

V Международный  
форум Финансового  
университета

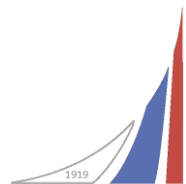


**КАК ПОПАСТЬ  
В ПЯТЕРКУ**

27-29 ноября 2018 г., Москва



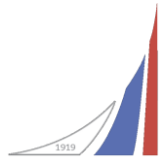
## Семантические технологии: новое качество обработки информации в бизнесе



# Финансовый университет

при Правительстве Российской Федерации

к.э.н. Сергей Вячеславович Макрушин,  
[SVMakrushin@fa.ru](mailto:SVMakrushin@fa.ru)



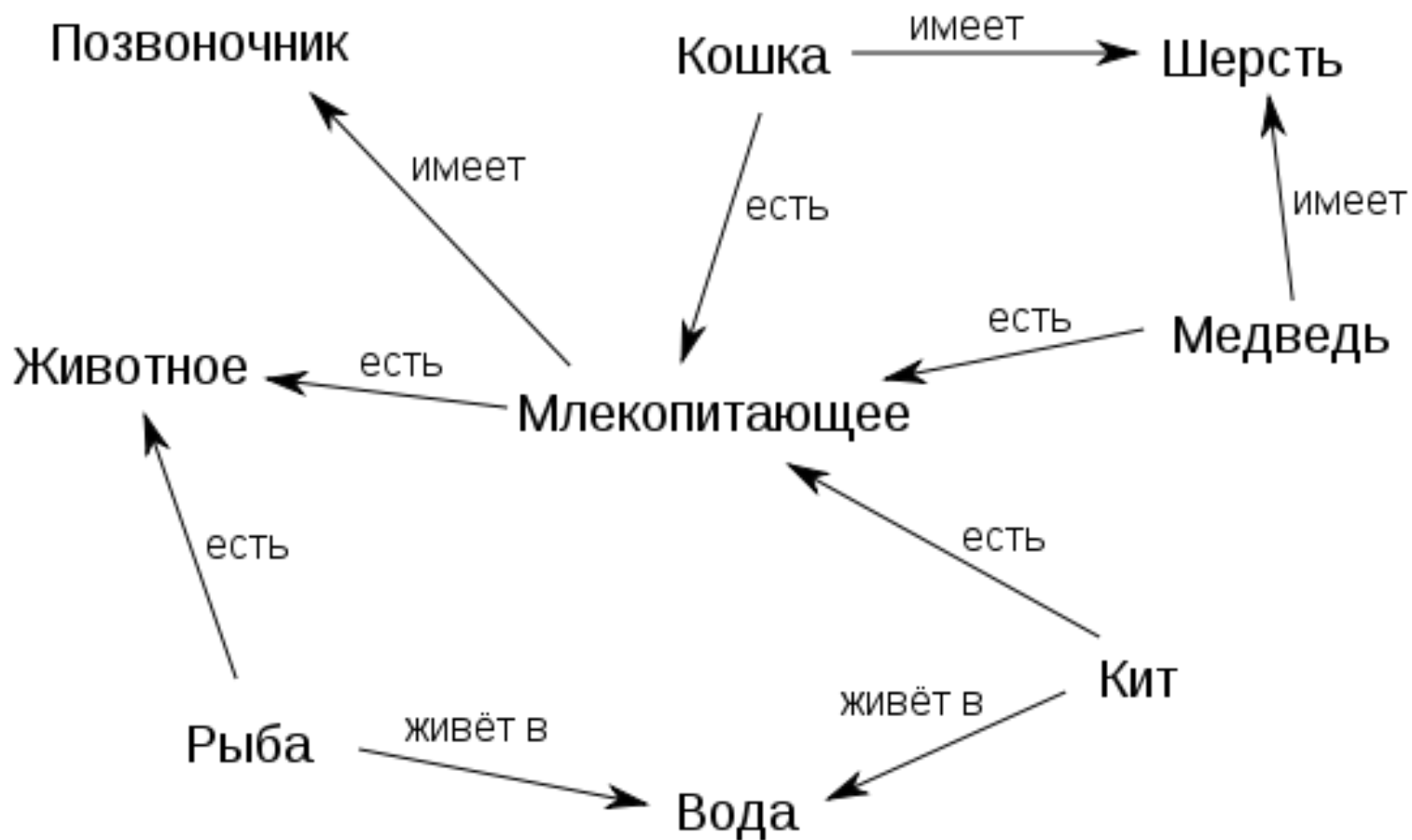
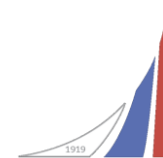
## Оперирование данными



## Оперирование смыслами

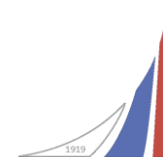










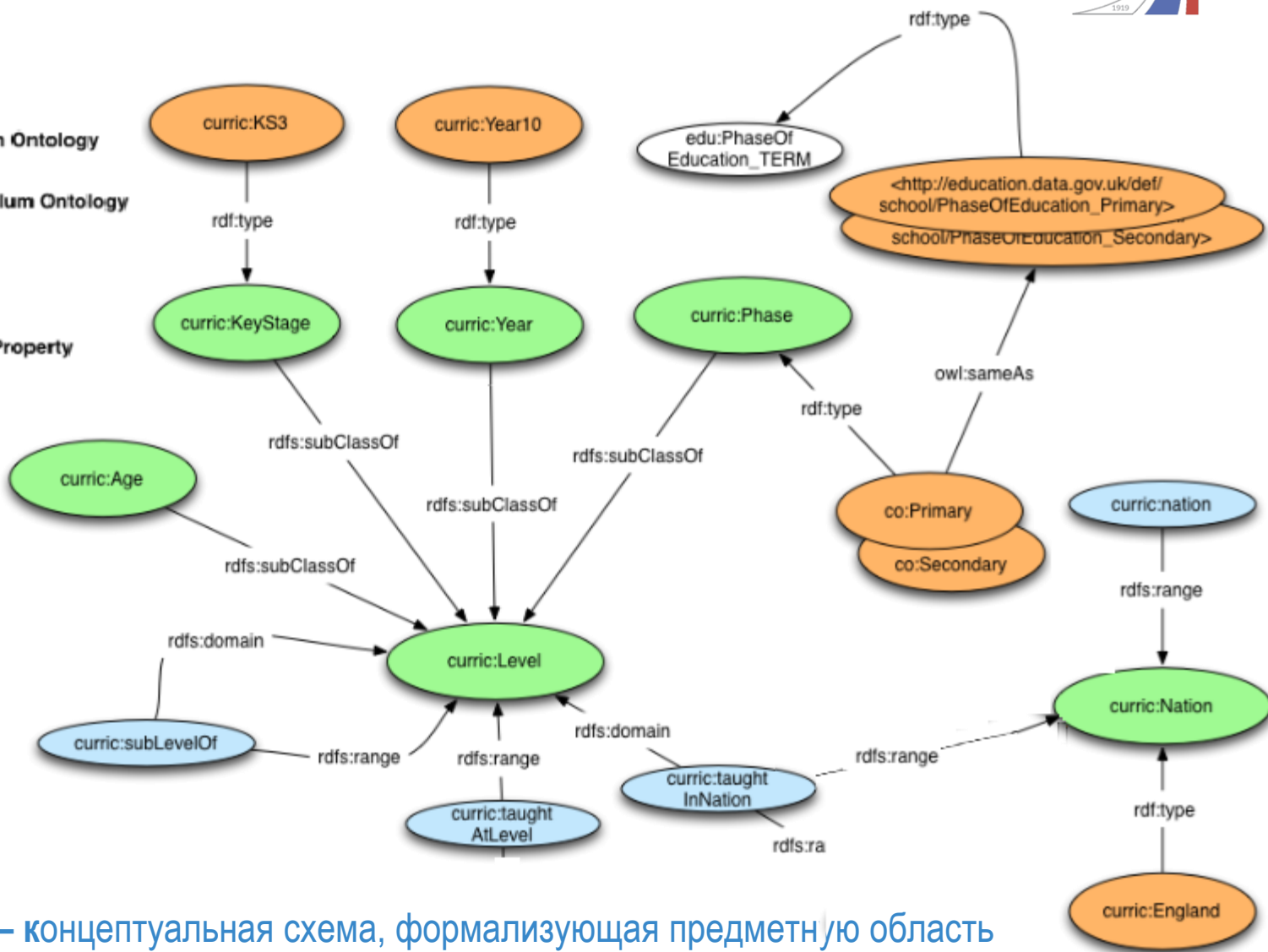
**Семантическая сеть** представляет знания в виде сети с ориентированными связями

- узлы сети – понятия (абстрактные или конкретные объекты)
- ориентированные связи – отношения между понятиями.



