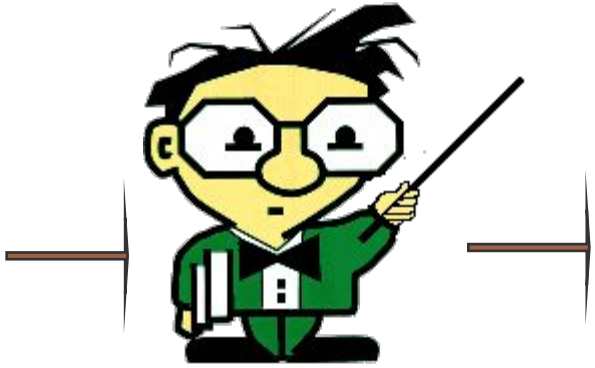
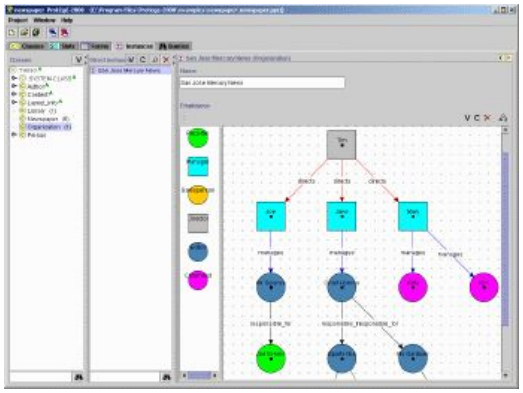
Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N





Web-сервис интеллектуальной обработки данных.

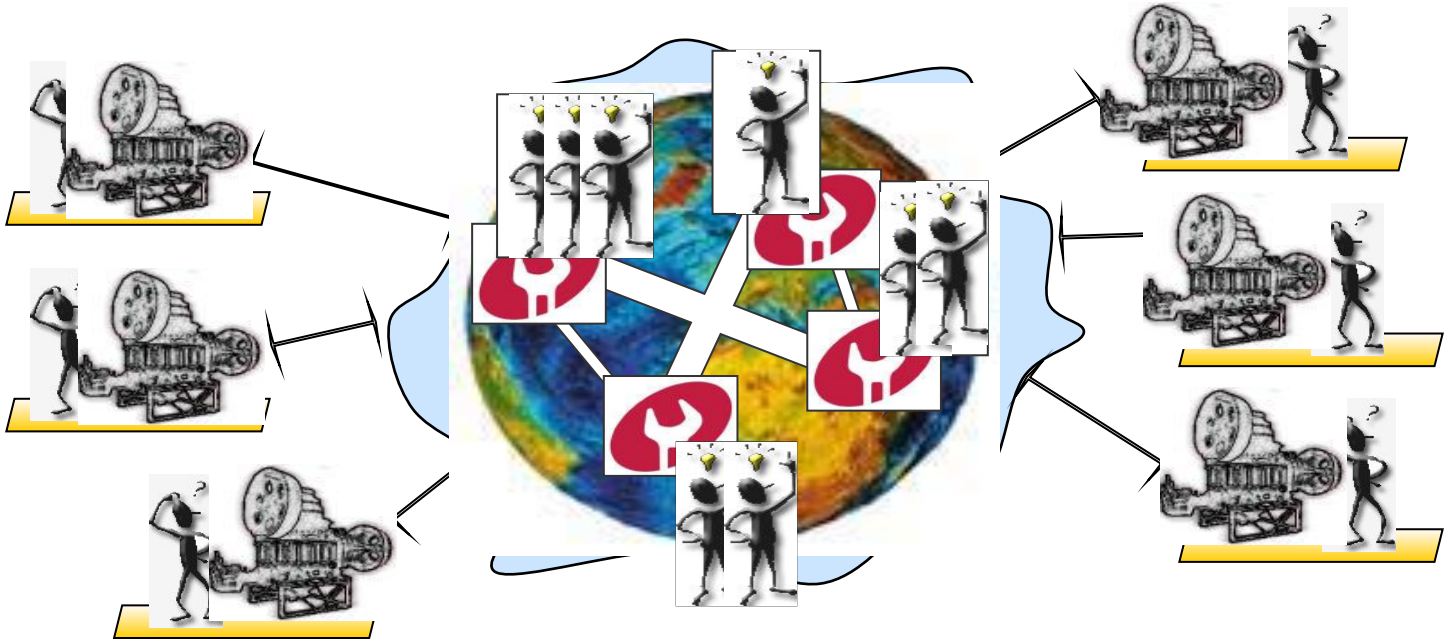
Разработка семантического программного адаптера

Терзиян А.В.

Студент гр. ИСПР-00-1

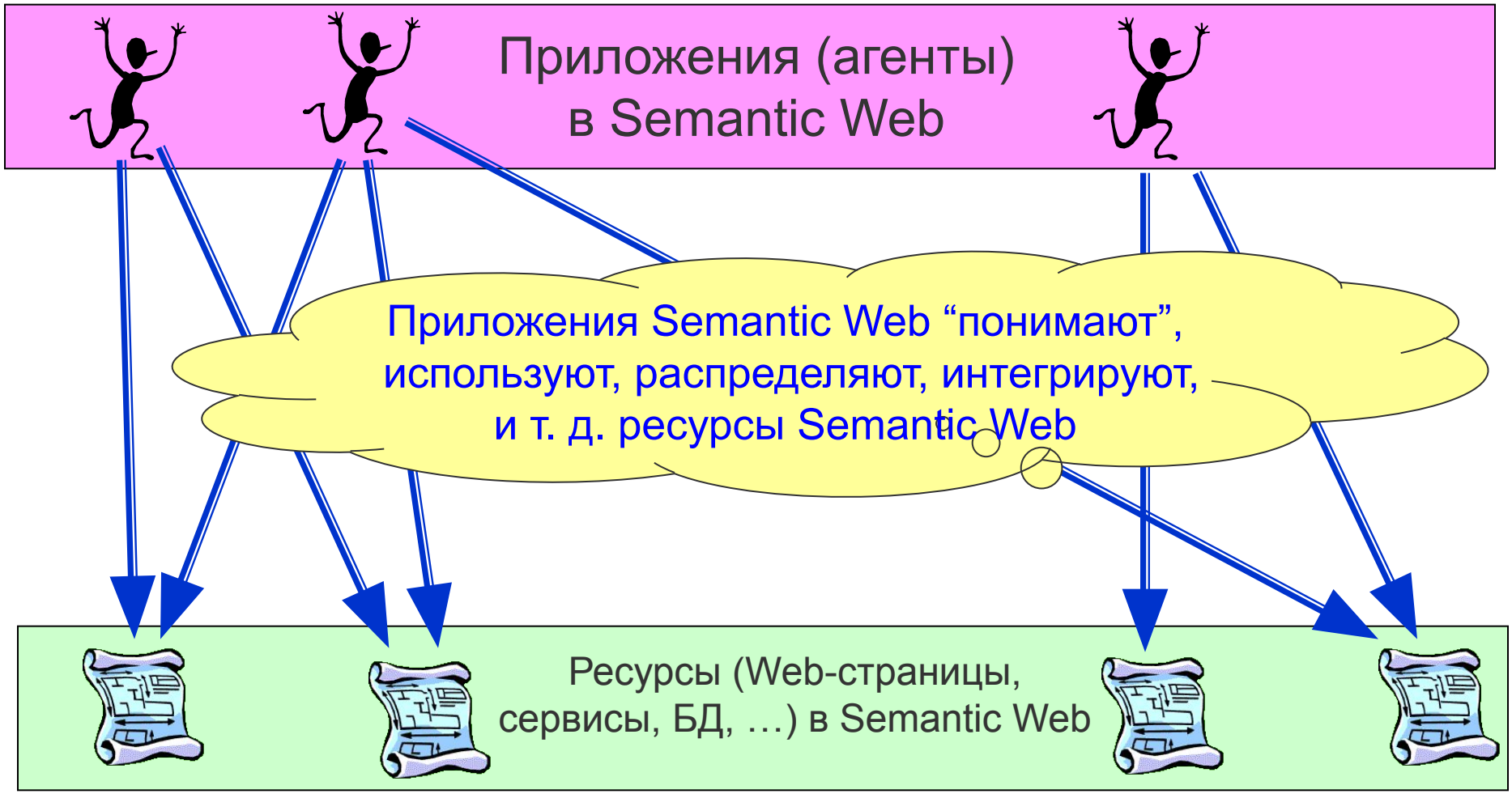
Научный руководитель: доцент каф. ИИ Витько А.В

Харьков, 2004

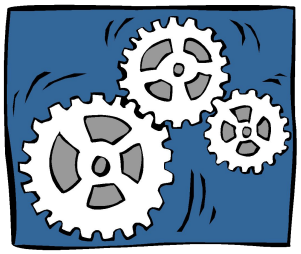


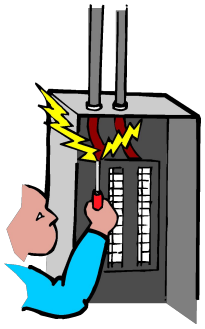
Мотивация и цель работы

Semantic Web: первоначальная концепция

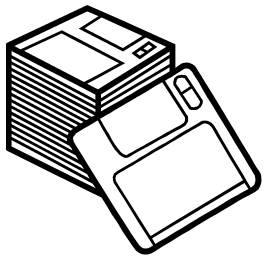
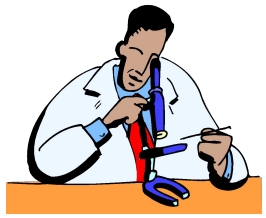


Новая концепция:
Global **U**nderstanding e**N**vironment (**GUN**)





