Основы обработки текстов

Лекция 10
Лексическая семантика
Извлечение информации

Возможные взгляды на семантику

- Лексическая семантика
  - –значение индивидуальных слов
- Композиционная семантика
  - –как значения комбинируются и определяют новые значения для словосочетаний
- Дискурс или прагматика
  - –как значения комбинируются между собой и другими знаниями, чтобы задать значение текста или дискурс

План

- Основные понятия
  - слова и отношения между ними
  - словари и тезаурусы
- Вычислительная семантика
  - Разрешение лексической многозначности
  - Привязка к базам знаний

Основные понятия

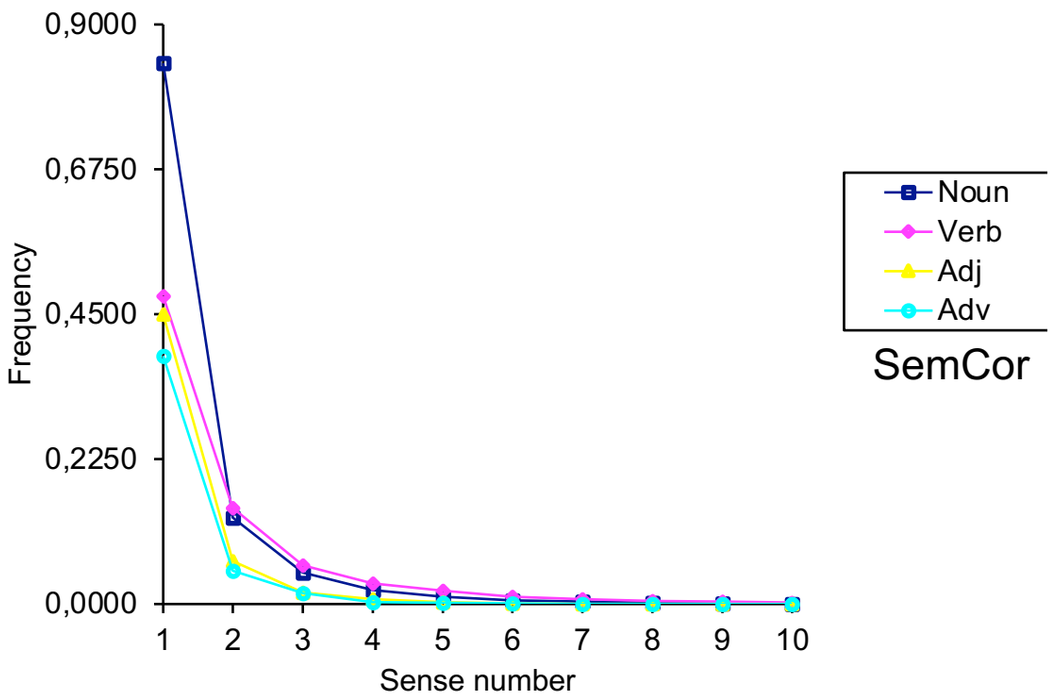
- Значение слова и многозначность
- Омонимия VS многозначность
  - –ключ
  - –платформа
- Метонимия
  - – Я три *тарелки* съел
- Зевгма
  - –За окном шел снег и рота красноармейцев
- Типы омонимов
  - –омофоны (луг-лук, плод-плот)
  - –омографы (м’ука - мук’а, гв’оздик-гвозд’ик)

Отношения между словами

- Синонимия
  - Машина / автомобиль
- Антонимия
  - большой / маленький, вверх / вниз, ложь / истина
- Обобщение и детализация (hyponym and hypernym/superordinate)
  - машина - транспорнтое средство
  - яблоко - фрукт
- Меронимы (партонимы) и холонимы
  - колесо - машина

Многозначность на практике

- Text-to-Speech
  - омографы
- Информационный поиск
- Извлечение информации
- Машинный перевод
- Закон Ципфа (Zipf law)



Основные вопросы

- Что такое значение?
  - –Сколько значений у слова “платформа”?
- Что нужно для того, чтобы понять значение?
  - –Что такое контекст?
  - –Как определить связь значения с контекстом?
    - Разреженность языка

Базы знаний

- (В широком смысле) **База знаний** — база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области
- (В узком смысле) **База знаний** — некоторое структурированное описание предметной области

WordNet

- База лексических отношений
  - содержит иерархии
  - сочетает в себе тезаурус и словарь
  - доступен on-line
  - разрабатываются версии для языков кроме английского (в т.ч. для русского)

Категория	**Уникальных форм**
Существительные	**117,097**
Глаголы	**11,488**
Прилагательные	**22,141**
Наречия	**4,601**

- http://http://wordnet.princeton.edu/
- http://wordnet.ru/

Формат WordNet

The noun "bass" has 8 senses in WordNet.

1. bass[1] - (the lowest part of the musical range)
2. bass[2], bass part[1] - (the lowest part in polyphonic music)
3. bass[3], basso[1] - (an adult male singer with the lowest voice)
4. sea bass[1], bass[4] - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass[1], bass[5] - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass[6], bass voice[1], basso[2] - (the lowest adult male singing voice)
7. bass[7] - (the member with the lowest range of a family of musical instruments)
8. bass[8] - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective "bass" has 1 sense in WordNet.

1. bass[1], deep[6] - (having or denoting a low vocal or instrumental range)
   "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"

WordNet: отношения между словами

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Has-Instance		From concepts to instances of the concept	$composer^1 \rightarrow Bach^1$
Instance		From instances to their concepts	$Austen^1 \rightarrow author^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Opposites	$leader^1 \rightarrow follower^1$

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	$increase^1 \Longleftrightarrow decrease^1$

Иерархии WordNet

```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living thing, animate thing,
                     => whole, unit
                        => object, physical object
                           => physical entity
                              => entity
                  => causal agent, cause, causal agency
                     => physical entity
                        => entity

Sense 7
bass --
(the member with the lowest range of a family of
musical instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

Как “значение” определяется в WordNet

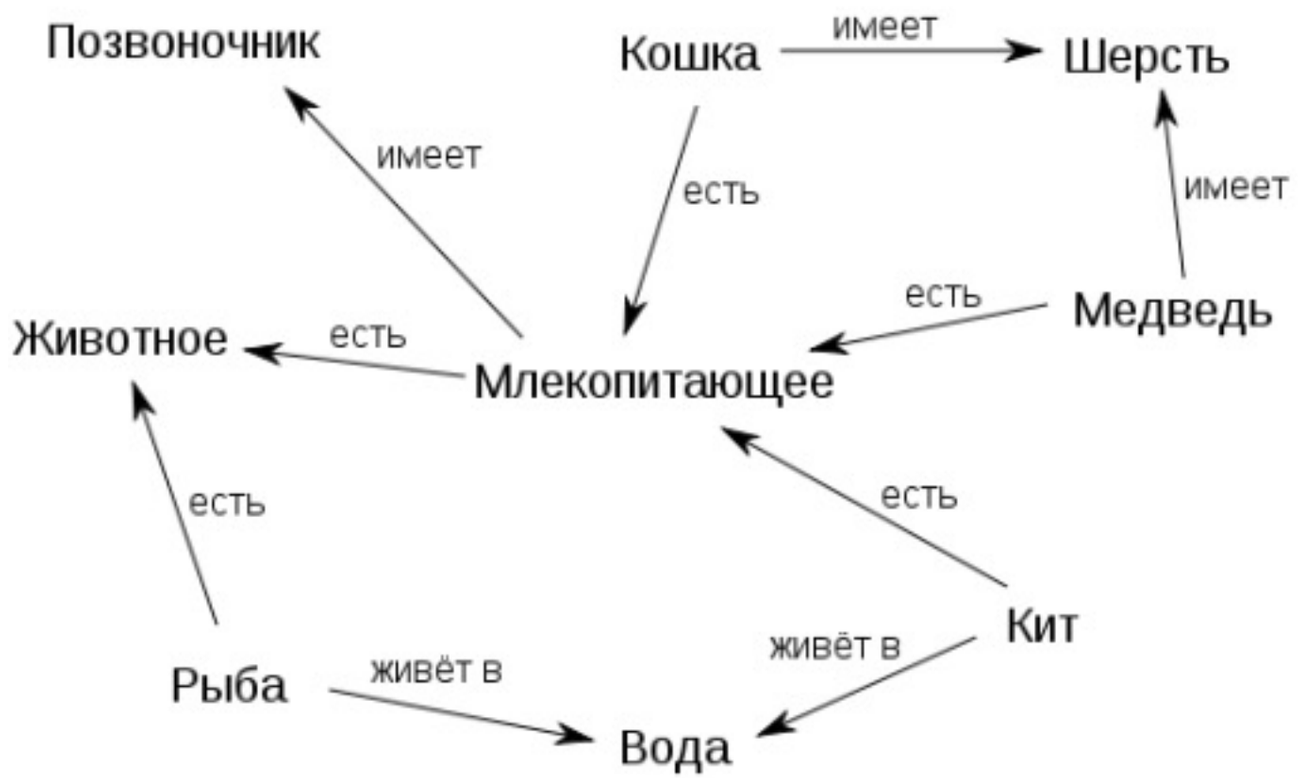
- Множество синонимов называется синсет
- Пример

