и разделяемая база данных знаний о мире или значительная по объему и кооперативно создаваемая и редактируемая база взаимосвязанных данных».

Что такое Freebase сегодня? Во-первых, это структурированные данные (специальная БД), во-вторых, кооперативно редактируемые каталоги тем и кооперативно формируемая семантика (фольксономии) и, наконец, это открытый API + открытые данные.

Уже сейчас «внутри» Freebase содержится более 3,3 млн объектов, включая примерно 750 тыс. физических лиц, 450 тыс. геоимен, 50 тыс. организаций и 40 тыс. фильмов, а также более 1000 типов и 3000 свойств (Рис. 4).

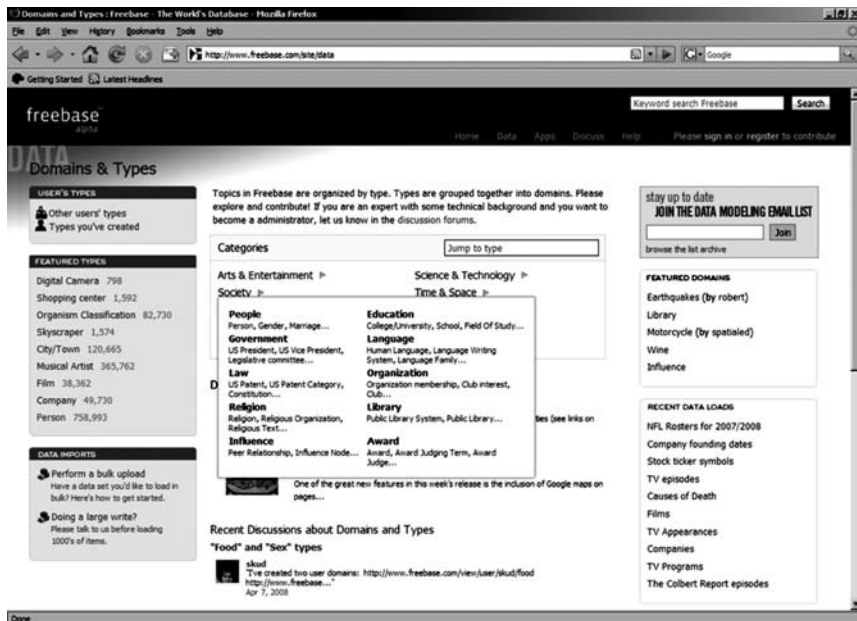


Рис. 4. Общие сведения о типах объектов Freebase

С учетом того, что Freebase позиционируется как система поддержки кооперативно формируемой семантики, здесь имеются достаточно развитые средства спецификации объектов и отношений между ними на онтологическом уровне (Рис. 5).

При этом, согласно идеологии Freebase, онтология – это не магия, панacea или изменение мира, а просто артефакт, API и в конечном счете социальный контент. Разработчики Freebase даже вводят понятие семантики «участия» (совместная семантика), иллюстрируя его схемой, представленной на Рис. 6.

Примером простой фольксономии объектов шоу-бизнеса и отношений между ними может быть фрагмент семантической сети, представленной на Рис. 7.

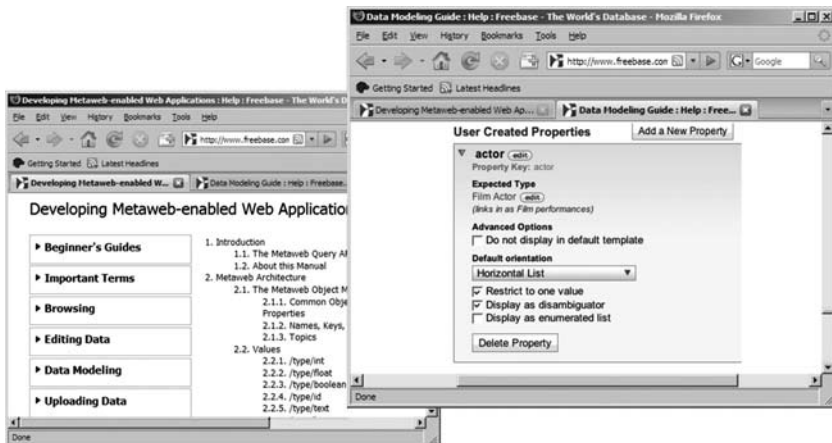


Рис. 5. Средства спецификации объектов и отношений для погружения в базу Freebase



Рис. 6. Семантика «участия» из базы данных Freebase

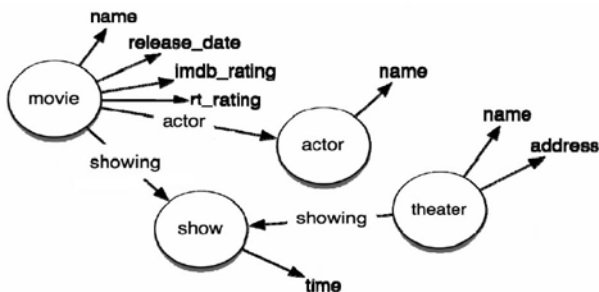


Рис. 7. Фрагмент фольксономии объектов шоу-бизнеса и отношений между ними

Однако в общем случае средства спецификации объектов во Freebase значительно богаче. Для примера на Рис. 8 показаны скриншоты списка атрибутов объектов типа ФизЛицо, которые дают представление о разнообразии информации, представленной в базе данных Freebase.

Запросы и сервисы сервера Metaweb, где расположена база Freebase, специфицируются с использованием языка MQL (Metaweb Query Language). Для облегчения реализации имеется библиотека шаблонов, написанных на языках Perl, Python, PHP и JavaScript.

Для примера на Рис. 9 представлены результаты запроса к базе знаний Freebase за информацией о русских физических лицах.

Таким образом, Freebase является успешным примером возможности кооперативного создания и поддержки достаточно сложных и больших баз знаний.

### *2.3.2. Проекты, связанные с использованием микроформатов*

Цель микроформатов – «внедрение» базисной семантики непосредственно в HTML-страницы. Пока микроформаты не столь выразительны, как другие средства семантизации контента (например, RDF или OWL), но они очень компактны и используют возможности XHTML для добавления семантики к страницам Интернета.

В настоящее время Facebook и Yahoo! Local, Google Social Graphs и другие проекты разных компаний используют микроформаты для аннотирования событий на своих страницах [Allsopp, 2006].

Основными примерами активно используемых микроформатов являются:

- hCard (для аннотирования HTML-страниц таким образом, чтобы понимающие данный формат веб-браузеры или поисковые машины могли вывести контактную информацию о человеке);
- hCalendar (для аннотирования событий на странице);
- hReview – (для создания аннотаций обозревателями);
- hResume – (для аннотирования резюме) и др.

Анализ литературы по разработке и использованию микроформатов показывает, что данный подход имеет серьезные перспективы в качестве альтернативы более мощным, но и более сложным методам «внедрения» семантики в рамках Интернета.





Рис. 8. Скриншоты списка атрибутов объектов типа Физ.Лицо

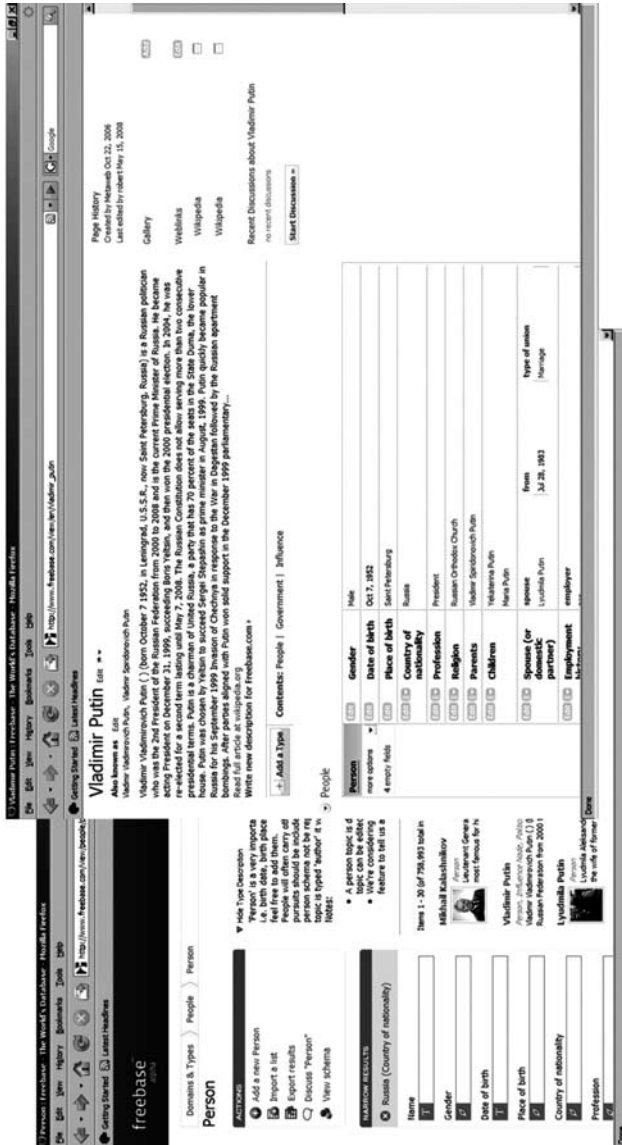


Рис. 9. Информация о русских физических лицах в базе знаний Freebase

## **2.4. Семантические технологии для пространств знаний и Semantic Web**

### **2.4.1. Онтологический инжиниринг и Semantic Web**

В последнее время, в связи с существенным повышением активности сообщества из области Semantic Technologies, наблюдается увеличение числа проектов, ориентированных на решение проблем формирования пространств знаний и развития концепции Semantic Web.

Концепция Semantic Web (SW), которую на международной конференции XML-2000 выдвинул Тим Бернерс-Ли (Tim Berners-Lee) – один из основоположников WWW и нынешний председатель WWW-консорциума (W3C), как известно, заключается в организации такого представления информации в сети, чтобы допускалась не только ее визуализация, как это происходит сейчас, но и эффективная автоматическая обработка [Berners-Lee, et al., 2001]. По определению W3C, Semantic Web представляет собой расширение WWW, в рамках которого информация (веб-контент) представляется в форматах, обеспечивающих ее использование программными агентами, позволяя им искать, разделять и интегрировать информацию значительно легче, чем это происходит сейчас.

С момента появления SW-концепции прошло больше 10 лет и специалисты уже говорят о семантической волне (Semantic Wave), которая, по оценке руководителя проекта Project 10X Миллса Дэвиса [Davis, 2006], существенным образом изменит характер работы с информацией. В аналитическом отчете Gartner Group [Cearley, et al., 2007] дается прогноз, что к 2012 г. в 80% общедоступных веб-сайтов будет в той или иной степени использоваться семантический гипертекст для создания семантических веб-документов (с вероятностью 0.7), а к 2012 г. в 15% общедоступных веб-сайтов будут использоваться развитые веб-онтологии для создания семантических баз данных (с вероятностью 0.6).

Вместе с тем следует отметить, что SW-эра, в отличие от интернет-эпохи, еще только приближается, и на этом пути существует значительное число научных, технических, технологических и чисто человеческих проблем, основными из которых являются [Benjamins, et al., 2002]:

- доступность семантического контента;
- доступность онтологий и средств их разработки, а также эволюция онтологий;

- масштабируемость;
- мультязыковость;
- визуализация;
- стабильность.

Доступность семантического контента является основной проблемой на пути формирования и использования пространств знаний, так как основная масса информации на Web не представлена в SW-форматах и нет надежды, что эта работа может быть выполнена вручную.

Онтологии, по мнению практически всех специалистов, являются ключевым компонентом в решении проблемы семантизации Web-контента. В связи с этим особое значение приобретают проблемы онтологического инжиниринга (методы и средства разработки и эволюции онтологий), а также доступность уже существующих онтологий.

Значительные усилия должны быть предприняты для хранения, обработки и поиска семантического контента, причем решения в этой области должны обеспечивать эффективную работу с огромными объемами знаний.

Проблема мультязыковости контента существует и в классическом Web, но для SW, который, по сути, должен поддерживать эффективный доступ к информации независимо от того, на каком языке она представлена изначально, эта проблема является одной из основных. И решение ее специалисты в значительной мере связывают с решением проблем онтологического инжиниринга и автоматической обработкой ЕЯ-текстов.

Представление информации для пользователей (визуализация контента) также должно претерпеть существенные изменения и обеспечить свободную ориентацию в огромном количестве фактов, которые отвечают его потребностям.

Последняя по счету, но не по важности, проблема связана с обеспечением стабильности SW, а это, в свою очередь, предполагает, что серьезные усилия должны быть предприняты в области стандартизации, обеспечивающей создание технологий, необходимых для формирования пространств знаний.