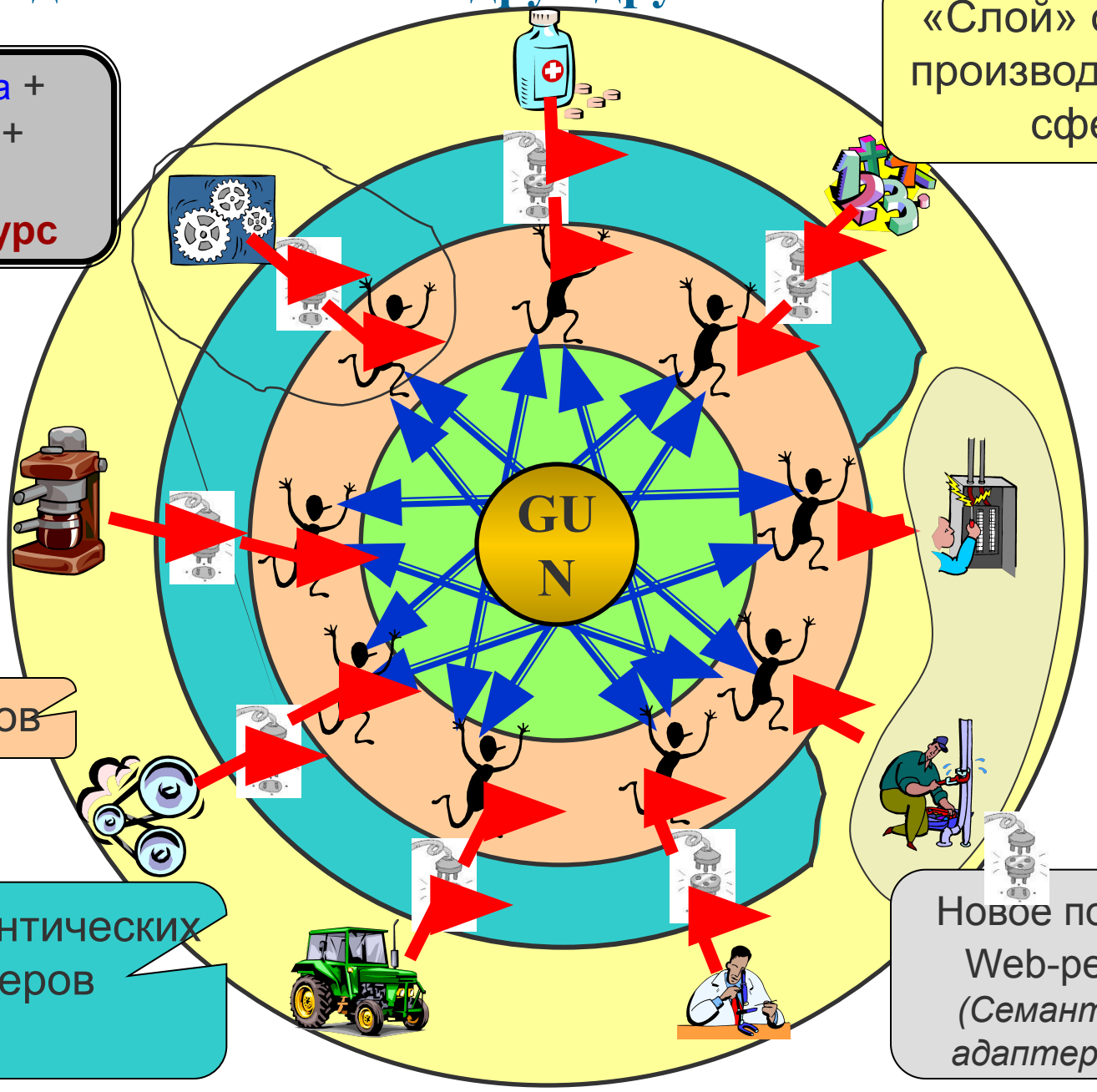
Как сделать так, чтобы объекты окружающего мира (например, производственной сферы) стали доступными из WWW и «понимали» друг-друга в случае необходимости ?

Все GUN-ресурсы проактивны и используют семантические адаптеры для “понимания” друг-друга

Объект мира +
+ Адаптер +
+ Агент =
= **GUN-ресурс**

«Слой» объектов производственной сферы



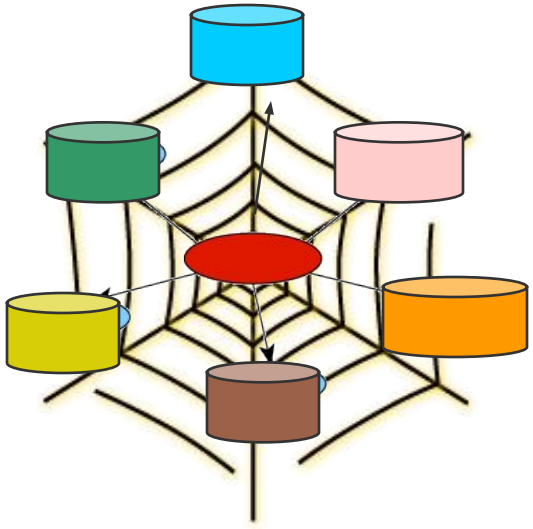
Слой агентов

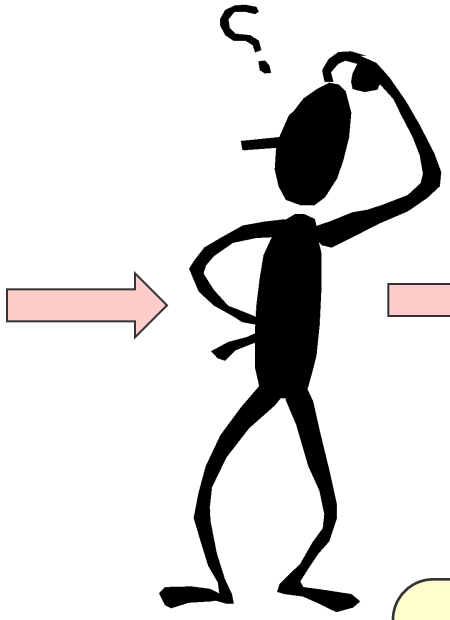
Слой семантических адаптеров

Новое поколение Web-ресурсов *(Семантический адаптер внутри)*

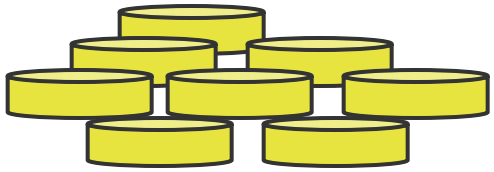
Извлечение знаний из распределенных и разнородных массивов данных о ресурсах в Web- или GUN-среде – одна из основных задач Web-сервисов по интеллектуальной обработке данных

Распределенные и разнородные массивы данных



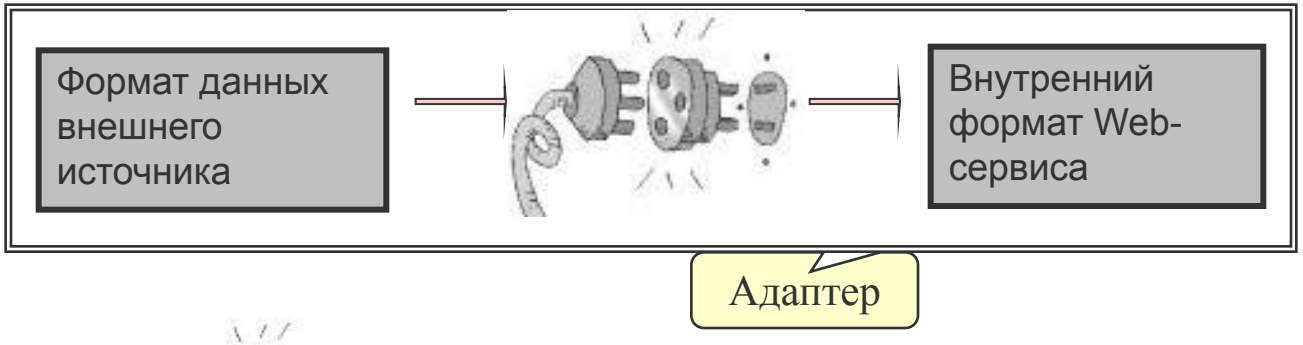


Знания
(например, модели диагностики)

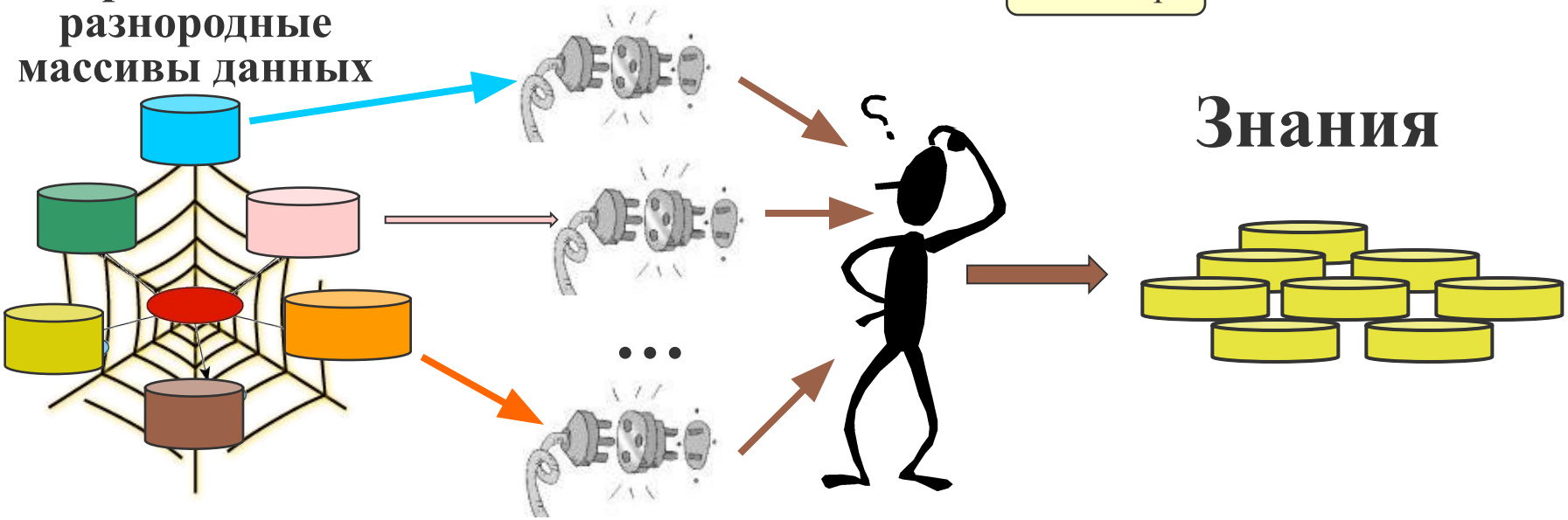


Web-сервис по интеллектуальной обработке данных

Однако, если Web-сервис понимает только один формат входных данных, тогда ему необходим адаптер для работы с каждым новым массивом данных