```
from nltk.corpus import wordnet
for synset in wordnet.synsets(‘platform'):
    print(synset.definition)
    print([lemma.name for lemma in synset.lemmas])
```

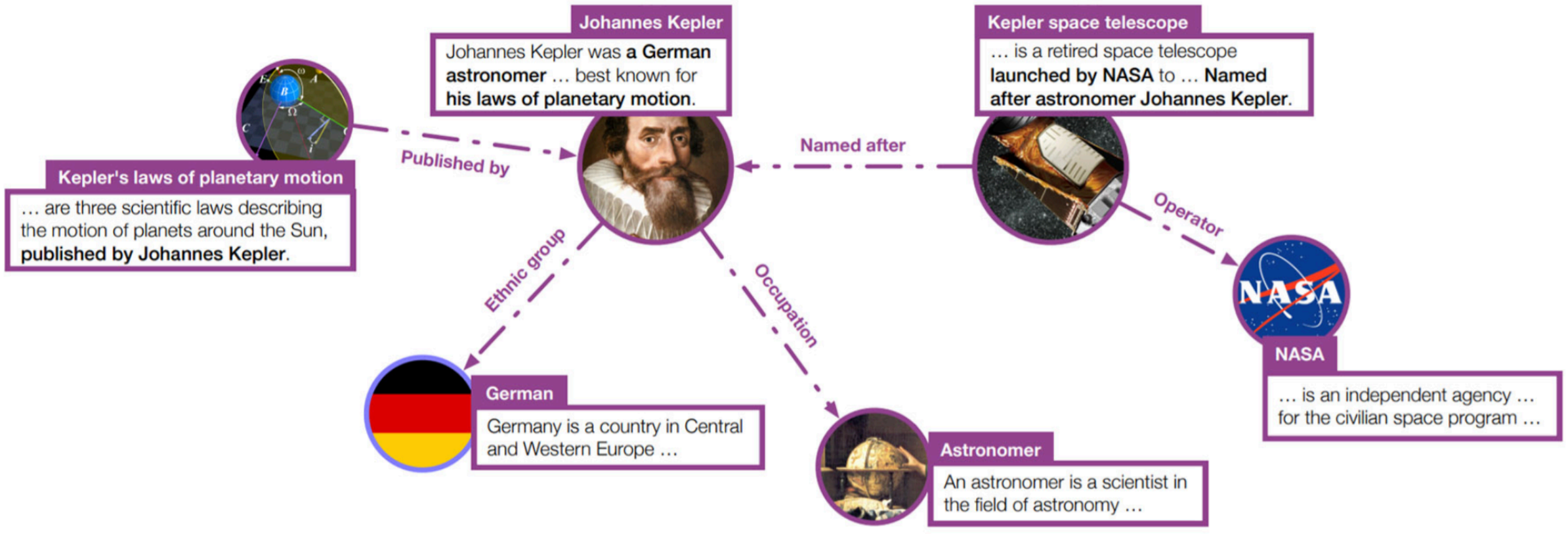
```
A document stating the aims and principles of a political party
['platform, 'political platform, 'political program, 'program']
Any military structure or vehicle bearing weapons
[platform, weapons platform]
…
```

Семантическая сеть

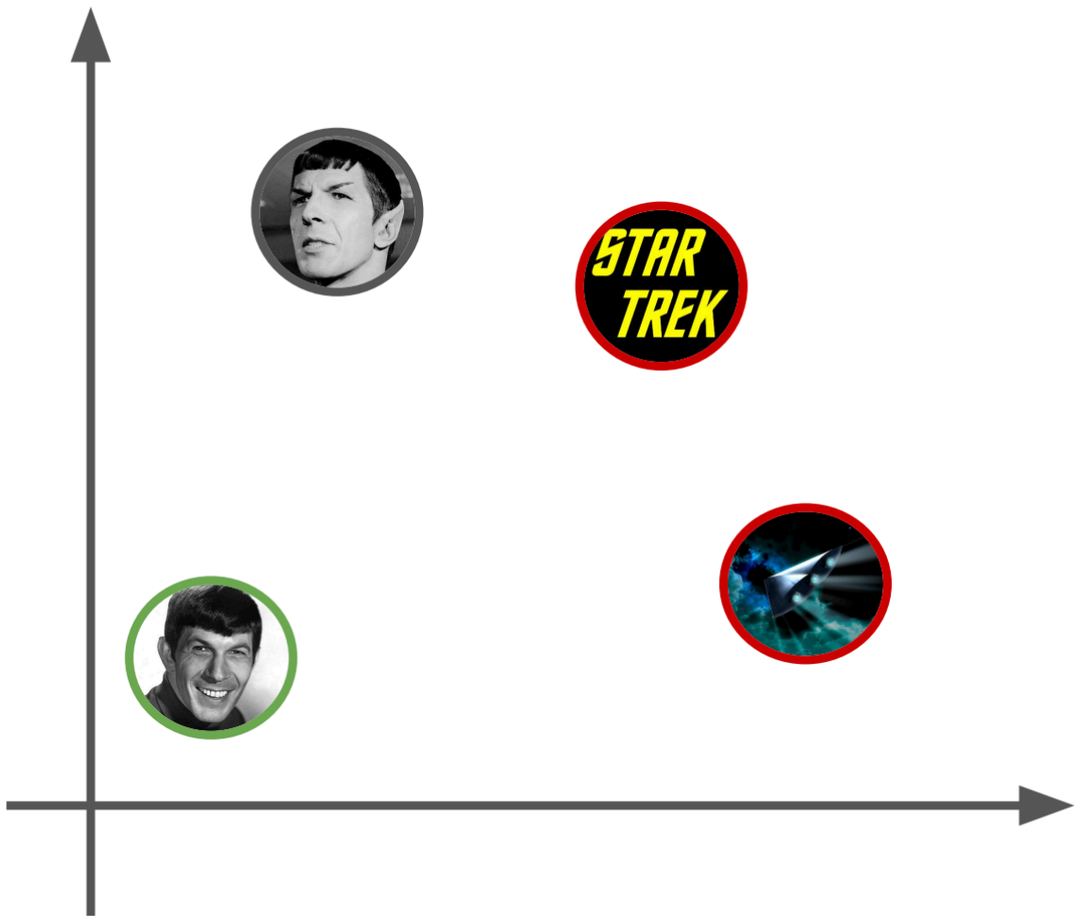
- **Семантическая сеть** — информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними.



Граф знаний

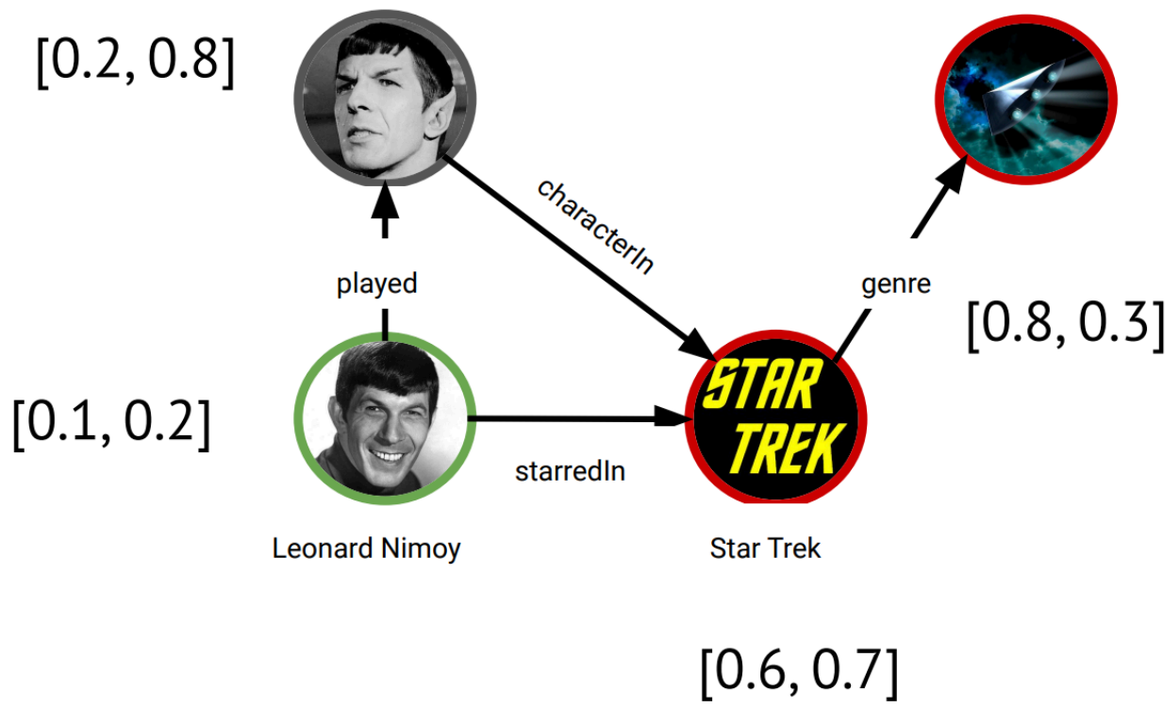

Johannes Kepler
Johannes Kepler was a German astronomer … best known for his laws of planetary motion.
Kepler space telescope
… is a retired space telescope launched by NASA to … Named after astronomer Johannes Kepler.
Kepler's laws of planetary motion
… are three scientific laws describing the motion of planets around the Sun, published by Johannes Kepler.
Published by
Named after
Operator
Ethnic group
Occupation
NASA
… is an independent agency … for the civilian space program …
German
Germany is a country in Central and Western Europe …
Astronomer
An astronomer is a scientist in the field of astronomy …

Векторное представление

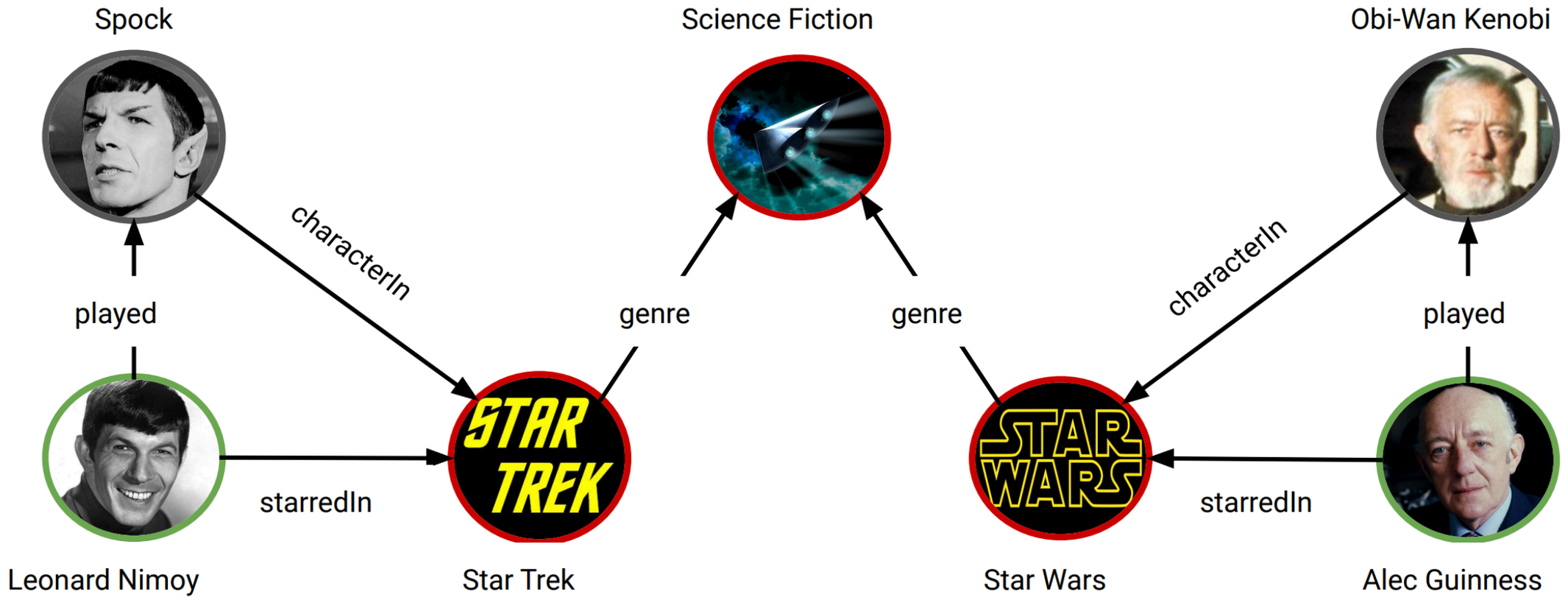


$$E \in \mathbb{R}^{N_e \times d}$$

$$R \in \mathbb{R}^{N_r \times d}$$



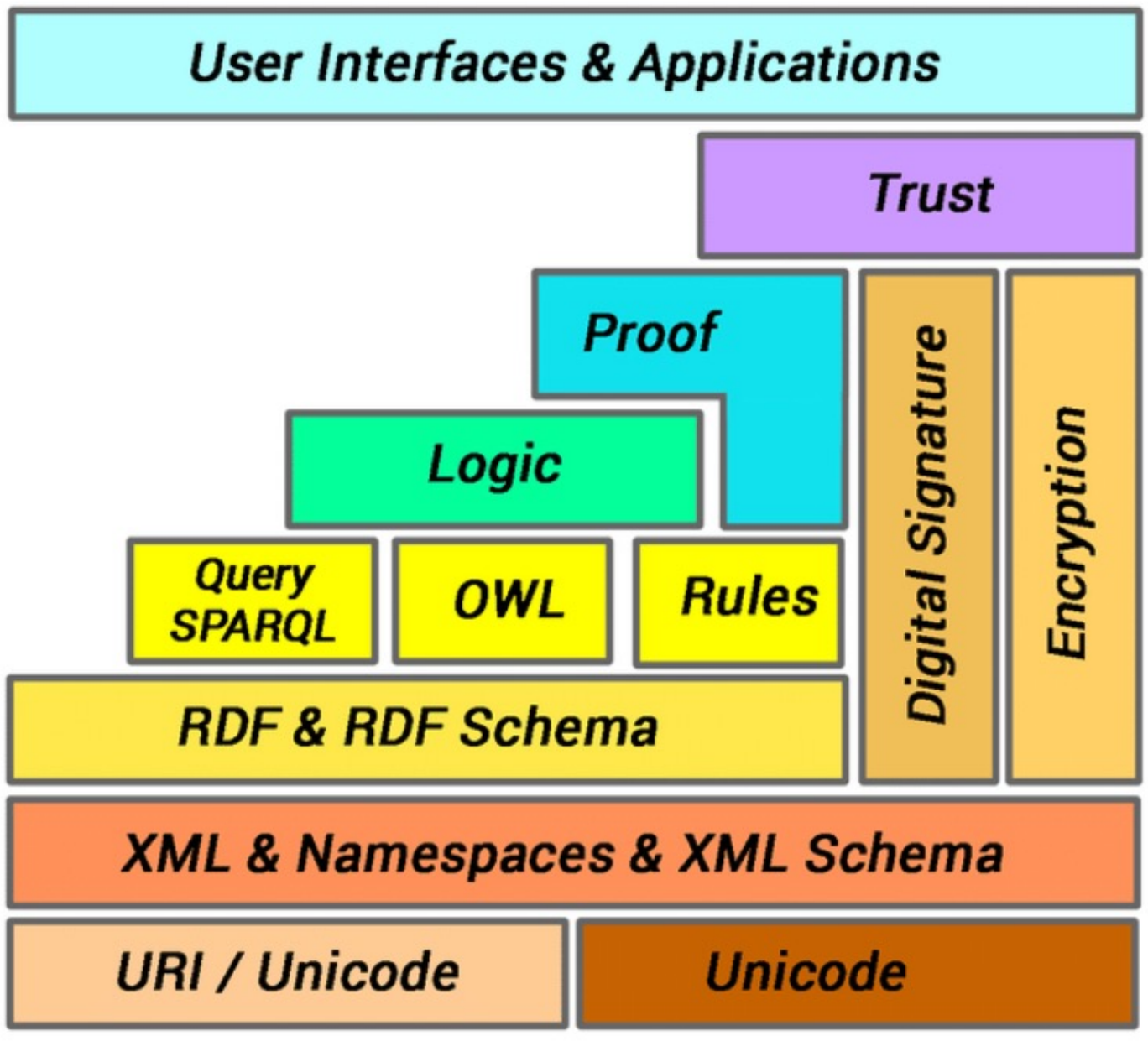
Символьное представление