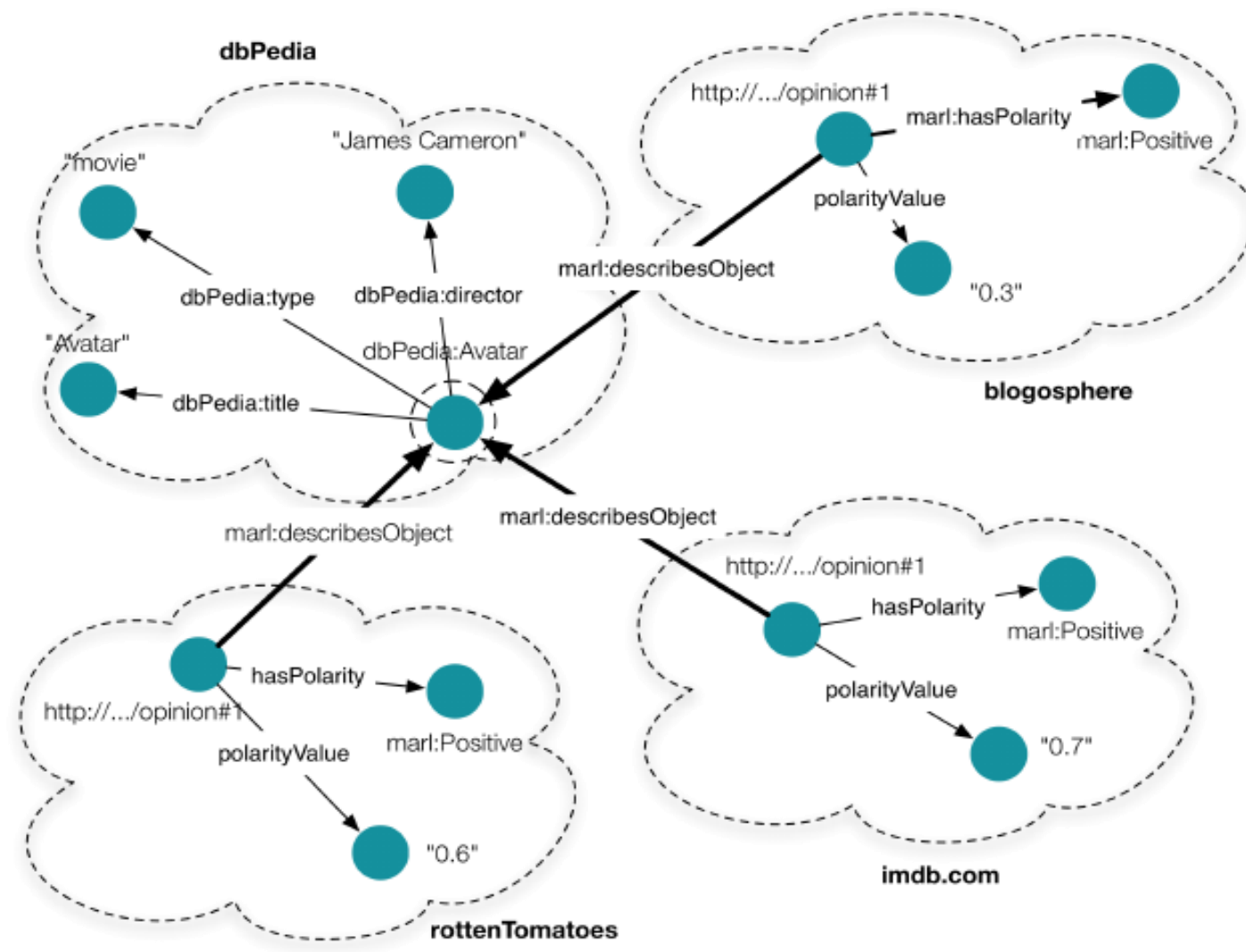
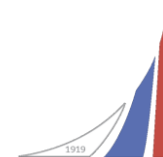
## BBC Curriculum Ontology

-  Class of Curriculum Ontology
-  Property of Curriculum Ontology
-  Instance
-  External Class or Property



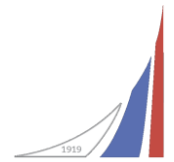
Онтология – концептуальная схема, формализующая предметную область

- содержит все основные классы объектов и связи между ними
- состоит из: понятий, экземпляров, атрибутов и отношений между ними

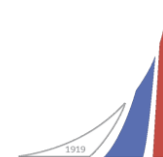


Семантическая паутина – WWW, преобразованная в глобальную семантическую сеть, пригодную для машинной обработки





Несмотря на активные изыскания в области создания семантической паутины в 2000х годах на данный момент вместо реализации глобального проекта развитие пошло по пути реализации прикладных семантических технологий.



## Why we Moved to Genesis Framework and ... - Crunchify

crunchify.com > Genesis Framework ▾

★★★★★ Rating: 5 - Review by App Shah

Crunchify Runs on Genesis Framework ... So far I'm loving it and I've fully customized my Crunchify Theme. ... Genesis Framework Review ... Rating: 5 out of 5. You've visited this page 2 times. Last visit: 4/16/16

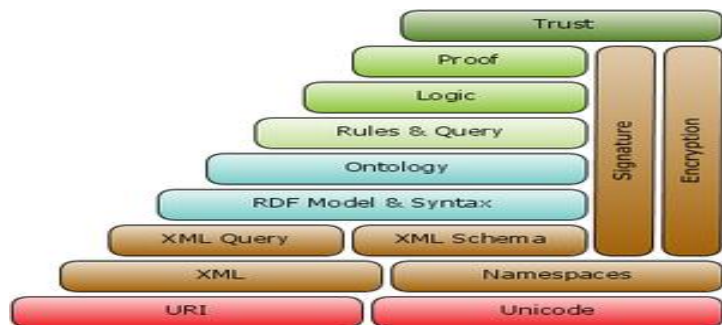
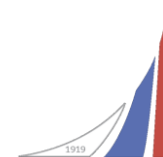
Rating widget  
Google Structured Schema.org Markup

```
<div class="crunchify-rating" itemscope itemtype="http://schema.org/Review">
  <span itemprop="itemReviewed" itemscope itemtype="http://schema.org/Thing" style="font-size: 14px;">
  <a itemprop="url" href="https://crunchify.com/refer/genesis" target="_blank"><span itemprop="name">
  <strong>Genesis WordPress Theme Framework Review</strong></a></span>
  <p></p>
  <strong>Description: </strong><span itemprop="description">One of the best WordPress Theme
  Framework which we use on Crunchify.com. Default comes with Google's rich schema.org snippet microdata.
  </span>
  <span itemprop="author" itemscope itemtype="http://schema.org/Person"><strong>Written by:</strong>
  <span itemprop="name">App Shah</span></span>
  <span itemprop="reviewRating" itemscope itemtype="http://schema.org/Rating"><strong>Rating:</strong>
  <span itemprop="ratingValue">5</span> out of <span itemprop="bestRating">5</span></span>
  <span itemprop="publisher" itemscope itemtype="http://schema.org/Organization">
  <meta itemprop="name" content="Crunchify, LLC."/>
  </span>
</div>
```