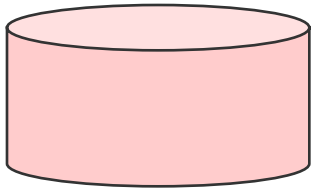


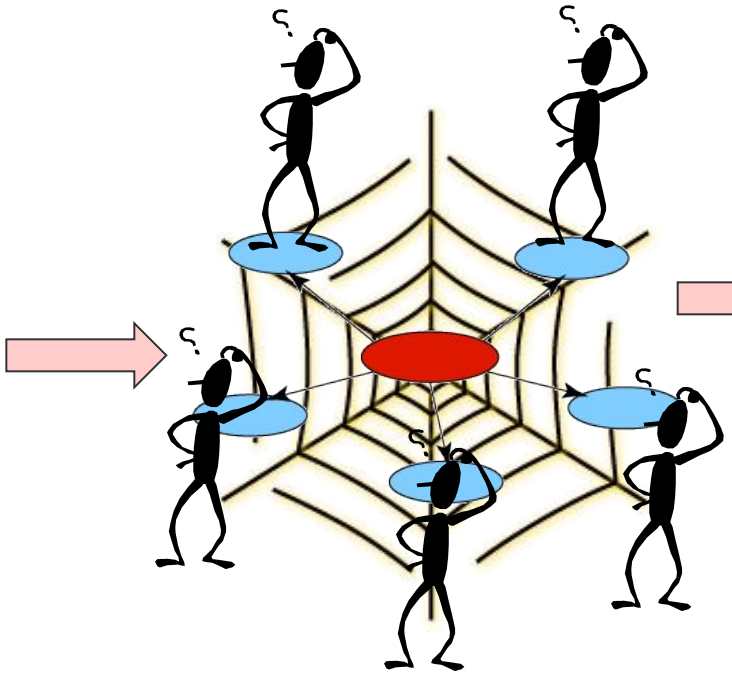
Распределенные и разнородные массивы данных



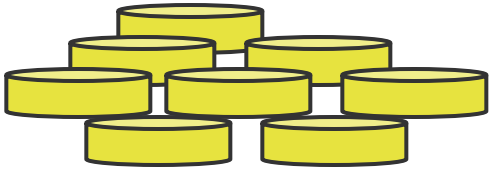
С другой стороны, данные могут браться из одного массива данных, но обрабатываться несколькими распределенными и разнородными Web-сервисами…

Web-массивы данных



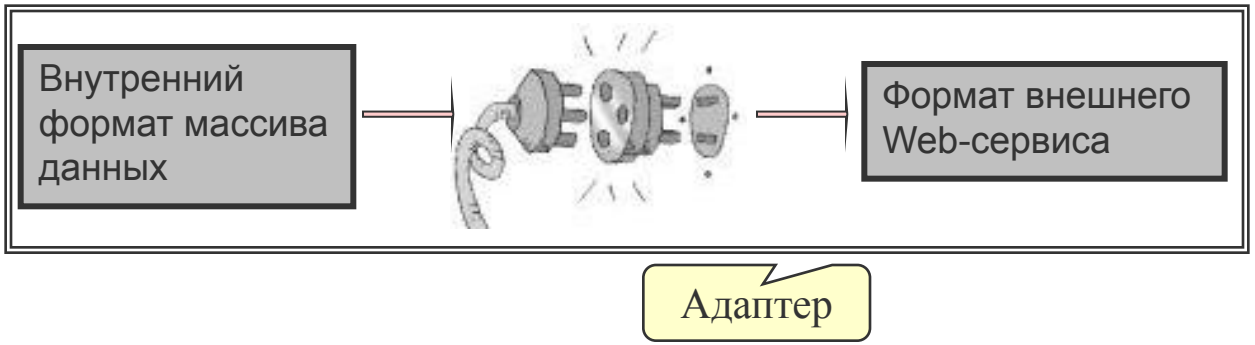


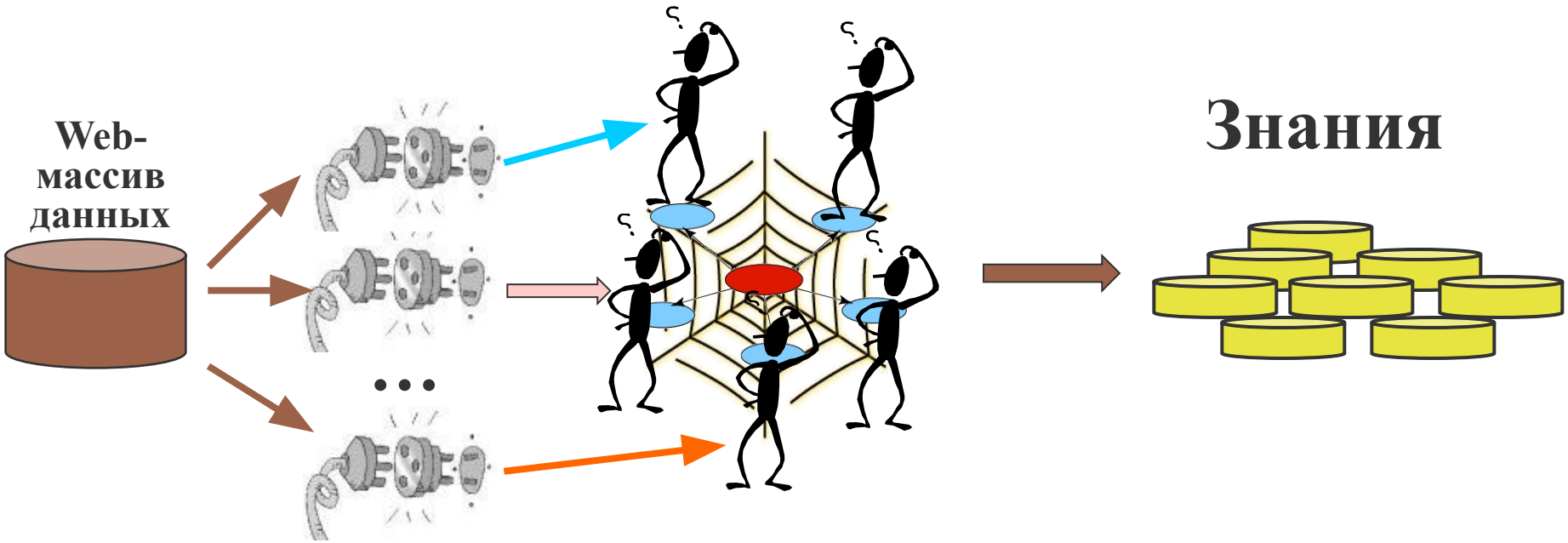
Знания



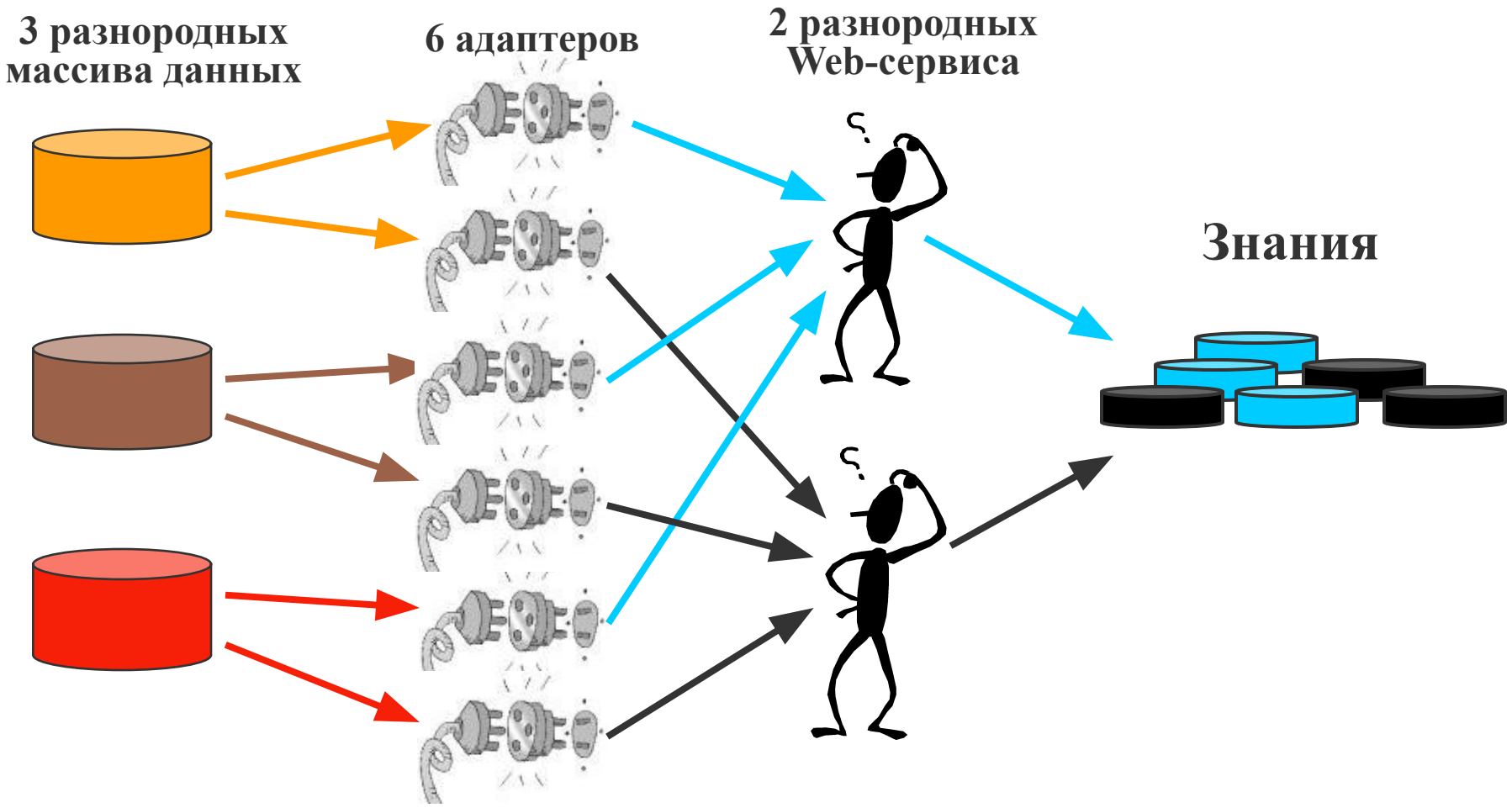
Сеть разнородных Web-сервисов по интеллектуальной обработке данных