```
Leonard Nimoy  played       Spock
Leonard Nimoy  starredIn    Star Trek
Spock          characterIn  Star Trek
Star Trek      genre        Science Fiction
```

```
Alec Guinness    starredIn    Star Wars
Alec Guinness    played       Obi-Wan Kenobi
Obi-Wan Kenobi   characterIn  Star Wars
Star Wars        genre        Science Fiction
```

Semantic Web



Проект Викидата

WIKIDATA

Item | Discussion | Read | View history | Search Wikidata

Moscow (Q649)

capital and most populous city of Russia

Moskva | Moscow, Russia | Moskva Federal City, Russia | Moscow, USSR | Moskva, Russia | City of Moscow | Moscow, Russian Federation | Moscow, Soviet Union | Moscow, Russian SFSR | Muscovite | Moscovite

▾ In more languages

Configure

Language	Label	Description	Also known as
English	Moscow	capital and most populous city of Russia	Moskva Moscow, Russia Moskva Federal City, Russia Moscow, USSR Moskva, Russia City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR Muscovite Moscovite
Russian	Москва	столица и крупнейший город России	Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

Wikipedia (267 entries)

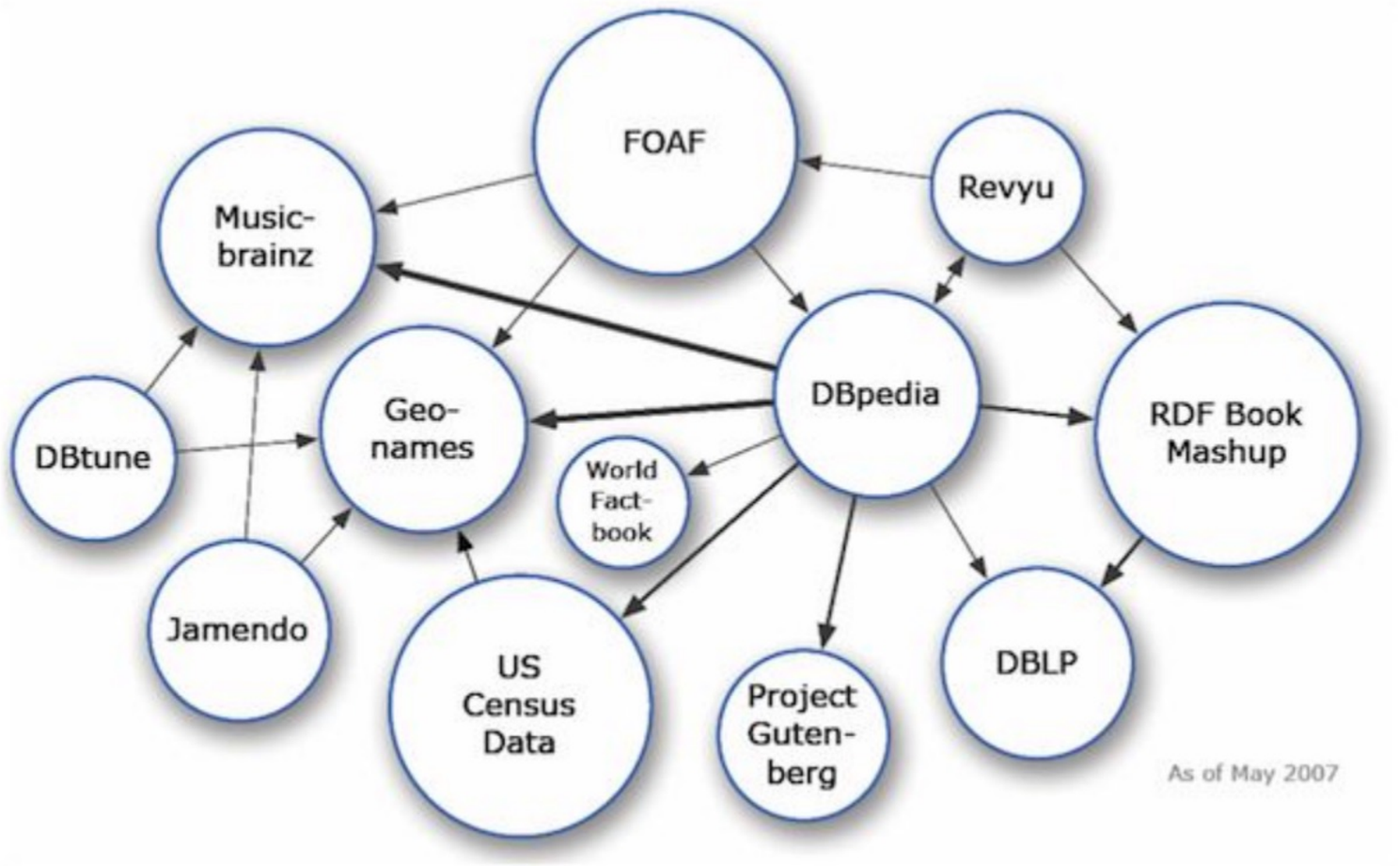
ab Москва
ace Moskôw
ady Москва
af Moskou
ak Moscow
als Moskau
alt Москва
am ሞስኮ
ang Moscow
an Moscú
arc ܡܘܣܩܘ
ar موسكو
ary موسكو
arz موسكو
ast Moscú
avk Moskva
av Москва
awa मस्को
ay Mosku

Statements

instance of	capital of Russia	
	start time	1918
	▸ 2 references	
	federal city of Russia	
	start time	1991
	▸ 1 reference	

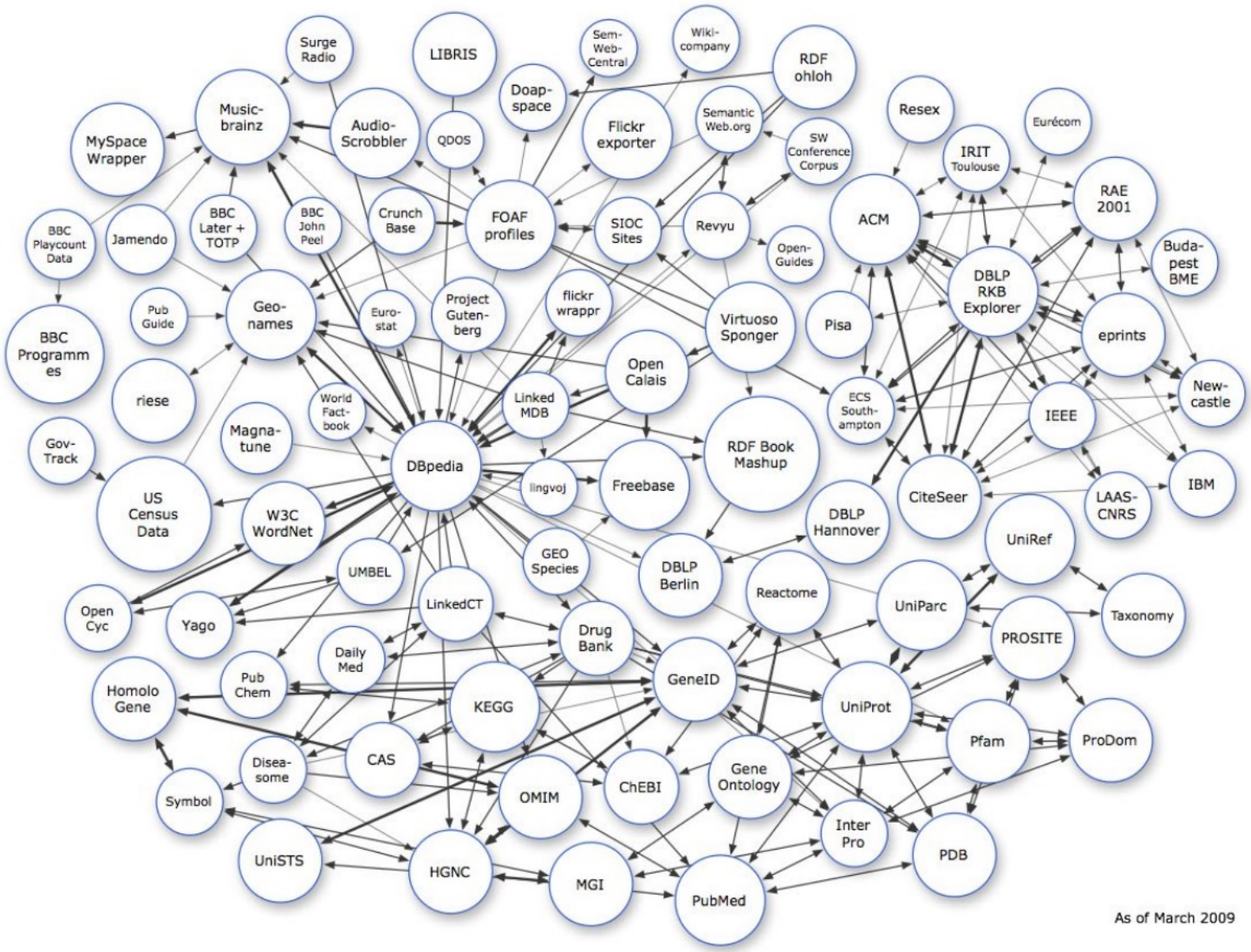
Link Open Data

2007



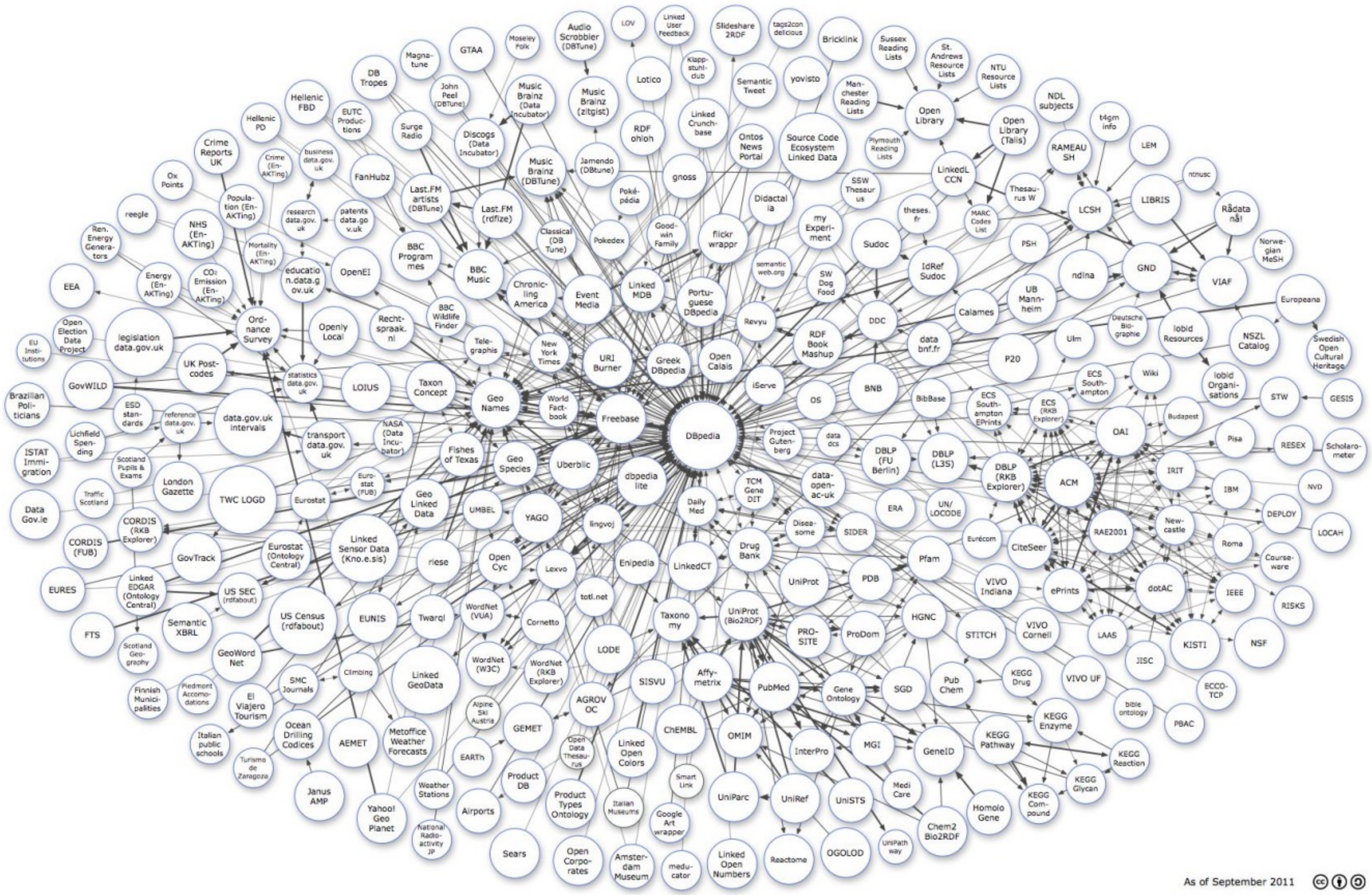
Link Open Data

2009

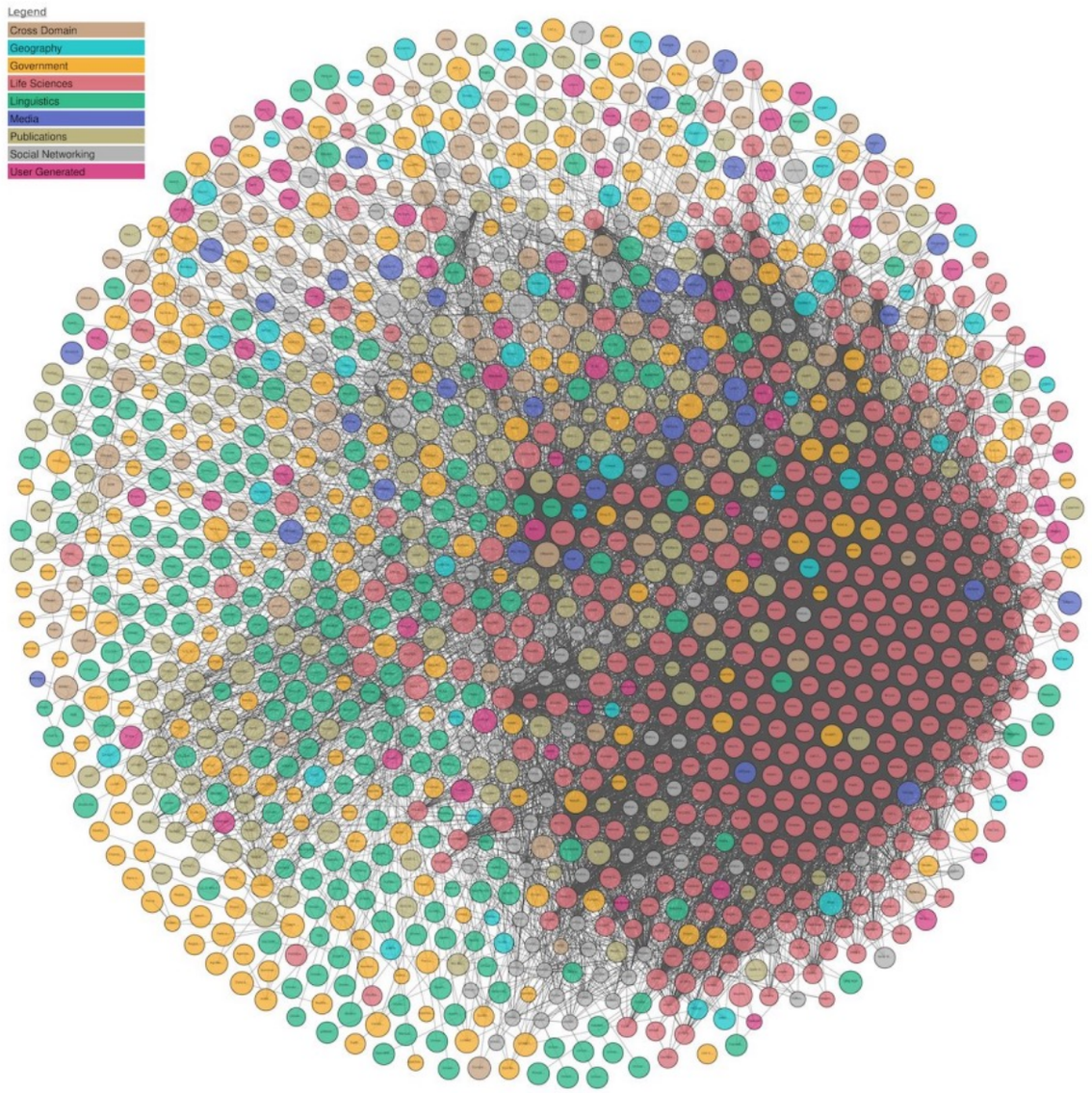


Link Open Data

2011

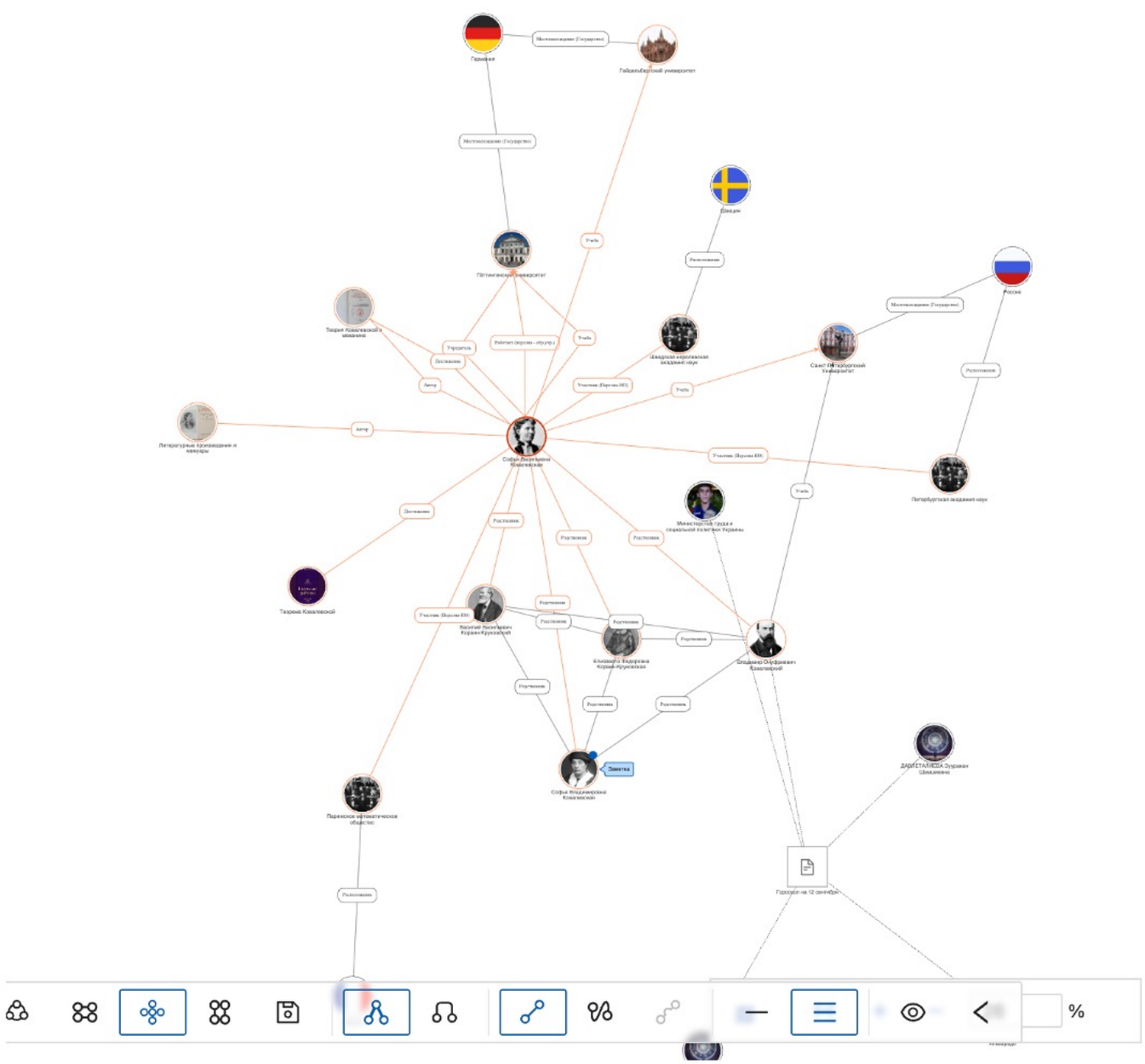


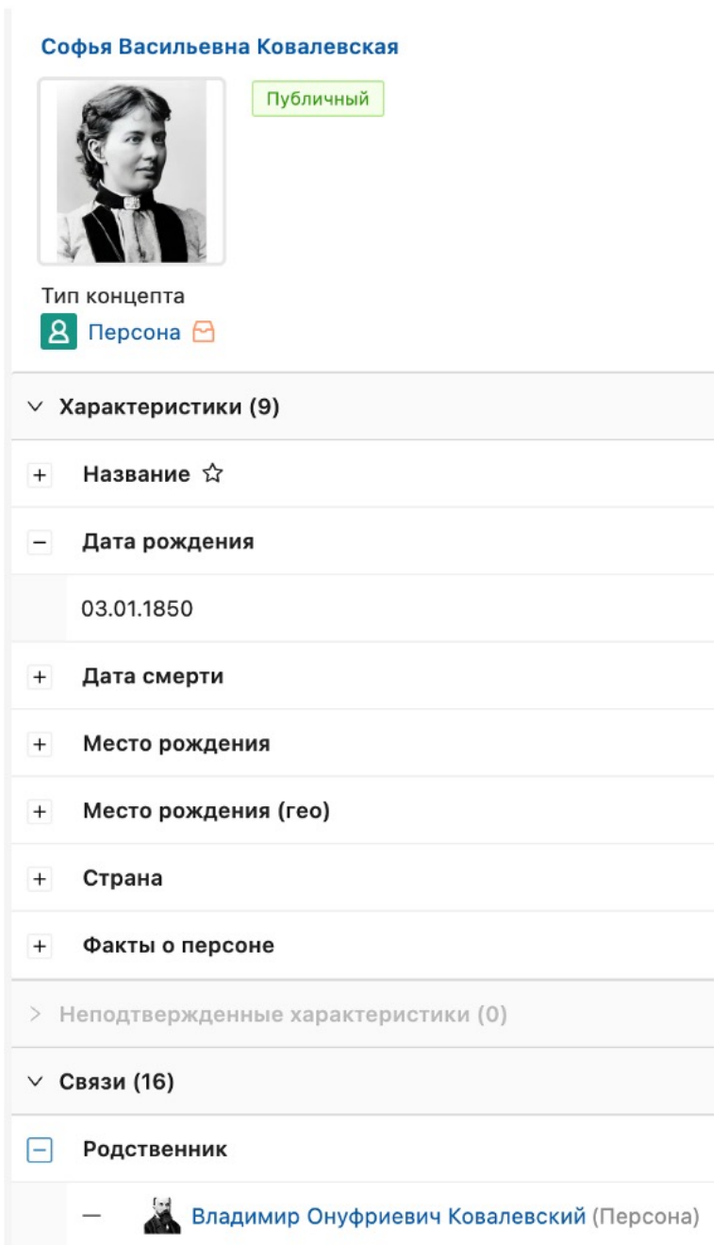
Link Open Data 2020



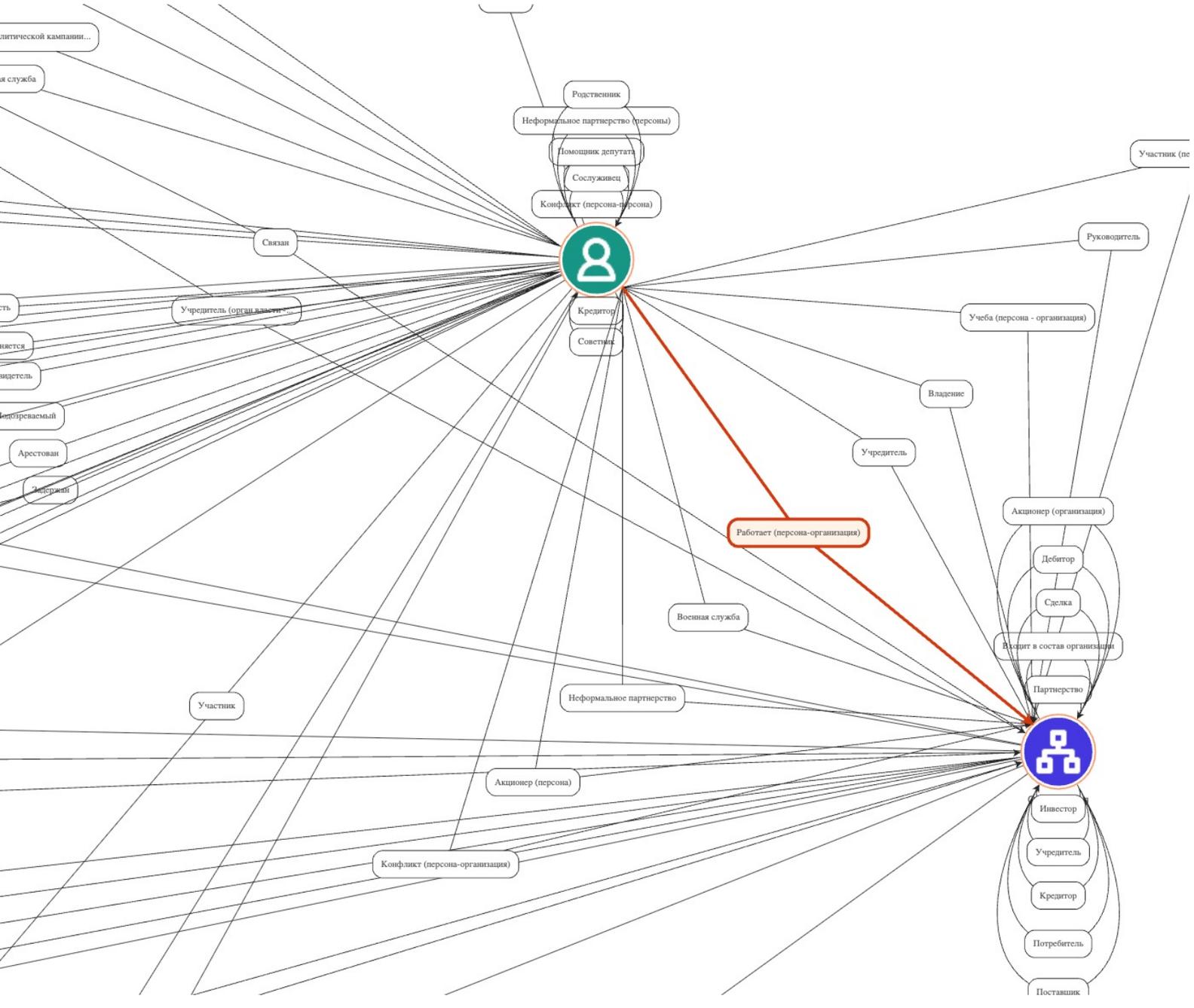
https://lod-cloud.net

Label Property Graph





Типизация графа



Тип связи № 786 Работает (персона-организация)

6

Тип сущности (из)

Персона

Тип сущности (в)

Организация

Направленная

Иерархическая

Типы характеристик (6)

ID факта	Строка		
Комментарий	Строка		
Характер деятельности	Строка		
Дата начала	Дата		
Дата окончания	Дата		
Должность	Работа		

Сценарии использования графов знаний

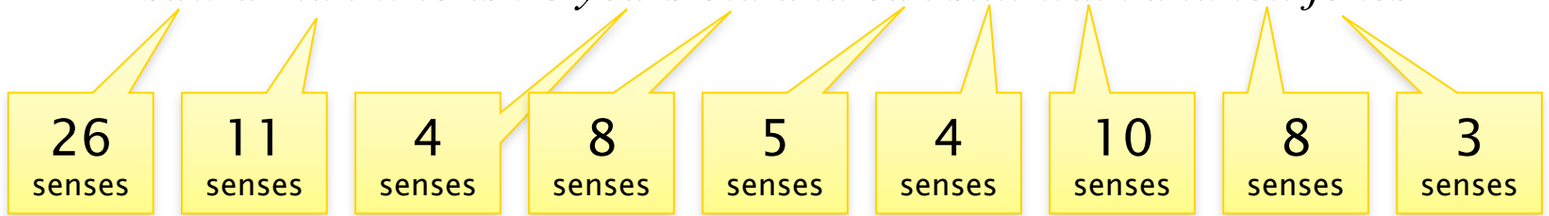
- Информационный поиск, расширение запросов, GraphRAG
- Проверка фактов, в том числе поиск противоречий (например, несовместимость лекарств)
- Интеграция структурированных и неструктурированных данных (корпоративные системы)
- Рекомендательные и вопросно-ответные системы
- Гибридный и объяснимый ИИ
- Аналитика сложных сетей и прогнозирование (поиск аномалий, вывод неочевидных закономерностей)

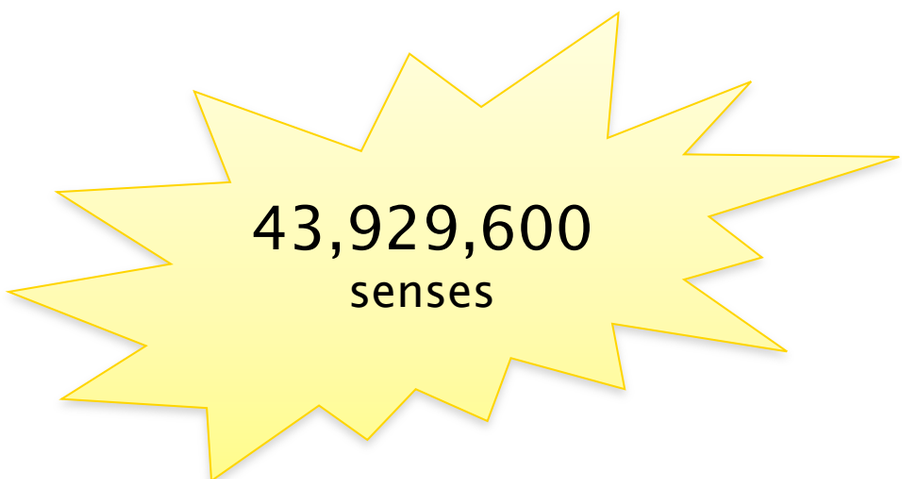
Вычислительная лексическая семантика