Как представляется, проблемы стандартизации являются одной из ключевых позиций в проблематике SW, и с этой точки зрения представляет интерес хотя бы краткое обсуждение схемы, представленной на Рис. 10, которую специалисты называют «слоеным пирогом» Тима Бернерс-Ли [Berners-Lee, 2003].

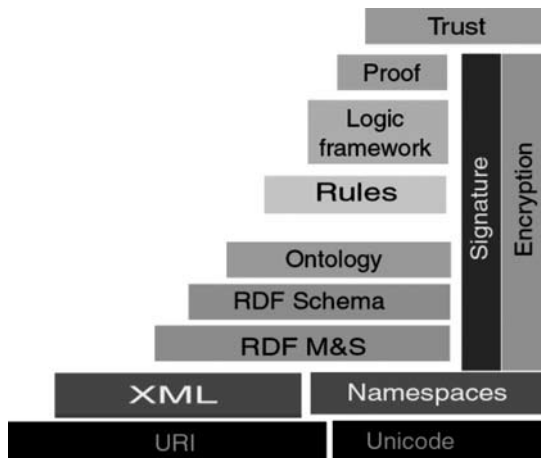


Рис. 10. «Слоеный пирог» Тима Бернерс-Ли

Для нижних уровней этой схемы, которые можно объединить в слой «RDF-данные», общая цель, сформулированная W3C, была следующей: разработка форматов сериализации данных и интероперабельность приложений. В результате усилий исследовательских коллективов и консорциума W3C для этого слоя были разработаны и реализованы рекомендации по форматам XML, Namespace (пространства имен) и RDF, которые в настоящее время существуют на уровне стандартов de facto. И можно констатировать, что результаты по данному направлению уже перешли из стадии исследований в стадию использования, в том числе и в коммерческих системах. На уровне RDF-схем предложены и поддерживаются W3C стандарты RDFS (RDF-схем), которые позволяют специфицировать словари используемых терминов, и разрабатываются соответствующие спецификации для существующих и новых приложений. Здесь, как и в предыдущем направлении, результаты уже перешли из стадии исследований в стадию использования.

На онтологическом уровне (Ontology) «слоеного пирога» SW ситуация несколько иная. По сути дела, в этом направлении был достаточно мощный задел в рамках исследований по представлению знаний (в частности, общие подходы к представлению знаний типа фреймов и семантических сетей, конкретные формализмы, языки и системы представления знаний, которые обсуждались выше (например, Frame Logics

[Kifer, et al., 1995], SHOE [Heflin, et al., 1998] и др). Вместе с тем работа по стандартизации средств представления знаний онтологического уровня далеко не закончена, а создание соответствующих средств онтологического инжиниринга в настоящее время – одна из «горячих точек» в данной области. Основным направлением исследований и разработок здесь является создание более мощных средств спецификации онтологий, обеспечивающих вывод на знаниях и проверку целостности знаний, средств поддержки целостности онтологических спецификаций в процессе эволюции как спецификаций самих моделей, так и стандартов, а также средств спецификации перекрестных ссылок между словарями и конвертирования спецификаций. Основным результатом в этом направлении можно считать «выравнивание» средств спецификации онтологий, разработанных в США (DAML) и в Европе (OIL), до общего формализма (DAML+OIL [DAML+OIL, 2001]), а также разработку консорциумом W3C стандарта de facto на спецификацию онтологий – языка OWL [OWL, 2004].

Понятно, что с переходом к верхним слоям схемы, представленной на Рис. 10, ситуация смещается от уже разработанных стандартов и их использования к исследованиям в соответствующих направлениях. И в этом смысле показателен промежуточный слой – «Слой правил» (Rules), проработка которого потребовала внесения изменений в слои предыдущих уровней (так, например, были добавлены средства спецификации переменных в RDF-слои), а также поиска новых выразительных и простых средств спецификации отношений и средств для спецификации запросов к базам знаний аналогичных SQL [Клайн, 2006] с возможностью фильтрации получаемых результатов. Наряду с этим в данном направлении ведутся исследования и разработки по теории монотонных и немонотонных систем вывода на правилах, а также работы по созданию новых приложений с использованием технологий типа «webized rule engine technology». Сложившуюся ситуацию в этом направлении можно охарактеризовать следующим образом: уже существуют разные системы спецификации правил и требуется их сравнительный анализ, «вебизация» и стандартизация, разработан язык SPARQL [SPARQL, 2008], который постепенно становится стандартом de facto на язык запросов к базам знаний.

На уровнях «Логические Основы» (Logic Framework) и «Подтверждение» (Proof) ситуация еще больше смещается в область фундаментальных исследований. Здесь предлагаются подходы к спецификации

аксиом для систем, основанных на правилах, исследуются различные логики, причем основное внимание уделяется системам, в которых не выполняется аксиома «замкнутого мира», предлагаются средства валидации доказательств. Но пока нет основы для стандартизации систем, основанных на правилах, а существующие системы правил легко экспортируются из разных систем, но плохо импортируются в другие системы.

Что касается уровня «Доверие» (Trust), то здесь только формируются направления фундаментальных и прикладных исследований, поскольку все утверждения в Web-среде существуют в некотором контексте и приложения должны учитывать эти контексты, так как нельзя считать, что все факты, полученные из сети, являются истинными. Таким образом, в данном направлении существует широкое поле для исследований и последующей стандартизации.

Однако, оценивая ситуацию по стандартизации в рамках SW в целом, можно констатировать, что уже создан базис стандартизации в виде XML-, RDF(S)-, OWL- и SPARQL-спецификаций, на которые могут опираться работы по семантизации контента и его использованию в различных приложениях.

Как отмечалось выше, одной из основных целей создания SW является «семантизация» контента, существующего в настоящее время в рамках классического WWW, и вновь создаваемого контента, что должно обеспечить интеллектуальное его использование программными агентами для решения задач, значимых для пользователей. По сути дела, это требует семантического аннотирования контента на базе соответствующих онтологий, где фиксируется смысл отдельных его элементов и связей между ними, что, в свою очередь, предполагает эволюцию взглядов на формирование пространств знаний, связанную с расширением сообщества разработчиков, усложнением разрабатываемых моделей и необходимостью средств коллективной разработки таких пространств.

Как показывает анализ литературы, а также с учетом вышесказанного, можно констатировать, что в настоящее время основные исследования и разработки в перечисленных выше направлениях сосредоточены в следующих областях:

- методологии и методы построения онтологий «с нуля» (from scratch);
- методы реинжиниринга онтологий (ontology reengineering);

- методы коллективной разработки онтологий (collaborative ontology development);
- методы и методологии объединения и выравнивания онтологий (ontology merging & alignment).

В каждой из указанных областей в настоящее время ведутся интенсивные исследования и разработки в рамках десятков и даже сотен проектов. Не имея возможности рассмотреть здесь даже наиболее известные из таких работ, а также с учетом того, что в научных публикациях и в Интернете представлены достаточно представительные обзоры по данной тематике, например, [Добров и др., 2006; Овдей и др., 2004; Загорюлько, 2008], здесь мы ограничимся лишь краткой характеристикой основных результатов западных специалистов и выполняемых на Западе проектов, а в дальнейших публикациях остановимся подробнее на работах и проектах российских специалистов и коллективов, которые в доступной литературе отражены слабее.

#### 2.4.2. Построение онтологий «с нуля»

Среди методологий и методов построения онтологий «с нуля» (from scratch, on fly) можно выделить следующие:

- подход Сус;
- метод Усколда и Кинга;
- методология Грюнингера и Фокса;
- методология METHONTOLOGY.

#### **Подход Сус**

Как известно [Lenat, et al., 1989], подход Сус сформировался в процессе выполнения амбициозного проекта по созданию сверхбольшой базы знаний «здорового смысла» (common sense knowledge), выполнявшегося в 80-х годах прошлого века под руководством Д. Лената (D. Lenat), одного из самых известных в США ученых в области баз знаний и экспертных систем. Оценивая этот проект с позиций сегодняшнего дня, можно сказать, что он закончился неудачей, как, впрочем, и еще более амбициозные проекты, например, японский проект «Вычислительные системы V поколения» и аналогичные этому проекты других стран. Вместе с тем именно в рамках проекта Сус были разработаны первые инструментальные средства инженерии знаний и, в частности, язык представления знаний СусL, базой которого послужило, с одной



стороны, исчисление предикатов высших порядков, а с другой – основной в то время язык систем искусственного интеллекта Лисп. Именно в рамках этого проекта впервые была поставлена и, безусловно, в ограниченном объеме, решена задача формирования больших баз знаний, валидации и верификации таких баз, а также задача вывода на знаниях. Именно в рамках проекта Сус была предложена и апробирована концепция Сус-агентов, имеющих общее ядро знаний из БЗ Сус и знания из своей специфической области. Именно в рамках проекта Сус была предложена идея структуризации БЗ в виде микротеорий, включающих знания из разных ПО, представленных с разных точек зрения.

Построение базы знаний (а сейчас мы говорим онтологии) в проекте Сус предполагает наличие следующих фаз:

- «Ручное» кодирование явных и неявных знаний, содержащихся в источниках знаний.
- «Ручное» кодирование знаний с помощью программных средств, используя знания, уже присутствующие в БЗ Сус.
- Полуавтоматическая фаза, когда разработчик «рекомендует» программным средствам источники знаний для обработки и «объясняет» им наиболее сложные места обрабатываемых текстов.

При этом на каждой из фаз, как правило, решаются две основные задачи:

- Разработка системы представления знаний и онтологии верхнего уровня, содержащей наиболее абстрактные понятия.
- Представление знаний, оставшихся вне формализации после решения первой задачи, на основе использования примитивов, разработанных и реализованных в процессе решения первой задачи.

После завершения проекта в конце XX в. Д. Ленат создал частную некоммерческую компанию Suscorp, Inc. [Suscop, 2010], основной задачей которой он видит развитие БЗ Сус и приложений, использующих эту базу знаний.

В настоящее время БЗ Сус содержит уже несколько сотен тысяч терминов и базовых утверждений и миллионы утверждений «здорового смысла», выведенных на их основе. Фрагмент БЗ Сус недавно был выпущен в открытое использование под именем OpenCus (одна из версий OpenCus 2006 г. содержит около 50 тыс. концептов и 300 тыс. фактов), которая доступна для исследователей в области искусственного интеллекта под лицензией ResearchCus.

В качестве приложений подхода Сус можно отметить:

- систему интеграции гетерогенных баз данных, в рамках которой словарь Сус отображается в схемы баз данных, в результате чего данные из БД интерпретируются в соответствии с терминами Сус-онтологии;
- интеллектуальный «движок» поиска изображений по информации, содержащейся в подписях к ним;
- модуль интеграции структурированной терминологии, обеспечивающий импорт сложных тезаурусов, их интеграцию и поддерживающий соответствующие процессы управления;
- модуль поиска информации в Интернете для расширения БЗ Сус.

### **Метод Усколда и Кинга**

Метод Усколда и Кинга (Uschold and King's method) был предложен по результатам разработки одной из самых известных онтологий моделирования бизнес-процессов – онтологии Enterprise Ontology [Uschold, et al., 1998]. По сути дела, при этом была предложена методология проектирования онтологий, где были зафиксированы следующие этапы:

- *Определение цели.* Спецификация того, зачем строится онтология, и как она будет использоваться.
- *Разработка онтологии.* Этап, выполнение которого осуществляется в рамках следующих фаз:

• Фиксация онтологии, где происходит:

- выявление (идентификация) ключевых понятий и отношений,
- разработка точных текстовых определений для каждого понятия и отношения;
- выявление терминов, относящихся к каждому понятию и отношению;
- согласование всех знаний, полученных в процессе фиксации разрабатываемой онтологии.

• Кодирование онтологии, в процессе которого осуществляется ее формальное представление на выбранном языке представления знаний.

• Интеграция. На этой фазе рассматриваются возможности использования уже существующих онтологий и интеграции их во вновь создаваемую онтологию.

– *Оценка созданной онтологии.* Этап, на котором осуществляется оценивание созданной онтологии с точки зрения ее соответствия исходным целям и задачам, а также с точки зрения эффективности использованных программных средств.

– Документирование.

Наиболее важным проектом, где использовался разработанный Усколдом и Кингом метод и методология, был Enterprise Project, выполнявшийся Artificial Intelligence Applications Institute Эдинбургского университета с такими партнерами, как IBM, Lloyd's Register, Logica UK Limited и Unilever, а наиболее важным приложением метода – разработка Enterprise Ontology, которая является коллекцией терминов и определений, относящихся к (торгово-промышленным) предприятиям. Интересно и то, что с использованием Enterprise Ontology был создан инструментарий Enterprise Toolset, который использовал агентную архитектуру для интеграции стандартных серийно выпускаемых программных средств в стиле plug-and-play.

