



**DATA
FABRIC**



Интеллектуальные информационные системы на основе семантических технологий

**или семантическая интерпретация BigData
или системы искусственного интеллекта на основе инженерии знаний**

Исаев Сергей
www.DataFabric.cc

Про что это вообще?

#Семантические сети (Семантика данных)

#Графы знаний

#Графовые базы данных (RDF-триплеты)

#Linked Data

#Онтологическое моделирование

Информационно-справочные системы

Главное функциональное назначение:

- **Возможность поиска информации в данных**

Полнотекстовый поиск по содержанию текстов или различные комбинации фильтров.

Информационно-справочные системы

Главная фундаментальная проблема:

- Система не может понимать смысл запроса

Для системы все равно – 0 или 1, А или Б

Система делает только то, что в нее заложено (запрограммировано), не понимая смысла.

Проблема работы с информацией

- Чаще всего человек сам не знает что он ищет.
- Получение информации происходит в процессе исследования.

Яндекс

Google

IBM Watson™


WolframAlpha

Интеллектуальные информационные системы

Система должна понимать смысл запроса Понимать в понятиях человека

Обращение в систему на логическом уровне:

- Какая налоговая ставка у меня будет, если я стану резидентом Сколково?

Ответ – Одно число, а не все документы с «налоговая ставка»

Работать со **знаниями**, а не с данными

Чем знания отличаются от данных?