- Разрешение лексической многозначности
- Привязка к базам знаний
- Наполнение графов знаний

Трудность разрешения лексической многозначности

I saw a man who is 98 years old and can still walk and tell jokes





Разрешение лексической многозначности

- Word Sense Disambiguation (WSD)
  - определение значения слова в контексте
  - обычно предполагается фиксированный список значений (например WordNet)
- Сводится к задаче классификации
- Отличается от задачи разграничения значений (word sense discrimination)

Разрешение лексической многозначности: варианты

- Определение значений только заранее выбранных слов (lexical sample task)
  - line - hard - serve; interest
  - Ранние работы
  - Обучение с учителем
- Определение значений всех слов (all-word task)
  - Проблема разреженности данных
  - Невозможно натренировать отдельный классификатор для каждого слова

Признаки

- Должны описывать контекст
- Предварительная обработка текста
  - параграфы, предложения, части речи, леммы, синтаксический разбор?
- Признаки в словосочетаниях с позициями
- Множества соседей

- Проблема разреженности языка
  - Использовать *семантическую близость.* *(длина пути в иерархии WordNet, косинус между*

Пример

An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

Collocational features	
word_L3	electric
POS_L3	JJ
word_L2	guitar
POS_L2	NN
word_L1	and
POS_L1	CC
word_R1	player
POS_R1	NN
word_R2	stand
POS_R2	VB
word_R3	off
POS_R3	RB

Bag-of-words features	
fishing	0
big	0
sound	0
player	1
fly	0
rod	0
pound	0
double	0
runs	0
playing	0
guitar	1
band	0

Алгоритмы

- Любые методы классификации

Вопрос на засыпку

- Как сделать классификатор для задачи определения значений всех слов (all-word task)?

Методы оценки

- Внешние (in vivo)
  - Машинный перевод с/без WSD
- Внутренние (in vitro)
  - Применение к размеченным данным (SemCor, SENSEVAL, SEMEVAL)
  - Измерение точности и полноты в сравнении со стандартными значениями
- Нижняя граница
  - Выбор случайных значений работает плохо
  - Более сильные границы: наиболее частое значение, алгоритм Леска
- Верхняя граница: согласие экспертов
  - 75-80 для задачи определения значений всех слов со значениями из WordNet
  - до 90% с менее гранулированными значениями

Методы основанные на словарях и тезаурусах

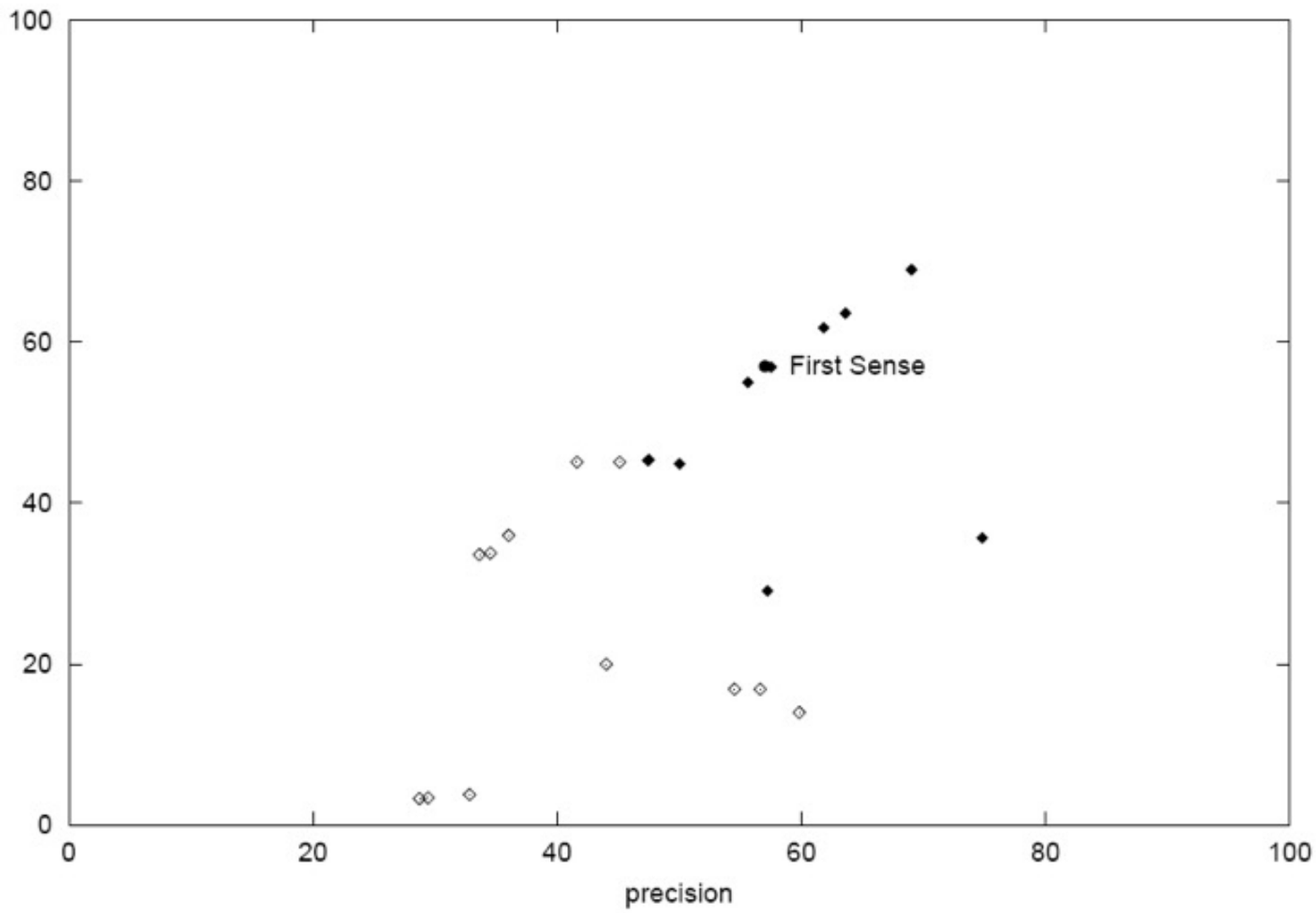
- Алгоритм Леска (1986)
  - –Взять все определения целевого слова из словаря
  - –Сравнить с определениями слов в контексте
  - –Выбрать значение с максимальным пересечением
- Пример
  - *pine*
    1. a kind of evergreen tree with needle-shaped leaves
    2. to waste away through sorrow or illness
  - *cone*
    1. A solid body which narrows to a point
    2. Something of this shape, whether solid or hollow