… и это подразумевает, что нам необходимы адаптеры для работы с разными Web-сервисами

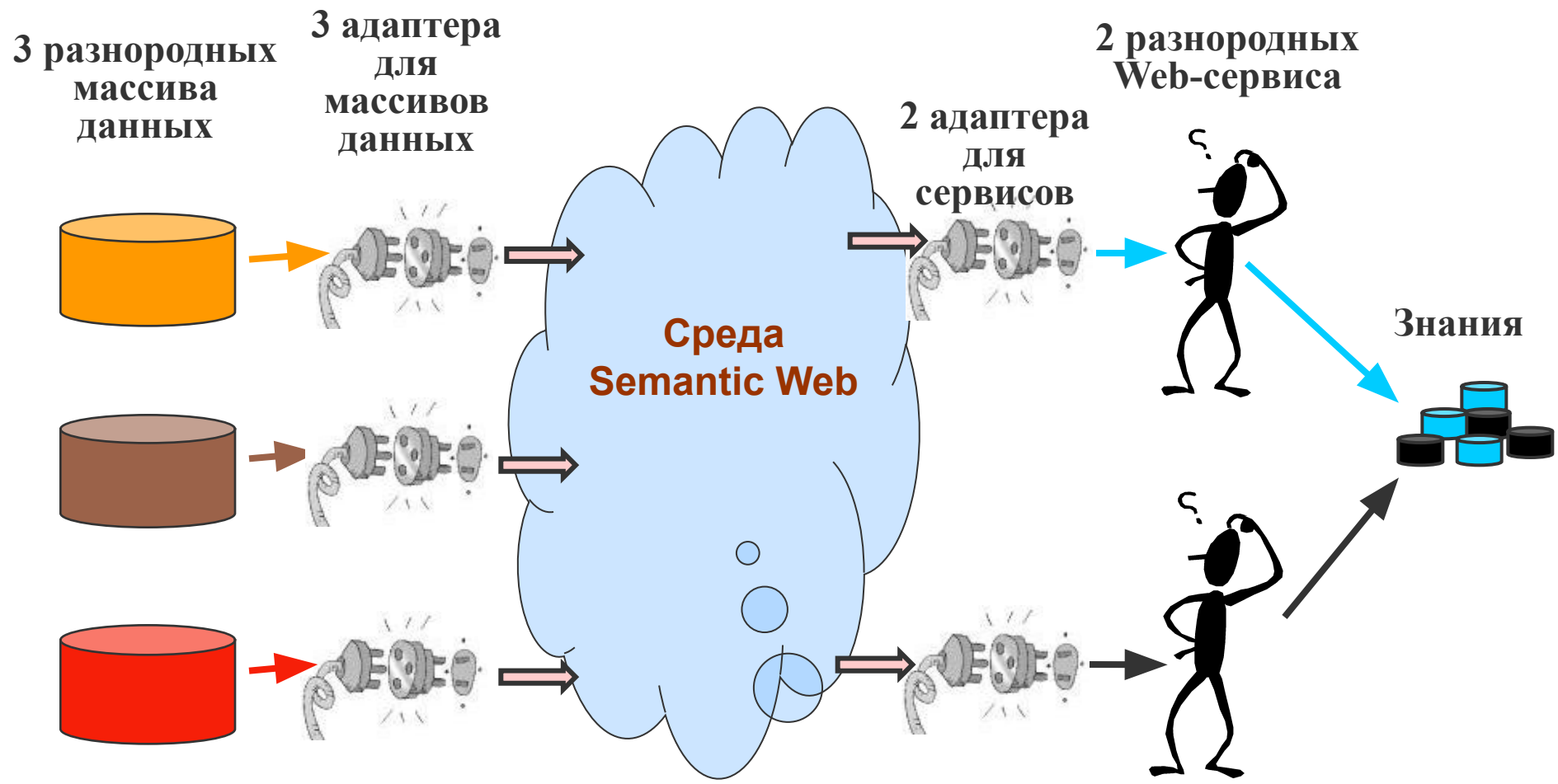




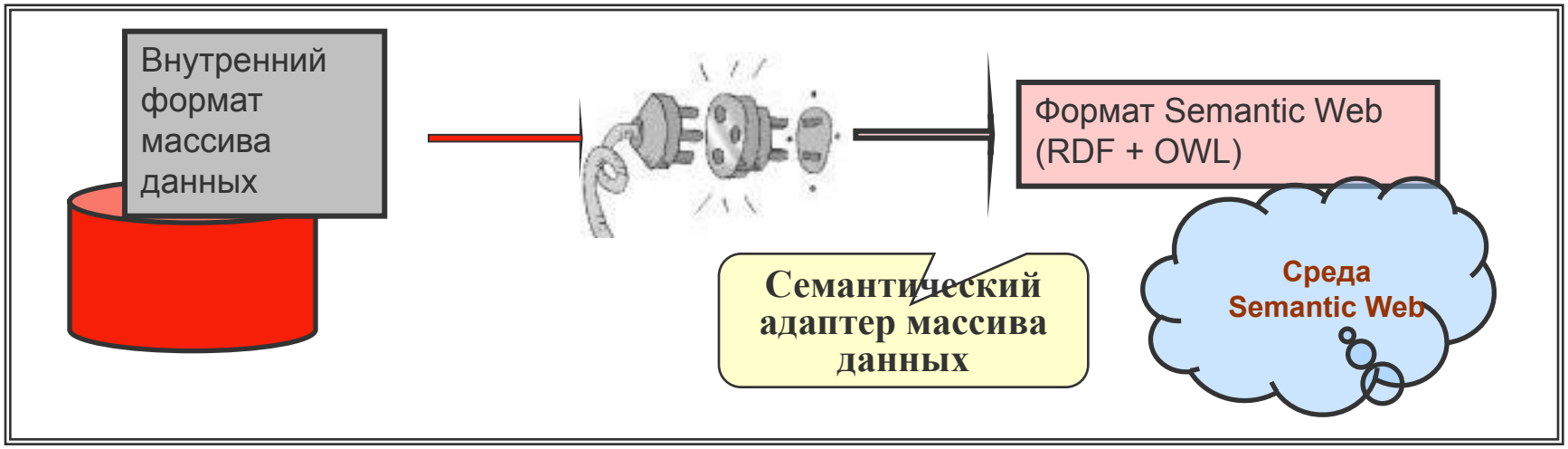
В общем случае для оперирования с *N* разнородными массивами данных и *M* разнородными Web-сервисами нам необходимо *M* × *N* адаптеров



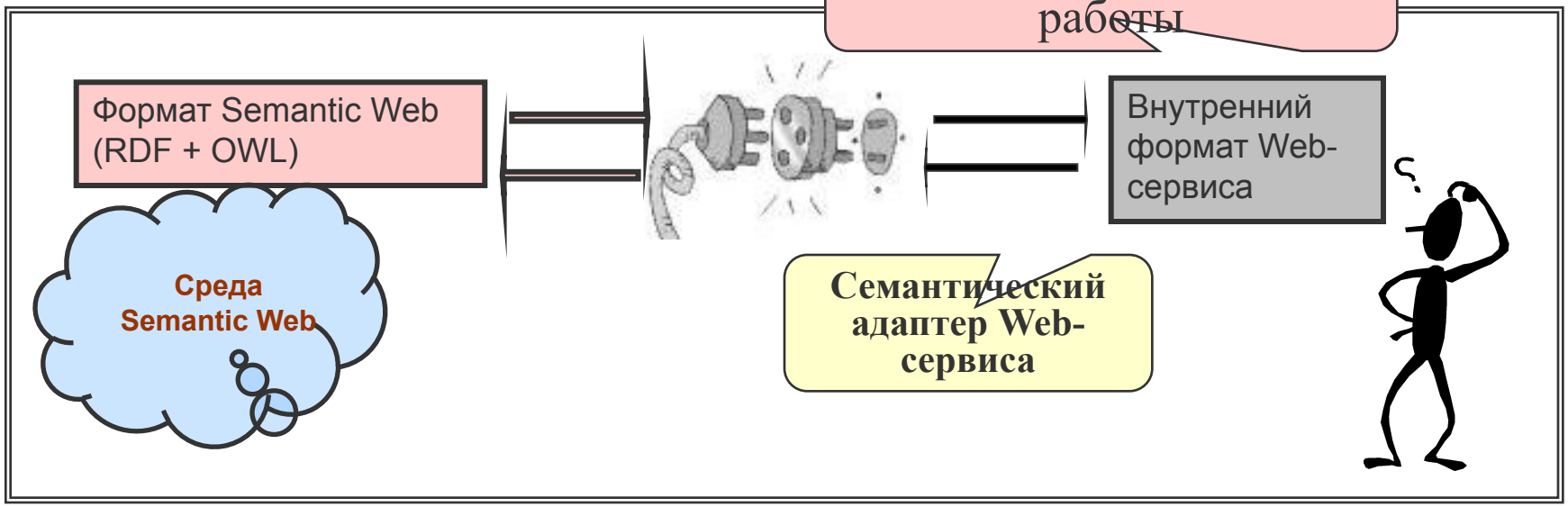
Для уменьшения стоимости адаптации решением может быть использование стандартных форматов данных (например, стандарты W3C для Semantic Web : RDF, RDFS, OWL) в качестве посредника между источниками данных и сервисами, что в худшем случае будет требовать *M* + *N* (семантических) адаптеров



Таким образом, нам необходимо реализовать два типа семантической адаптации: (1) преобразование массивов данных в среду Semantic Web и (2) наоборот



Задача бакалаврской работы



Основные цели работы

- Разработка способов семантического представления обучающих выборок
- Разработка программного адаптера, реализующего функции:
  - преобразования входных данных для Web-сервиса из стандартизованного формата OWL в массив данных, соответствующий внутреннему представлению информации в Web-сервисе;
  - преобразование выходных данных, полученных в результате работы Web-сервиса из внутреннего формата в стандартизованный;
- Экспериментальное тестирование адаптера на примере данных, подготовленных в среде Protégé.

