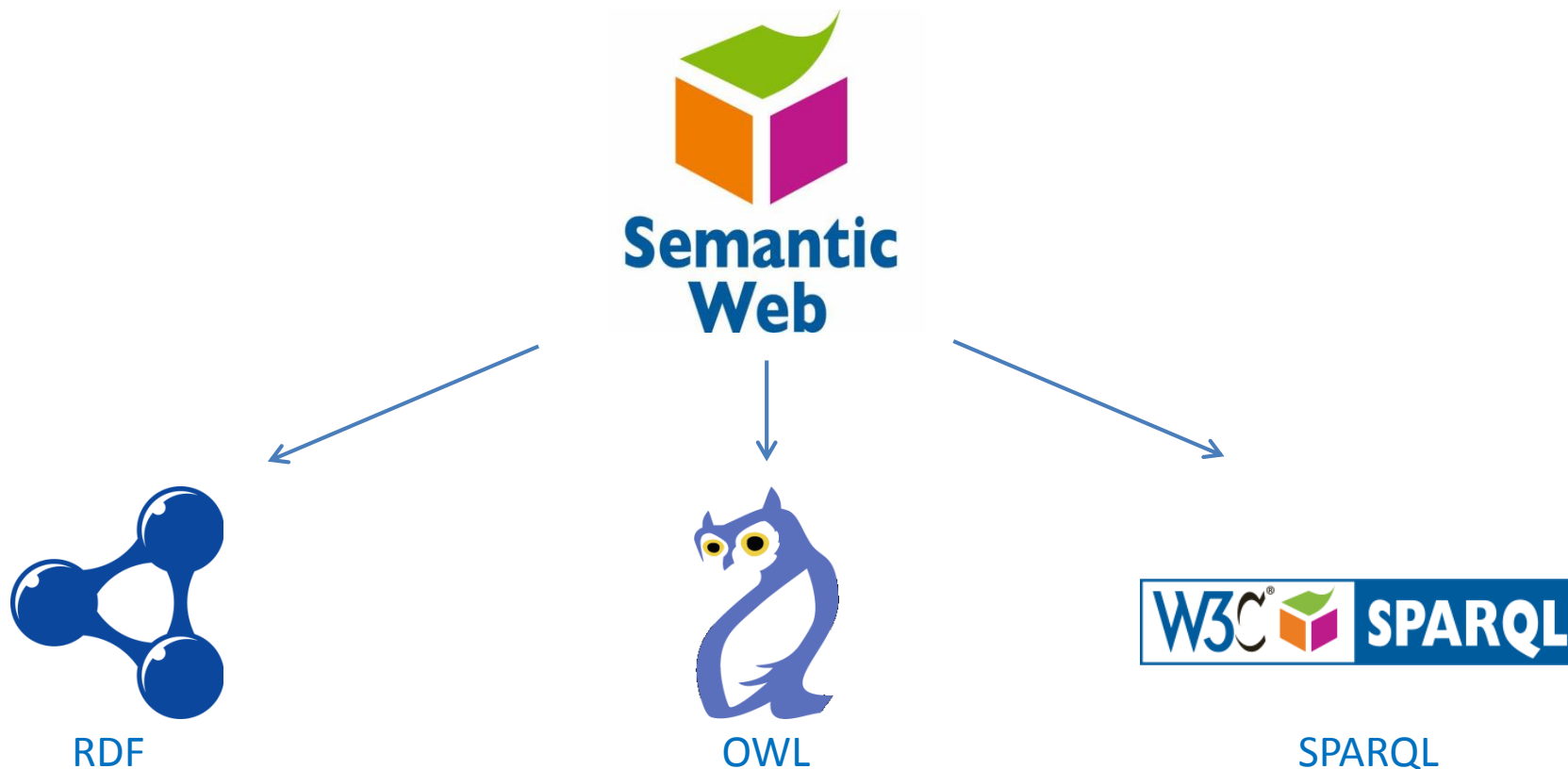


# Технологии Semantic Web



Алик Кириллович

Казанский Федеральный Университет

<http://www.alik.su>

Ver. 2.0

# Содержание курса

- I. Введение в семантику и Semantic Web
- II. Формат представления данных RDF
- III. Формальные онтологии
- IV. Язык описания онтологий OWL
- V. Язык запросов SPARQL
- VI. Semantic Web в дикой природе:  
популярные ресурсы, онтологии и сервисы



# World Wide Web



- Крупнейшее в мире хранилище информации

IDC, 2008:

Объем информации в интернете составляет 5444 петабайт.