    3. Fruit of certain evergreen trees
  - Определить значение: *pine cone*

Варианты алгоритма Леска

- Упрощенный (Simplified Lesk)
  - Взять все определения целевого слова из словаря
  - Сравнить со ~~определениями~~ словами в контексте
  - Выбрать значение с максимальным пересечением
- Корпусный (Corpus Lesk)
  - Включить предложения из размеченного корпуса в сигнатуру каждого значения
  - Взвесить слова через IDF
  - IDF(w) = -log P(w)
  - Показывает лучшие результаты
  - Использовался как нижняя граница на SENSEVAL

Наиболее частое значение

- Сравнение методов на SENSEVAL-2

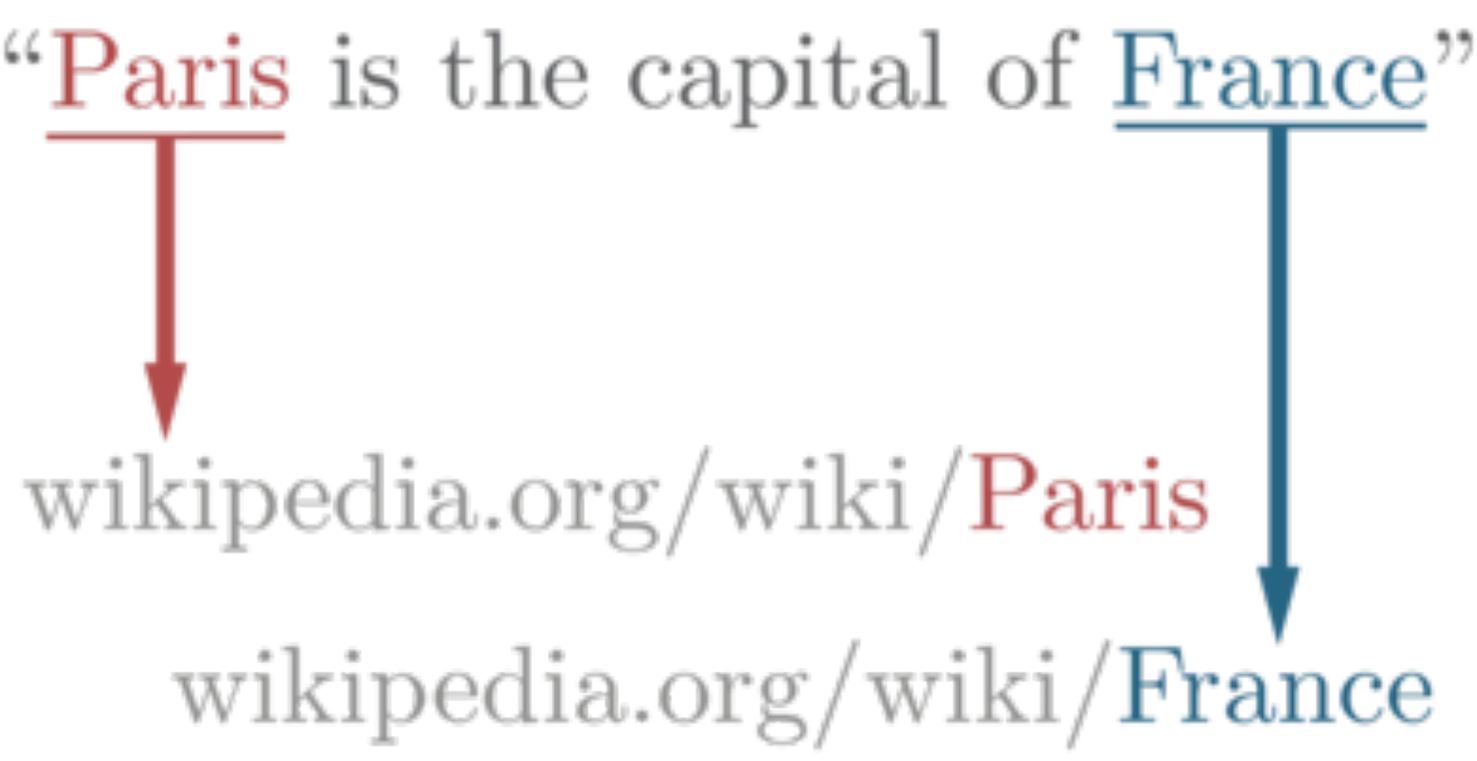


- McCarthy et. al. 2004 ACL - поиск наиболее частого значения по неразмеченному корпусу

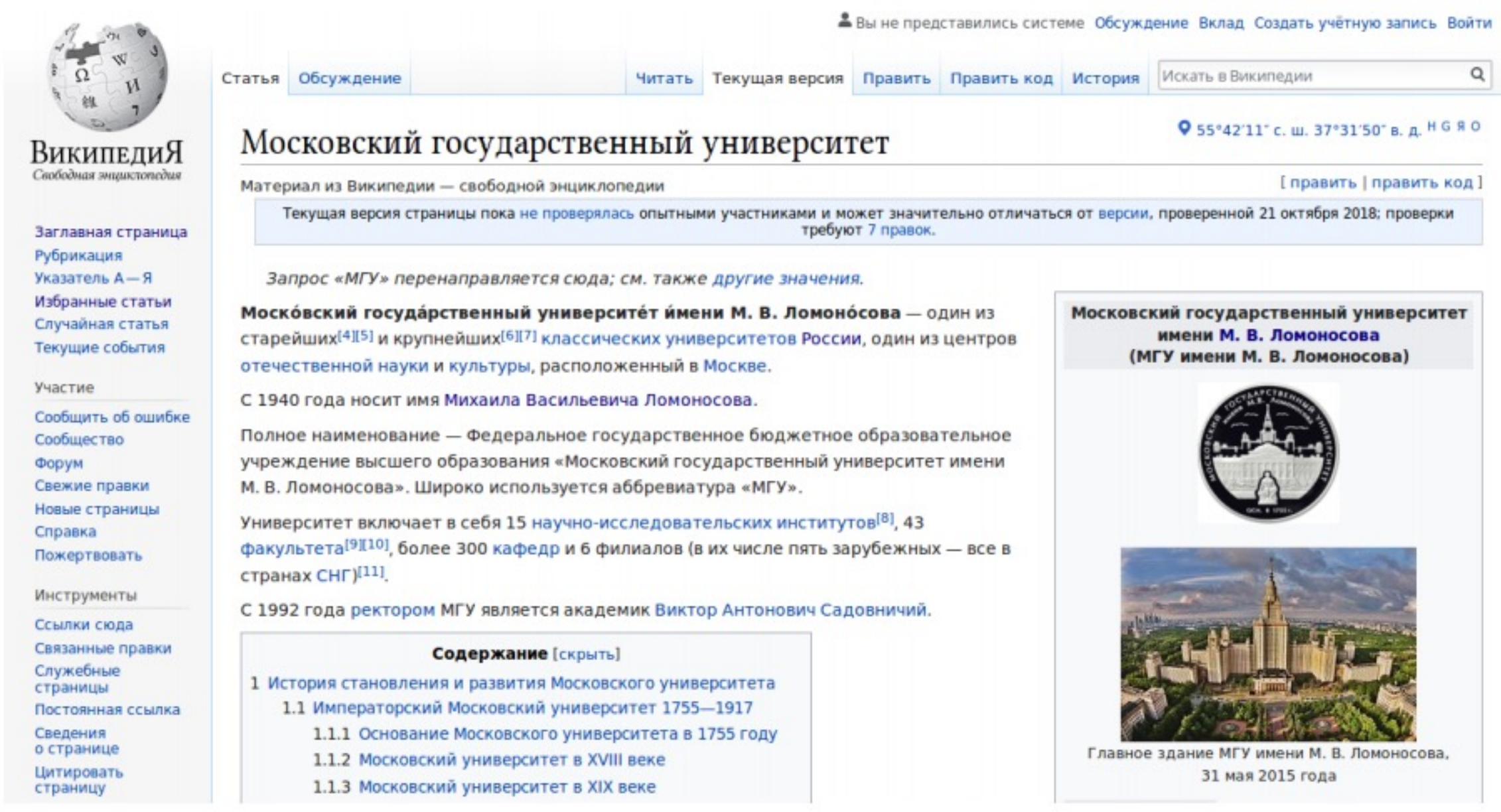
Привязка к базам знаний

Связывание (именованных) сущностей

- Вход: неструктурированный текст
- Выход: Значение для каждой (именованной) сущности из базы знаний



База знаний. Википедия



Вы не представились системе Обсуждение Вклад Создать учётную запись Войти

ВикипедиЯ
Свободная энциклопедия

Статья | Обсуждение | Читать | Текущая версия | Править | Править код | История | Искать в Википедии

Заглавная страница
Рубрикация
Указатель А — Я
Избранные статьи
Случайная статья
Текущие события

Участие
Сообщить об ошибке
Сообщество
Форум
Свежие правки
Новые страницы
Справка
Пожертвовать

Инструменты
Ссылки сюда
Связанные правки
Служебные страницы
Постоянная ссылка
Сведения о странице
Цитировать страницу

Московский государственный университет

55°42′11″ с. ш. 37°31′50″ в. д. H G R O

Материал из Википедии — свободной энциклопедии

[править | править код]

Текущая версия страницы пока не проверялась опытными участниками и может значительно отличаться от версии, проверенной 21 октября 2018; проверки требуют 7 правок.

Запрос «МГУ» перенаправляется сюда; см. также другие значения.

Моско́вский госуда́рственный университе́т и́мени М. В. Ломоно́сова — один из старейших[4][5] и крупнейших[6][7] классических университетов России, один из центров отечественной науки и культуры, расположенный в Москве.

С 1940 года носит имя Михаила Васильевича Ломоносова.

Полное наименование — Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова». Широко используется аббревиатура «МГУ».

Университет включает в себя 15 научно-исследовательских институтов[8], 43 факультета[9][10], более 300 кафедр и 6 филиалов (в их числе пять зарубежных — все в странах СНГ)[11].

С 1992 года ректором МГУ является академик Виктор Антонович Садовничий.

Содержание [скрыть]

1 История становления и развития Московского университета
1.1 Императорский Московский университет 1755—1917
1.1.1 Основание Московского университета в 1755 году
1.1.2 Московский университет в XVIII веке