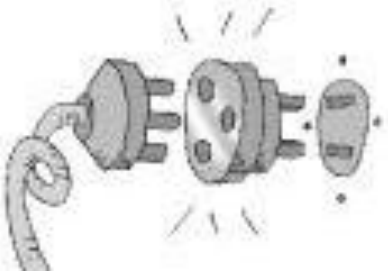
Описание обучающих выборок в семантическом формате

- целевой атрибут представляется самым высшим классом в иерархии классов (owl:Class);
- возможные значения целевого атрибута представляются подклассами (owl:SubClassOf) класса целевого атрибута;
- атрибуты описываются свойствами класса целевого атрибута. Возможны два варианта:
  - использовать простые свойства (owl:DatatypeProperty), значения атрибутов задаются простыми значениями стандартных типов данных;
  - использовать объектные свойства (owl:ObjectProperty), значения атрибутов задаются экземплярами заранее определенных «дополнительных» классов значений;
- примеры в выборке представляются в виде экземпляров класса целевого атрибута.

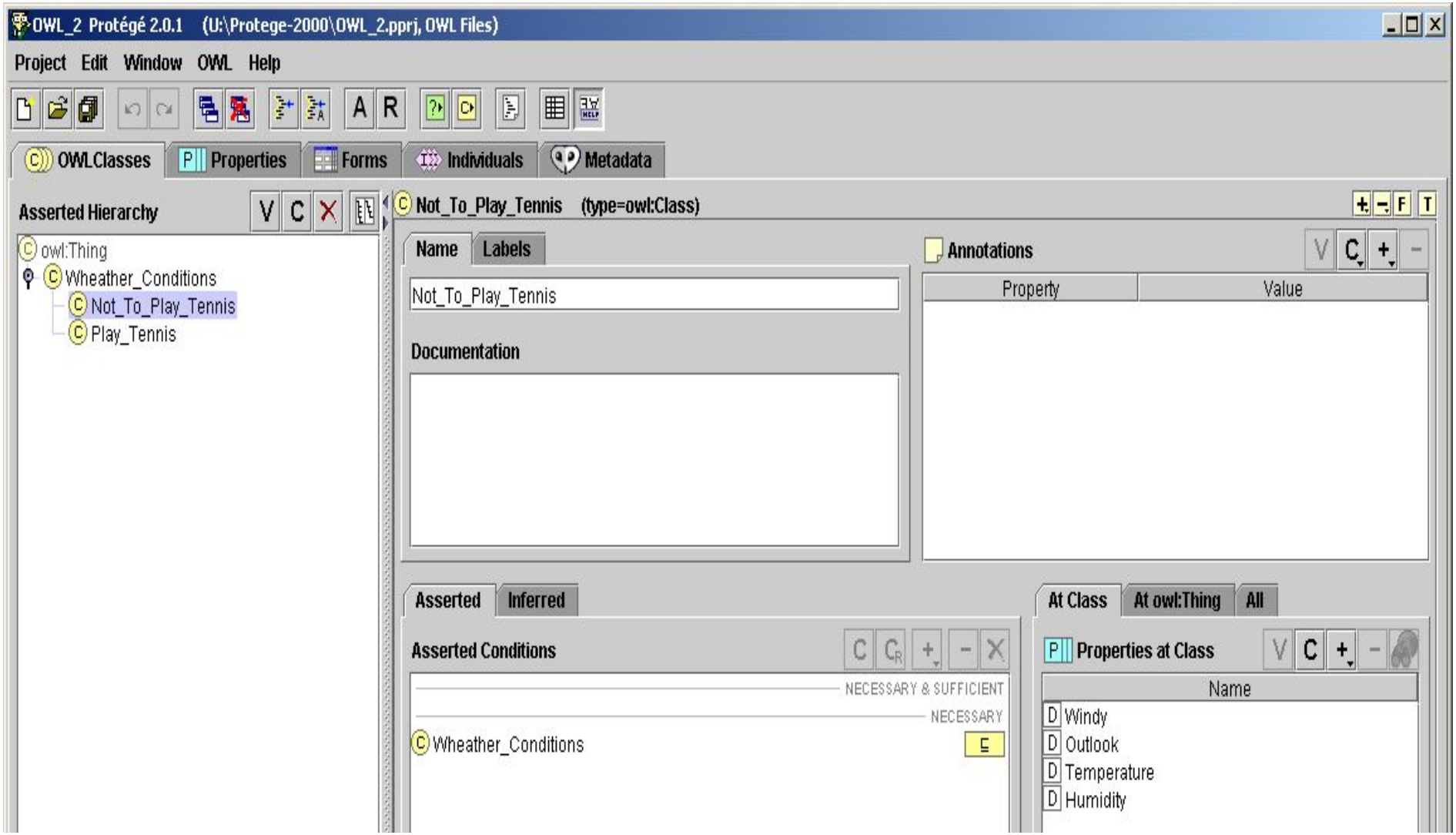
Схема преобразования OWL-Table I

Вариант 1: значения атрибутов в онтологии являются символьными строками

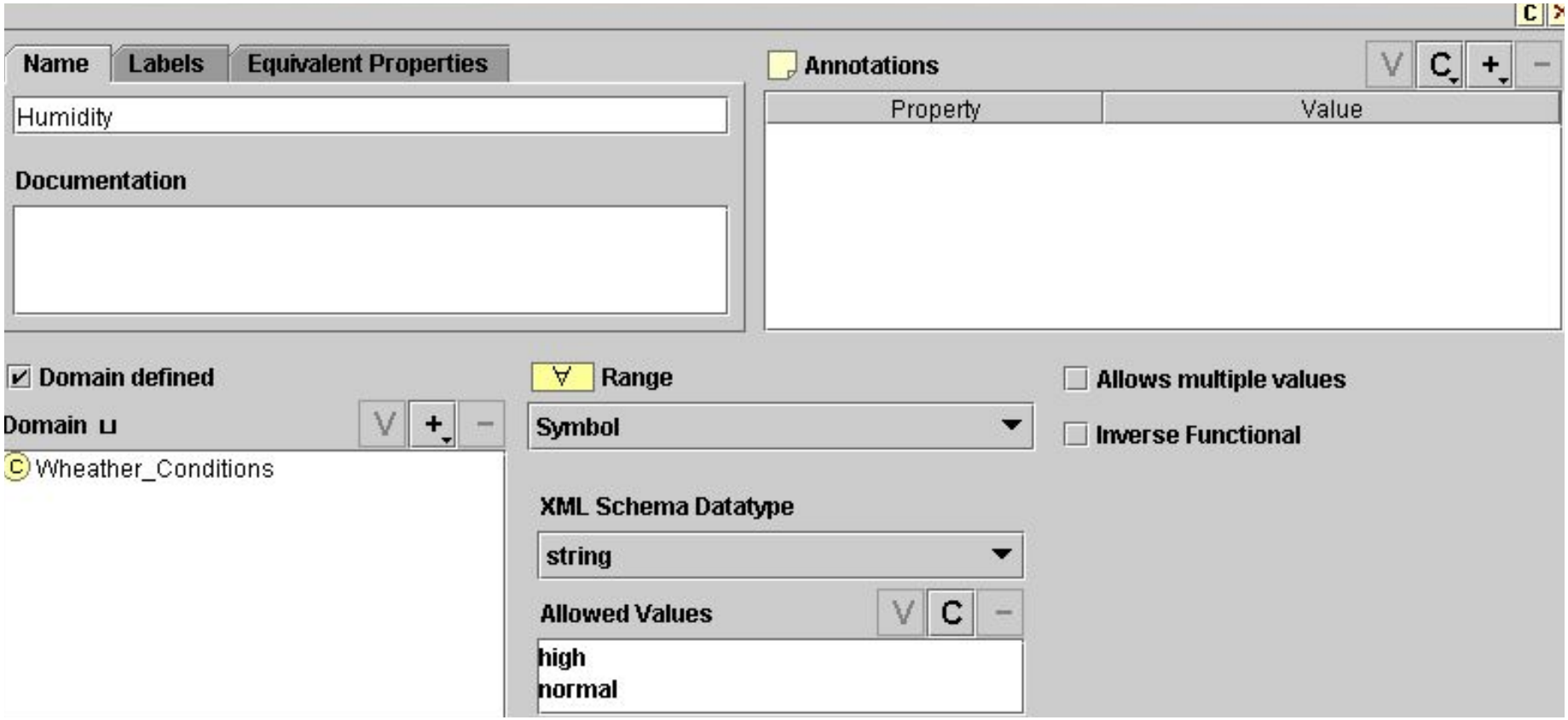
Проиллюстрировано в среде Protégé-2000



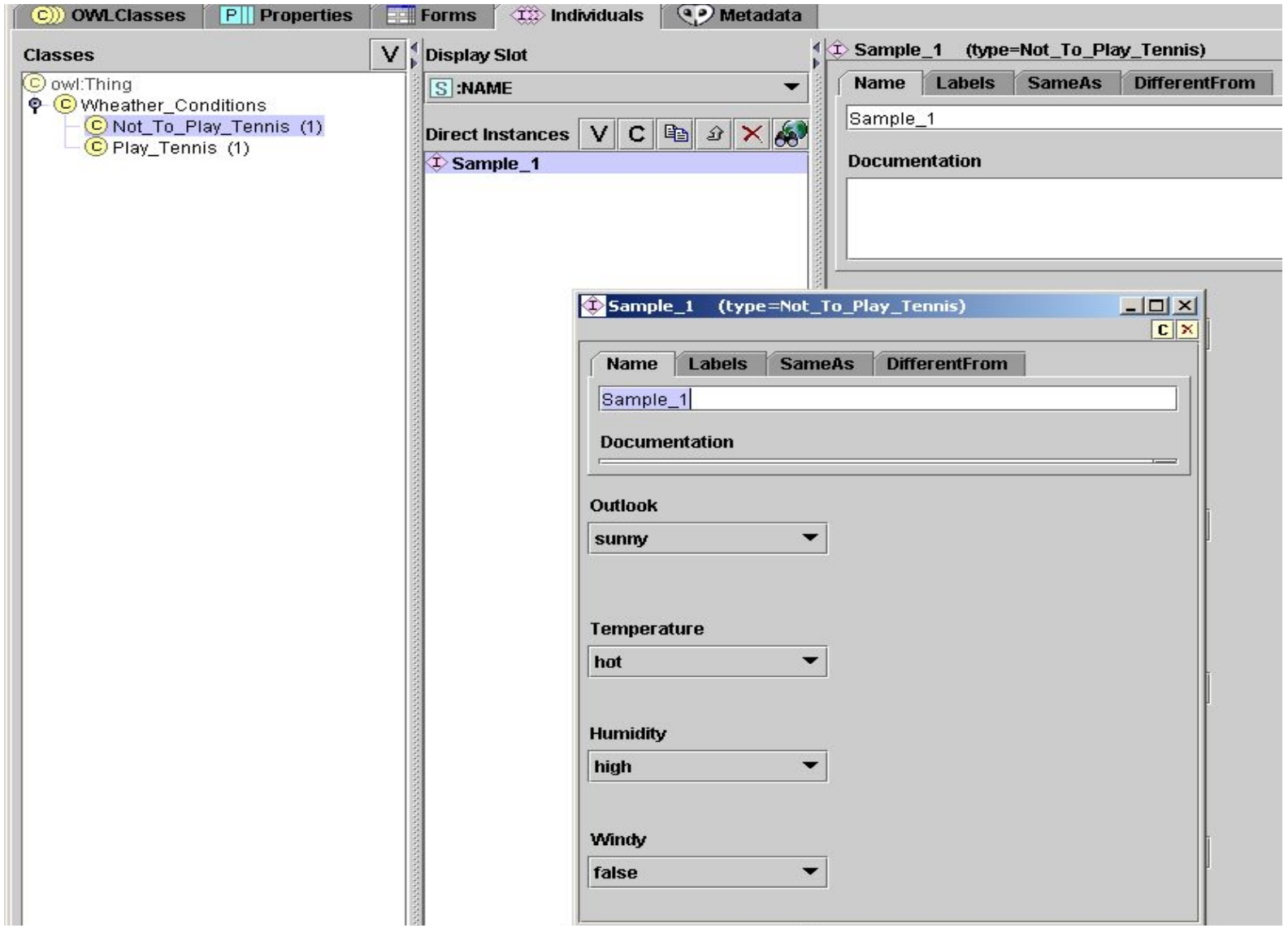
Задание атрибутов в Protégé 2000



Задание типов и возможных значений атрибутов



Задание примеров из обучающей выборки



Преобразование из OWL в таблицу

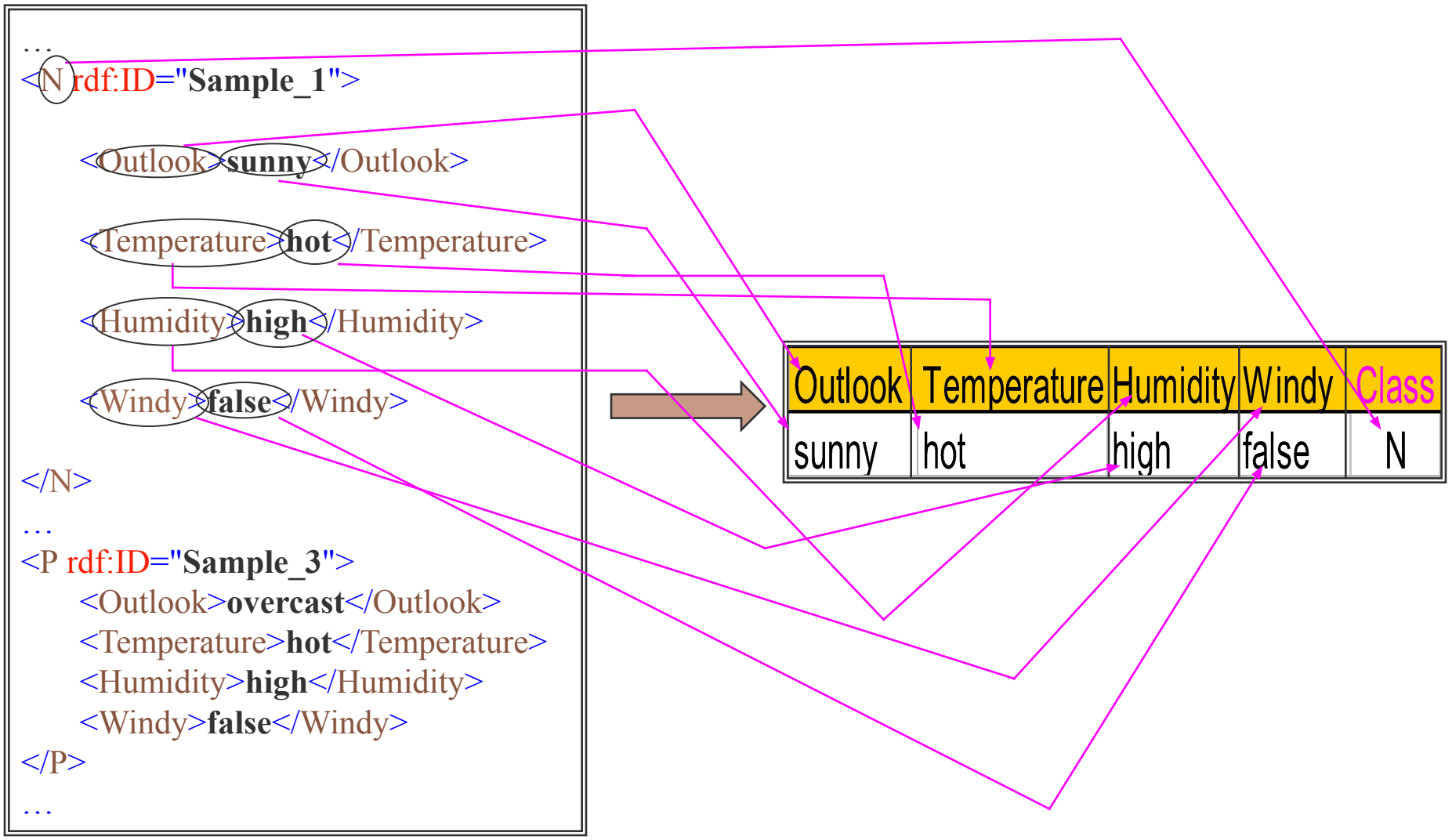
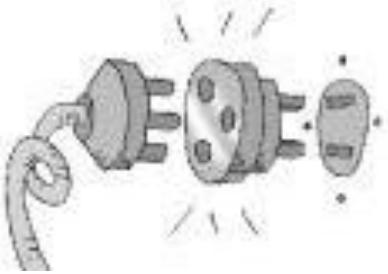