Для сравнения: объем информации во всех книгах, изданных в истории человечества, составляет только 200 петабайт, т.е. почти в 30 раз меньше.

Яндекс, «Контент Рунета», 2009:

Объем Рунета (6% мировой паутины) соответствует 35 миллионам книг.

- Мгновенный доступ, мощный поиск по ключевым словам

**Но:**

- Информация понятна людям, но непонятна машинам
- Компьютеры обрабатывают документы как набор символов, но не понимают их смысл

**Решение:** Semantic Web

# Semantic Web



---

**Semantic Web** ( $\approx$  Linked Data, Linking Open Data, Web of Data) — надстройка над WWW, позволяющая сделать информацию в Web понятной не только людям, но и машинам

---

Возможности:

- Семантический поиск
- Объединение данных
- Логический вывод
- Агенты

# Семантический поиск

Семантический поиск — поиск не по ключевым словам,  
а по смыслу

Ноутбуки с объемом памяти > 8GB по цене до \$1000

10 самых длинных рек России

Какие американские президенты родились в Техасе?

Насколько подорожала стоимость тарифа электроэнергии за 2014 год?

Сколько мужей было у Аллы Пугачевой?

# Объединение данных



Объединение данных помогает находить ответы на вопросы, которых нет ни в одном источнике по отдельности, но есть в их совокупности

Какие побочные эффекты у лекарства от туберкулёза?

Источники: *DRUGBANK*,  *Diseasome*,  *Sider*

Какие конгрессмены голосовали против закона о защите окружающей среды, и, при этом, представляют округа с самым высоким уровнем вредных выбросов?

Источники: *govtrack.us*,  *GeoNames*,  *US Census*

# Логический вывод



**Логический вывод** — вычисление новых знаний из уже имеющихся

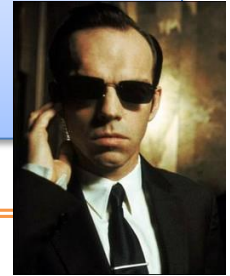
Найти российские университеты:

КФУ находится в Казани, Казань находится в России → КФУ — российский универ

Найти праправнуков королевы Виктории:

Виктория — мать Эдуарда VII, Эдуард VII — отец Георга V, Георг V — отец Георга VI, Георг VI — отец Елизаветы II → Елизавета II — праправнучка Виктории

# Агенты



**Интеллектуальный агент** — программа, способная автономно выполнять указанное человеком задание по поиску и обработке информации

Пример задания:

Siri, закажи мне столик в лучшем итальянском ресторане в Пало-Альто.

См.: Adam Cheyer, Tom Gruber. Siri: A virtual personal assistant for iphone, an ontology-driven application for the masses // Open, International, Virtual Community of Practice on Ontology, Ontological Engineering and Semantic Technology, 2010

В настоящее время широкое распространение агентов — только мечты.

Пока основное предназначение Semantic Web — поиск и интеграция данных.

**Что же такое семантика?**



# Семантика

- Цель коммуникации — передать от отправителя получателю информацию об **объекте** реального мира или ситуации
- Однако передать информацию напрямую («прямо в мозг») невозможно
- Поэтому, информация записывается на некотором **языке** с определенным **синтаксисом** в виде последовательности **знаков** (напр., знаков или букв)
- Последовательность символов можно легко передать (в виде речи, письма)
- Получатель восстанавливает исходную информацию с помощью **семантики**



«Земля вращается вокруг Солнца»



Семантика



«Земля вращается вокруг Солнца»