1.1.3 Московский университет в XIX веке

Московский государственный университет имени М. В. Ломоносова (МГУ имени М. В. Ломоносова)

Главное здание МГУ имени М. В. Ломоносова, 31 мая 2015 года

Значения

Основание Московского университета в 1755 году [править | править код]

Создание университета было предложено И. И. Шуваловым и М. В. Ломоносовым. Первоначально открытие университет
подготовительных работ, связанных в первую очер
подписан императрицей Елизаветой Петровной 24
утверждён проект об учреждении Московского Уни
XX—XXI веках — 25 января) отмечается как дата
прочитаны 26 апреля (7 мая) 1755 года[15]. Иван Ив
директором. Ни в официальных документах, предс
упомянуто[16]. По мнению историка М. Т. Белявско
испортил ломоносовский проект, внеся в него ряд

Ива́н Ива́нович Шува́лов — русский государственный деятель, генерал-адъютант (1760), фаворит императрицы Елизаветы I Петровны, меценат, основатель Московского университета и Академии художеств. Почётный

Платформа

Материал из Википедии — свободной энциклопедии [править | править код]

Текущая версия страницы пока не проверялась опытными участниками и может значительно отличаться от версии, проверенной 6 ноября 2016; проверки требуют 8 правок.

Платфо́рма (фр. *plateforme* от «плоский» + *forme* «форма») может означать:

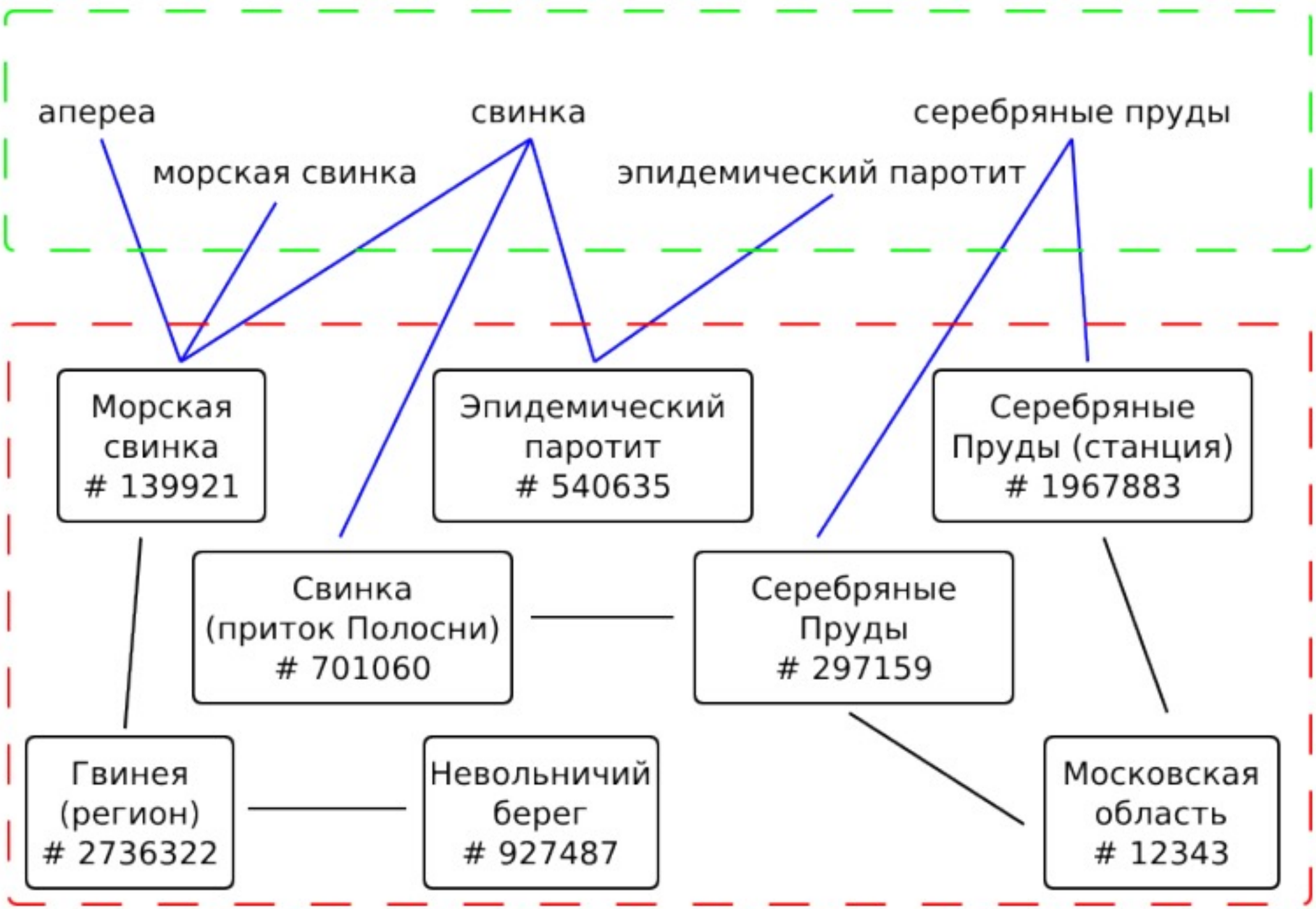
Значения в Викисловаре

- Платформа — возвышенная площадка, помост[1] для посадки пассажиров в вагоны, для нагрузки и выгрузки товаров.
  - Остановочный пункт — предназначенный для остановки поездов участок железнодорожного пути, часто ошибочно метонимически называющийся *платформой*
- **Платформа** — в Англии и Америке публичная трибуна для произнесения речей, и в переносном смысле политическая речь — программа кандидата в депутаты пред выборами.
- Платформа — совокупность основных компонентов, набор комплектующих, типовые конструктивные и технологические решения, применяемое оборудование в конструкции автомобиля.
- Платформа — железнодорожный подвижной состав: род вагона, товарный вагон без крыши.
- Платформа — крупный участок континентальной земной коры, характеризующийся относительно спокойным тектоническим режимом.
- Платформа — один из компонентов обуви: заметно утолщённая подошва обуви.

Название [править | править код]

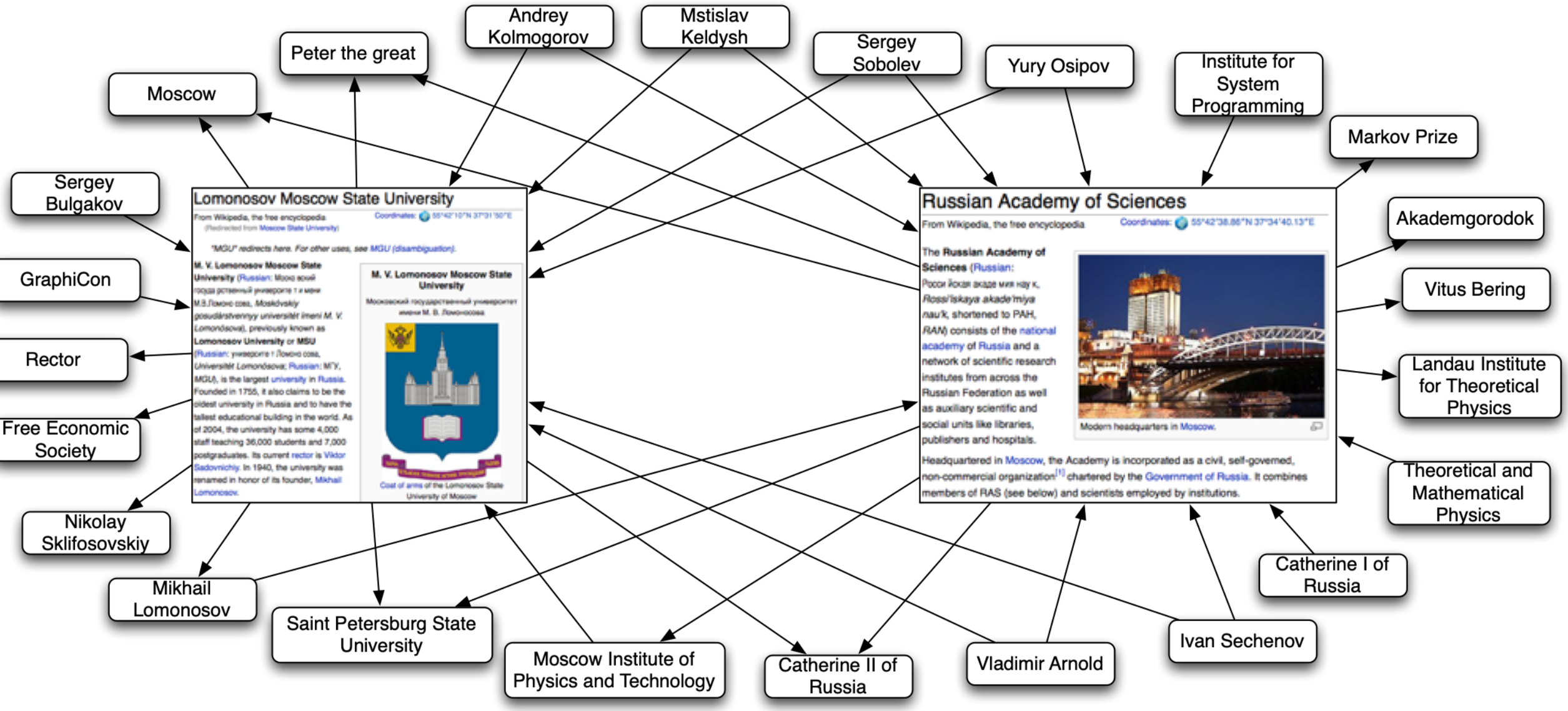
- «Платформа» — музыкально-литературный клуб в Санкт-Петербурге.
- «Платформа» — популярная американская стрижка.
- «Платформа» — роман французского писателя Мишеля Уэльбека, опубликованный в 2001 году.

База знаний. Википедия



Использование Википедии: семантическая близость

- Нормализованное количество общих соседей



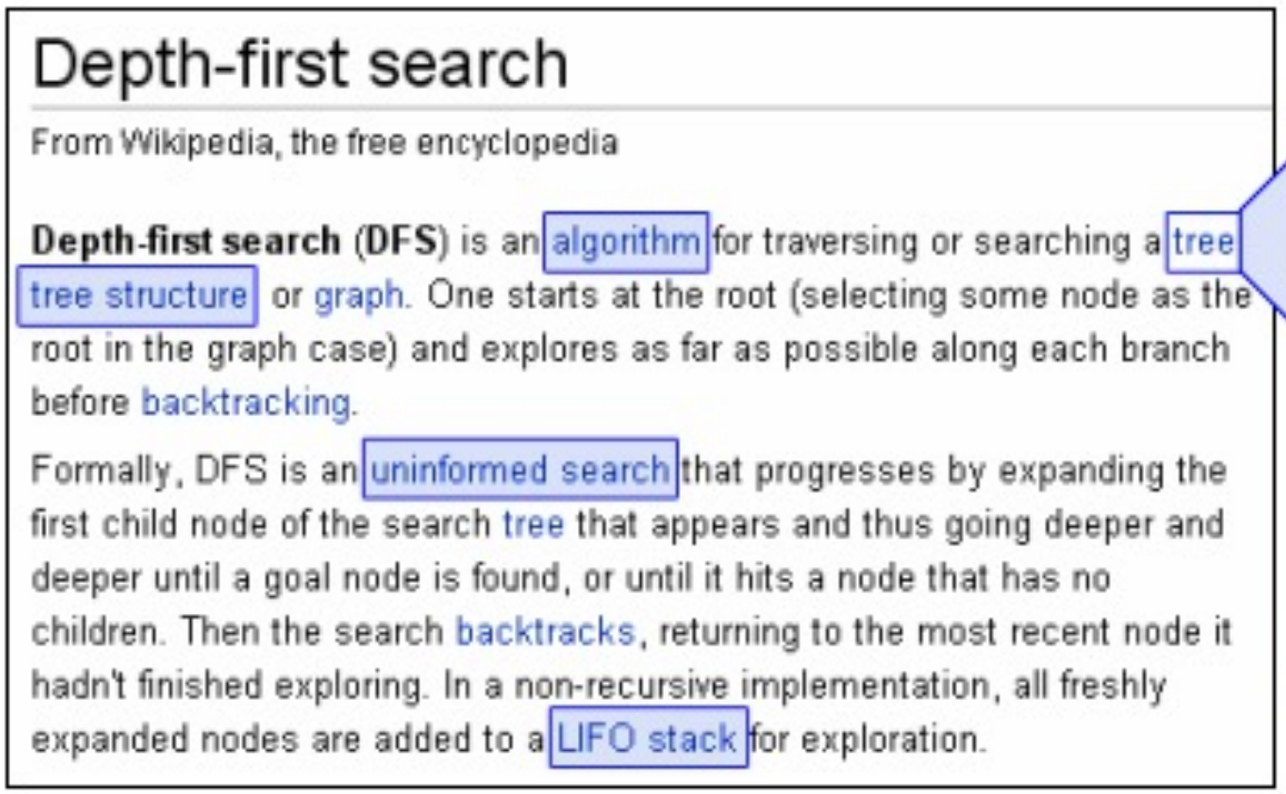
- Близкие концепты чаще встречаются вместе

Основные проблемы

