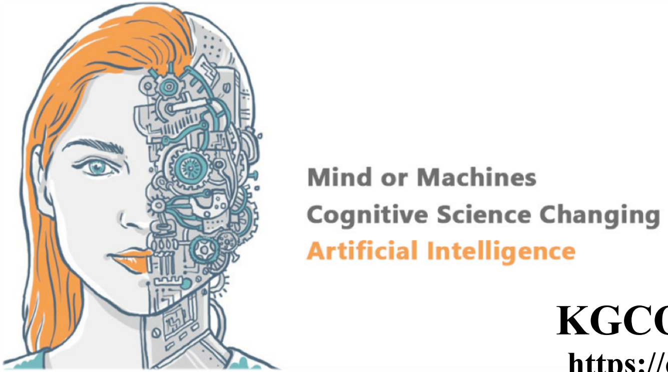
# Knowledge Graph: From Theory, Techniques, Applications to Challenges

**World Wide Web – Information Network – Social Network – Knowledge Network**

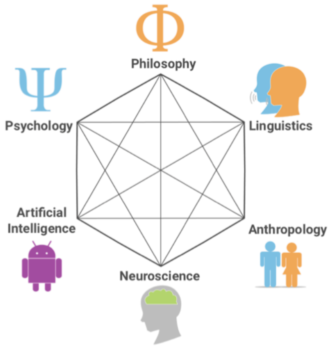



**Peng Wang**

pwang@seu.edu.cn

**KGCODE Laboratory, Southeast University**

https://cse.seu.edu.cn/2019/0105/c23024a257526/page.htm




School of Computer Science and Engineering/School of Artificial Intelligence, Southeast University

# AI in the past decade



Cognitive intelligence

NLP
ML
IR
QA
Knowledge graph
……


2011 IBM Watson

2011 Palantir

Cognitive intelligence

Big data analytics
Data visualization
Knowledge graph
……

Perceived intelligence

Deep learning
Monte Carlo Tree


2016 Google AlphaGo

2017 CMU Libratus

General intelligence

Game theory
Nash equaliser
……

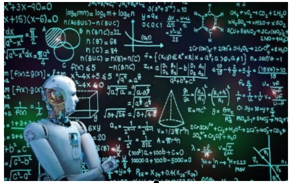
# Computing Intelligence VS Perceived Intelligence VS Cognitive Intelligence

difficulty/
value

our work

focus — Cognitive intelligence

## Understand, reasoning, using knowledge



more knowledge;
Not as good as humans at reasoning and understanding

combine — Perceived intelligence

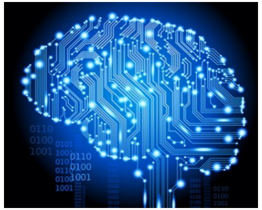
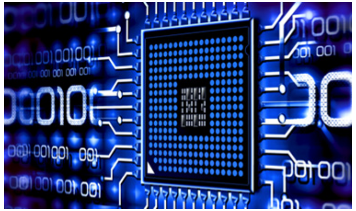
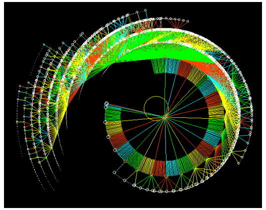
## visual, auditory and other perceived intelligence




Machines are close to or beyond humans in some perceptive tasks

Computing intelligence

## quick processing

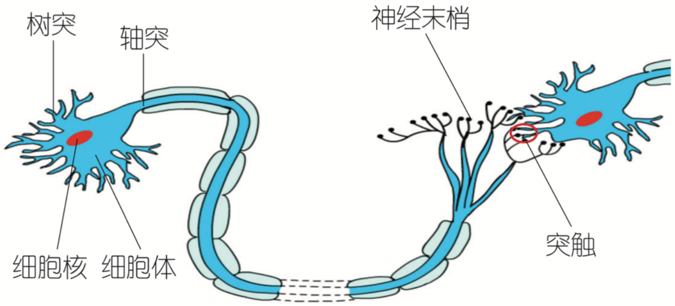




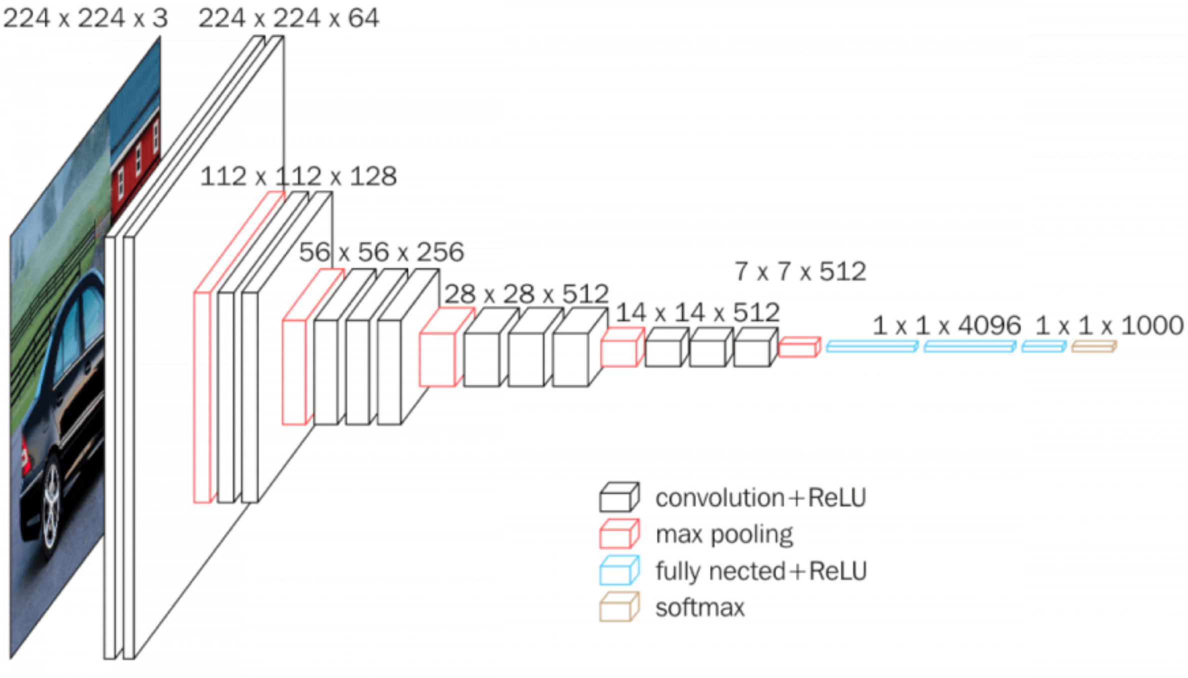
Machines have overtaken humans in computing tasks