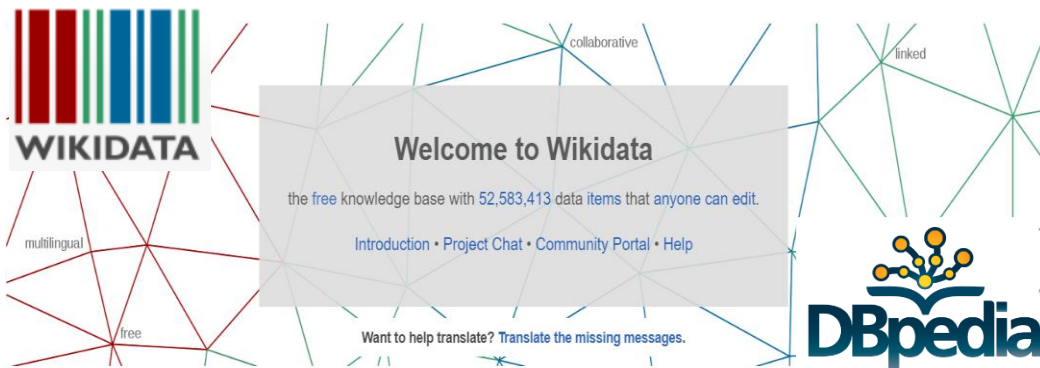


Использование метаданных в Web – пример реализации прикладных семантических технологий.

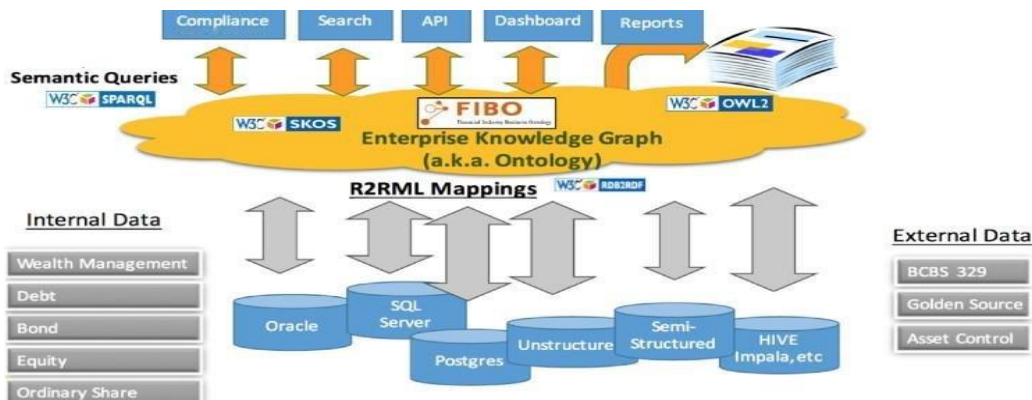




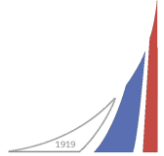
Стек семантических технологий



Крупные универсальные онтологии



Отраслевые онтологии

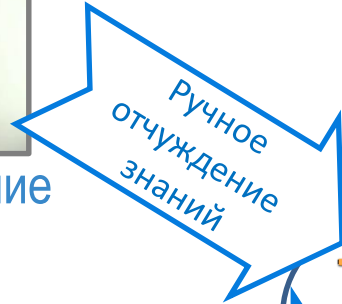


## СОЗДАНИЕ

## ИСПОЛЬЗОВАНИЕ



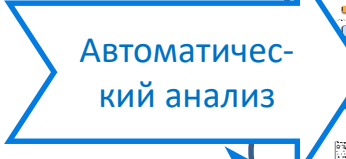
Экспертное знание



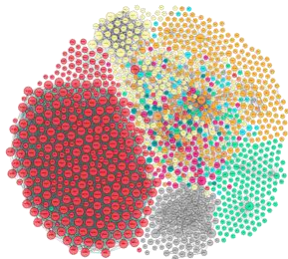
Ручное отчуждение знаний



Информационные массивы



Автоматический анализ

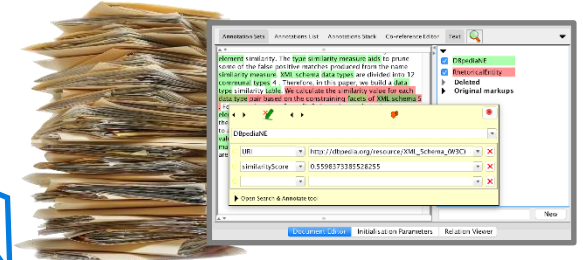


Внешние семантические модели

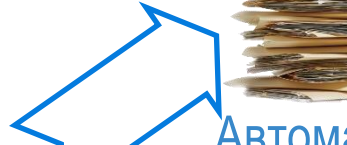


Обогащение

### Семантическая модель



Автоматическая разметка и семантическая интерпретация

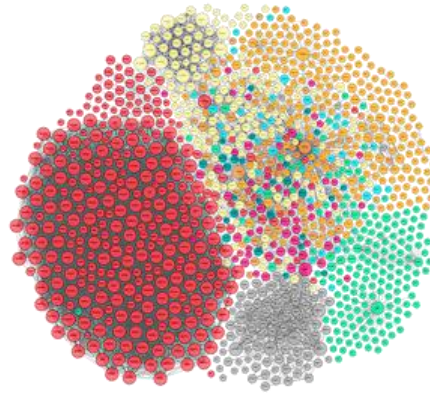
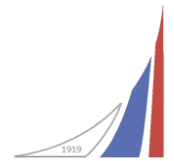