RDF-триплеты

RDF Triple Graph Example



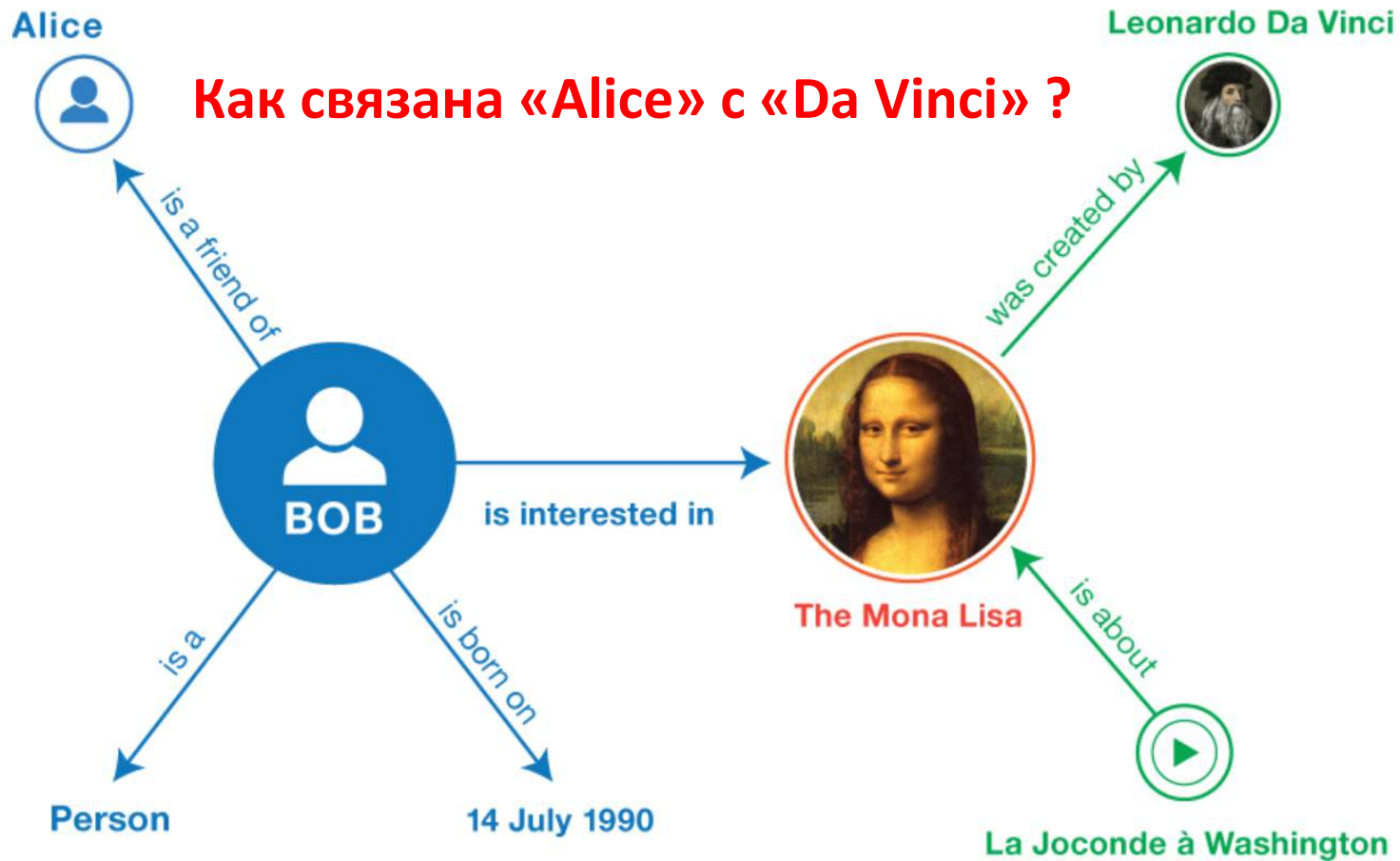
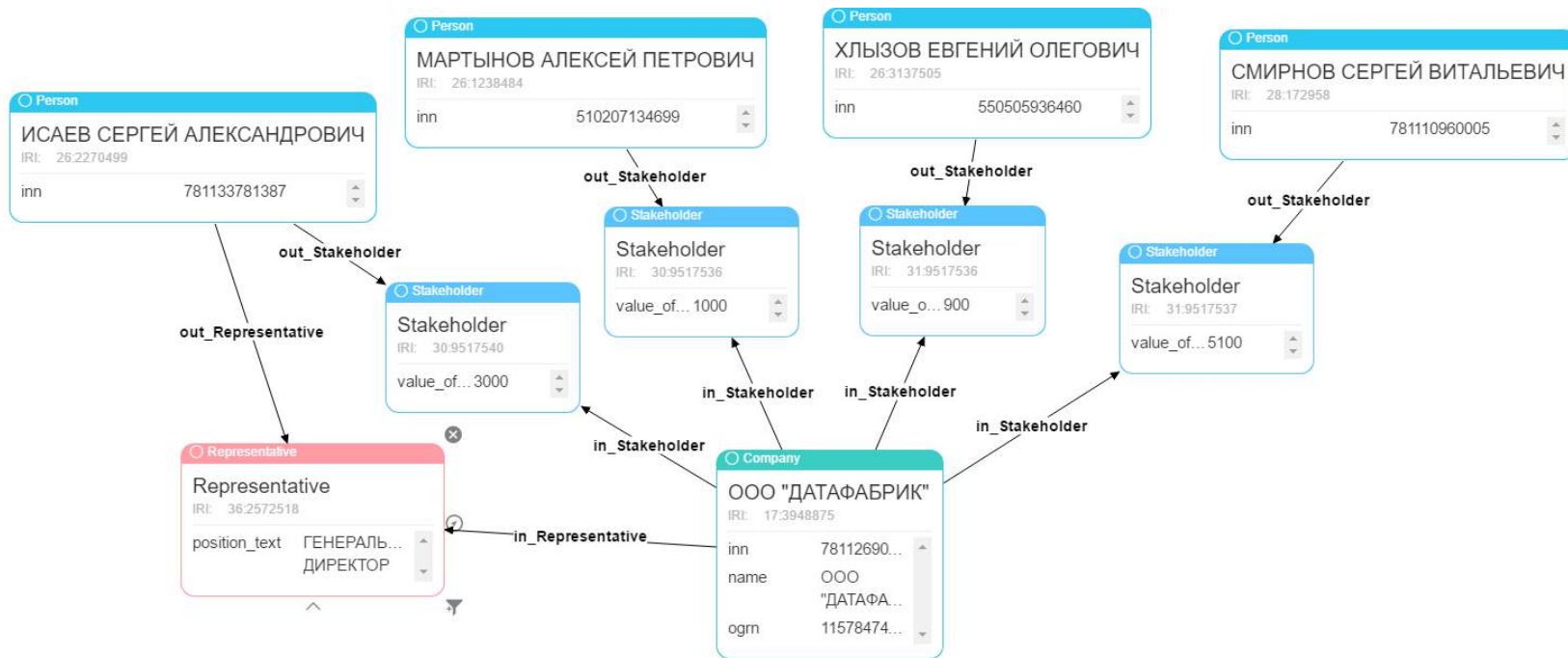


Fig. 1 Informal graph of the sample triples

Пример хранения знаний ФНС



Онтологическое моделирование предметной области

Обучение системы через описание области компетенций – делаем структуру мира

Мы сами рассказываем машине что означают те или иные данные.

Обучая машину, мы перекладываем на нее задачу понимать смысл данных. С человека на машину.

Machine Learning

У нас есть данные -> Мы обучаем систему

Онтологическое моделирование

Мы обучаем систему -> Система понимает данные



Онтология

Онтологии позволяют формально описать предметную область, связать множество таких описаний друг с другом, использовать описание для аннотирования данных.