# Intelligence simulation



## Simulates the work of neurons in the brain

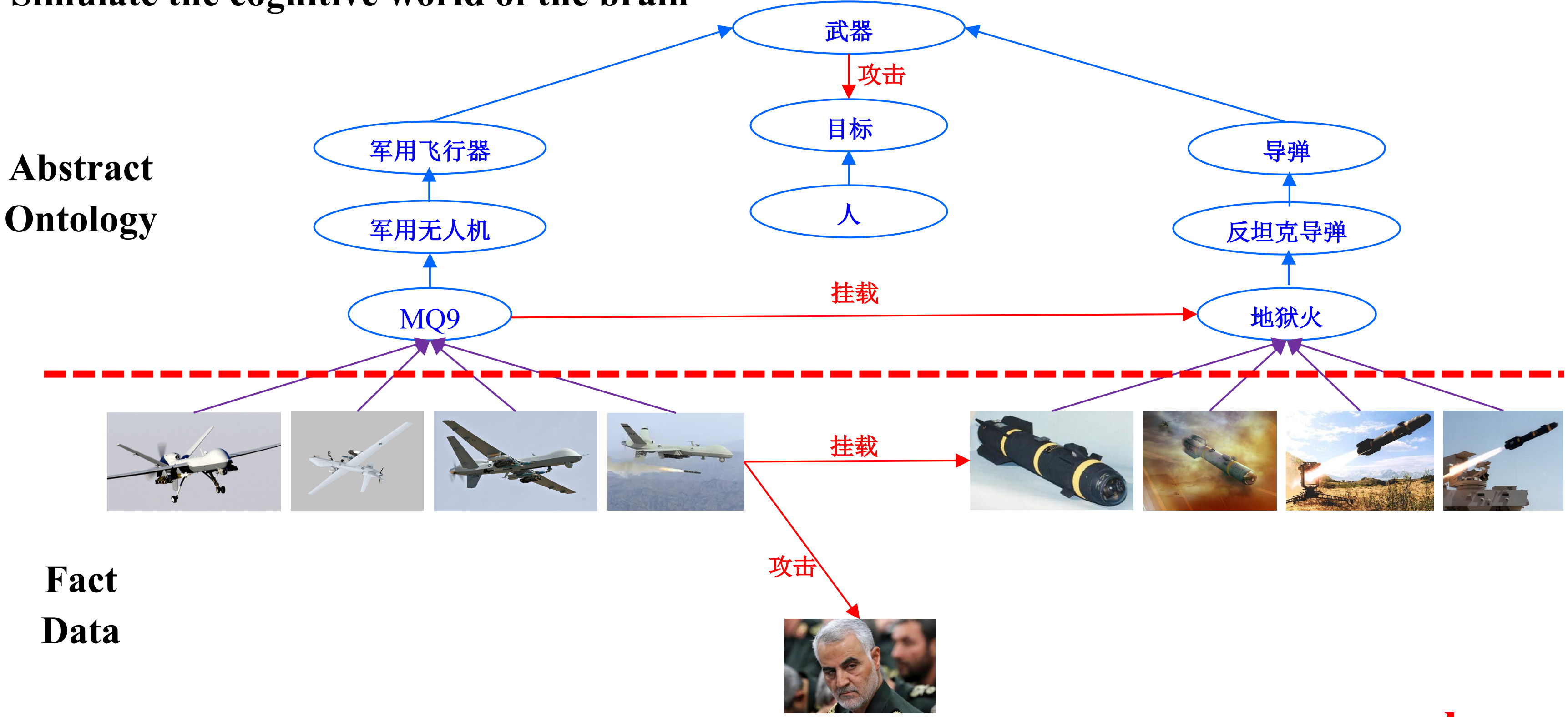



**difficult to explain**

# Intelligence simulation

Simulate the cognitive world of the brain



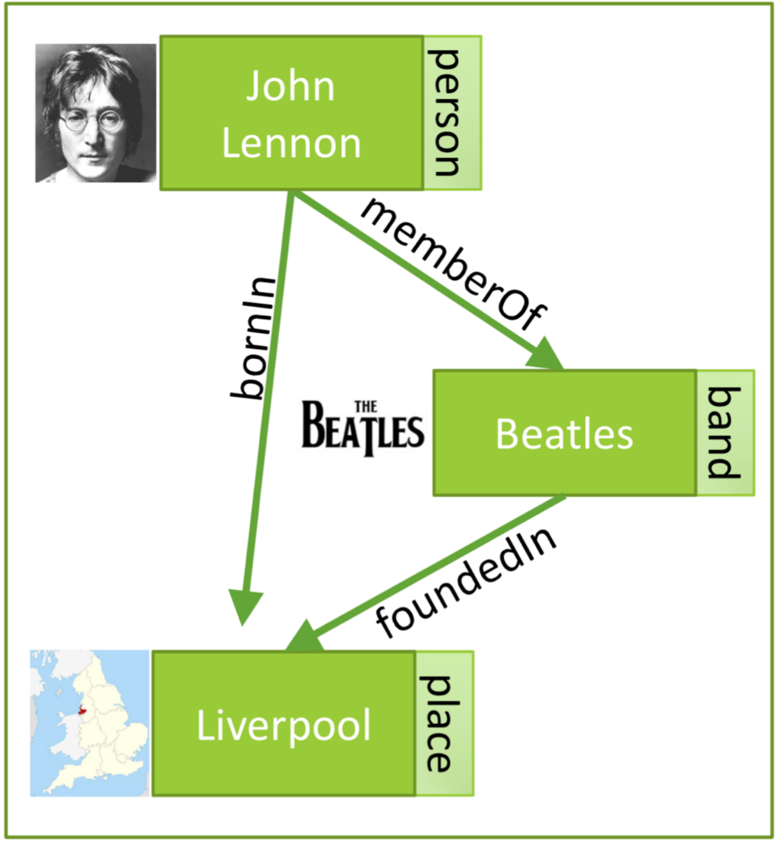

**can be explained**

# Knowledge Graph



A graph representation of knowledge

- Concepts, attributes, entities, relationships
- A node is an entity
- The node has a property label
- The edges of two nodes are relationships between entities
- Emphasize entities, but can also describe concepts