Мета модель для интеграции

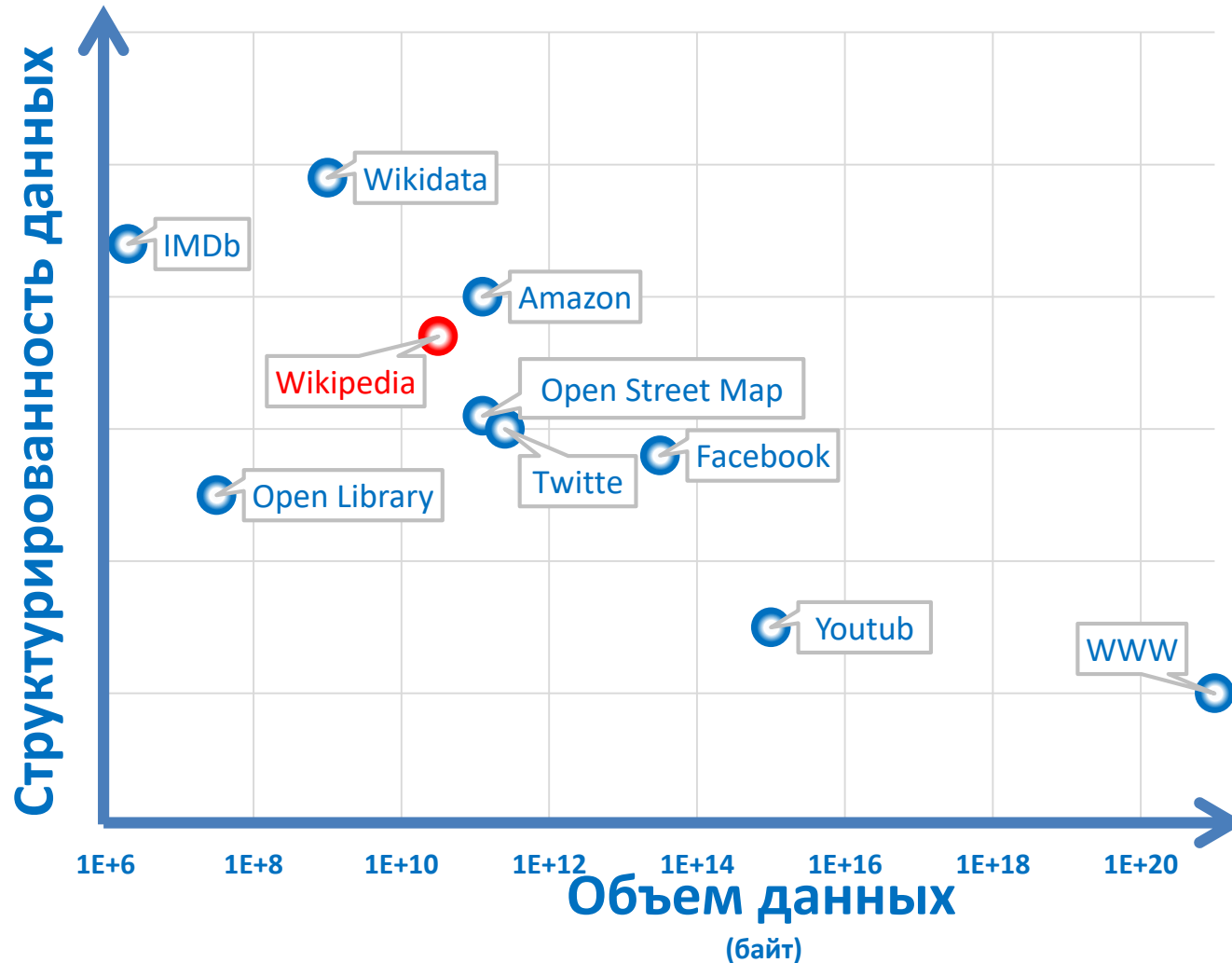
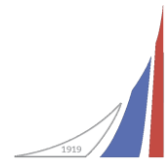


Автоматический вывод

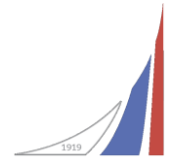




## Использование глобальных семантических моделей (на примере Википедии)



Википедия – крупнейший информационный артефакт современной эпохи, обладающий уникальным сочетанием доступности, полноты информации о мире и структурированности

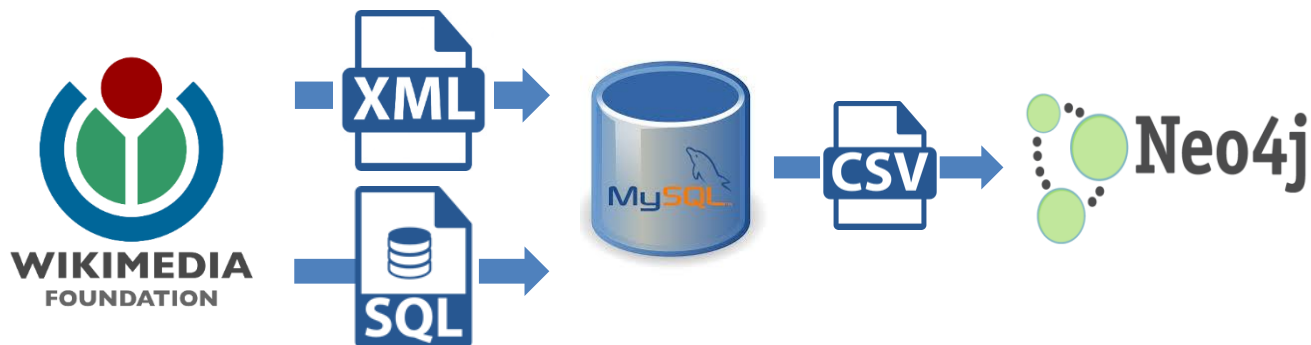
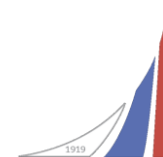


## Преимущества работы с Википедией:

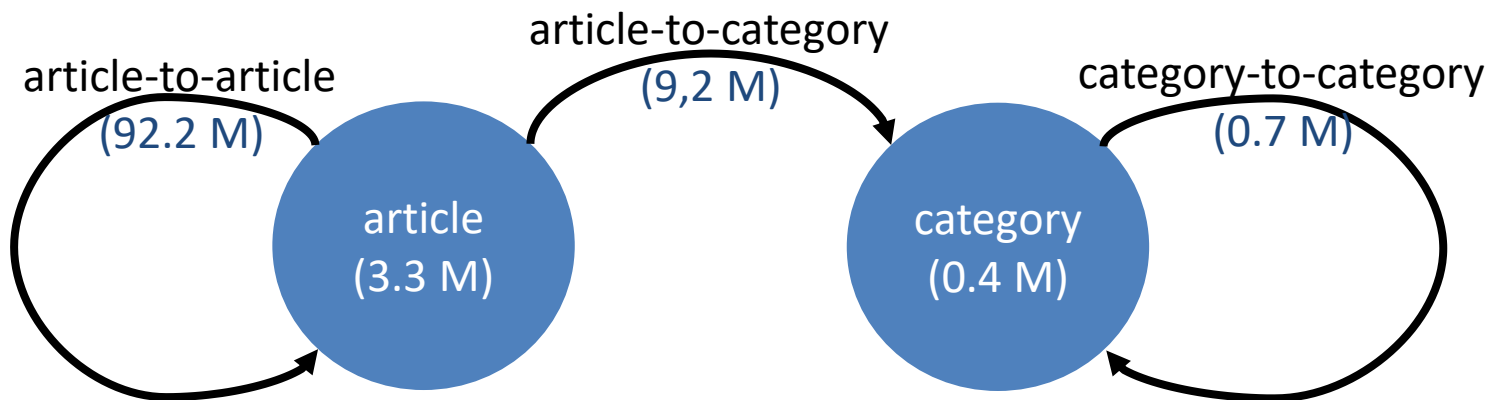
- контент распространяется свободно
- достаточно полно описывает окружающую реальность
- объем не является чрезвычайно большим
- информация является частично структурированной

На базе Википедии можно создавать эффективные прикладные семантические решения, прежде всего в области **семантического анализа документов** на естественном языке.