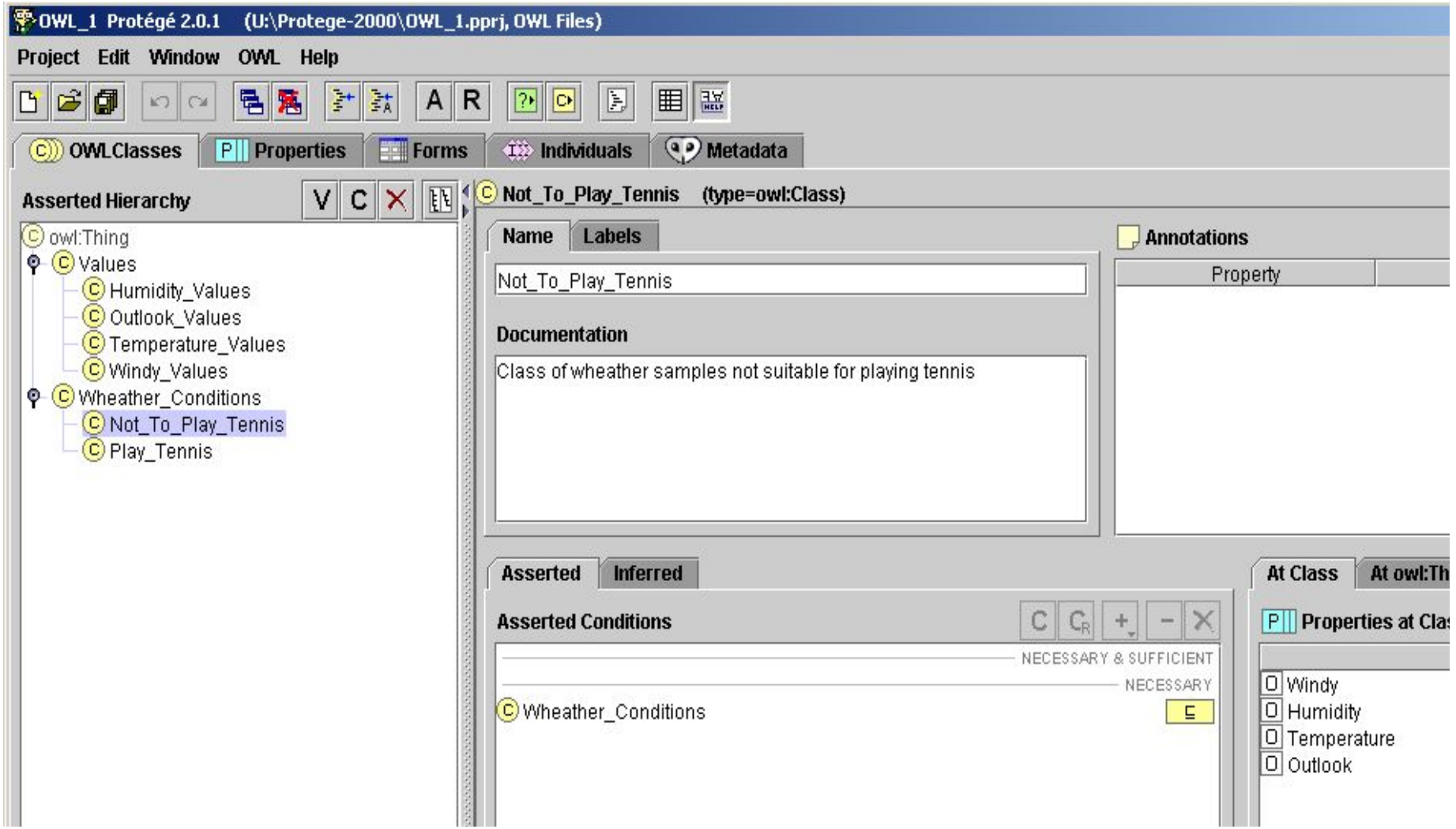
Схема преобразования OWL-Table II

Вариант 2: значения атрибутов в онтологии являются элементами некоторых классов

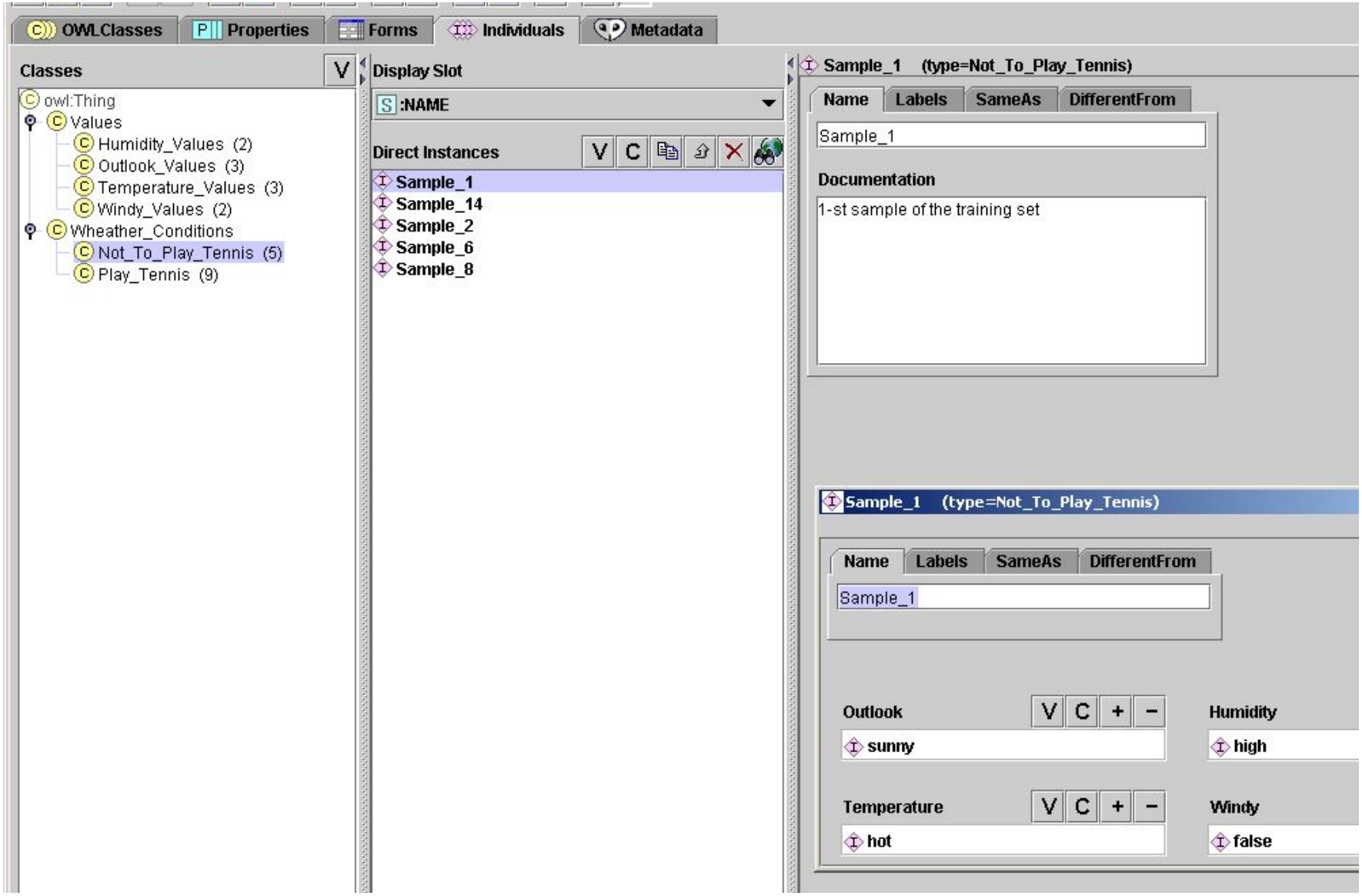
Проиллюстрировано в среде Protégé-2000



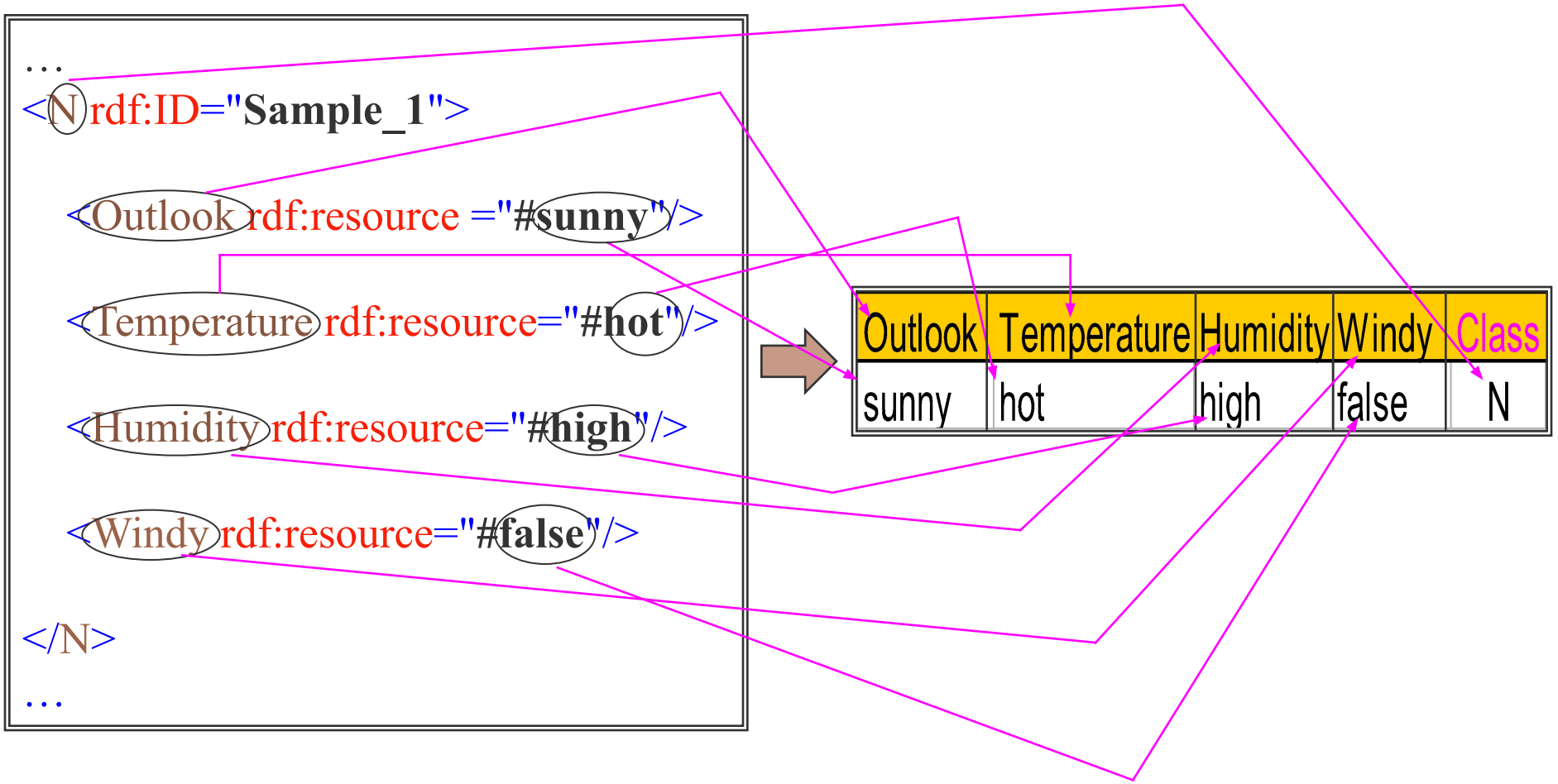
Создание дополнительных классов для возможных значений атрибутов



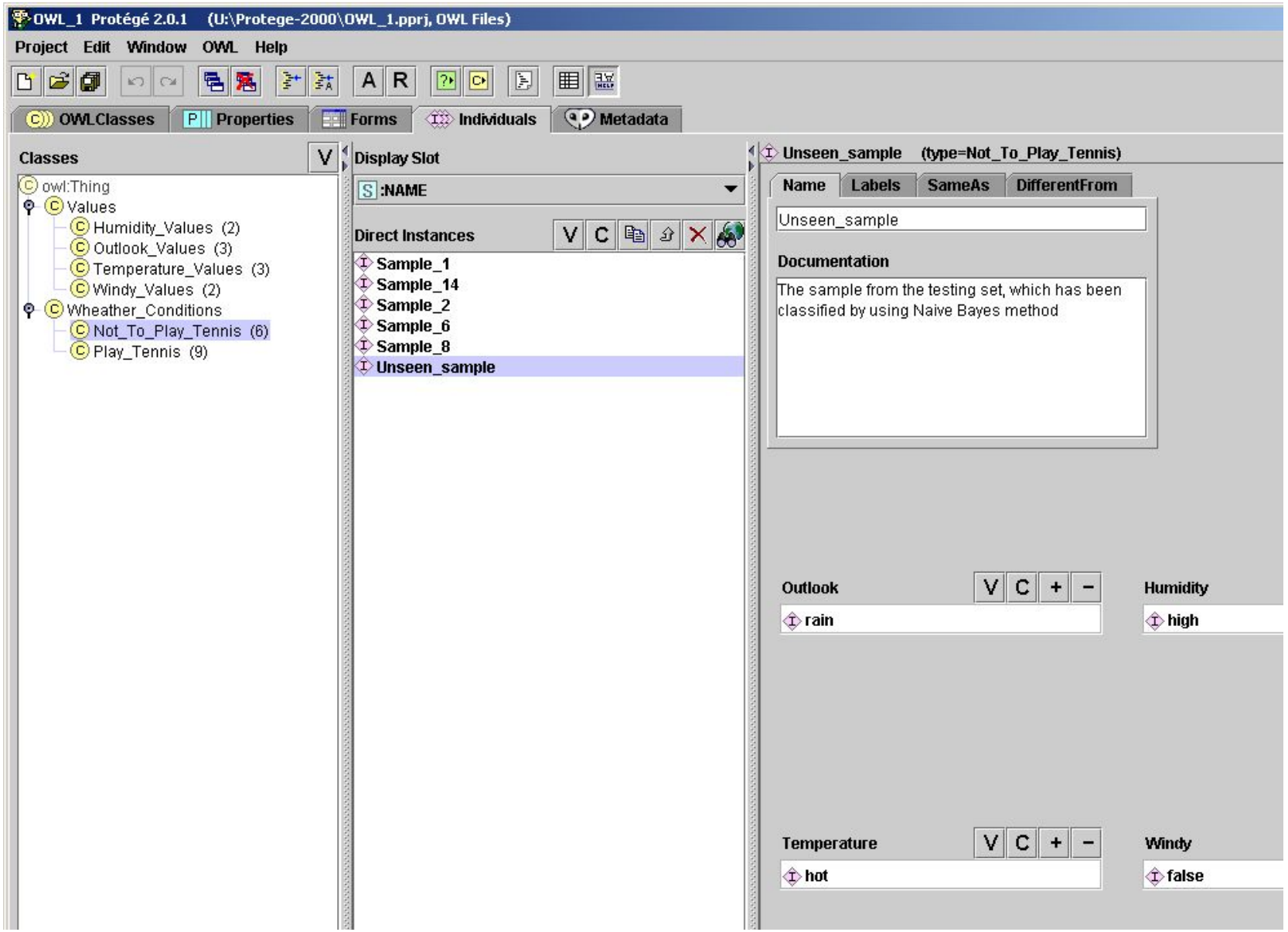
Задание примеров со значениями атрибутов, являющимися экземплярами классов



Преобразование из OWL в таблицу



Добавление классифицированного примера в онтологию



Укрупненный алгоритм работы семантического адаптера

Шаг 1. Создаем онтологическую модель для подаваемого на вход адаптера OWL-документа.