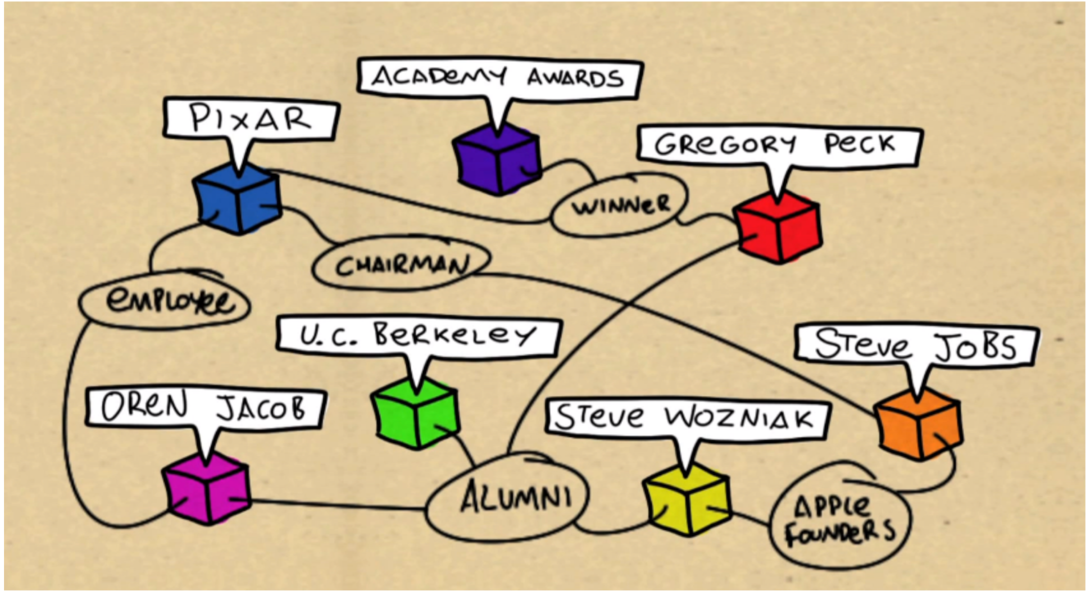


# The nature of the knowledge graph

東南大學 SOUTHEAST UNIVERSITY

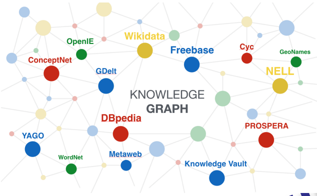
**The world is not made of strings ,**
**but is made of things**

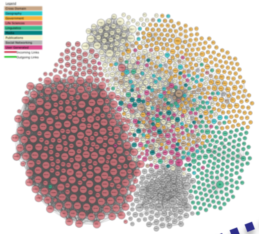
**万物及其联系的网络**

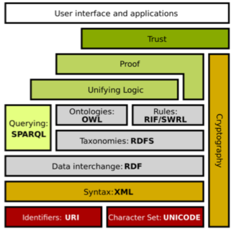


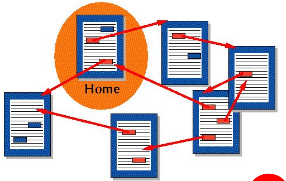
# Revolution of Information Organization

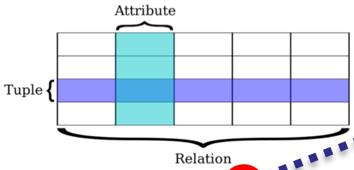







**1970**
**RDB**
**(Structural Data)**

**1989**
**WWW**
**(Doc+Link)**

**1999**
**Semantic Web**

**2006**
**Linked Data**

**2012**
**KG**
**(Entity+Relation)**

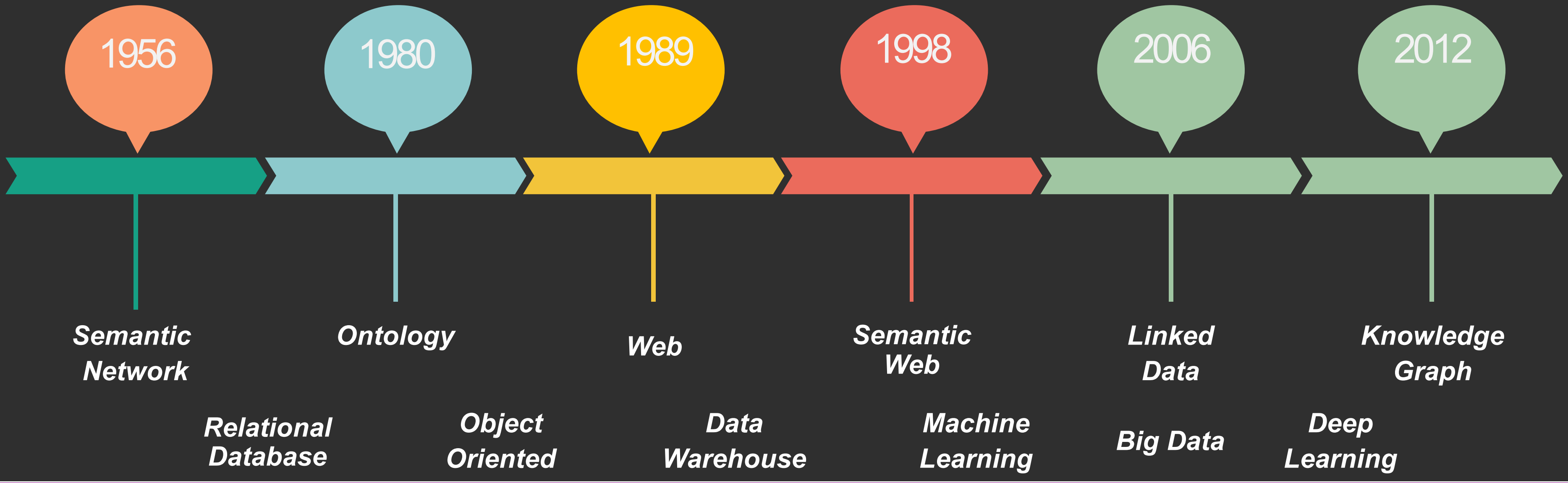
unstructured data/Scattered data
→
high-value knowledge mining
→
Cognitive intelligence applications

# Evolution of knowledge graph



**Knowledge graph is the technology of artificial intelligence that has been tried, failed and developed by academic and industrial circles for nearly 50 years**

## Knowledge Graph = Data + Knowledge



# The value of knowledge

“110”


**machine perspective**

对计算机而言110是数据，没有任何多余和特别的意义

**Non-knowledge perspective**

借助复杂的机器学习、数据挖掘等模型，从大量的数据中可分析出110可能的意义

**knowledge perspective**

构建一条知识：

“110是公安报警电话号码”

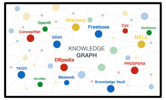
任何知道该知识的系统都完全知道110的意义

0101010101110101110111011 [0,0.5,0.1,0.05,0.6,0.9,0.3,…] (110，is-a, 公安报警电话号码)