Таким образом, данные становятся связанными как с общими знаниями, так и с другими данными, обеспечивается интерпретация данных на машинном уровне.

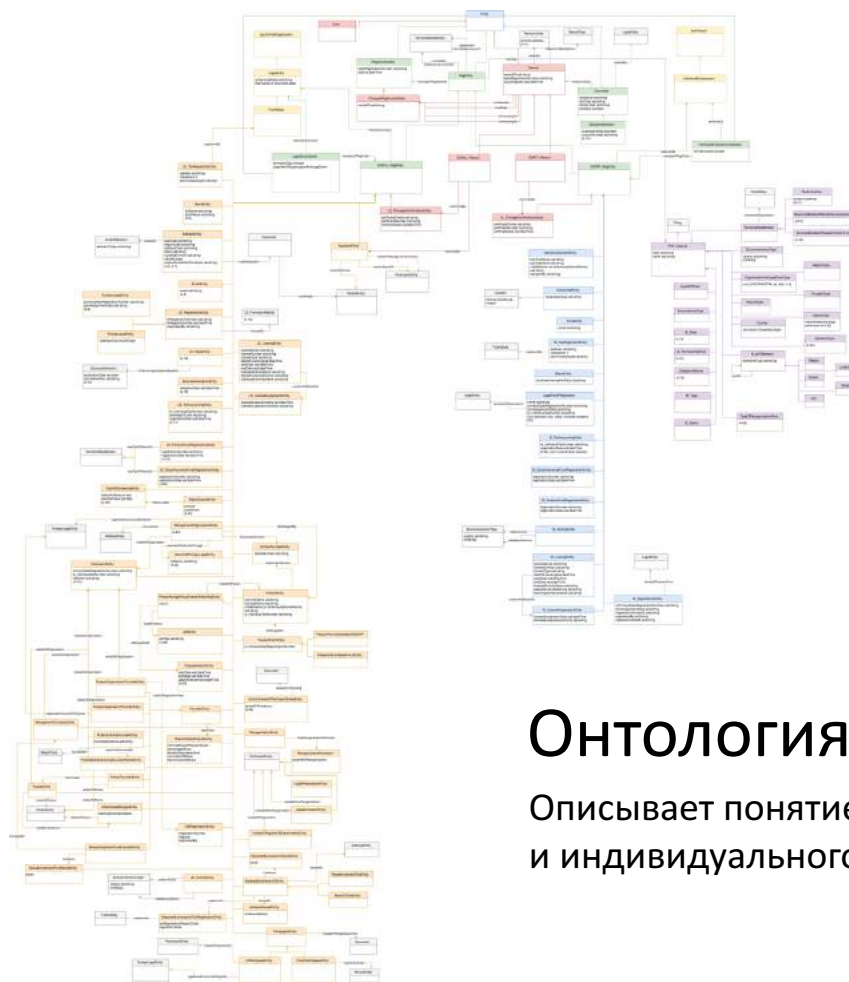
Linked Data – Связанные данные

Онтология

Использование стандартов позволяет комбинировать различные информационные системы на уровне понятий не прибегая к явному сопоставлению вида:

**"Колонка 1 таблицы А является тем же самым
что колонка 3 в Таблице Б"**





Онтология ЕГР ФНС

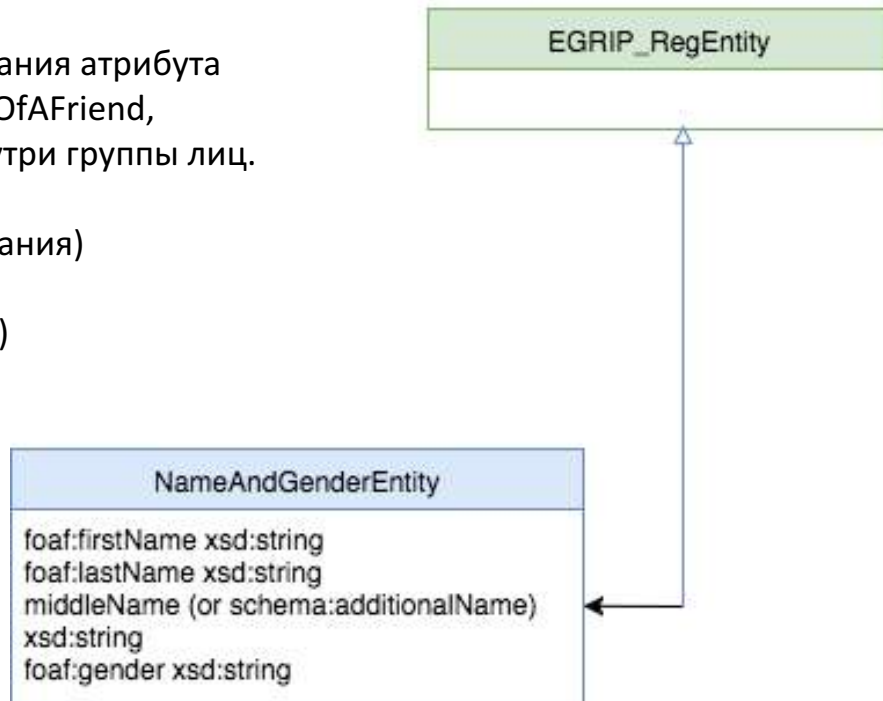
Описывает понятие юридического лица
и индивидуального предпринимателя