- Синонимия (вариации имен): New York, NY, Big Apple
- **Многозначность**: NY ? New York / New Year
- Отсутствие значения
- Скорость обработки
- Постоянное изменение базы знаний
- Много языков

Алгоритм Milne-Witten

- Использует однозначные термины в качестве контекста



Depth-first search

From Wikipedia, the free encyclopedia

Depth-first search (**DFS**) is an algorithm for traversing or searching a tree tree structure or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Milne, David, and Ian H. Witten. "Learning to link with wikipedia." *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008.

Алгоритм Milne-Witten

- Commonness (популярность значения)

$$commonness(e, m) = \frac{count(e, m)}{count(e)}$$

- Relatedness (расстояние до контекста)

$$relatedness(a, b) = \frac{log\ max(|A|, |B|) - log|A \cap B|}{log|W| - log\ min(|A|, |B|)}$$

a, b - концепты;
A, B - множества концептов, ссылающихся на e и c соответственно;
W - множество всех концептов.

- Вероятность термина быть ссылкой

$$link\_prob(t) = \frac{count(t\ is\ link)}{count(t)}$$

Алгоритм Milne-Witten

- Семантическое расстояние от концепта *a* до контекста *S*:

$$distance(a, S) = \frac{\sum_{s \in S} w_s \times relatedness(a, s)}{\sum_{s \in S} w_s}$$

- Вес w_s концепта s:

$$w_s = \frac{1}{2}(link\_prob(term(s)) + \frac{1}{|S|}\sum_{c \in S} relatedness(s, c))$$

- Качество контекста S:

$$quality(S) = \sum_{s \in S} w_s$$

Алгоритм Milne-Witten

- Признаки:
  - популярность значения - commonness(e, m);
  - расстояние до контекста - distance(e, S);
  - когерентность контекста - quality(S).
- Алгоритмы машинного обучения:
  - Naive Bayes;
  - дерево решений C4.5;
  - метод опорных векторов;
  - дерево решений C4.5 с бэггингом.
- Обучение / применение:
  - положительный пример — правильное значение термина;
  - отрицательные примеры — все остальные возможные значения термина;
  - на этапе применения выбирается концепт, с максимальной уверенностью классифицированный как правильный.

Извлечение информации

Параметр	Классический IE	Open IE
Схема отношений	Фиксированная, заранее определённая (например: «родился в», «работает в»)	Не фиксирована, извлекает любые отношения из текста
Гибкость	Низкая, работает только для конкретных доменов и заранее заданных отношений	Высокая, может извлекать новые, неожиданные факты и связи
Необходимость разметки данных	Требует ручной разметки для обучения моделей	Может работать с минимальной или автоматической разметкой
Тип входного текста	Обычно структурированные или доменные тексты	Любой текст, в том числе новостные статьи, веб-страницы, соцсети
Формат вывода	Обычно фиксированные шаблонные отношения (relation, entity1, entity2)	Тройки (subject, relation, object), где relation формируется автоматически из текста