**data** **information** **knowledge**

# KG VS DL






| | Deep Learning | Knowledge Graph |
|---|---|---|
| principle | • Implicit simulation of human intelligence | • Explicit simulation of human intelligence<br>• |
| scenario | • fields of perceptive intelligence<br>• Go, voice, images, video | • Widely used in a variety of tasks<br>• Search, human-computer interaction |
| characteristics | • huge training data<br>• powerful computing<br>• hard to explain | • A great amount of knowledge<br>• Explainable<br>• Understandable |
| progress | • surpass humans on some missions | • More knowledge than humans<br>• cannot reasoning |
| trend | • deep fusion in cognitive intelligence | |

# KG VS KB VS DB

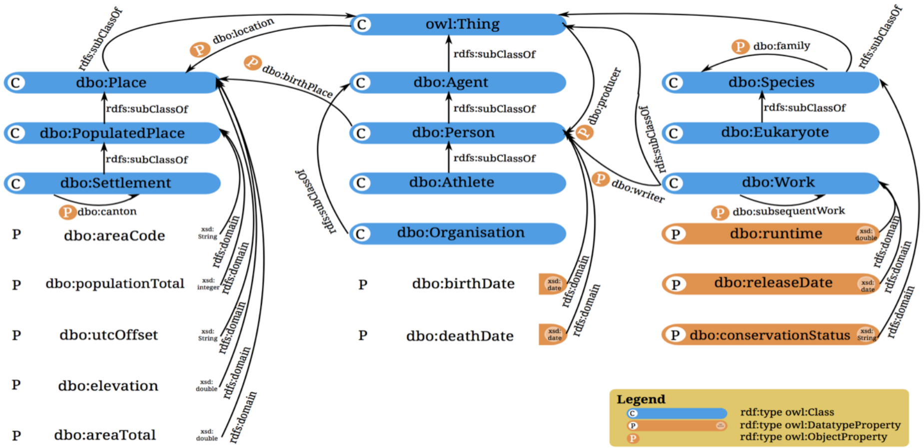


| | semantics | data | application | |
|---|---|---|---|---|
| **DB** | no semantics | rich data | easy | **Success** |
| **KB** | rich semantics | few fact | difficult | **Bottleneck** |
| **KG** | few semantics | rich fact | easy | **Success** |

**Knowledge Graph = Data + Knowledge**

# KG VS KB VS DB






| | |
|---|---|
| #Triples | 1.6B (now 2.8B) |
| #Subjects (Entities) | 43M |
| #Types | 1.1K |
| #Predicates | 4.5K |
| #Objects | 102M |

**DBPedia**
**C:685 P:2795 I:4.23 million**

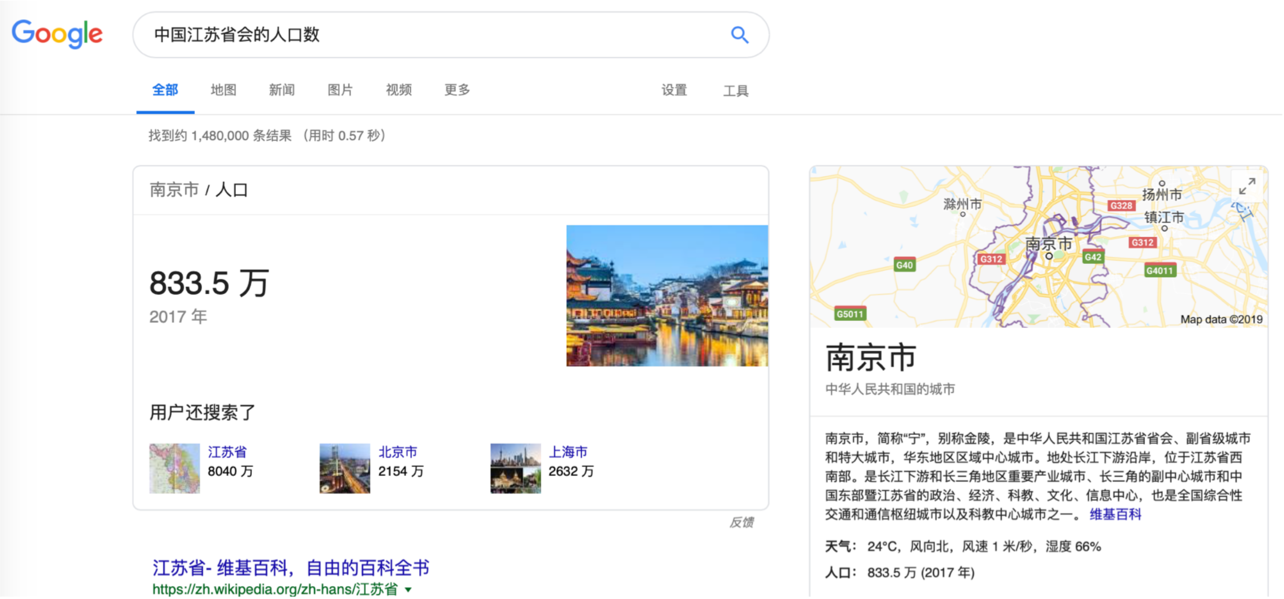
**Google KG**
**C:1100 P:4500 I:100 million**

**few semantics, rich facts/instances**

# KG application scenarios



- **Search**
  - Improve search accuracy
  - Semantic search
  - intent understanding
  - Multimodal search



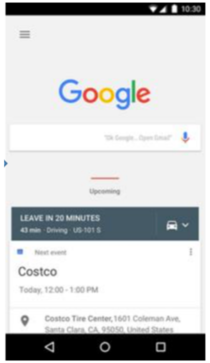

**Knowledge graph is high-quality data, the best way to apply it is to use it directly!**

# KG application scenarios

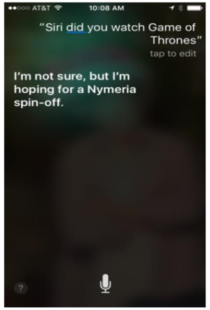


- **QA**
  - Improve the accuracy of QA
  - Improve the QA experience
  - Guide conversation
  - Multiple rounds of QA