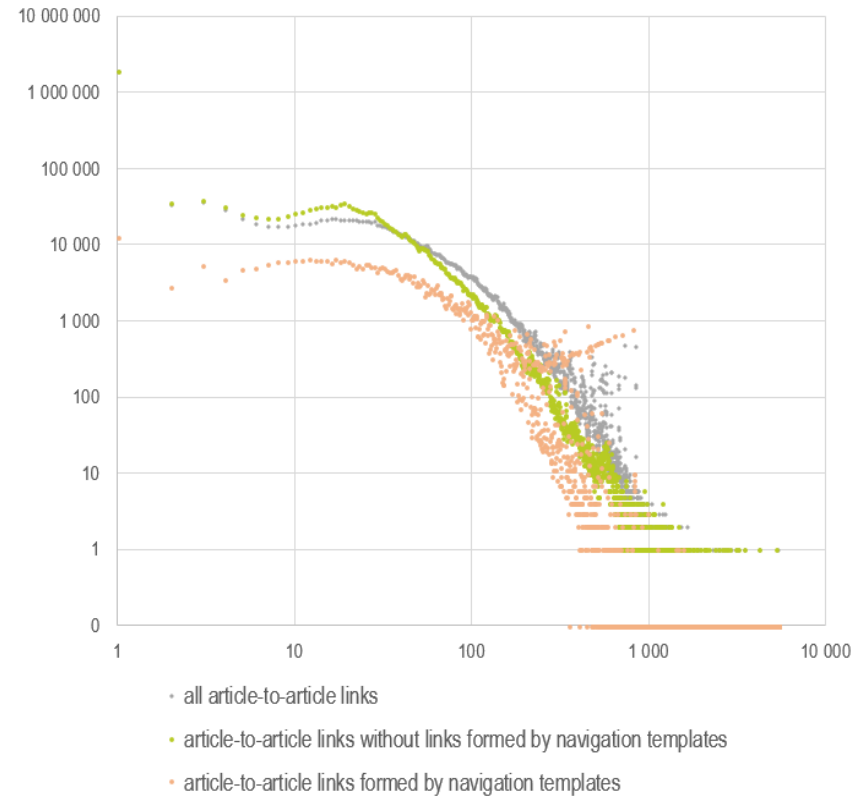
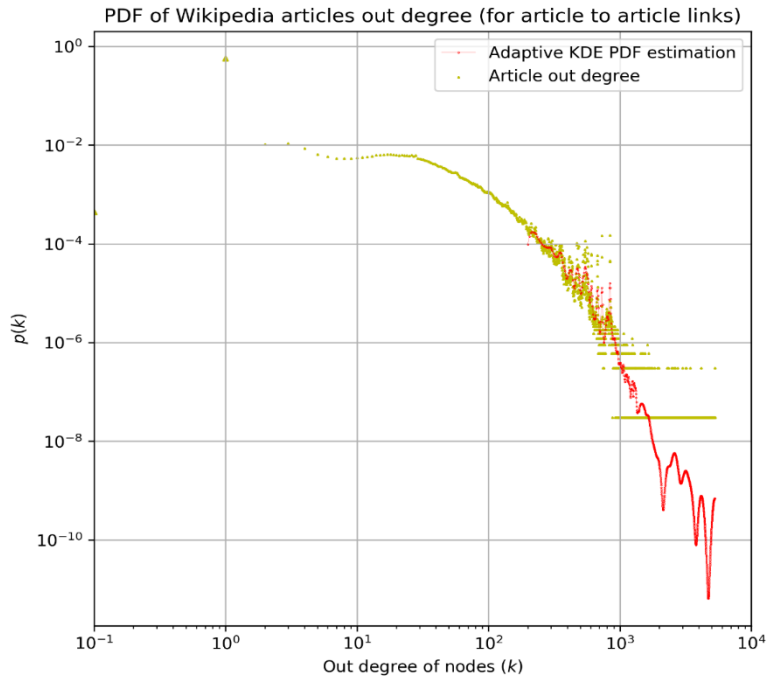
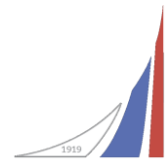
*Реализованная технология создания Вики-графа*



*Статистика Вики-графа русскоязычного сегмента Википедии*

**Вики-граф** – сеть (граф) из статей и категорий, представляющий основную семантическую информацию из Википедии

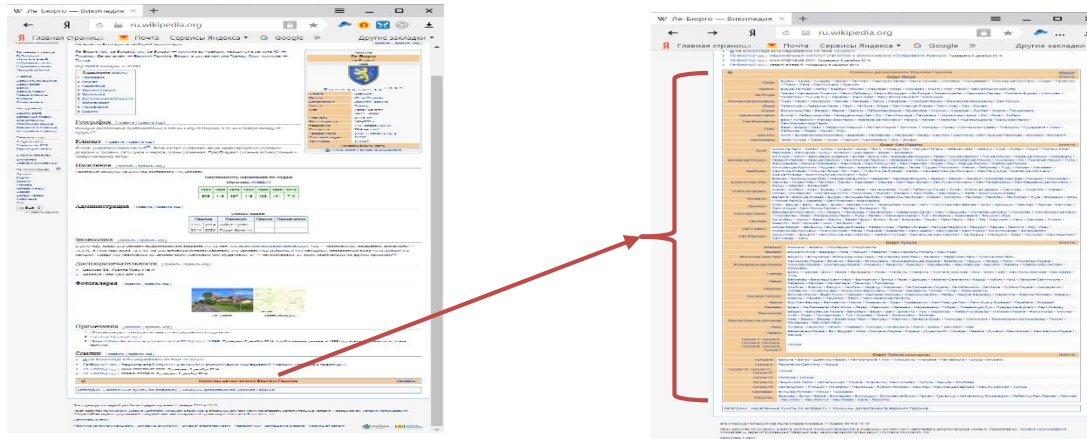
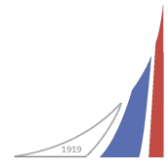
- **Узлы сети** – статьи и категории Википедии
- **Ориентированные связи** – ссылки между страницами и принадлежность к категориям



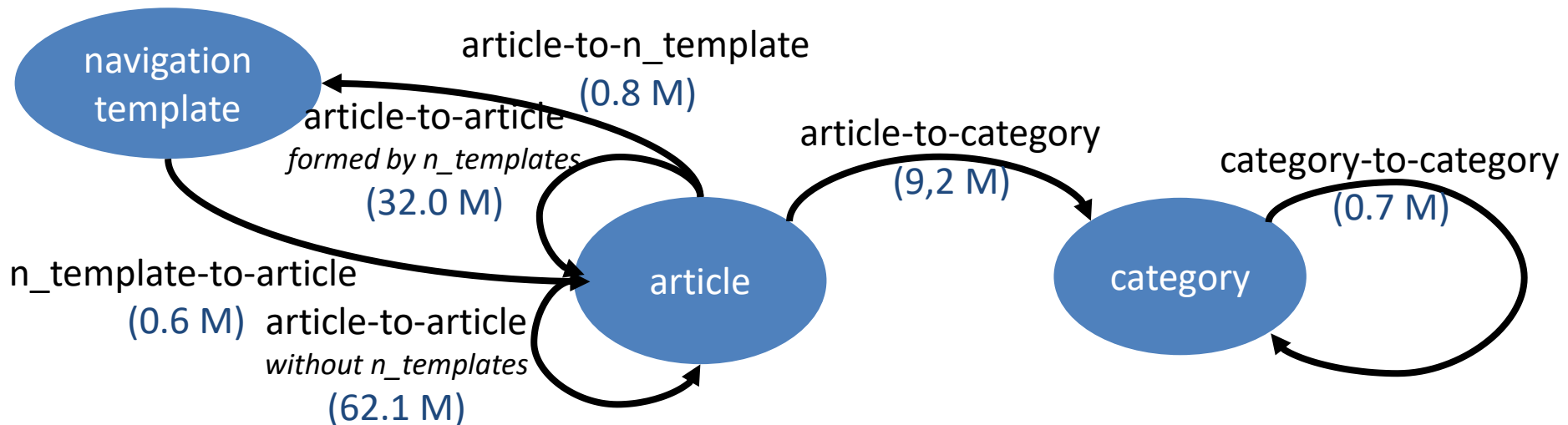
*График частоты статей с различным количеством исходящих ссылок (распределение исходящих степеней узлов) в построенном вики-графе\**

Применение методологии теории сложных сетей для анализа Вики-графа позволило выявить наличие аномалий в его структуре и избавиться от них скорректировав схему сети.