### **Методология Грюнингера и Фокса**

Методология Грюнингера и Фокса (Grüniger and Fox's methodology), как и многие другие методологии и методы онтологического инжиниринга первого поколения, была сформирована на основе опыта разработки конкретной онтологии, в данном случае TOVE [TOVE, 2010], ориентированной на предметную область моделирования бизнес-процессов. По существу, эта методология предполагает создание онтологии как логической модели знаний и включает следующие этапы:

– *Фиксация мотивационного сценария.* В рамках данной методологии постулируется, что создание любой онтологии мотивируется некоторыми сценариями, возникающими в конкретной прикладной области, которые специфицируют множество интуитивно возможных решений для указанных в сценарии проблем.

– *Формулирование неформальных вопросов проверки компетенции.* Вопросы оценки компетентности создаваемой онтологии основаны на мотивационных сценариях и рассматриваются как требования к изобразительной мощности онтологии и возможности с ее помощью решать задачи, обозначенные в мотивационных сценариях.

– *Спецификация терминологии онтологии на формальном языке,* которая базируется на выполнении фаз:

• *Получение неформальной онтологии.* Из множества неформальных вопросов проверки компетенции онтологии выделяется множество терминов, которые должны быть базисом для спецификации на формальном языке.

• *Спецификация формальной терминологии.* Выделенное в предыдущей фазе множество терминов специфицируется на формальном языке (в данном случае использовался ЯПЗ KIF [KIF, 2010]).

– *Формулировка вопросов оценки компетенции с использованием терминологии онтологии.* По сути дела, на этом этапе происходит спецификация запросов на формальном языке для оценки компетентности онтологии.

– *Спецификация аксиом для терминов онтологии на формальном языке.* Здесь специфицируется семантика терминов онтологии и ограничения на их интерпретацию в виде утверждений исчисления предикатов первого порядка.

– *Задание условий полноты онтологии.* На данном этапе задаются условия, при выполнении которых решение вопросов, связанных с компетенцией онтологии, будет полным.

Наиболее значимыми приложениями методологии Грюнингера и Фокса были:

– Enterprise Design Workbench (APM для проектирования предприятия) – среда проектирования, позволяющая пользователю анализировать проекты предприятия. Важной функциональной особенностью APM Enterprise Design Workbench была поддержка сравнительного анализа альтернативных проектов предприятий и руководство проектировщиком.

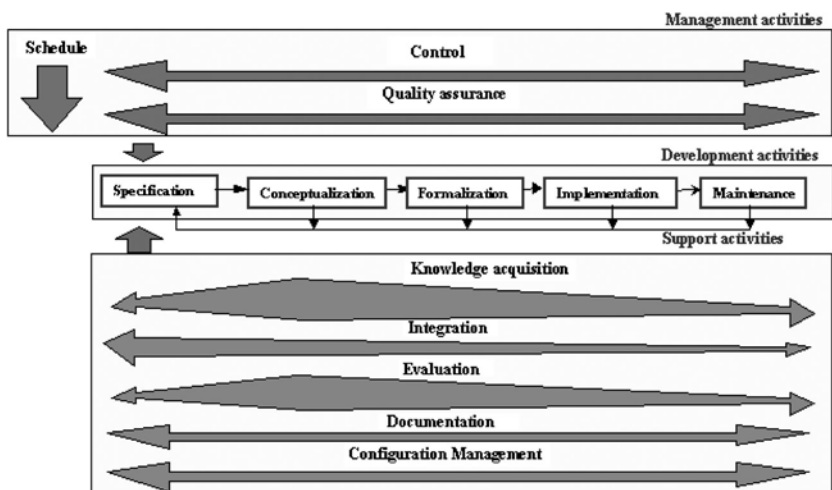
– Integrated Supply Chain Management Project agents – организация цепочки снабжения как сети взаимодействующих интеллектуальных агентов, каждый из которых выполняет одну или несколько функций в цепочке снабжения и координирует свои действия с другими агентами.

### **Методология METHONTOLOGY**

Методология METHONTOLOGY [Fernandez, et al., 2006] разработана в лаборатории ИИ Мадридского политехнического университета. Отличительным признаком этой методологии является то, что METHONTOLOGY сформирована на базе анализа и переосмысления основных видов деятельности, свойственных процессам разработки программного обеспечения и инженерии знаний. Таким образом, METHONTOLOGY интегрирует опыт проектирования сложных объектов из двух областей знаний.

Основные процессы разработки онтологии и ее жизненный цикл в METHONTOLOGY представлены на Рис. 11.

Как следует из представленной выше схемы, данная методология включает в себя идентификацию процесса разработки онтологий, жиз-



**Рис. 11.** Процессы разработки и жизненный цикл METHONTOLOGY

ненный цикл, основанный на эволюции прототипов, и отдельные приемы для выполнения каждой деятельности. В свою очередь, жизненный цикл включает такие стадии, как спецификация, концептуализация, формализация, реализация и сопровождение, а также основные процессы, такие как управление, контроль качества, приобретение знаний, интеграция, оценивание, документирование и управление конфигурациями. Подход, использованный при создании METHONTOLOGY, хорошо коррелирует со спиральной моделью Бозма [Larman, et al., 2003], широко используемой в технологиях разработки программного обеспечения.

Методология METHONTOLOGY активно использовалась в ассоциации FIPA (Foundation for Intelligent Physical Agents) [FIPA, 2010], проекты и разработки которой были ориентированы на создание мульти-агентных систем и интеллектуальных агентных приложений.

Примеры онтологий, разработанных с использованием методологии METHONTOLOGY, перечислены ниже. Это:

- CHEMICALS (содержит знания в области химических элементов и кристаллических структур).
- Monatomic Ions (собирает информацию об одноатомных ионах).

- Онтологии загрязняющих веществ окружающей среды (представляют методы обнаружения различных загрязняющих компонентов в воде, воздухе, земле и максимально допустимые концентрации этих веществ с учетом действующих законов).
- The Reference-Ontology (базовая онтология для описания онтологий справочников типа «желтых страниц»).
- Silicate ontology (моделирует свойства минералов и, в частности, силикатов).
- Онтологии, разработанные в проекте IST-1999-10589 МКВЕЕМ (путешествия, каталоги тканей, жилье; использованы в мультязычной платформе для электронной коммерции).
- OntoRoadMap (онтология об онтологиях, методологиях разработки онтологий, средствах разработки онтологий, событиях, связанных с онтологиями (конференции, семинары и т.п.)).

Примеры приложений, которые используют некоторые из перечисленных выше онтологий:

- (Onto)<sup>2</sup>Agent (брокер онтологий, который использует Reference-Ontology в качестве источника знаний и находит описания онтологий, удовлетворяющих заданному множеству ограничений).
- Приложение OntoRoadMap application (разработано как развитие (Onto)<sup>2</sup>Agent, является основанным на онтологии веб-приложением, которое позволяет сообществу регистрировать, просматривать и находить онтологии, методологии, программные средства и языки для построения онтологий, приложения в таких областях, как Semantic Web, e-commerce, КМ, NLP, и т.п., а также об основных конференциях, семинарах и событиях в этой области).
- Ontogeneration (система, использующая (прикладную) онтологию CHEMICALS и лингвистическую онтологию GUM, чтобы генерировать тексты на испанском языке в ответ на запрос в области химии).

#### *2.4.3. Реинжиниринг онтологий*

Как известно [Gomez-Perez, et al., 1999], реинжиниринг онтологий – одно из направлений онтологического инжиниринга, в рамках которого происходит анализ концептуальной модели уже реализованной онтологии и отображение ее в другую, более подходящую для решения новых задач концептуальную модель, которая реализуется как новая онтология.

Следует сразу отметить, что реинжиниринг онтологий в значительной мере опирается на методы реинжиниринга программного обеспе-

чения и, по существу, адаптирует их к области проектирования онтологий.

Так, например, метод реинжиниринга, применяемый онтологической группой уже упоминавшегося выше Мадридского политехнического университета, явно использует схемы реинжиниринга программного обеспечения и сводится к выполнению следующих стадий:

– *Инженерный анализ (reverse engineering)*, целью которой является экстракция концептуальной модели онтологии на основе ее кода на определенном формальном языке.

– *Реструктурирование (restructuring)*. На данной стадии выполняется преобразование исходной концептуальной модели в новую концептуальную модель с учетом использования реструктурированной онтологии других онтологий или приложений. Реструктуризация, как правило, включает две фазы:

- **Анализ** (проверка того, что новая онтология (ее классы, экземпляры, отношения и функции) полна, непротиворечива (нет конфликтов), минимальна (нет явных или неявных дублирований) и синтаксически корректна.

- **Синтез** (реализуется корректная онтология и документируются все сделанные изменения).

– *Прямая разработка (forward engineering)*. Цель выполнения этой стадии заключается в получении новой реализации новой онтологии на базе новой концептуальной модели.

В качестве примеров использования метода реинжиниринга онтологий можно указать онтологию Standard Units, где в процессе реинжиниринга потребовались единицы измерений международной системы, чтобы построить онтологию Monatomic Ions, а это, в свою очередь, привело к переиспользованию онтологии Standard Units, которая входит в библиотеку онтологий Ontolingua Server.

#### 2.4.4. Коллективная разработка онтологий

В онтологическом инжиниринге любая онтология рассматривается как результат консенсуса группы специалистов о модели некоторой области знаний. Поэтому с развитием методов и средств в этой области все большее внимание стало уделяться инструментальной поддержке коллективной разработки онтологий, в рамках которой существует несколько основных проблем – управление взаимодействием и коммуникациями между разработчиками; контроль за доступом к текущим

результатам совместного проектирования; фиксация авторских прав на знания экспертов, переданные в общее пользование; обнаружение ошибок проектирования и управление коррекцией ошибок; конкурентное управление изменениями и некоторые другие. Для решения этих проблем в настоящее время уже предложено несколько достаточно хорошо проработанных подходов и соответствующих инструментальных средств, среди которых можно особо отметить:

- проект кооперативного Protege;
- проект NeON;
- инфраструктуру совместной разработки согласованных баз знаний Co<sub>4</sub>.

### **Проект кооперативного Protege**

Команда разработчиков Стэнфордского исследовательского института (SRI) давно известна среди специалистов в области онтологического инжиниринга своей системой поддержки разработки онтологий Protege [Protégé, 2010]. А на одной из всемирных конференций по семантическим технологиям SemTech-2008 они выступили с докладом о новой версии этой системы – кооперативной Protege [Tudorache, et al., 2008].

Основными требованиями к такому инструментарию, по мнению разработчиков, являются:

- Наличие средств для дискуссий и достижения консенсуса, включая аннотирование компонент и (возможно) изменений, а также включение их в процесс разработки.
- Обеспечение контекста для дискуссий в процессе моделирования решений.
- Поддержка истории изменений и ассоциированных с ними дискуссий.
- Обеспечение доверия за счет поддержки истории концептов и обеспечения путей установления доверия.
- Персонализация взглядов на онтологию на основе ролей пользователя и его задач, уровня компетенции пользователя и области доверия к пользователю.
- Контроль доступа путем взвешенного контроля процессов редактирования и просмотра.
- Поддержка ролей пользователя.



– Гибкая поддержка схем разработки, включая конфигурируемые схемы разработки и совмещение исполнения схемы разработки и процесса создания онтологий.

Основные характеристики кооперативной версии Protege:

- Расширение существующей версии.
- Поддержка, в частности:
  - аннотирования компонент онтологий и изменений,
  - линий дискуссий,
  - предложений и голосования,
  - поиска и фильтрации,
  - регистрации пользователей, групп и политик.
- Использование в новой версии инструментария таких языков и систем представления онтологических знаний, как OWL и Frames.
- Возможность многопользовательской и автономной работы.
- Распределенная инсталляция Protégé.

Наиболее интересным и важным аспектом новой версии, на наш взгляд, является использование онтологии ChAO (Changes & Annotation) для управления процессами совместного проектирования онтологий, фрагмент которой представлен на Рис. 12. По существу, при этом новая система Protégé сама становится системой, управляемой онтологией проектирования.

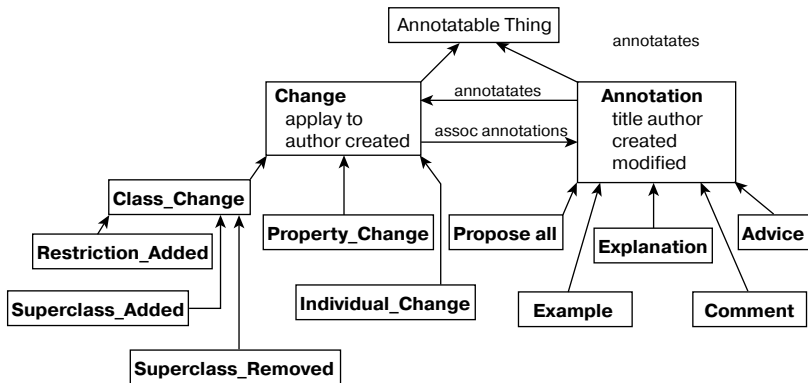


Рис. 12. Фрагмент онтологии ChAO кооперативной версии Protege

Остальные расширения, сделанные в новой версии, с учетом вышесказанного, более «традиционны» и предсказуемы. Это:

- Новые функционалы редактора онтологий.
- Компонент аннотирования, где онтология пользователя аннотируется примерами из аннотационной онтологии.
- Компонент поддержки трека изменений, в рамках которого изменения сохраняются как примеры в аннотационной онтологии.
- Архитектура клиент – сервер, в которой приложения и компоненты на стороне клиента используют общий Changes & Annotations API для работы с аннотациями и изменениями, связанными с онтологией.
- Веб-версия системы, поддерживающая:
  - просмотр онтологий через Веб-браузер,
  - «легкое» редактирование,
  - совместную работу,
  - разные схемы взаимодействия,
  - простоту кастомизации и расширения.