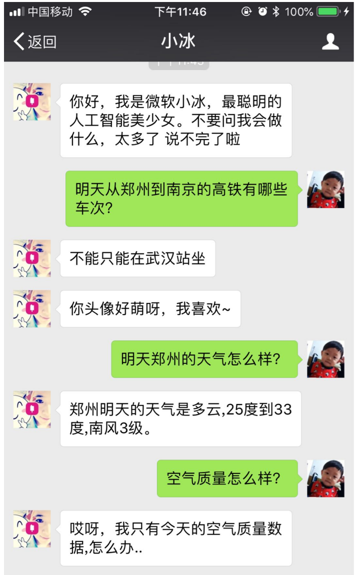
Google Now




Apple Siri




Amazon Alexa

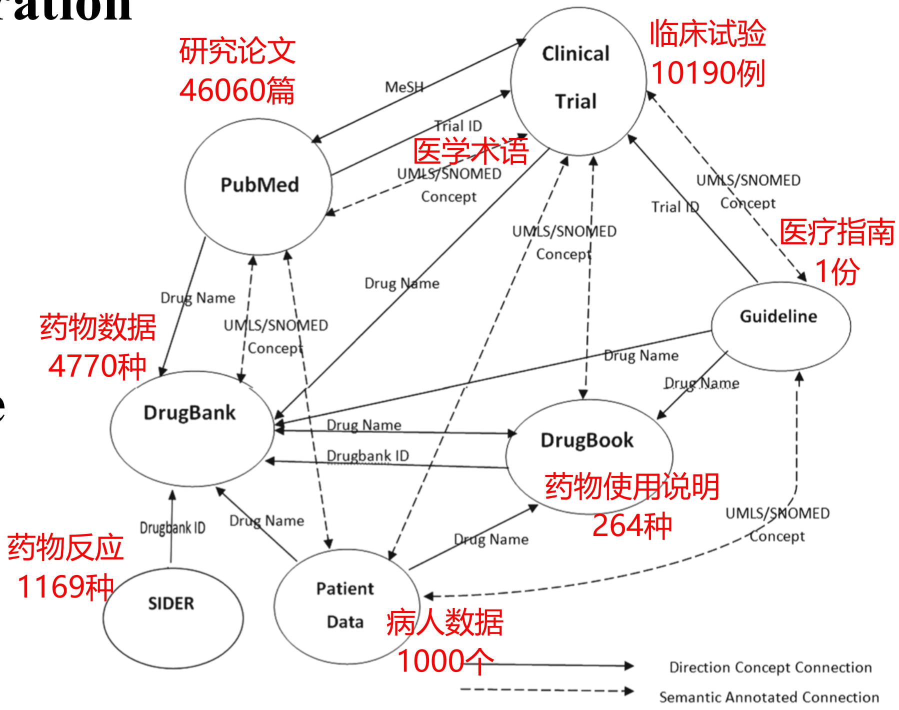


**QA is an important application scenario for AI**

# KG application scenarios



- **Data integration**
  - Large-scale heterogeneous data integration
  - Mining relations of data
  - Versatility
  - Scalability
  - Flexible
  - Knowledge reuse
  - Non-interference with the data source

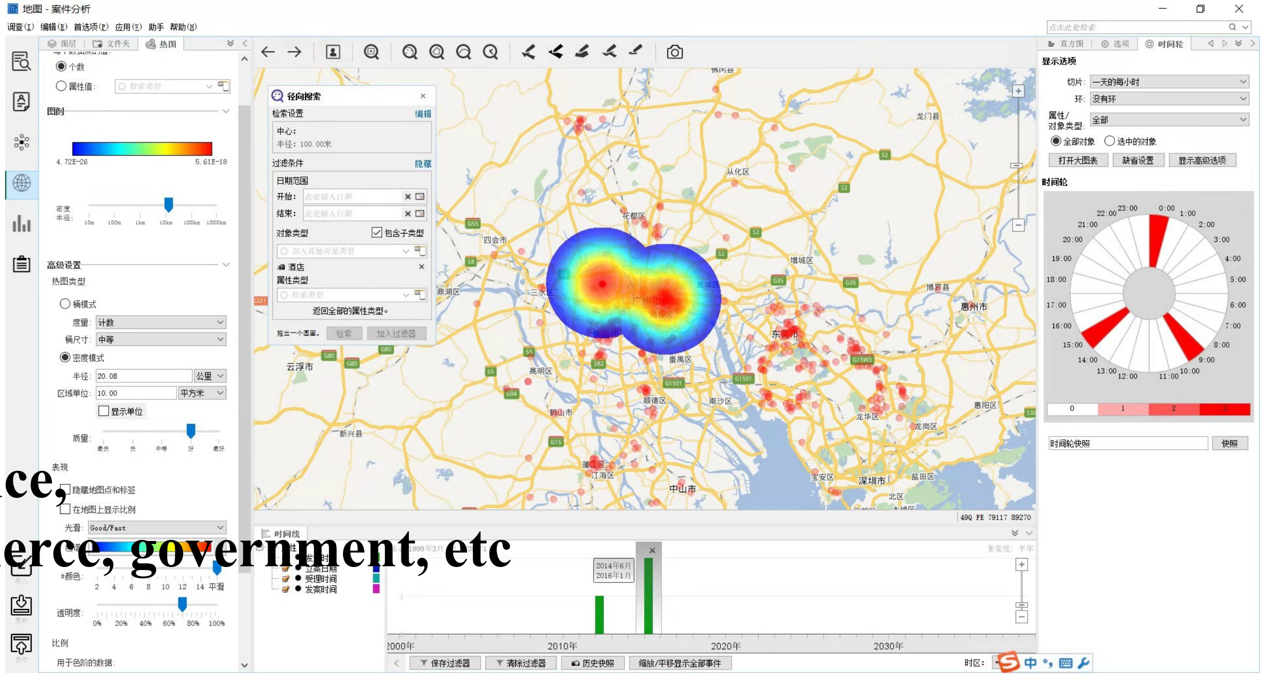



**KG realizes the intelligent data integration mechanism of data-information-knowledge!**

# KG application scenarios



- **Decision**
  - Organize data
  - Link data
  - Knowledge mining
  - Knowledge discovery
  - Knowledge reasoning
  - NLU
  - Avariety of AI Apps
  - national defense, finance, Manufacturing, commerce, government, etc