На фрагменте видно, что класс EGRIP_RegEntity является суперклассом для класса NameAndGenderEntity, Который, в свою очередь, имеет несколько атрибутов (например, foaf:name).

Префикс foaf в данном случае означает, что для описания атрибута используется стандартизированная онтология FriendOfAFriend, используемая для описания человека и связей внутри группы лиц.

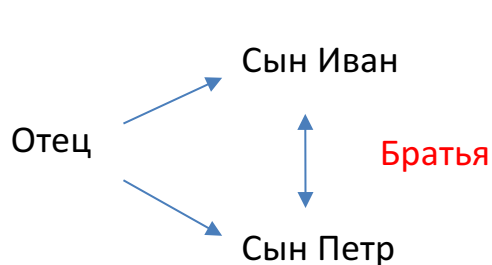
Таким образом, онтология определяет

- а) Классы (понятия) - (человек, записьВРеестре, компания)
- б) Атрибуты – (пол, имя, юридическоеНазвание)
- в) Отношения - (основал, являетсяПравопреемником)

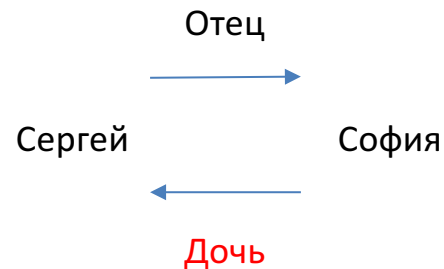


SPARQL – логический API

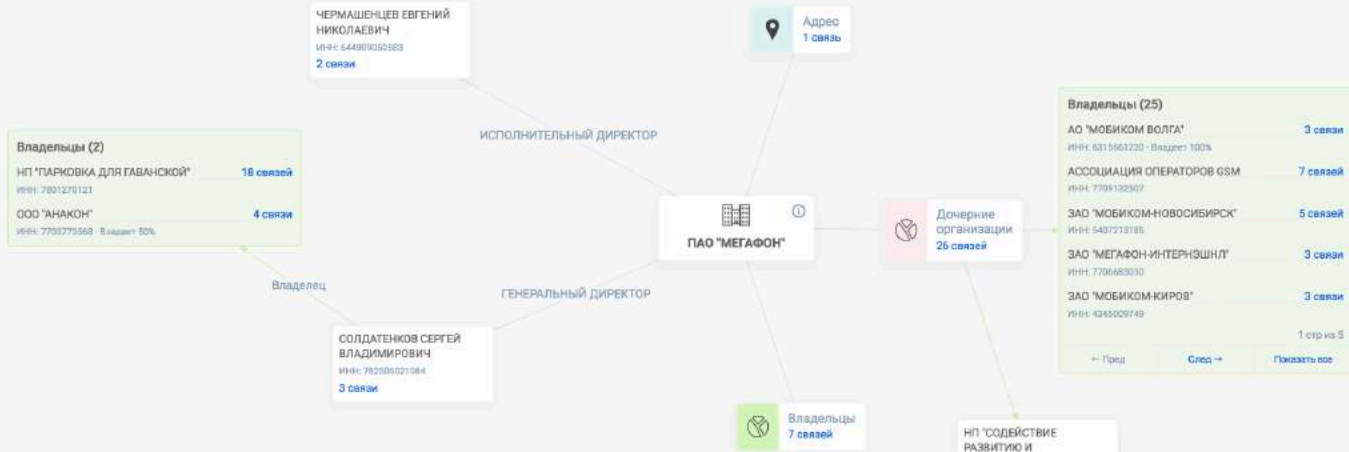
Доступ к данным на логическом уровне



Вычисляемые связи



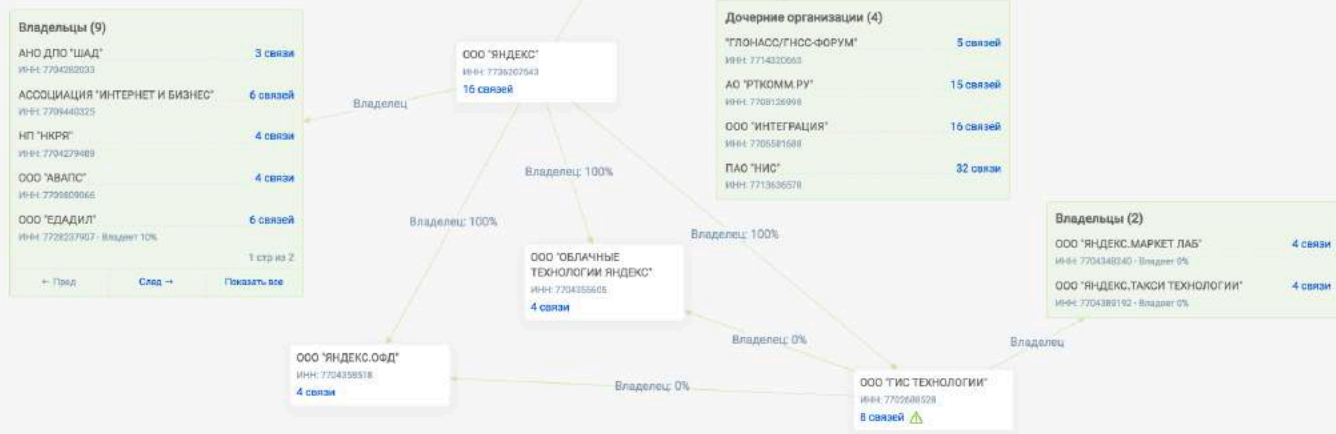
Обратное тождество

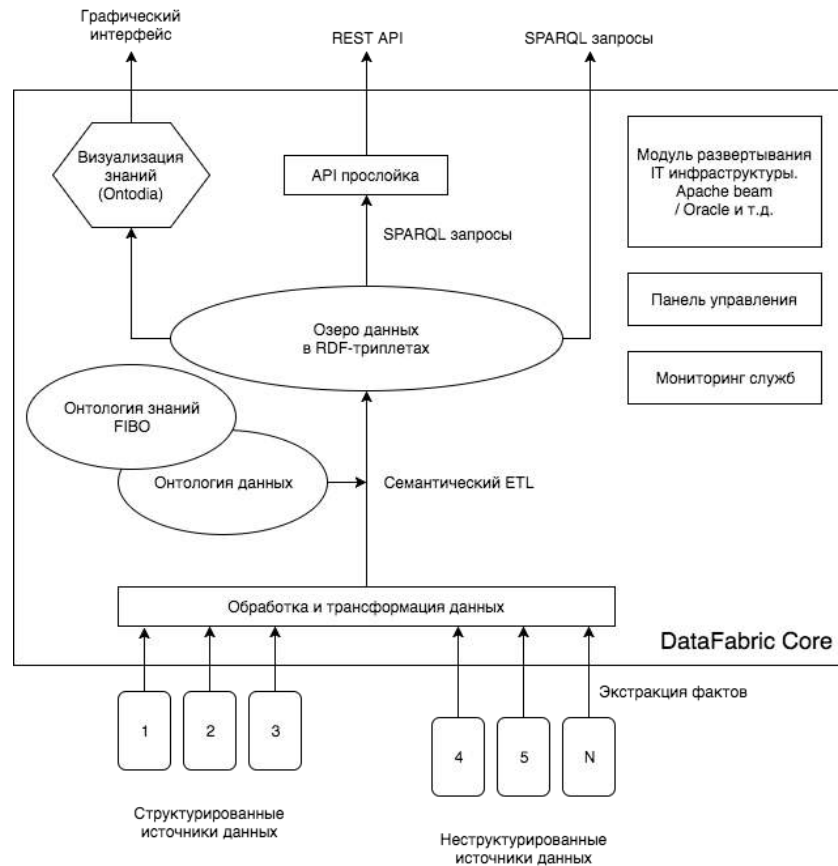


Кейс – связи компаний из ФНС

tree.datafabric.cc

Semantic Tech inside





Системы **визуальной аналитики** бизнес информации

Исаев Сергей
ООО «ДатаФабрик»
Сайт: DataFabric.cc
Тел: +7 (921) 345-65-43
Почта: isaev@datafabric.cc