В целом направление движения разработчиков кооперативной версии Protege понятно и представляется, что это новый шаг в инструментарии онтологического инжиниринга.

### **Проект NeON**

Европейский проект NeON (Networked Ontologies), выполняемый в соответствии с 6-й Рамочной программой по направлению «Semantic-based Knowledge and Content Systems» (бюджет 14,7 млн евро, 4 года) в целом ориентирован на достижение тех же целей, что и кооперативная версия Protege.

Действительно, как указано на сайте этого проекта [Gomez-Perez, et al., 2009], платформа NeOn – это:

- поддержка «жизненного цикла», включая взаимодействие активностей периода разработки и исполнения;
- ориентация на онтологический инжиниринг и использование онтологий;
- расширяемость архитектуры на всех уровнях.

Инструментарий NeON поддерживает онтологический инжиниринг и управление, полный «жизненный цикл» и сетевую работу с онтологиями (модульность, отображение и т.д.). Разработан этот инструментарий на платформе Eclipse и расширяет базовую архитектуру за счет механизма плагинов Eclipse и веб-сервисов. Общая структура платформы NeON представлена на Рис. 13.

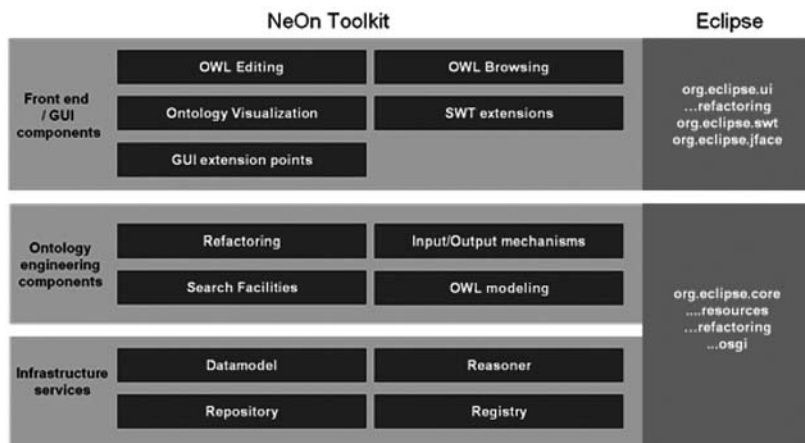


Рис. 13. Общая структура платформы NeON

Оценивая проект NeON в целом, можно констатировать, что он находится в общей «струе» исследований и разработок в этой области. А специфика проекта в том, что это международная кооперация специалистов из разных европейских стран, причем отдельные участники уже изначально вошли в этот проект со своими наработками. По сути дела, эта платформа конкурирует с кооперативной версией Protégé, а также с коммерческими инструментариями онтологического инжиниринга, например, с инструментарием фирмы Altova [Altova, 2010].

#### Совместная разработка согласованных баз знаний $Co_4$

$Co_4$  [Euzenat, 1996] (for Collaborative construction of consensual knowledge bases) является инфраструктурой, обеспечивающей совместное конструирование базы знаний в интернет-среде. При этом разработка баз знаний трактуется как социальный процесс, в который вовлечено сообщество множества агентов, а система  $Co_4$  поддерживает разработку с помощью «знающих» людей, которые являются равноправными участниками.

Одним из основных требований, которые должна поддержать инфраструктура, является требование консенсуса (модификация принимается только после согласования со всеми участниками процесса проектирования). Для того чтобы такой подход работал, участники проекта не

должны непосредственно модифицировать базу знаний, а только свое личное рабочее пространство.

В рамках  $Co_4$  каждый участник рассматривается системой как база знаний. Для построения согласованной базы знаний индивидуальные базы знаний должны быть связаны. Поэтому базы знаний в общем случае организуются в дерево, чьи листья являются пользовательскими базами знаний, а промежуточные – групповыми базами знаний. При этом каждая групповая БЗ представляет знания, согласованные между его сыновьями. Групповая БЗ, функционирующая как агент мультиагентной системы, посылает своим подписчикам сообщения об изменениях, принятых всеми, и просьбу прислать комментарии (должны ли эти изменения быть зафиксированы или нет). Когда подписчики достаточно уверены в своих порциях знаний, они могут предоставить их своей групповой БЗ. Это предложение затем представляется другим подписчикам как запрос на комментарий. В ответ пользователи должны дать один из ответов: «принять» (ассерт), когда они считают, что предложенные знания должны быть интегрированы в согласованную БЗ, «отвергнуть» (режет) – в противном случае, и «запрос, вызов» (challenge), когда они предлагают другое изменение.

В целом инфраструктура  $Co_4$  представляет магистральное направление онтологического инжиниринга на современном этапе. А ее успех, на наш взгляд, будет определяться качеством и эффективностью реализации соответствующих инструментальных средств.

#### *2.4.5. Объединение и выравнивание онтологий*

Вопросы слияния и выравнивания онтологий стали ключевыми направлениями исследований и разработок в области онтологического инжиниринга в последние годы. Такая ситуация определяется естественным развитием технологий онтологического инжиниринга и потребностями практики. Так, создание объединенных онтологий на этапе проектирования новых онтологий может оказаться критическим фактором успеха при слиянии компаний или организаций в одну структуру или вызвана необходимостью получения расширенной онтологии с новой областью охвата и/или более высоким качеством существующей онтологии. Не менее важно и то, что в настоящее время онтологии становятся все более значимым ресурсом для семантического аннотирования веб-страниц в интересах Semantic Web, где разнородность и разноформатность информации требуют использования различных онтологий

даже в близких предметных областях. При этом разнообразие используемых онтологий должно естественным образом сосуществовать в рамках разных прикладных систем и их взаимодействия между собой. Все вышесказанное определяет важность слияния и выравнивания различных онтологий как в процессе их проектирования, так и на этапах их практического использования и появления новых методологий, методов и средств объединения и выравнивания (merging & alignment) онтологий, некоторые из которых рассматриваются ниже.

Среди многих работ в этой области можно выделить следующие методологии объединения и выравнивания онтологий:

- методология ONIONS;
- система PROMPT;
- метод FCA-Merge.

### **Методология ONIONS**

Методология ONIONS (ONtological Integration Of Naïve Sources) [Gangemi, et al., 1999] разработана для поддержки процессов концептуального анализа, онтологической интеграции и слияния онтологий. Она ориентирована на обеспечение аксиоматизации, спецификацию «прозрачной» семантики и необходимой онтологической глубины терминологии объединенной предметной области, для которой уже существуют частные онтологии.

Онтологический анализ и слияние терминологий в рамках методологии ONIONS выполняются в две фазы:

- реинжиниринг данных, по которым строится новая объединенная онтология;
- слияние онтологий, которые входят в новую онтологию .

Фаза реинжиниринга включает извлечение, форматирование, анализ и формализацию концептуально релевантных данных из источников. При этом извлечение и форматирование выполняется путем свертки релевантных данных после предварительной оценки типов их данных, а анализ состоит в выявлении подразумеваемых значений всех используемых терминов. Как результат, проанализированные данные новой онтологии формализуются на некотором логическом языке.

Слияние означает создание новой онтологии, имеющей такое же экстенциональное покрытие (объем), что и сливаемые онтологии, но разное количество и/или распределение свойств и аксиом (различное содержание). Если онтологии, которые нужно объединить, строились и/

или совершенствовались вне определенной базовой онтологии, трудоемкость их слияния достаточно велика, так как в этом случае в качестве стратегий слияния можно использовать лишь структурное распределение аксиом и «похожесть» наименований понятий и отношений.

Учитывая вышесказанное, методология ONIONS предполагает использование фундаментальных и ядерных онтологий, в свою очередь, построенных в рамках методологии ONIONS, что обеспечивает логическую интеграцию онтологий путем объединения их словарей и сравнения аксиоматизаций [Pisanelli, et al., 1997]. При этом предполагается наличие репозитория понятий предметной области, который будет использоваться для управления синонимией, категоризацией и полисемией, а собственно методология обеспечивает руководство по их оцениванию.

Стратегия слияния ONIONS использует приемы из анализа предметной области для выделения аксиом из «голых» таксономий. В ситуации с объединением «легковесных» онтологий и простых таксономий, связывающих множества понятий с одним и тем же родительским понятием, стратегия слияния очевидна. В общем случае методология ONIONS предлагает следующие приемы извлечения синонимов, обогащения таксономии и управления полисемией:

— *Извлечение синонимов.* Синонимия встречается, когда два понятия с разными именами имеют одну и ту же подразумеваемую модель и область значений. В таких случаях работают несколько эвристик для выделения кандидатов в синонимы: эвристики, основанные на подобию имен, эвристики лексической смежности (совместной встречаемости), эвристики разделения свойств на базе использования доступных лексических ресурсов. При этом потенциальные синонимы обычно включаются в множество лексических реализаций понятия.

— *Обогащение таксономии.* Любые два родственные понятия, полученные в результате логической интеграции двух онтологий, должны быть разделены явно, чтобы они не дали ложных синонимов. В таких случаях методология ONIONS требует, чтобы для любого множества родственных понятий из онтологий разнородных областей концептуальное различие между каждым из них было выведено и формализовано аксиомами, которые используют, если это возможно, понятия и отношения, уже существующие в библиотеке ONIONS.

— *Управление полисемией.* Полисемия появляется тогда, когда два понятия с перекрывающимися или отдельными моделями имеют одно и

то же имя или являются уже принятыми синонимами. На этом уровне наиболее интересным случаем является множественная категоризация (понятие  $C$  является синонимом понятия  $C1$  в  $L$ , при этом  $C$  относится к категории понятия  $D$  в  $L$ , а  $C1$  относится к категории понятия  $D1$  в  $L$ ). Этого не может быть, если синонимия в новой онтологии порождается на базе общих свойств определенных понятий. В таких случаях методология ONIONS предлагает схему аксиом строить заново или получить ее путем уточнения существующей аксиомы в базовой онтологии. Таким образом  $C$  и  $C1$  будут соотнесены с различными понятиями внутри схемы аксиом.

Разработка методологии ONIONS началась в 1993 г. в качестве альтернативы к методологии конструирования онтологий снизу-вверх. Позже методология ONIONS применялась при создании ядра медицинской онтологии ON9, интеграции клинических управленческих стандартов, создании онтологий по правовому регулированию и банковскому обслуживанию, слиянию терминологии в онтологии по рыбной ловле, а также «переиспользованию» WordNet для построения словарей понятий и соответствующих таксономий.

В последние годы в рамках методологии ONIONS было уточнено использование фундаментальных и ядерных онтологий за счет разрешения широкой аксиоматизации верхнего уровня, в результате чего была создана известная онтология DOLCE [DOLCE, 2010] (Descriptive Ontology for Linguistic and Cognitive Engineering), которая была первой в библиотеке базовых онтологий проекта WonderWeb и которую предполагалось применять для согласования «мнений» в среде интеллектуальных агентов, использующих разные терминологии.

### **Система PROMPT**

Изначально PROMPT [Noy, et al., 2000] – это алгоритм слияния и выравнивания онтологий, который был реализован как плагин к системе Protégé-2000. В процессе реализации системы PROMPT, пожалуй, впервые были разработаны инструментальные средства для полуавтоматического слияния и/или выравнивания онтологий. При этом некоторые задачи выполнялись автоматически, а в тех случаях, когда это было невозможно, система выступала как консультант пользователя при выполнении задач, где требовалось вмешательство человека-эксперта. В частности, автоматически определяются возможные противоречия (несовместности) в текущем состоянии онтологии, которые могут быть

результатом действий пользователя, и предлагаются пути исключения этих противоречий.

PROMPT основывается на достаточно общей модели знаний и принимает во внимание различные характеристики исходных онтологий для поиска противоречий и рекомендаций по их исключению. Основными эвристиками, работающими в системе PROMPT, являются:

- *Эвристика имен.* Если классы и слоты имеют похожие имена и одинаковый тип, они являются хорошими кандидатами на слияние.
- *Эвристика иерархии классов.* Если пользователь сливает два класса, а PROMPT уже обнаружил, что их суперклассы похожи, тогда для выполнения слияния имеется больше оснований.
- *Эвристика присоединения слота к классу.* Если два слота из различных онтологий присоединяются к сливаемому классу и их имена, фасеты и значения фасетов подобны, эти слоты являются кандидатами на слияние.
- *Эвристика фасет и значения фасет.* Если пользователь сливает два слота, их ограничения на область значений являются кандидатами на слияние.