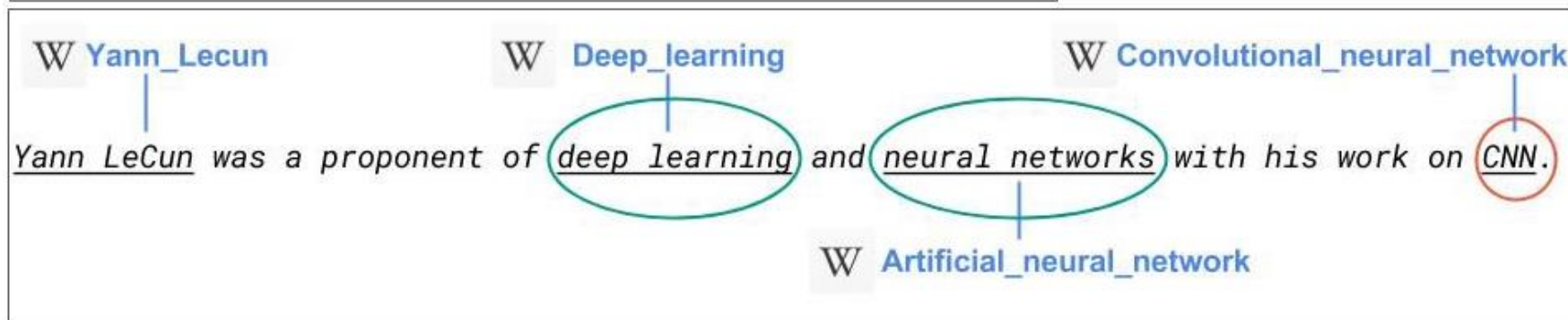
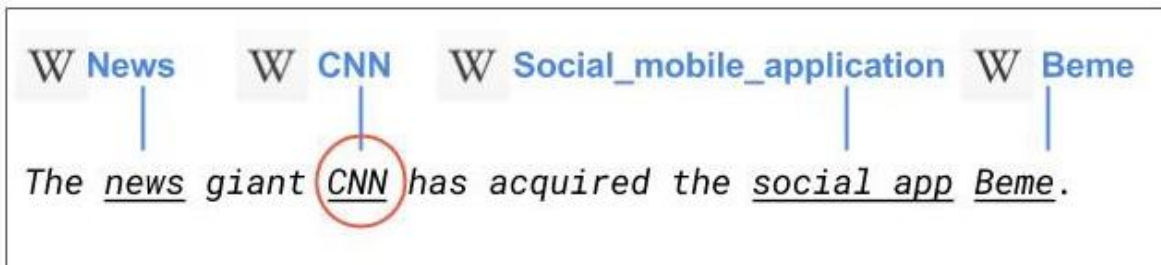
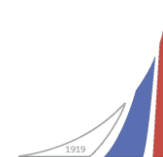
\* в log-log координатах



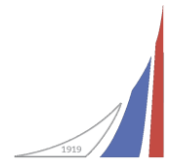
Навигационные блоки встроены в статьи с помощью шаблонов Википедии – страниц предназначенных для вставки стандартных блоков текста в другие страницы.



Модифицированная структура Вики-графа: **более 30% всех ссылок в Википедии сформировано в навигационных блоках (!)**

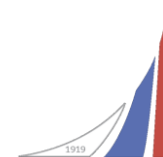


- Вики-граф позволяет **оценивать близость различных понятий** (за счет использования категорий и с помощью других подходов)
- Используя Вики-граф можно «викифицировать» произвольный текст – выявить в нем сущности, описанные в Википедии, и тем самым **построить семантический портрет документа и автоматически связать его с другими документами**
- Вики-графом можно обогатить корпоративную базу знаний и создать **специализированный инструмент семантического анализа текстов**



# Семантический анализ больших информационных массивов (на примере Инстаграм)

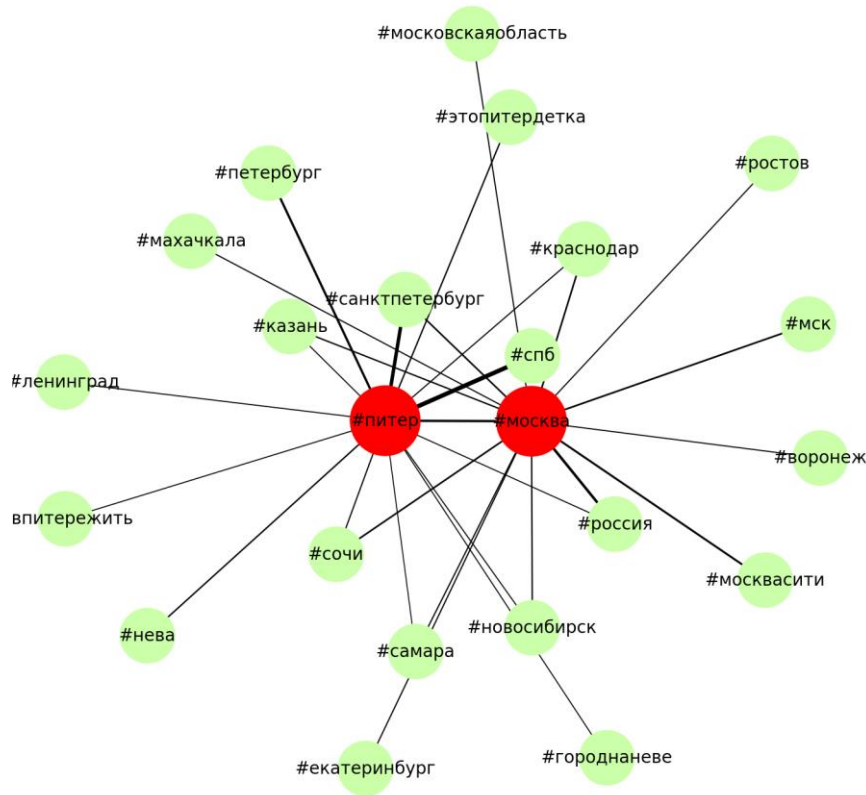
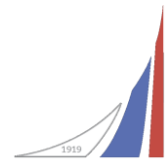




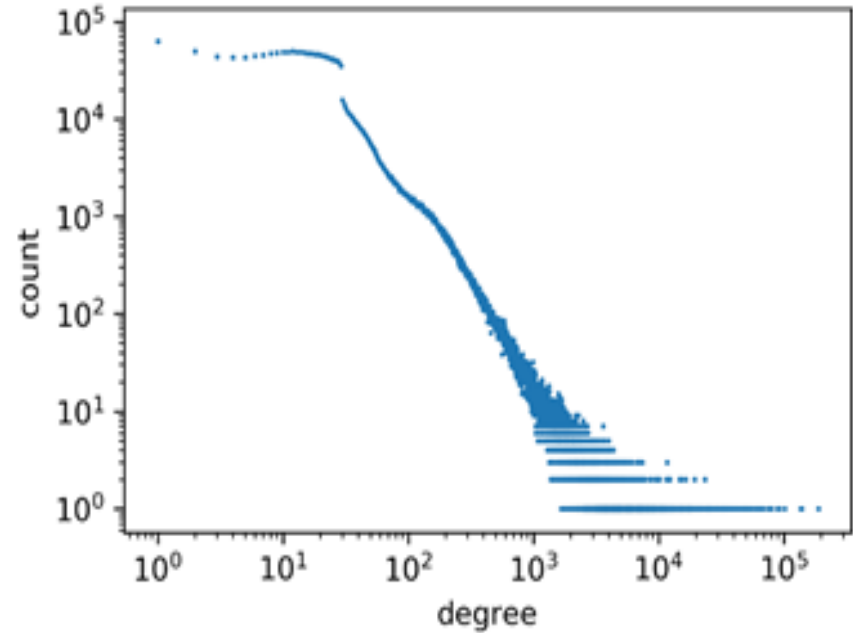
ibragimovme Пишу статью о популярных хеш тегах в #instagram #follow #followme #followforfollow #followme #follower #followers #alwaysfollowback #followbackalways #teamfollowback #followbackteam #f4f #ifollow #followall #followhim #follower #pleasefollow #pleasefollowme #me #likeforlike #like4like #likeback #instagood #tweegram #photooftheday #iphonesia

**Хэштеги** – новый инструмент семантической навигации, позволяющий искать информацию, привлекать читателей, пользующийся большой популярностью в современных коммуникационных средах.