Шаг 2. Обрабатываем свойства модели

Цикл по всем свойствам: {

если свойство типа данных или объектное свойство,

то свойство является атрибутом;

если атрибут типа данных,

то обрабатываем возможные значения;

если атрибут объектный,

то находим класс возможных значений, переходим к этому классу и обрабатываем его экземпляры как возможные значения атрибута

}

Укрупненный алгоритм работы семантического адаптера (продолжение)

Шаг 3. Обрабатываем классы и экземпляры модели

Цикл по всем классам: {

если класс неабстрактный и не является классом значений,

то класс соответствует значению целевого атрибута;

Цикл по всем экземплярам класса:{

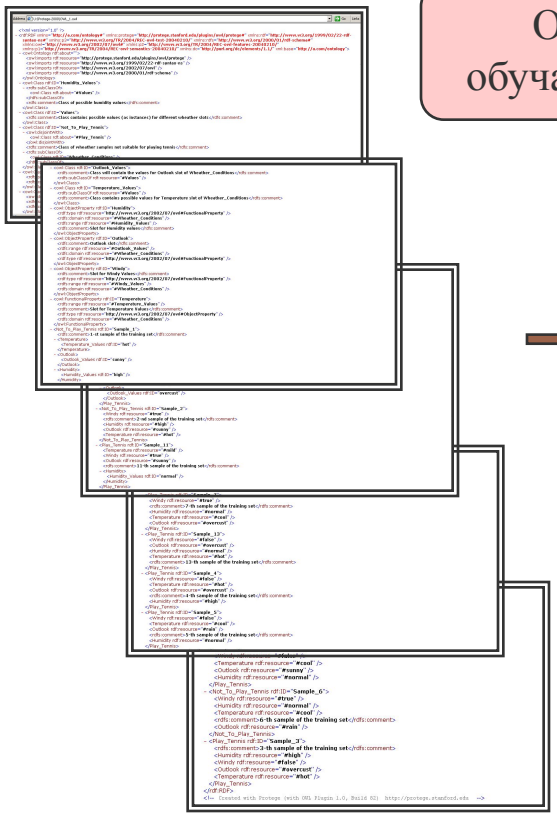
заполняем выборку, представленную во внутреннем формате Web-сервиса:

- класс экземпляра заносим в целевой атрибут;

- значения всех свойств заносим в определяющие атрибуты.}

}

Задания, выполненные адаптером (1): преобразование OWL-файла в массив



OWL онтология с обучающими примерами

Выборка в табличной форме

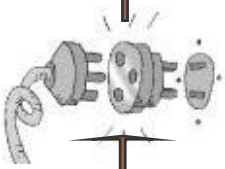
Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Задания, выполненные адаптером

(2): Обновление OWL-файла классифицированными примерами

```
- <Not_To_Play_Tennis rdf:ID="Unseen_sample">
    <Humidity rdf:resource="#high" />
    <Outlook rdf:resource="#rain" />
    <Temperature rdf:resource="#hot" />
    <Windy rdf:resource="#false" />
    <rdfs:comment>The sample from the testing set, which has been classified by using Naive Bayes method</rdfs:comment>
  </Not_To_Play_Tennis>
```

расширенная OWL онтология



Классифицированный пример в OWL формате

Outlook	Temperature	Humidity	Windy	Class
rain	hot	high	false	N

Классифицированный пример в табличной форме

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

+

Outlook	Temperature	Humidity	Windy	Class
rain	hot	high	false	?

Результаты работы (1)

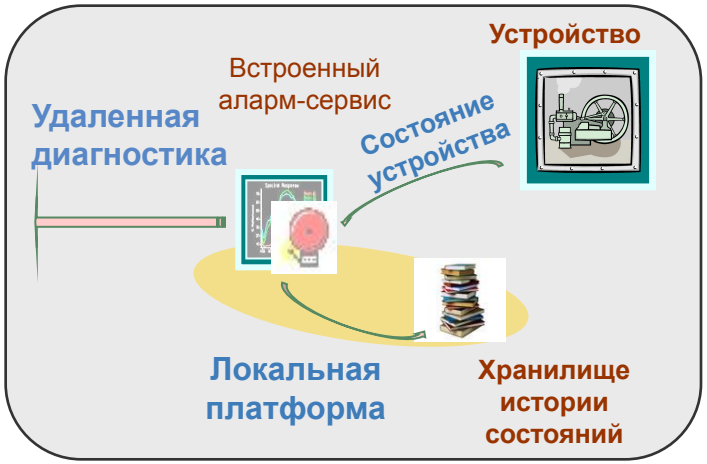
- Исследованы современные **Интернет-технологии**, связанные с Semantic Web, Web-сервисами, мультиагентными системами, интеллектуальной обработкой данных, автоматизированным промышленным мониторингом и удаленной диагностикой.
- Взята за основу концепция **GUN**, предполагающая проактивность Web-ресурсов и адаптеры для привязки ресурсов к Web.
- Был обоснован выбор стандартов представления информации в глобальных информационных системах, определенные W3C-консорциумом, таких как **RDF, RDFS и OWL**.

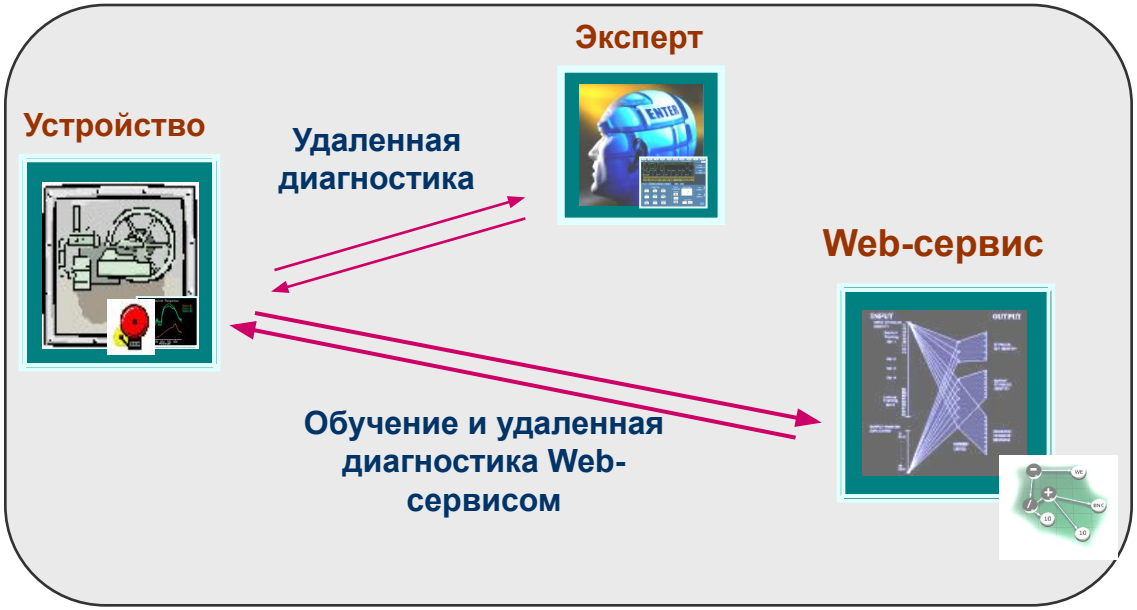
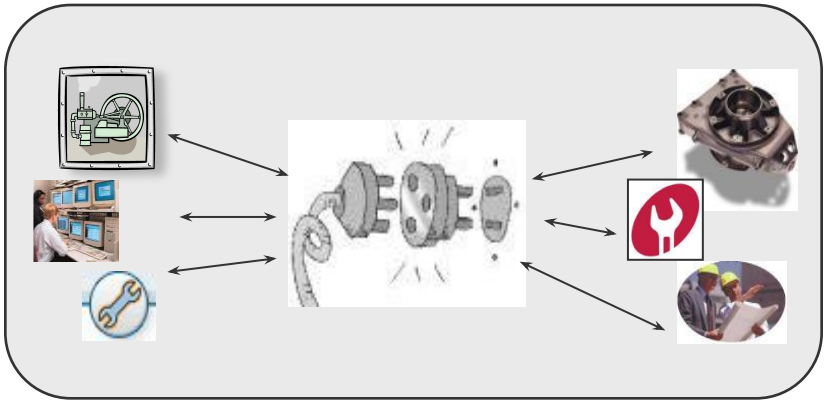
Результаты работы (2)

- Был предложен способ **универсального представления обучающих и тестирующих выборок** для Web-сервиса в формате онтологий, закодированных средствами языка OWL.
- Был разработан **механизм семантической адаптации**, реализующий преобразование данных из OWL в табличное представление и обратно и позволяющий в перспективе совместно использовать (интегрировать) разнородные сервисы диагностики.

Результаты работы (3)

- Был разработан **семантический адаптер** как программный модуль на языке Java™ 2 SDK Standard Edition Version 1.4.2_03 в среде NetBeansTM IDE 3.6. Была использована библиотека **JENA** Semantic Web Framework Version 2.1.
- Адаптер был экспериментально протестирован на примерах данных, подготовленных в Protégé-2000.
- Работа выполнялась как часть комплексной работы (5 человек).
- По результатам работы были опубликованы тезисы доклада на **8-м Международном Молодежном Форуме** «Радиоэлектроника и молодежь в XXI веке» и готовится статья.





Спасибо за внимание !