Дополнительно, чтобы обеспечить помощь пользователю при слиянии и/или выравнивании онтологий, PROMPT идентифицирует противоречия, некоторые из которых перечислены ниже:

- конфликты имен (более чем один фрейм с одним и тем же именем);
- висячие ссылки (фрейм ссылается на несуществующий фрейм);
- избыточность в иерархии классов (более чем один путь от класса к его родителю);
- ограничения на значения слота, которые противоречат наследованию на классах.

Для идентификации противоречий, перечисленных выше, PROMPT опирается на графовую структуру онтологии, а чтобы ограничить пространство поиска противоречий; поддерживает перемещение от конкретной вершины графа к связанным с ней вершинам только на один или два уровня.

Пример экранной формы, иллюстрирующей работу плагина PROMPT в составе Protégé, показан на Рис. 14.



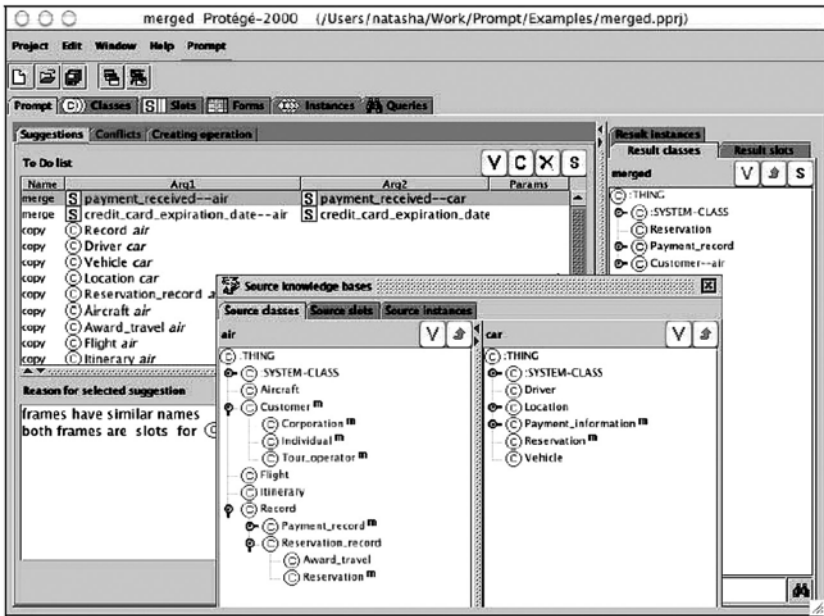


Рис. 14. Работа плагина PROMPT в составе Protégé

Расширением PROMPT является AnchorPROMPT [Noy, et al., 2001], где возможно более масштабное сравнение графовых структур объединяемых онтологий. Как и в случае PROMPT, AnchorPROMPT трактует онтологию как граф с классами в качестве вершин и слотами в качестве связей (дуг). На вход алгоритма подается множество указателей-якорей – пар связанных термов, определенных пользователем, а AnchorPROMPT автоматически выполняет лексическое их сопоставление и анализирует пути в подграфе, ограниченном указателями. В результате такого анализа определяется, какие классы появляются в подобных позициях на подобных путях и делается вывод о том, что, вероятно, эти два класса и представляют семантически подобные понятия.

К сожалению, задача слияния и выравнивания онтологий в полном объеме не решена до сих пор, а все инструментальные средства, ориентированные на ее решение, по существу, являются более или менее интеллектуальными помощниками человека-эксперта.

## Метод FCA-Merge

Метод слияния онтологий FCA-Merge использует подход снизу-вверх. Как утверждается в работе [Stumme, et al., 2001], здесь предлагается глобальное структурное описание самого процесса слияния. Идея этого метода состоит в том, что сначала из множества текстовых документов, релевантных предметным областям сливаемых онтологий, извлекаются экземпляры понятий. При этом используются методы обработки ЕЯ-текстов и, в частности, методы извлечения информации из текстов, обсуждавшиеся выше.

Для обработки извлеченных из текстов экземпляров понятий применяются математически корректные методы формального анализа понятий (Formal Concept Analysis), с помощью которых строится решетка понятий. Полученный результат исследуется разработчиком онтологий и транслируется в новую онтологию.

Общая схема слияния онтологий по методу FCA-Merge представлена на Рис. 15.

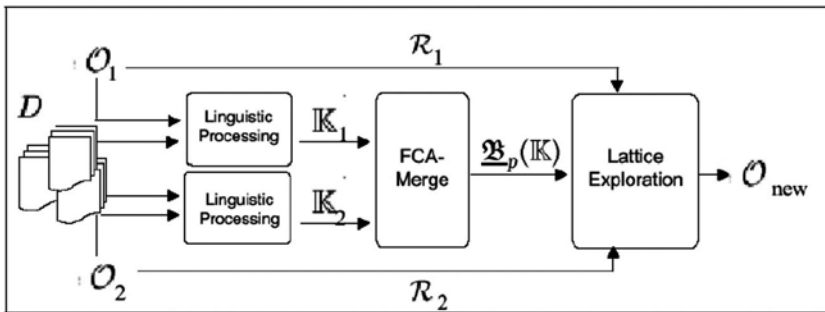


Рис. 15. Общая схема слияния онтологий по методу FCA-Merge

Как следует из представленной схемы, процесс слияния двух онтологий состоит из следующих этапов:

- Извлечение экземпляров и вычисление двух формальных контекстов  $K_1$  и  $K_2$ .
- Применение базового алгоритма FCA-Merge, который выводит общий контекст и вычисляет решетку понятий.
- Интерактивная генерация новой онтологии, основанная на использовании полученной на предыдущем этапе решетки понятий.

В рамках метода FCA-Merge первые два этапа выполняются автоматически, а последний требует участия человека-эксперта. Для поддержки последнего этапа используются графические средства редактирования онтологий.

Следует отметить, что для получения хороших результатов слияния онтологий по методу FCA-Merge входные данные должны удовлетворять нескольким условиям.

Первое из них состоит в том, что документы, на базе которых выполняется первый этап (извлечение экземпляров понятий), должны быть релевантны предметной области каждой из исходных онтологий. Второе – документы должны покрывать все понятия исходных онтологий, а понятия, которые не покрыты входным корпусом документов, должны интерпретироваться вручную уже после выполнения процедуры слияния, или множество документов должно быть расширено, а процедура выявления понятий должна быть повторена. И наконец, третье условие состоит в том, что документы должны разделять понятия достаточно хорошо, поскольку если два различных понятия всегда появляются в одних и тех же документах, алгоритм FCA-merge будет отображать их в одно и то же понятие в целевой онтологии (если это решение не блокируется инженером знаний). Если такая ситуация случается слишком часто, инженер знаний должен добавить в корпус дополнительные документы, которые бы способствовали разделению понятий.

Оценивая метод FCA-merge в целом, можно констатировать, что он является разумным компромиссом между математическими и экспертными методами слияния онтологий и, по-видимому, будет развиваться.

## ***2.5. Предварительные выводы***

Исследования и разработки в области онтологического моделирования и инжиниринга активно развиваются во всем мире. При этом трудно выделить лидеров по географическому признаку – примерно одинаковые научно-технические позиции занимают исследовательские коллективы разных стран. При этом результаты стран Старого Света в целом несколько отстают от результатов США, хотя отдельные европейские коллективы находятся на мировом уровне и активно участвуют в международных проектах по данной проблематике.

Анализ результатов, полученных в области проектирования пространств знаний, показывает, что очень многие идеи, выдвинутые и

реализованные в рамках инициативы (КА)<sup>2</sup>, проекта SHOЕ и других аналогичных проектов, сейчас «переоткрываются» снова.

Оценивая формализм представления онтологических знаний SHOЕ и поддержку процессов аннотирования веб-ресурсов в данном проекте, можно отметить, что это достаточно мощная система методов и средств, которая сложнее для пользователя, чем Ontobroker. В своей реализации проект SHOЕ «тяготел» к существующим в то время информационным технологиям обработки интернет-страниц больше, чем реализация в рамках инициативы (КА)<sup>2</sup>, и в этом смысле был более прагматичным. Вместе с тем мощность средств представления знаний, а следовательно, и возможности формирования пространств знаний здесь были ниже, чем у инициативы (КА)<sup>2</sup>. Тем не менее можно констатировать, что именно рассмотренные выше два проекта, по сути дела, и заложили научно-технические основы формирования пространств знаний.

Исследования и разработки в области информационных технологий тоже вносили и вносят свой вклад в формирование пространств знаний. При этом одним из самых успешных проектов в области Web 2.0 уже несколько лет остается база данных Freebase, реализованная компанией Metaweb Technologies, которая позиционируется как «открытая и разделяемая база данных знаний о мире или значительная по объему и кооперативно создаваемая и редактируемая база взаимосвязанных данных», где имеются достаточно мощные средства спецификации объектов и развитые веб-сервисы для пользователей.

Вместе с тем в рамках информационных технологий активно развиваются и средства «внедрения» базисной семантики непосредственно в HTML-страницы на уровне микроформатов, которые пока не столь выразительны, как другие средства семантизации контента (например, RDF или OWL), но очень компактны и используют возможности XHTML для добавления семантики к страницам Интернета. В настоящее время Facebook, Yahoo! Local, Google Social Graphs и другие проекты разных компаний используют микроформаты для аннотирования событий на своих страницах.

Как показано в данном разделе, онтологическое моделирование и инжиниринг тесно связаны с исследованиями в области Semantic Technologies и Semantic Web. И, оценивая ситуацию в целом, можно констатировать, что уже создан базис стандартизации в виде XML-, RDF(S)-, OWL- и SPARQL-спецификаций, на которые могут опираться работы по семантизации контента и его использованию в различных приложениях.

Наряду с этим, поскольку, как следует из материалов данного раздела, создание онтологий остается сложной и наукоемкой задачей, активно развиваются методологии и методы построения онтологий «с нуля» (from scratch), методы реинжиниринга онтологий (ontology reengineering) и коллективной разработки онтологий (collaborative ontology development), а также методы и методологии объединения и выравнивания онтологий (ontology merging & alignment). При этом на «передний план» все больше выходят методы и средства коллективного онтологического инжиниринга.

### **3. Базисные онтологические теории и модели**

#### ***3.1. Классификации онтологий***

Онтологический инжиниринг имеет уже достаточную историю исследований и разработок [Ushold, Gruninger, 1996; Guarino, 1998; Gavrilova, 2007; Гаврилова и др., 2001; Добров и др., 2006; Хорошевский, 2008; Гаврилова, 2009; Хорошевский, 2009], которая уже привела к построению широкого спектра различных онтологических моделей, требующих определенной классификации. В работе [Oberle, 2006] рассматривается одна из таких классификаций, пространство признаков которой представлено на Рис. 16.

В рамках другой классификации [Добров и др., 2006] онтологии различаются:

- по степени формальности;
- наполнению, содержанию;
- цели создания.

Существуют и другие классификации онтологических моделей и создаваемых на их базе онтологий.

Так, например, спектр онтологий по степени формальности спецификации [Oberle, 2006] представлен на Рис. 17, где каждая точка соответствует наличию некоторых ключевых структур, отличающих ее от других точек на спектре, а косая черта условно отделяет онтологии от других ресурсов, имеющих, в той или иной мере, онтологический характер.

В рамках классификации онтологий по цели создания часто выделяют четыре уровня: онтологии представления, онтологии верхнего



**Рис. 16.** Классификация онтологий по цели, выразительности и уровню детализации

уровня, онтологии предметных областей и прикладные онтологии. При этом целью создания онтологий представления является описание области представления знаний и разработка языков для спецификации более низких уровней онтологий. Примером онтологии этого уровня является описание понятий языка OWL средствами RDF/RDFS.

Онтологии верхнего уровня, по определению IEEE Standard Upper Ontology. SUO WG [SUO WG, 2003], – суть онтологии понятий, которые являются метапонятиями (базовыми, абстрактными, философскими) и, следовательно, достаточно общими для того, чтобы их можно было использовать в широком спектре предметных областей. В настоящее время уже существует несколько серьезных проектов, направленных на создание таких онтологий, например, Cyc, DOLCE, SUMO, Sowa's Ontology и других [Lenat, et al., 1995; Masolo, et al., 2003; Cohen, et al., 1999; Sowa, 1995], но в целом попытки создать онтологию верхнего уровня для всех предметных областей, которая к тому же признавалась бы всеми специалистами, работающими в области онтологического инжиниринга, пока не привели к ожидаемым результатам.