**Цель проекта:** реализовать рекомендательную систему по подбору хэштегов

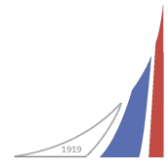


*Пример графа совместной встречаемости*



*График частоты хэштегов с различным количеством связей*

На базе основе анализа десятков миллионов сообщений из Инстаграм был построен **граф совместной встречаемости для хэштегов**: сеть из **6,13 миллионов узлов (хэштегов)** и **63,89 миллионов связей** между ними.



## word2vec

Input:  
one document

Lorem ipsum dolor  
sit amet, consele-  
tur sadipecing elit,  
sed diam nonumy  
eirmod tempor  
invidunt ut labore  
et dolore magna  
aliquyam erat, sed  
diam voluptua. At  
vero eos et

word  
vectors  
↓

Model:



most\_similar('france'):

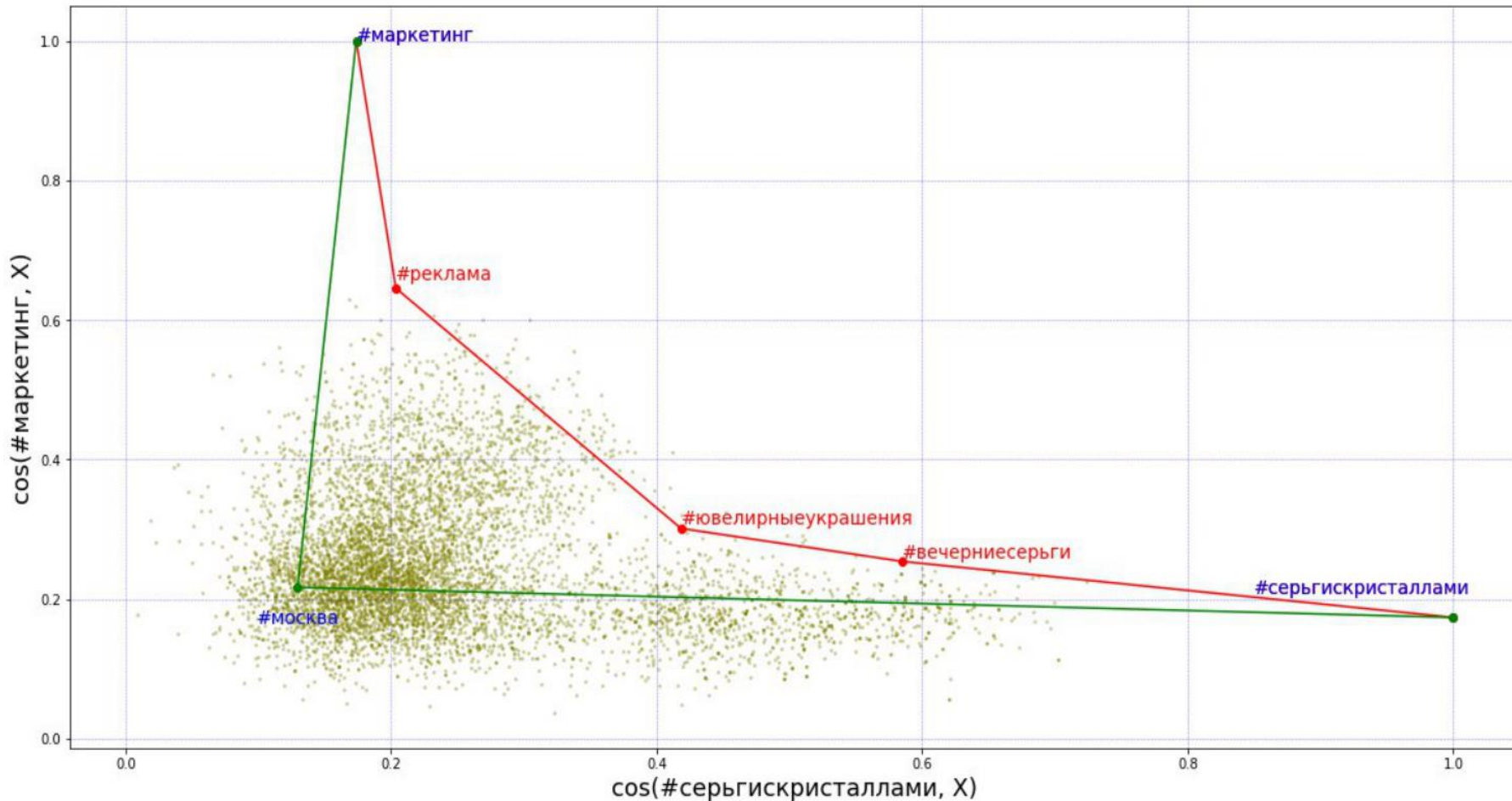
spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130

highest cosine  
distance values  
in vector space  
of the nearest  
words

*Принцип векторного представления слов (embedding), получаемого по технологии word2vec*

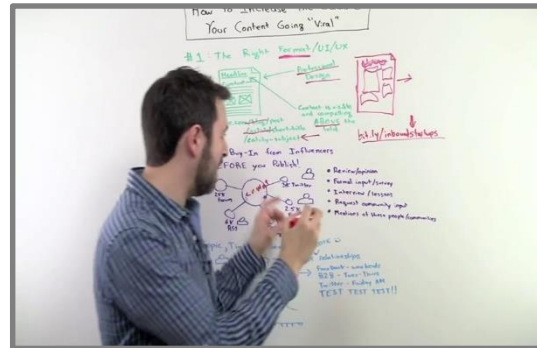
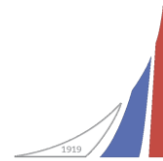
Используя технологию word2vec для каждого узла сети совместной встречаемости была определена **координата в пространстве** большой размерности. Расстояние в этом пространстве можно рассматривать как **метрику семантической близости** слов.

Используя информацию о связности хэштегов в сети и расстояние между ними, полученное от word2vec, был **разработан алгоритм быстрого поиска «семантического пути»** (ассоциативной цепочки) – последовательности семантически связанных тегов, соединяющих два произвольных хэштега.



*Пример построения семантических путей между произвольными хэштегами*

- Решение подходит для автоматической рекомендации дополнительных хэштегов
- На базе созданных технологий можно строить специализированные решения автоматического семантического связывания



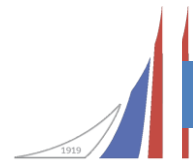
## Работа с экспертными знаниями (на примере создания метамодели предметной области)





## Цели проекта:

- Формирование корпоративной базы знаний объектных моделей КИС.
- Формирование методологии идентификации элементов объектной модели в формате, позволяющем ее автоматизированное использование.
- **Поддержка идентификации сущностей и связей** на этапе анализа при развитии КИС.