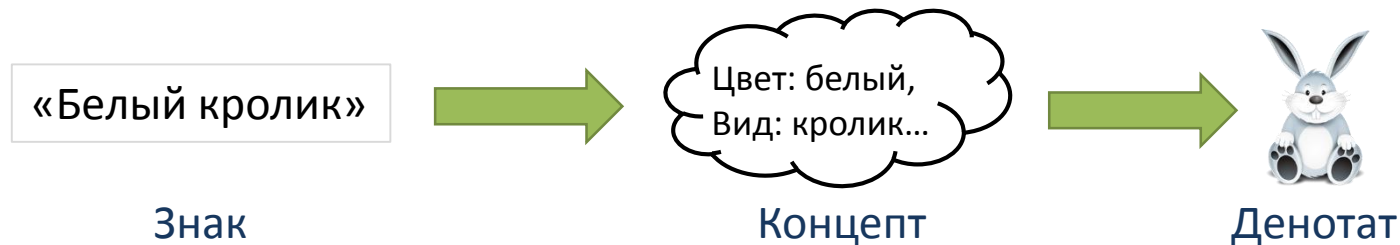
# Денотат

- **Семантика** языка — это способ соотнесения **знака** с объектом реального мира или ситуацией из некоторой предметной области (**домена**)
- **Денотат** — объект или ситуация, которые соотносятся со знаком



# Концепт

- Знак соотносится с объектом благодаря наличию у знака концепта
- **Концепт** ( $\approx$  понятие, десигнат, интенционал, смысл) — это некоторый набор условий, которые должен удовлетворять объект или ситуация, чтобы соотноситься со знаком
- Т.о. если объект или ситуация удовлетворяют условиям концепта, то они соотносятся с его знаком, т.е. становятся денотатом



Не во всех языках концепты явно и формально определены

# Неявная семантика

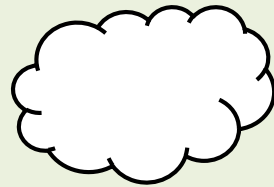
Языки с неявной семантикой — концепты знаков не определены

Естественный язык

Язык танца:



Знак: движения



Концепт



Денотат: эмоция

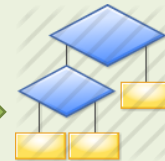
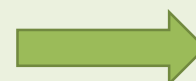
Люди понимают эти языки благодаря тому, что интуитивно «чувствуют» их концепты (но не всегда могут объяснить их машинам).

# Явная неформальная семантика

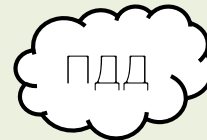
Языки с явной семантикой — концепты знаков явно определены

Языки программирования и разметки

```
«for (i=0; i<10; i++)»
```



Язык дорожных знаков



Машины могут понимать эти языки.

Но понимание каждого концепта надо программировать *вручную*.

Браузер понимает HTML.

Но: если появится новый тег в HTML → браузер надо перепрограммировать.

Беспилотные автомобили Google понимают язык дорожных знаков.

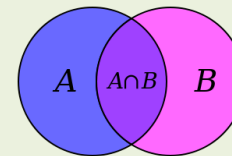
Но: появится новый знак в ПДД → беспилотники надо перепрограммировать.

# Формальная семантика

Языки с формальной семантикой — концепты знаков явно определены на метаязыке с явной семантикой

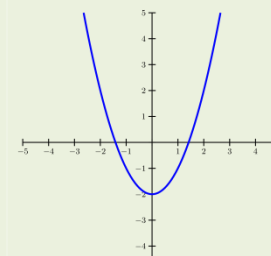
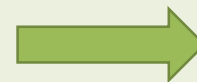
Язык логики и математики

« $A = B \cap C$ »



Язык логики и математики

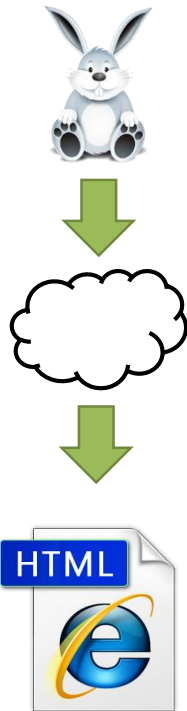
« $y = x^2 + b$ »