Назначение онтологий предметных областей схоже с назначением онтологий верхнего уровня, но область их охвата ограничена опреде-

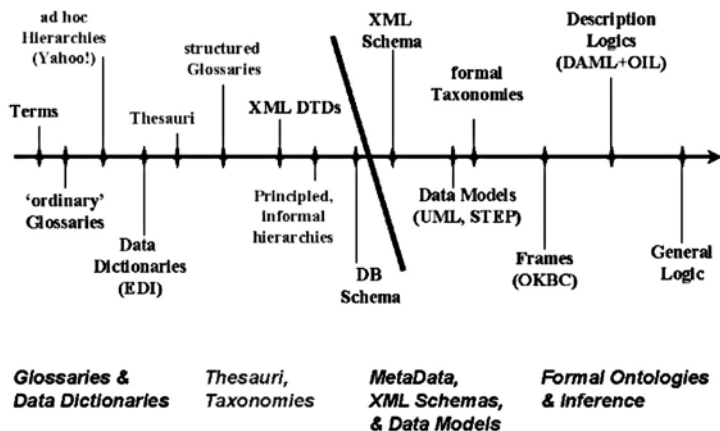


Рис. 17. Онтологии и «онтологии»

ленной предметной областью (медицина, бизнес, культура и т.п.), а прикладные онтологии создаются для описания концептуальных моделей конкретных задач и/или приложений.

Классификация онтологий по содержанию похожа на предыдущую с тем отличием, что упор здесь делается на реальное содержимое онтологии, а не на абстрактную цель, преследуемую ее авторами. В рамках этой классификации онтологии верхнего уровня описывают общее знание о моделируемом мире, формируя общую для онтологий нижних уровней систему понятий. Как правило, в таких онтологиях выделяют онтологии времени, пространства, событий и некоторые другие.

По цели создания, как показано на Рис. 16, онтологии делятся на справочные и онтологии задач.

И наконец, в силу своего предназначения, онтологии должны быть разделяемыми (shared) ресурсами (содержимое онтологии одновременно используется несколькими лицами, группами или сообществами), хотя это в большинстве случаев относится к онтологиям верхнего уровня, которые к тому же часто содержат знания здравого смысла (common sense).

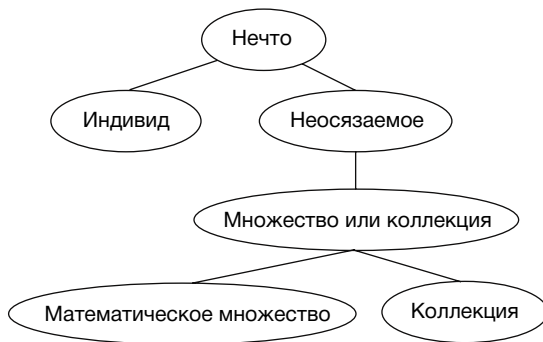
Учитывая вышесказанное, можно констатировать, что в проекте, который послужил основой для данной серии препринтов, наибольший интерес представляют (в соответствии с классификацией, приведенной

на Рис. 16) онтологии всех уровней детализации, которые покрывают весь спектр целей и при этом являются машинно-ориентированными.

### 3.2. *Онтологическая модель СУС*

Онтологическая модель СУС [Lenat, et al., 1995] – одна из первых моделей этого класса, которая была разработана в рамках проекта Сус под руководством одного из самых известных в мире специалистов в области инженерии знаний Д. Лената (D. Lenat), и продолжает разрабатываться в настоящее время в проекте OpenСус как онтология в области общего знания (<http://www.opencyc.com>).

Ключевым понятием в данной онтологии является понятие коллекции, которая может содержать подколлекции и экземпляры, которые, в свою очередь, могут быть любыми терминами онтологии. В вершине иерархии коллекций находится универсальная коллекция с именем «Нечто», которая, по определению, содержит все, что существует в рамках описываемой области. При этом коллекция «Нечто» является как экземпляром, так и подколлекцией самой себя, но не подколлекцией какой-либо другой коллекции. На первом уровне иерархии коллекция «Нечто» делится на 116 подколлекций (фрагмент иерархии коллекций верхних уровней представлен на Рис. 18).



**Рис. 18.** Фрагмент иерархии коллекций верхних уровней онтологии СУС

Коллекция «Индивид» содержит сущности, которые не являются ни множествами, ни коллекциями. При этом индивиды могут быть абстрактными или конкретными, включать физические объекты, события,



отношения, числа, группы, состоять из частей, иметь сложную структуру, но ни один экземпляр этой коллекции не может иметь элементов или подмножеств. Так, например, некоторая фирма (1), группа, содержащая всех ее работников (2), коллекция всех работников фирмы (3) и множество всех работников фирмы (4) – четыре разных понятия и только первые два из них – индивиды.

Структурно база знаний Сус состоит из констант (терминов) и правил (формул), оперирующих этими константами. Правила делятся на два вида: выводимые утверждения и аксиомы, понимаемые как утверждения, которые были явно и вручную введены в базу знаний экспертами, а не появились там (или могут появиться) в результате работы машины вывода. Все утверждения или формулы в базе знаний Сус фиксируются на специальном языке представления знаний СусL, эквивалентном по мощности исчислению предикатов первого порядка.

Как представляется, в настоящее время онтология Сус и ее часть OpenСус, доступная для некоммерческого использования, в прикладных проектах используются слабо.

### ***3.3. Онтология верхнего уровня DOLCE***

Онтология верхнего уровня DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) – первая из онтологий в библиотеке проекта WonderWeb (<http://www.loa-cnr.it/DOLCE.html>), которая была разработана для SemanticWeb с целью согласования «мнений» между интеллектуальными агентами, использующими разную терминологию.

С одной стороны, онтология DOLCE не претендует на звание универсальной, стандартной или общей, поскольку основная цель ее разработчиков состояла в том, чтобы создать модель, помогающую при сравнении и объяснении связей с другими онтологиями библиотеки WFOL (базовой библиотеки онтологий WonderWeb), а также для выявления скрытых допущений, лежащих в основе существующих онтологий и лингвистических ресурсов, таких как WordNet.

С другой стороны, в основу проектирования онтологии DOLCE положено фундаментальное философское разделение всех сущностей на универсалии (сущности потенциально или реально имеющие экземпляры) и индивиды (или частности), которые не имеют и не могут иметь экземпляров, и при этом DOLCE – онтология индивидов, в том смысле, что область описания ограничена только ими. Заметим, что при таком

подходе понятие «Время» тоже рассматривается как индивид, не имеющий экземпляров. Еще одна особенность онтологии DOLCE, заимствованная разработчиками из философии, – явное разделение на «Постоянные» и «Происходящие» сущности, различие между которыми состоит в том, что первые имеются в наличии целиком и неизменно в некотором фиксированном промежутке времени (например, компания в течение времени своего существования), а вторые разворачиваются во времени, и в каждый момент в некотором временном интервале они могут быть различными (например, ураган), но при этом сохранять свою идентичность. Такое разделение на «объект» и «процесс» весьма условно и привело, например, к тому, что в онтологии определены два типа отношения «ЧАСТЬ – ЦЕЛОЕ» (первое никак не зависит от времени, второе имеет временной индекс, определяющий в каких временных рамках отношение действует).

Фрагмент онтологии DOLCE, который показывает, что ее можно отнести к типу Upper Ontology, представлен на Рис. 19.



**Рис. 19.** Фрагмент онтологии DOLCE

Для представления своей онтологии авторы DOLCE избрали более гибкий, чем в проекте Сус, подход. Здесь онтология сначала фиксируется на бумаге с использованием исчисления предикатов первого порядка, затем для нее описывается та часть аксиом, которая может быть представлена на языке OWL, а оставшиеся аксиомы, выраженные на языке KIF [KIF, 1998], добавляются к OWL-описаниям в виде комментариев. Таким образом, достигается выразительность уровня KIF и совместимость с OWL, хотя приложения, не имеющие информации о действительной структуре OWL-документа, не смогут получить доступ к «закомментированным» знаниям.

Следует отметить, что в библиотеке онтологий WonderWeb в настоящее время имеется спектр онтологий разного уровня для разных предметных областей. И в этом смысле онтология DOLCE жива и используется на практике.

### ***3.4. Онтология верхнего уровня SUMO***

SUMO (Suggested Upper Merged Ontology) – онтология верхнего уровня, разработанная в рамках проекта рабочей группы IEEE SUO (IEEE Standard Upper Ontology Working Group) и Teknowledge SUO WG, 2003]. При этом SUMO претендует на статус стандарта для онтологий верхнего уровня.

Онтология SUMO содержит наиболее общие и самые абстрактные концепты, имеет исчерпывающую иерархию фундаментальных понятий (около тысячи) и впечатляющий набор аксиом (примерно четыре тысячи), определяющих эти понятия. Назначение SUMO – содействовать улучшению интероперабельности данных, извлечению и поиску информации, автоматическому выводу (доказательствам), обработке естественного языка и др. Онтология охватывает следующие области знания: общие виды процессов и объектов, абстракции (теория множеств, атрибуты, отношения), числа и единицы измерения, временные понятия, части и целое, агенты и намерения.

SUMO является «канонической» онтологией верхнего уровня, поскольку содержит обозримое число концептов и аксиом, имеет ясную иерархию классов, легко расширяется, является итогом объединения различных общедоступных онтологий верхнего уровня (в том числе онтологии Джона Соуи (J. Sowa's ontology), о которой речь пойдет ниже. К преимуществам SUMO можно отнести и возможность трансляции

описания онтологии на любой из основных языков представления знаний, наличие онтологии среднего уровня (MLO), «бесшовно» интегрированной с SUMO, несколько десятков примеров практического применения, а также связь с WordNet [WordNet, 2011] – наиболее известным на данный момент тезаурусом, в котором содержится около 150 тыс. слов повседневного английского языка.

Иерархия классов в SUMO, фрагмент которой представлен на Рис. 20, менее «запутана», чем в Сус, и более удобна для практического применения, чем в DOLCE.



**Рис. 20.** Фрагмент иерархии классов онтологии SUMO

Основными концептами SUMO, как и во многих других онтологиях верхнего уровня, являются «Сущность» и ее категории – «Физический» и «Абстрактный». Первая категория включает все, что имеет положение в пространстве-времени, а вторая – все остальное (а точнее, только то, что существует в воображении). Категория «Физический» делится на «Объект» и «Процесс», что соответствует подходу, реализованному в DOLCE. Непосредственно под концептом «Объект» находятся два непересекающихся понятия – «СвязныйОбъект» и «Коллекция», первое из которых обозначает любой объект, все части которого непосредственно или косвенно связаны друг с другом. Ниже, на диалекте языка KIF [SUO-KIF, 2003], специфицированы аксиомы, определяющие различие между концептами «НепрерывныйОбъект» и «ДискретныйОбъект».

A1.

(=>

(and

(subclass-of ?OBJECTTYPE НепрерывныйОбъект)

(instance-of ?OBJECT ?OBJECTTYPE)

(part-of ?PART ?OBJECT)

)

(instance-of ?PART ?OBJECTTYPE)

)

A2.

(equal ДискретныйОбъект (ComplementFn НепрерывныйОбъект))

Аксиома A1 формализует утверждение «Если (PART) является частью объекта (OBJECT), являющегося, в свою очередь, экземпляром некоторого подкласса (OBJECTTYPE) класса НепрерывныйОбъект, то эта часть (PART), так же как и OBJECT, является экземпляром класса OBJECTTYPE», а аксиома A2 постулирует тот факт, что классы ДискретныйОбъект и НепрерывныйОбъект являются взаимодополняющими.

Онтология SUMO активно используется в разных прикладных проектах, где применяются онтологические модели.

### ***3.5. Онтология Джона Соуы***

Онтология Дж. Соуы, предложенная в работе [Sowa, 1995], определяет базовые онтологические категории логики, лингвистики, философии и искусственного интеллекта.

Интересным подходом к онтологиям верхнего уровня в данном случае является то, что для сохранения «открытости» такая онтология, по мнению Соуы, должна быть основана не на фиксированной иерархии концептов, а на каркасе, описывающем различия, по которому иерархия генерируется автоматически. Ниже, в Табл. 1, приведены онтологические категории Дж. Соуы верхнего уровня.

Кроме категорий, приведенных в Табл. 1, в онтологии Дж. Соуы есть еще два понятия: «Сущность», которое не определяет никаких отличительных признаков или различий и является надтипом для всех других концептов, и понятие «Абсурдный» – тип, наследующий все возможные, в том числе противоречащие различия, у которого нет ни одного экземпляра. Кроме того, в онтологии Дж. Соуы проводится различие

Таблица 1. Онтологические категории Дж. Совы верхнего уровня

	Физический		Абстрактный	
	Континуальный	Происходящий	Континуальный	Происходящий
Независимый	<i>Объект</i>	<i>Процесс</i>	<i>Схема</i>	<i>Скрипт</i>
Относительный	<i>Слияние</i>	<i>Участие</i>	<i>Описание</i>	<i>История</i>
Опосредованный	<i>Структура</i>	<i>Ситуация</i>	<i>Причина</i>	<i>Цель</i>

между абстрактным и физическим (именно в таком виде оно заимствовано разработчиками SUMO), а отдельно выделяются категории независимости, относительности и опосредованности.