### Стандартные поля:

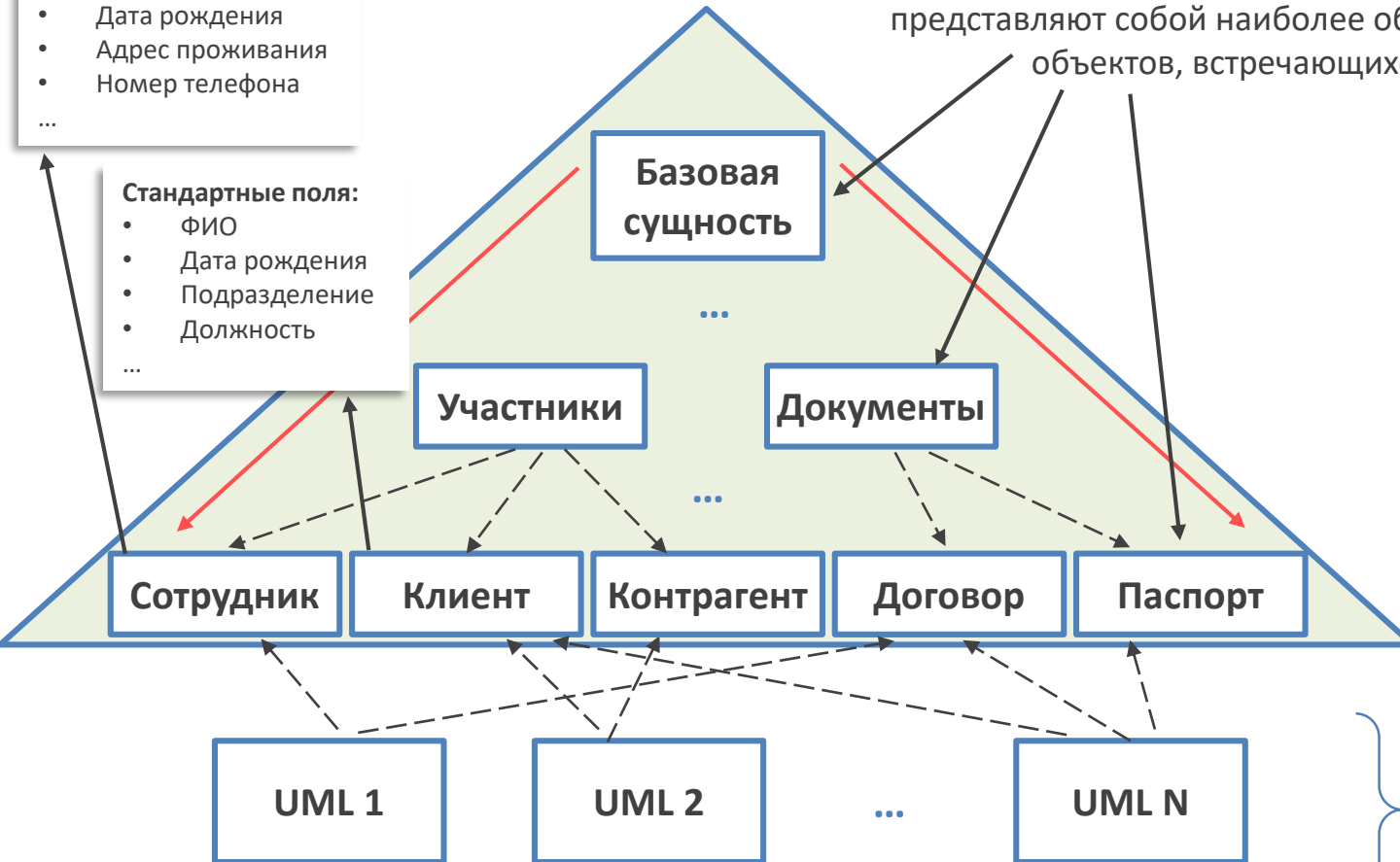
- ФИО
- Дата рождения
- Адрес проживания
- Номер телефона
- ...

### Стандартные поля:

- ФИО
- Дата рождения
- Подразделение
- Должность
- ...

### Метаклассы

представляют собой наиболее общее описание типовых объектов, встречающихся в разных ИС

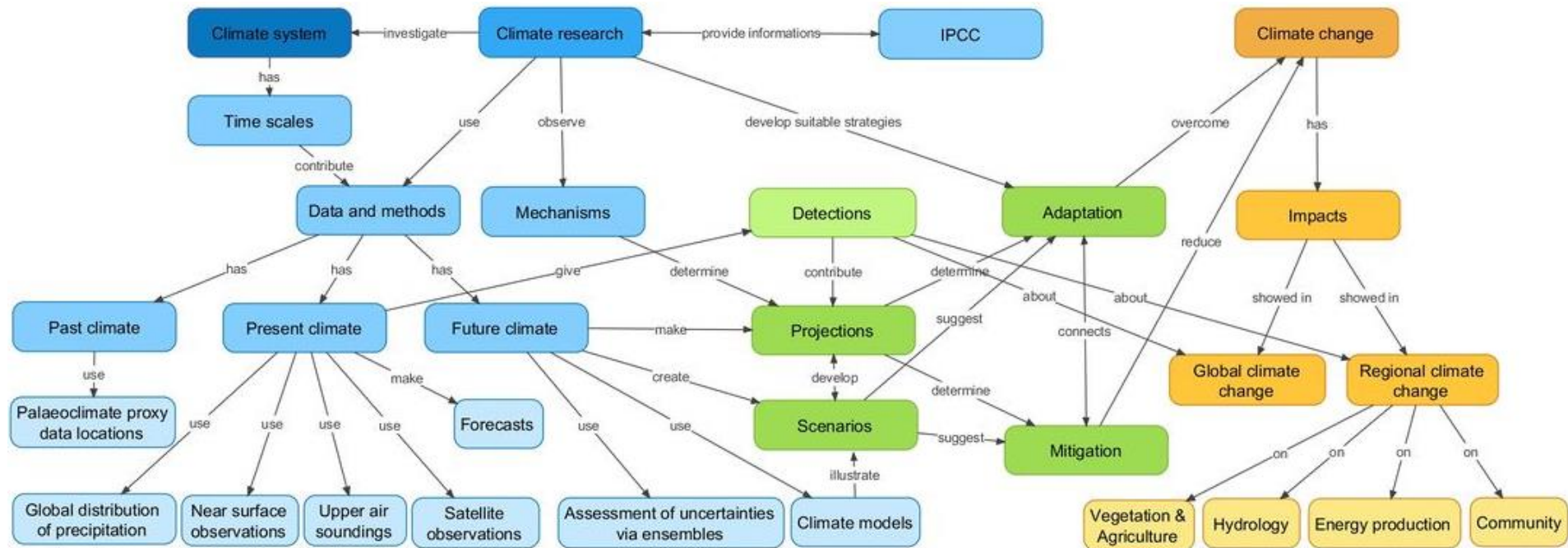
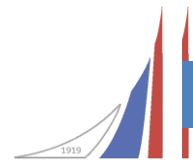


**Метамодель**  
система типовых объектов, взаимосвязей между ними и их свойств

**UML уровня конкретных ИС**  
(напр. диаграммы классов ИС)

### Требования к метамодели:

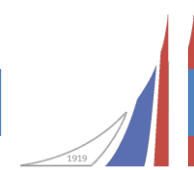
- Максимально гибкая структура
- Метаклассы можно соотнести с элементами UML-диаграмм



## Суть механизма Concept map:

- Максимально гибкая структура
- Любой узел графа может содержать произвольный набор свойств
- Любая связь может иметь произвольный тип и набор свойств

В частности, сущности и связи могут быть элементами нотации UML



### Jurisdiction (external)

30th January 2018 at 3:19pm

he limits or territory within which authority may be exercised; the power, right, or authority to interpret and apply the law

*Описание элементов диаграммы*

### monetary measure

30th January 2018 at 2:47pm

some measure of some sum of money

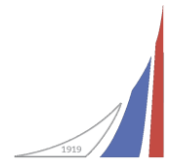
### Currency

30th January 2018 at 2:46pm

medium of exchange value, defined by reference to the geographical location of the authorities responsible for it



- Используется Вики-движок с поддержкой создания Concept map
- Используются стандартные сценарии добавления новых сущностей в метамодель
- UML в виде XML переводится в RDF XML, RDF загружается в Вики-движок



Спасибо за внимание!



к.э.н. Сергей Вячеславович Макрушин  
[SVMakrushin@fa.ru](mailto:SVMakrushin@fa.ru)