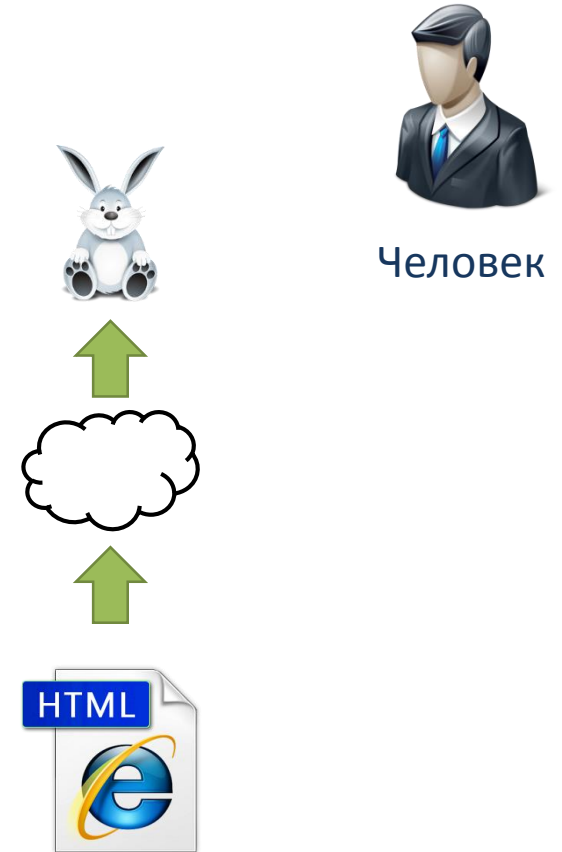
Машины могут самостоятельно прочитать и понять определение концептов.

Что же происходит с семантикой в World Wide Web?

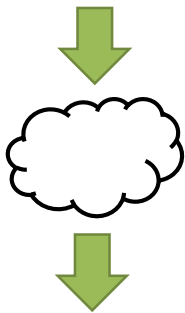
# Семантика в WWW



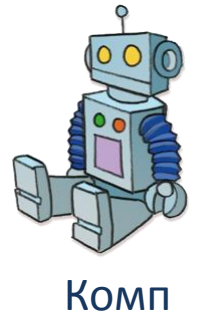
Люди интуитивно чувствуют концепты



# Семантика в WWW



Машины *не знают*  
концепты естественного языка





# Что делать?

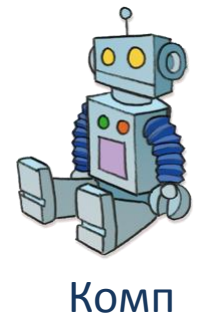
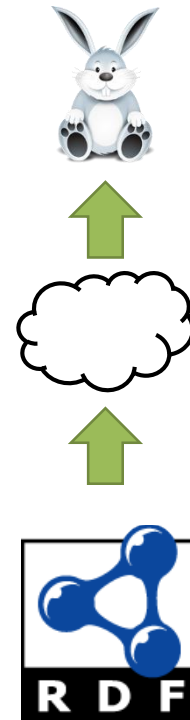
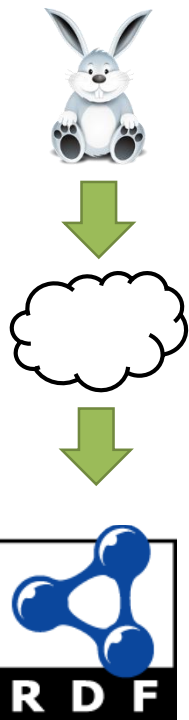
Два подхода:

- **Подход NLP** — научить машину понимать синтаксис и семантику естественного языка

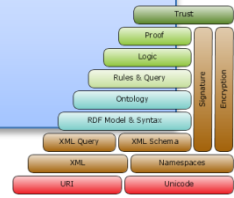
Решение задачи в общем виде сопоставимо по сложности с созданием искусственного интеллекта

- **Подход Semantic Web** — описать информацию на языке понятном машине

# Семантика в Semantic Web



# Архитектура Semantic Web



Технологии:

- **RDF** — язык, позволяющий **записать** утверждения
- **RDFs** и **OWL** — языки описания онтологий

**Онтология** определяют концепты из некоторой предметной области, с помощью чего машина и понимает семантику

- **SPARQL** — язык **запросов**

# Онтология

Онтология описывает понятия предметной области и связи между ними

- **Базовые понятия**

Человек, Мужчина, Женщина, супруг, ребенок

- **Сложные понятия на основе простых**

Мать одиночка = женщина  $\wedge$  не имеет супруга  $\wedge$  имеет  $> 3$  детей

племянник = брат  $\circ$  ребенок

- **Аксиомы**

Мужчина и Женщина не пересекаются

потомок – транзитивное свойство

Аксиомы позволяют проводить логический вывод