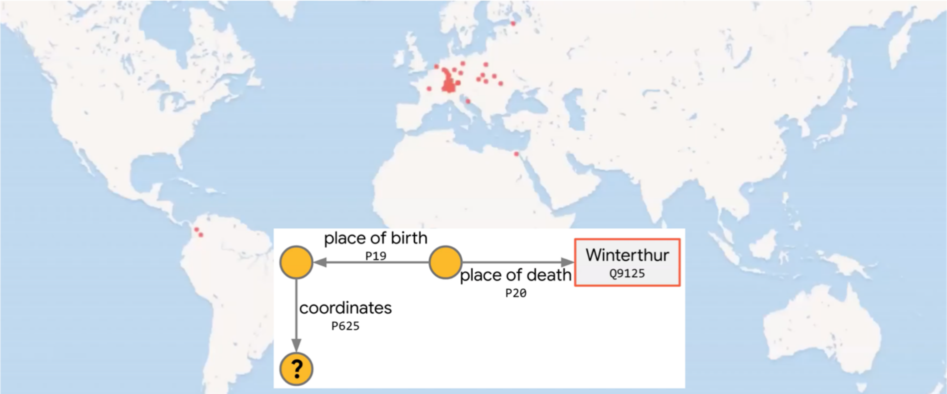



KG and a variety of AI technology combined

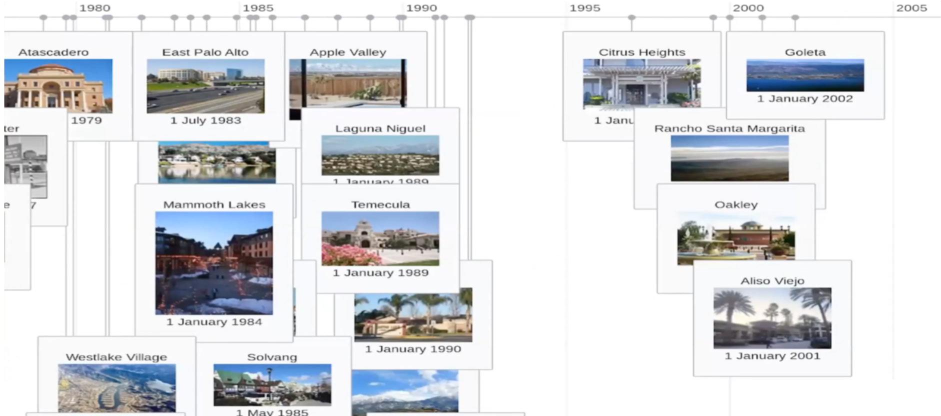
# Accurate and complex intelligent queries

東南大學 SOUTHEAST UNIVERSITY

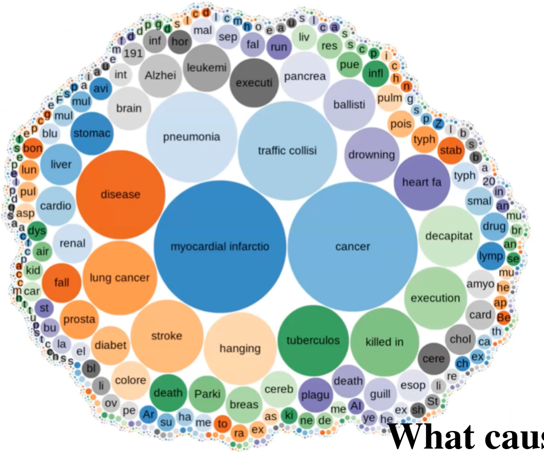

Where did the Winterthur-born man die?




Where were the people who died in Winterthur born?



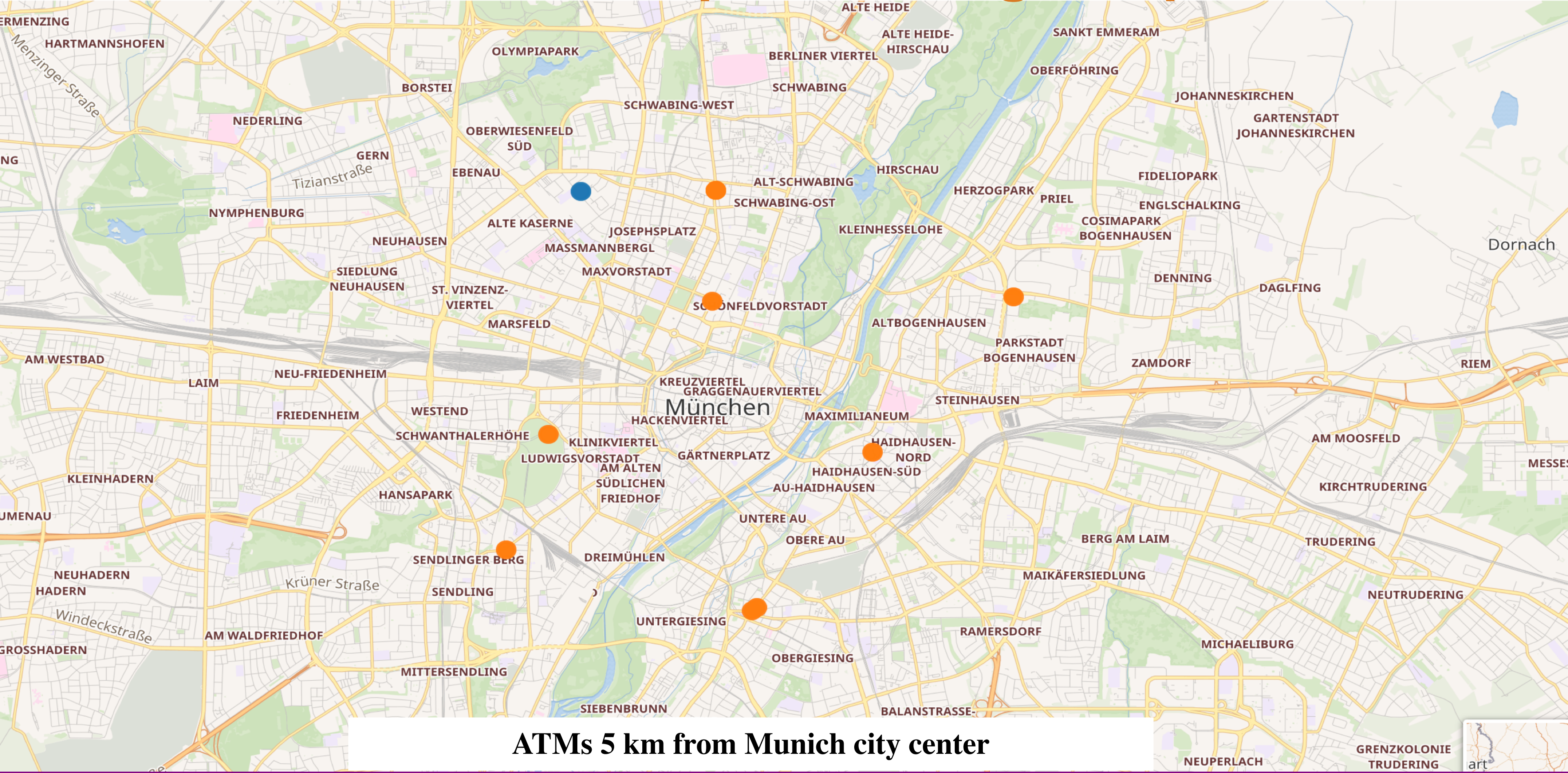

California city established in recent years?




What causes death?