Применение	Узкоспециализированные задачи, корпоративные системы, биомедицина, финансы	Построение графов знаний, QA системы, семантический поиск, аналитика больших корпусов
Обработка неопределённых или новых отношений	Сложно, требует добавления новых правил	Естественно извлекает новые отношения без дополнительной разметки
Сложность реализации	Средняя, требует настройки под домен	Высокая, требует NLP-парсинга, синтаксического анализа и фильтрации шумов
Примеры инструментов	GATE, IBM Watson IE, LingPipe	Stanford OpenIE, OpenIE5, ClausIE, MINIE

Извлечение знаний: LLM

В воскресенье, 20 июля, компания Microsoft выпустила срочные патчи, закрывающие две уязвимости в ПО для совместной работы Microsoft SharePoint. Речь идет о корпоративных инсталляциях SharePoint, облачная версия SharePoint Online, входящая в пакет Microsoft 365, не подвержена проблеме. Уязвимости с идентификаторами CVE-2025-53770 и CVE-2025-53771 являются вариантами проблем, обнаруженных еще в мае этого года и пропатченных (как выяснилось, не до конца) в июльском наборе обновлений от Microsoft.

Извлеки сущности и связи из следующего текста. Верни структурированные тройки <s,p,o> для сохранения в граф знаний neo4j.

<Microsoft	выпустила	срочные патчи>
<срочные патчи	закрывают	две уязвимости>
<уязвимости	обнаружены_в	Microsoft SharePoint>
<Microsoft SharePoint	является	ПО для совместной работы>
<SharePoint Online	входит_в	Microsoft 365>
<SharePoint Online	не_подвержен	уязвимостям>
<уязвимости	имеют_идентификатор	CVE-2025-53770>
<уязвимости	имеют_идентификатор	CVE-2025-53771>
<CVE-2025-53770	является_вариантом	проблемы_обнаруженной_в_мае_2025>
<CVE-2025-53771	является_вариантом	проблемы_обнаруженной_в_мае_2025>
<проблемы_обнаруженные_в_мае_2025	были_пропатчены	июльским_набором_обновлений>
<июльский_набор_обновлений	выпущен	Microsoft>
<июльский_набор_обновлений	не_устранил_полностью	уязвимости>
<патчи	выпущены_дата	2025-07-20>
<патчи	выпущены_день_недели	воскресенье>

Понимание слов предложения ≠ понимание предложения

- Во время интервью:
  — Ваша главная слабость?
  — Правильно интерпретирую семантику вопроса, но игнорирую его суть.
  — Не могли бы вы привести пример?
  — Мог бы.

Ссылки

- https://ods.ai/tracks/kgcourse2022
- https://dl.acm.org/doi/abs/10.1145/1458082.1458150

Learning to **link** with **wikipedia**

D Milne, IH Witten - Proceedings of the 17th ACM conference on …, 2008 - dl.acm.org

… unstructured text, and enrich it with **links** to the appropriate **Wikipedia** articles. The resulting … -**learning** approach to disambiguation that uses the **links** found within **Wikipedia** articles for …

★ Save Cite Cited by 1733 Related articles All 11 versions

Следующая лекция

- Машинный перевод