«Независимые» сущности не нуждаются в существовании каких-либо связей с другими сущностями, но любая «Относительная» сущность обязательно имеет хотя бы одну связь с некоторой другой сущностью. Для существования «Опосредованной» сущности необходимо наличие некоторого отношения, связывающего какие-то другие сущности, имеющие отношение также и к первой (например, партнерство). Онтология Джона Совы описывает роли и отношения, агентов, процессы, и т.д. Онтологический «бриллиант» таксономии объектов онтологии Дж. Совы представлен на Рис. 21.

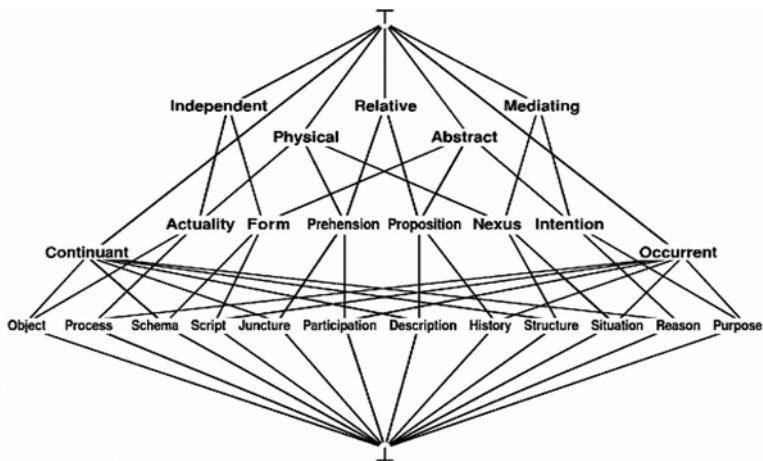


Рис. 21. Онтологический «бриллиант» Дж. Совы

Заметим, что онтология Дж. Соуи достаточно часто используется в проектах в качестве Upper Ontology.

### 3.6. Теория активностей и онтологические модели

Теория активностей является результатом исследований русской школы психологов [Vygotsky, 1978; Leont'ev, 1978], но наиболее активные работы по развитию и использованию этой теории ведутся в Скандинавии [Engestrom, 1987].

В рамках этой теории предполагается, что контекст определяется анализом действий, а ключевыми сущностями базового уровня здесь являются предмет (subject), объект (object) и сообщество (community), а понятия средств (tools), правил (rules) и разделения труда (divisions of labor) фиксируют другие артефакты, используемые активностями для фиксации контекста. При этом в активности предмет изменяет объект для генерации результата с помощью средств в контексте определенного сообщества с его правилами и разделением труда (Рис. 22).

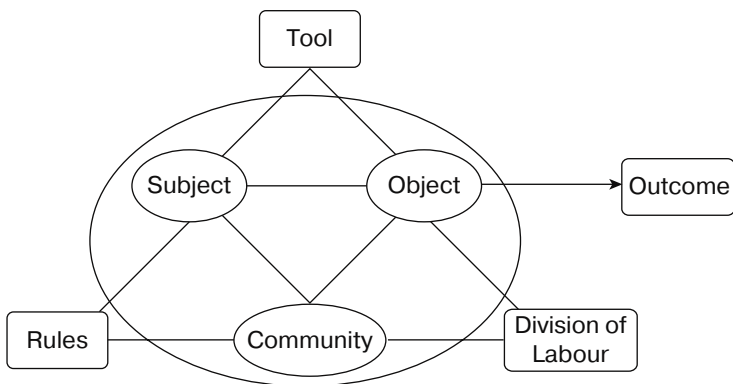


Рис. 22. Общая структура активности

В теории активностей существует разделение сущностей на людей (people) и вещи (things) [Nardi, 1996]. При этом люди являются не узлами или агентами системы, а ее предметами, которые (и только они) имеют мотивы и сознание, а собственно в теории определяются факторы контекста и взаимодействие между людьми и окружением. Интересно в данном случае то, что теория активностей это не только теория, но,

кроме того, это подход к спецификации моделей на основе использования шаблонов.

Учитывая вышесказанное, теория активностей часто используется в настоящее время как основа для генерации бизнес-онтологий с применением структур, представленных на Рис. 22, для спецификации конкретных активностей. Для примера, на Рис. 23 показана структура активности по сбору и авторизации знаний [Collins et al., 2002].

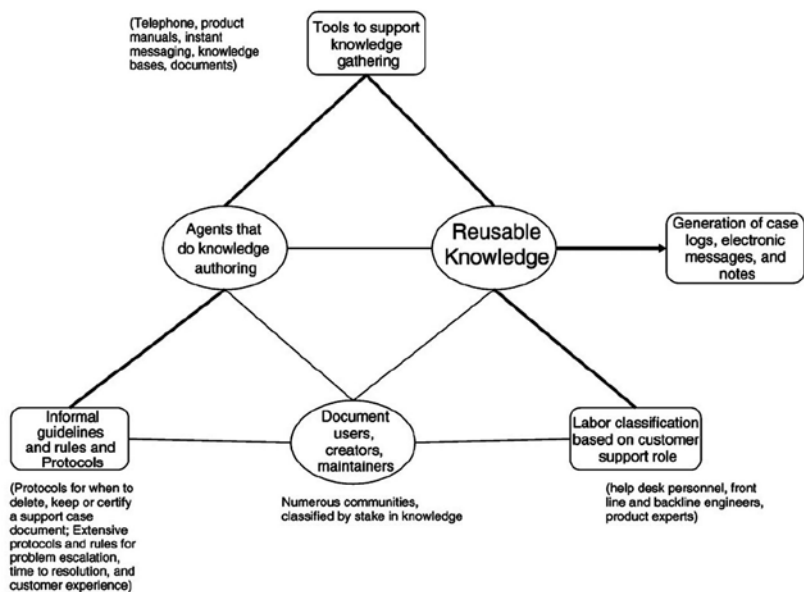


Рис. 23. Структура активности по сбору и авторизации знаний

Теорию активностей предполагается использовать и в рамках создания системы онтологических моделей предприятий и отраслей современной России.

### 3.7. Предварительные выводы

Как показано в данном разделе, онтологии верхнего уровня описывают самые общие знания об окружающем мире и они во многом похожи. Так, во всех онтологиях проводится разделение сущностей на абстрактные (такие сущности не могут занимать положения ни в пространстве,



ни во времени) и реально существующие (материальные, осязаемые). Во всех онтологиях так или иначе присутствует деление на постоянные и темпоральные сущности, деление на объект и процесс, и т.д.

В разделе представлены существующие классификации онтологий и на этой базе рассмотрены наиболее известные из онтологий верхнего уровня (Upper Ontology), среди которых особое внимание уделено онтологиям CYC, DOLCE, SUMO и онтологии Дж. Совы. Представлены новые теории создания онтологических моделей, в частности, теория активностей и ее связь с онтологическими моделями.

Показано, что даже на верхних онтологических уровнях наряду с общностью наблюдаются и существенные различия. Так, в онтологии SUMO первично разделение на абстрактные и материальные сущности, а разделение на постоянные и временные – вторично; в DOLCE на верхнем уровне производится разделение на постоянные, временные, абстрактные и качественные сущности, а в онтологии Дж. Совы иерархии сущностей в явном виде нет, поскольку в ней представлены только категории, по которым понятия разделяются или группируются. Однако, что очевидно и важно, онтологии верхнего уровня являются составляющей практически всех современных онтологических систем.

В разделе также представлена теория активностей, которая является результатом исследований русской школы психологов (Леонтьев, Выготский), хотя наиболее активные работы по развитию и использованию этой теории сейчас ведутся в Скандинавии. С точки зрения новых моделей предприятий и отраслей экономики интересно, что контекст в этой теории определяется анализом действий, а ключевыми сущностями базового уровня здесь являются предметы, объекты и сообщество. При этом понятия средств, правил и разделения труда фиксируют другие артефакты, используемые активностями для фиксации контекста. Учитывая вышесказанное, теория активностей часто используется в настоящее время как основа для генерации бизнес-онтологий.

## 4. Заключение

В настоящей работе представлены результаты анализа мирового опыта исследований и разработок в области онтологического моделирования. При этом основное внимание в данном препринте уделено изло-

жению результатов исследований по моделям и методам представления знаний, которые были получены в рамках исследований по искусственному интеллекту и в конечном счете стали базисом для развертывания работ по формированию пространств знаний и Semantic Web, обсуждению понятий, методов и средств все еще новой для России области онтологического инжиниринга и моделирования, а также описанию онтологий верхнего уровня и принципов теории активностей.

Как показывают материалы настоящей работы, исследования и разработки в области онтологического моделирования и инжиниринга активно развиваются во всем мире. При этом трудно выделить лидеров по географическому признаку – примерно одинаковые научно-технические позиции занимают исследовательские коллективы разных стран. Результаты европейских стран по общим вопросам онтологического инжиниринга в целом отстают от результатов США, хотя отдельные исследовательские коллективы находятся на мировом уровне и активно участвуют в международных проектах по данной проблематике.

Какие конкретные выводы и рекомендации для проекта, который послужил стимулом для подготовки настоящего и последующих препринтов, вытекают из проведенного выше анализа?

Прежде всего, убеждение в том, что в рамках проекта должна быть спроектирована сбалансированная система экономических онтологий, в которой:

- В качестве теоретического базиса, по-видимому, целесообразно использовать модели теории активностей, предложенной российскими учеными Л. Выготским и А. Леонтьевым, которые в настоящее время активно используются финской и итальянской школами онтологического инжиниринга, а также исследовательскими коллективами из ведущих университетов США и стран Юго-Восточной Азии.
- На верхнем уровне в проектируемой системе взаимосвязанных онтологий предпочтительно опираться на общепризнанные Upper Ontology уровня SUMO, DOLCE и/или онтологии Дж. Совы, выбирая из них те подмножества понятий и отношений между ними, которые отвечают целям и задачам проекта, и расширяя создаваемую в проекте онтологию верхнего уровня новыми концептами и отношениями, которые обычно не рассматриваются в существующих Upper Ontology.
- Уровень базовых онтологий, по-видимому, целесообразно формировать из хорошо проработанных онтологий данного уровня (например, из TOVE, Enterprise Ontology, REA и некоторых других), кото-

рые рассматриваются во втором препринте данной серии, дополняя и расширяя их онтологиями времени, пространства и некоторыми другими (например, онтологиями информационного пространства и моделей задач).

— Предметные онтологии в проекте, как представляется, будут в значительной мере разрабатываться заново, поскольку в настоящее время в мире таких онтологий нет, а исследования и разработки в области онтологий микроэкономики практически не ведутся.

И наконец, последнее, на что представляется важным обратить внимание исследователей и разработчиков в данной области. В настоящее время при проектировании сложных информационно-аналитических систем следует обращать самое серьезное внимание на вопросы стыковки проектируемой предметной системы онтологических моделей с лингвистическими моделями, обеспечивающими формирование и сопровождение пространств знаний.

## Литература

- [Гаврилова и др., 2001] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем: учебник. СПб.: Питер, 2001.
- [Гаврилова, 2009] Гаврилова Т.А. Инженерия знаний // Инновационное развитие: экономика, интеллектуальные ресурсы, управление знаниями / под ред. Б.З. Мильнера. М.: ИНФРА-М, 2009. С. 480–500.
- [Добров и др., 2006] Добров Б.В., Иванов В.В., Лукашевич Н.В. и др. Онтологии и тезаурусы: учебно-методическое пособие // Казань: Казанский государственный университет, 2006.
- [Загорулько, 2008] Загорулько Ю.А. Методы и методологии разработки, сопровождения и реинжиниринга онтологий: труды симпозиума «Онтологическое моделирование». М.: ИПИ РАН, 2008.
- [Клайн, 2006] Клайн К. SQL. Справочник. М.: Кудиц-Образ, 2006.
- [Клоксин др., 1987] Клоксин У., Меллиш К. Программирование на языке ПРОЛОГ. М.: Мир, 1987.
- [Овдей и др., 2004] Овдей О.М., Проскудина Г.Ю. Обзор инструментов инженерии онтологий // Электронные библиотеки. 2004. Вып. 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part4/op>.