# Accurate and complex intelligent queries

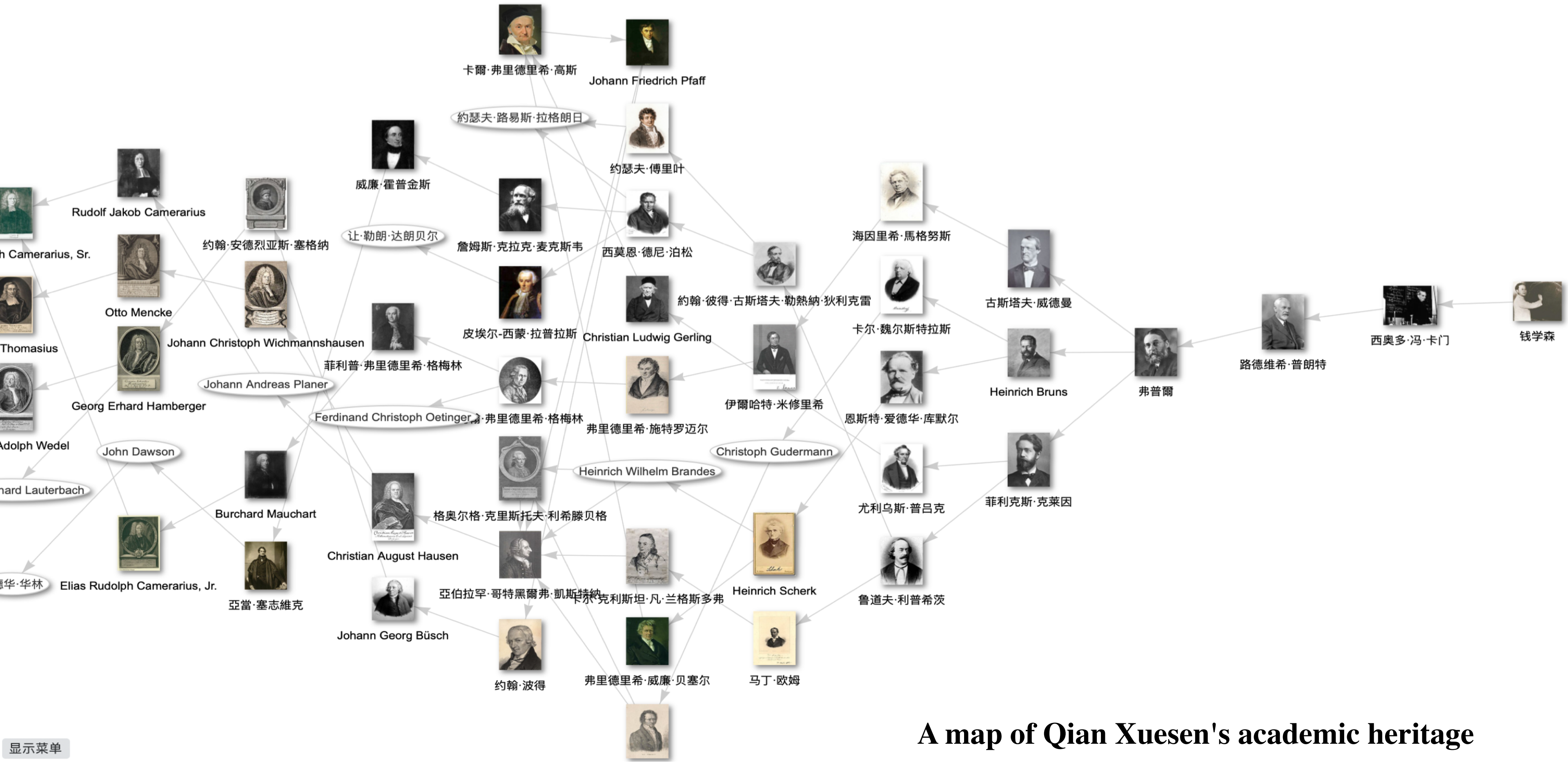





ATMs 5 km from Munich city center

# Accurate and complex intelligent queries

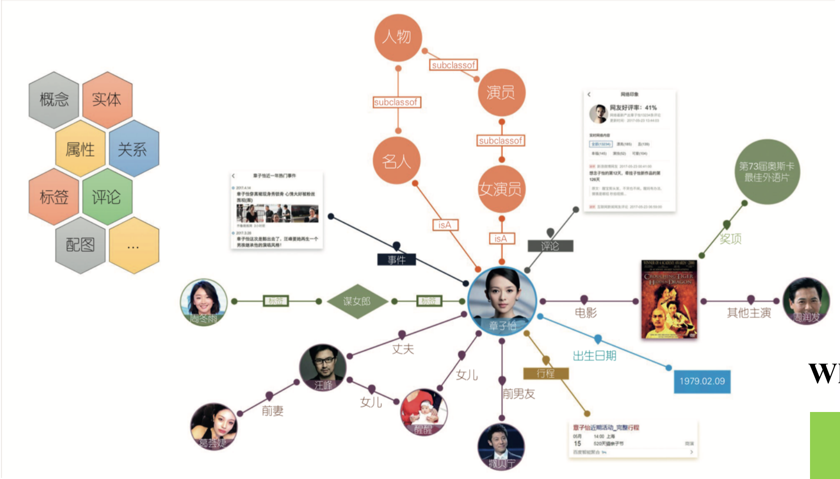
东南大学 SOUTHEAST UNIVERSITY



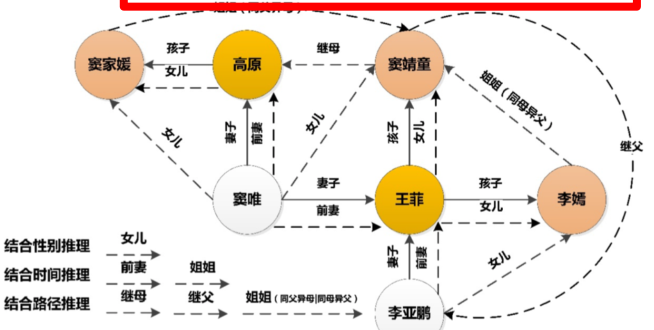

A map of Qian Xuesen's academic heritage

# Application example - search






Complex knowledge reasoning




Who is the ex-husband of Dou Jing Tong's father's ex-wife?