- [Хорошевский, 2008] Хорошевский В.Ф. Онтологические модели и Semantic Web: откуда и куда мы идем?: сборник трудов симпозиума «Онтологическое моделирование». М.: ИПИ РАН, 2008.
- [Хорошевский, 2009] Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic Web (Ч. 2) // Искусственный интеллект и принятие решений. 2009. № 4. С. 15–36.
- [Allsopp, 2006] Allsopp J. The Big Picture on Microformats, August 28, 2006, [http://www.digital-web.com/articles/the\\_big\\_picture\\_on\\_microformats/](http://www.digital-web.com/articles/the_big_picture_on_microformats/).
- [Altova, 2010] Internet site of Altova company, <http://www.altova.com/>
- [Benjamins, et al., 1999] Benjamins R., Decker S., Fensel D., Gomez-Perez A. “(KA)<sup>2</sup>: Building Ontologies for the Internet”: A Mid Term Report. International Journal of Human Computer Studies (IJHCS). 51(3). September 1999.
- [Benjamins, et al., 2002] Benjamins V.R., Contreras J., Corcho O. and Gomez-Perez A., Six Challenges for the Semantic Web, [http://www.cs.man.ac.uk/~ocorcho/documents/KRR2002WS\\_BenjaminsEtAl.pdf](http://www.cs.man.ac.uk/~ocorcho/documents/KRR2002WS_BenjaminsEtAl.pdf).
- [Berners-Lee, 2003] Berners-Lee T. The Semantic Web and Research Challenges, <http://www.w3.org/2003/Talks/01-sweb-tbl/slide1-0.html>.
- [Berners-Lee, et al., 2001] Berners-Lee T., Hendler J., Lassila O. (2001). The Semantic Web. Scientific American, 284(5), 34–43.
- [Cearley, et al., 2007] Cearley D.W., Andrews W., Gall N.: Finding and Exploiting Value in Semantic Technologies on the Web. 9 May 2007, ID №: G00148725. Gartner, Inc., 2007.
- [Cohen, et al., 1999] Cohen P., Chaudhri V., Pease A., and Schrag R. (1999), Does Prior Knowledge Facilitate the Development of Knowledge Based Systems, In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-1999). Menlo Park, Calif.: AAAI Press. <http://home.earthlink.net/~adampease/professional/cohen-aaai99.ps>
- [Collins et al., 2002] Collins P., Shukla S. and Redmiles D. “Activity Theory and System Design: A View from the Trenches.” Computer Supported Cooperative Work (CSCW) – Special Issue on Activity Theory and the Practice of Design 11(1–2): (2002), pp. 55–80.
- [Cycorp, 2010] Internet site of Cycorp, Inc. <http://cyc.com/cyc>.
- [DAML+OIL, 2001] DAML+OIL Reference Description. Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein // W3C Note 18 December 2001.

- [Davis, 2006] Davis M. Semantic Wave 2006. Executive Guide to Billion Dollar Markets. A Project10X Special Report. January 2006.
- [Decker, et al., 1999] Decker S., Erdmann M., Fensel D., Studer R. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In R. Meersman et al. (eds.): Semantic Issues in Multimedia Systems. Proceedings of DS-8. Kluwer Academic Publisher, Boston, 1999.
- [DOLCE, 2010] Internet site of Laboratory for Applied Ontology, <http://www.loa-cnr.it/DOLCE.html>.
- [Engestrom, 1987] Engestrom Y. Learning by expanding: an activity theoretical approach to developmental research. Orienta-Konsultit, Helsinki, Finland, 1987.
- [Euzenat, 1996] Euzenat J. Corporate memory through cooperative creation of knowledge bases and hyper-documents, In: Proc. of 10th Workshop on Knowledge Acquisition (KAW), Banff (CA), pp. 1–18, 1996.
- [Evet, et al., 1994] Evett M.P., Hendler J.A. and Spector L. Parallel Knowledge Representation on the Connection Machine // Journal of Parallel and Distributed Computing, 22: 168–184, 1994.
- [Fensel et al. 1999] Fensel D., Angele J., Decker S., Erdmann M., Schnurr H., Staab S., Studer R., Witt A. (1999). On2broker: Semantic-Based Access to Information Sources at the WWW. In Proceedings of the World Conference on the WWW and Internet (WebNet 99), Honolulu, Hawaii, USA, October 25–30, 1999.
- [Fernandez, et al., 2006] Fernandez M., Gomez-Perez A., Juristo N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI Technical Report SS-97-06, 2006.
- [FIPA, 2010] FIPA – Foundation for Intelligent Physical Agents, Geneva, Switzerland. <http://www.fipa.org>.
- [Freebase, 2010] Internet site Freebase, [http://wiki.freebase.com/wiki/Main\\_Page](http://wiki.freebase.com/wiki/Main_Page).
- [Gangemi, et al., 1999] Gangemi A., Pisanelli D.M., Steve G. An overview of the ONIONS project: Applying ontologies to the integration of medical terminologies. Data Knowl Eng 31: 183–220, 1999.
- [Gavrilova, 2007] Gavrilova T. Ontological Engineering for Practical Knowledge Work // Lecture Notes in Artificial Intelligence 4693, Proc. of 11th Int. Conf. Knowledge-Based Intelligent Information and Engineering Systems (KES 2007), Italy, Vietri sul Mare, Springer, 2007. pp. 1154–1162.

- [Gomez-Perez, et al., 1999] Gomez-Perez A., Rojas M.D. Ontological Reengineering and Reuse // In: Proc. of 11th European Workshop on Knowledge Acquisition, Modelling and Management, Springer Verlag, 1999.
- [Gomez-Perez, et al., 2009] Gomez-Perez A., Suarez-Figueroa M.C. Scenarios for Building Ontology Networks within the NeOn Methodology. Proceedings of the Fifth International Conference on Knowledge Capture (K-CAP 2009), 2009.
- [Guarino, 1998] Guarino N. Some Ontological Principles for Designing Upper Level Lexical Resources // Proceedings of First International Conference on Language Resources and Evaluation. Granada, Spain, 1998.
- [Heflin, et al., 1998] Heflin J., Hendler J., Luke S. Reading Between the Lines: Using SHOE to Discover Implicit Knowledge from the Web. In: AAAI-98 Workshop on AI and Information Integration, 1998.
- [Heflin, et al., 2000] Heflin J., Hendler J. Searching the Web with SHOE, In AAAI-2000 Workshop on AI for Web Search, 2000.
- [KIF, 1998] Knowledge Interchange Format draft proposed American National Standard (dpANS) NCITS.T2/98-004, 1998.
- [KIF, 2010] Knowledge Interchange Format. <http://logic.stanford.edu/kif/dpans.html>.
- [Kifer, et al., 1995] Kifer M., Lausen G., Wu J. Logical Foundations of Object Oriented and Frame Based Languages. Journal of ACM 1995. Vol. 42, pp. 741–843].
- [Lamping, et al., 1995] Lamping L., Rao R., Pirolli P. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 1995.
- [Larman, et al., 2003] Larman C., Basili V.R. Iterative and Incremental Development: A Brief History, Computer. Vol. 36. No. 6, pp. 47–56, June, 2003.
- [Lenat, et al., 1989] Lenat D., Guha R.V., Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project, Addison-Wesley, 1989.
- [Lenat, et al., 1995] Lenat D., Miller G., Yokoi T. CYC, WordNet, and EDR: critiques and responses. Communications of the ACM. Vol. 38, Issue 11 (November 1995), pp. 45–48.
- [Leont'ev, 1978] Leont'ev A.N. Activity, Consciousness, Personality. Englewood Cliffs, NJ, Prentice Hall, 1978.

- [Maedche, et al., 2001] Maedche A., Staab S. “Learning Ontologies for the Semantic Web”, Semantic Web WorkShop 2001, Hongkong, China, 2001.
- [Masolo, et al., 2003] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari, Luc Schneider, The WonderWeb Library of Foundational Ontologies. Preliminary Report, IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web, ISTC-CNR, 2003.
- [Nardi, 1996]. Nardi B. Ed. Context and Consciousness: Activity Theory and Human-Computer Interaction. Cambridge, MA, MIT Press, 1996.
- [Noy, et al., 2000] Noy N., Musen M. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment // Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000). Austin, Texas, 2000.
- [Noy, et al., 2001] Noy N.F., Musen M.A.: Anchor-Prompt: Using Non-local Context for Semantic Matching. In Proceedings of the Workshop on Ontologies and Information Sharing at IJCAI-2001, 2001.
- [Oberle, 2006] Oberle D. Semantic Management of Middleware, Springer, Berlin, 2006.
- [Ontolingua, 2010] Ontolingua Homepage, <http://ksl.stanford.edu/software/ontolingua/>.
- [OWL, 2004] OWL Web Ontology Language Guide. Smith K.M., Welty Ch., McGuinness D.L. (eds.) // W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.
- [Pisanelli, et al., 1997] Pisanelli D.M., Gangemi A., Steve G. „WWW-available Conceptual Integration of Medical Terminologies: the ONIONS Experience“, In: Proc. of AMIA97, Philadelphia, Hanley&Belfus, 1997.
- [Protégé, 2010] Protege Homepage, <http://protege.stanford.edu/>.
- [SHOE, 1999] The SHOE Knowledge Annotator, <http://www.cs.umd.edu/projects/plus/SHOE> (1999).
- [Sowa, 1995] Sowa J.F. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Boston, MA: PWS Publishing Co., 1995.
- [SPARQL, 2008] SPARQL Query Language for RDF // W3C Recommendation 15 January 2008.
- [Stumme, et al., 2001] Stumme G., Maedche A. FCA-MERGE: Bottom-Up Merging of Ontologies // In Proc of IJCAI-2001, 2001.

- [SUO WG, 2003] IEEE Standard Upper Ontology. SUO WG, <http://suo.ieee.org/2003>.
- [SUO-KIF, 2003] IEEE Standard Upper Ontology (SUO) Study Group Knowledge Interchange Format. <http://suo.ieee.org/SUO/KIF/suo-kif.html>.
- [TOVE, 2010] TOVE Ontology Project. <http://www.eil.utoronto.ca/enterprise-modelling/tove/>.
- [Tudorache, et al., 2008] Tudorache T., Noy N.F., Tu S.W., Musen M.A. Supporting collaborative ontology development in Protege // In Proc. Of Seventh International Semantic Web Conference, Karlsruhe, Germany, Springer, 2008.
- [Uschold, et al., 1998] Uschold M., King M., Moralee S., Zorgios Y. The Enterprise Ontology The Knowledge Engineering Review. Vol. 13, Special Issue on Putting Ontologies to Use / M. Uschold, A. Tate (eds.), 1998.
- [Ushold, Gruninger, 1996] Ushold M., Gruninger M. (1996). Ontologies : Principles, Methods and Applications. In: Knowledge Engineering Review, Vol. 11. No. 2. June 1996.
- [Vygotsky, 1978] Vygotsky L.S. Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press, pp. 1978.
- [WordNet, 2011] WordNet: a lexical database for the English language. <http://wordnet.princeton.edu>.



**Efimenko, I., Khoroshevsky, V.** Ontological Modeling of Enterprises and Markets of Modern Russia: Part 1. Ontological Modeling: Approaches, Models, Methods, Tools, Solutions : Working paper WP7/2011/08 (part 1) [Text] / I.V. Efimenko, V.F. Khoroshevsky ; National Research University "Higher School of Economics". – Moscow : Publishing House of the Higher School of Economics, 2011. – 76 p. – 150 copies.

The presented state-of-the-art review introduces the key issues of the ontology modeling and engineering and includes retrospective analysis of this domain. Results received in R&D in the domain are discussed in a context of Artificial Intelligence, Information Technologies, Semantic Technologies and Semantic Web. Models, methods and tools of ontology «on fly» creation, ontology reengineering, teamwork in ontology developing, as well as ontology merging and alignment, are considered. Fundamental ontology theories and models, existing classifications of ontologies, as well as known top-level ontologies (such as CYC, DOLCE, SUMO and Sowa-ontology) are presented. New theories of developing ontologies, in particular, the activity theory and its relation to ontological models are discussed shortly. Paper serves as an introduction and forms the basis for understanding the material presented in the subsequent preprints of the team of authors.

*Efimenko Irina* – National Research University "Higher School of Economics" (Moscow), [iefimenko@hse.ru](mailto:iefimenko@hse.ru).

*Khoroshevsky Vladimir* – Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, [khora@ccas.ru](mailto:khora@ccas.ru).

*Препринт WP7/2011/08 (ч. 1)*  
*Серия WP7*  
Математические методы анализа решений  
в экономике, бизнесе и политике

Ефименко Ирина Владимировна,  
Хорошевский Владимир Федорович

**Онтологическое моделирование экономики  
предприятий и отраслей современной России**

Часть 1. Онтологическое моделирование:  
подходы, модели, методы, средства, решения

Зав. редакцией оперативного выпуска *А.В. Заиченко*  
Корректор *Е.Е. Андреева*  
Технический редактор *Ю.Н. Петрина*

Отпечатано в типографии  
Национального исследовательского университета  
«Высшая школа экономики» с представленного оригинал-макета

Формат 60×84  $\frac{1}{16}$ . Тираж 150 экз. Уч.-изд. л. 4,5

Усл. печ. л. 4,4. Заказ № . Изд. № 1372

Национальный исследовательский университет  
«Высшая школа экономики»  
125319, Москва, Кочновский проезд, 3  
Типография Национального исследовательского университета  
«Высшая школа экономики»  
Тел.: (499) 611-24-15