**Answer**
**Not link**

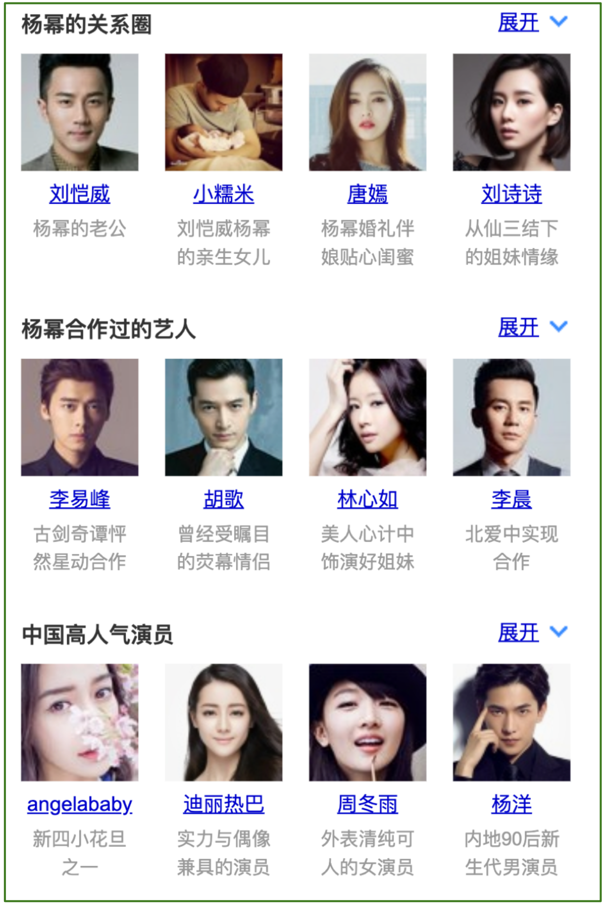
Explanatory knowledge reasoning



KGCODE LAB

# Application example - recommendation

东南大學 SOUTHEAST UNIVERSITY





# Application example – recommendation & ranking



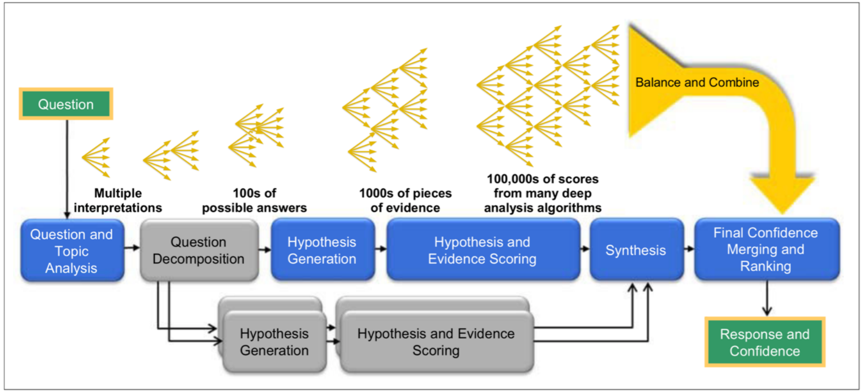
Which char more important?

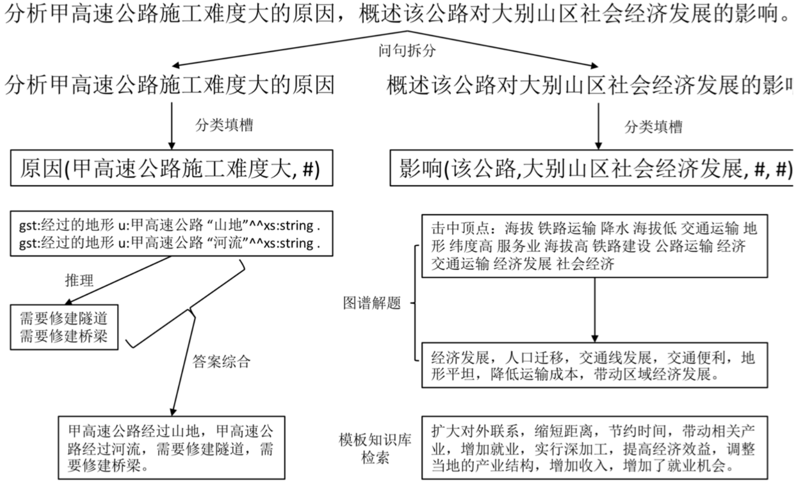

Why people who bought this lamp also bought this chair?

Trend of Darth Vader lamps?

# Application example - QA & search

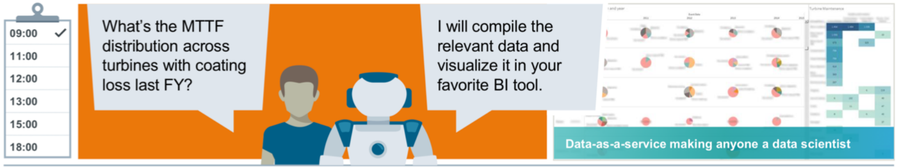





# Application example - manufacture






**Challenge**
- Required **data is distributed** across multiple databases
- Source systems have highly **complex schemas**
- Need to **include unstructured information** into analyses
- Reactivity and efficiency needs call for **end-user access to data**

\+

**Solution**
- **NLP** to make information from documents accessible for analytics
- **Physical and virtual data integration** to provide unified view
- Access using a **domain ontology** and intelligent **query construction support**
- **Connectors** providing data to existing tools in legacy format

=

**Value Generation**
- **Unified Data Hub:** All information accessible from one system, independent of source and type
- **Empower domain experts**: Subject matter experts can use domain language to access data
- **Enabler for analytics**: Foundation for fleet-level analytics

涡轮机去年平均故障间隔时间分布—涂层损耗?


**Challenge**
- Product configuration information is **scattered** across spreadsheets, **inconsistent and redundant**
- **Missing transparency** on technology interactions
- Information only on **HOW to design**, but **not on WHY** to do so

\+

**Solution**
- Use an Industrial Knowledge Graph to **store product configuration knowledge with rich semantics**
- For new order, **create constraint system on the fly** using knowledge graph information & **solve for feasible solutions**
- Use Industrial Knowledge Graph technology to **browse solutions**

=

**Value Generation**
- Introduce **knowledge management** into turbine configuration
- **Integrated design process** across all components and technologies
- Semantics allow **explain the WHY** behind a design decision
- **Speed-up** due to automation

如何匹配新涡轮机和用户的需求?


**Challenge**
- Siemens bank has a **wide range of investments** across industries
- Complex **networks of company relations** (own, partners, competitors)
- **Limited transparency on risks** due to external events, fraud, partner and competitor activities, etc

\+

**Solution**
- **Combine internal and purchased information** on companies and projects in Knowledge Graph
- Highly **flexible query interface** to support arbitrary queries (structured and natural language search)
- Interfaces to **support analytics** on top of integrated data

=

**Value Generation**
- **Highly agile analysis of risks** caused e.g. by unforeseen events
- **Improved transparency** over partners and competitors
- Identification of **potentially fraudulent behavior patterns**

我们的哪些资产直接或间接被这场风暴影响?


**Challenge**
- Large organizations have **huge bodies of rules & regulations**
- **Numerous facets of rule scope** (country, site, organization, …) complicate finding applicable rules
- Rules tell **WHAT to do, not HOW**
- Communication is done via **PDF-based circular documents**

\+

**Solution**
- Utilize **NLP technology** to extract subject, roles, scope, and activity information from circulars
- Industrial Knowledge graph to **integrate with organizational knowledge**, giving **contextualized descriptive guidance** (e.g. whom to call)

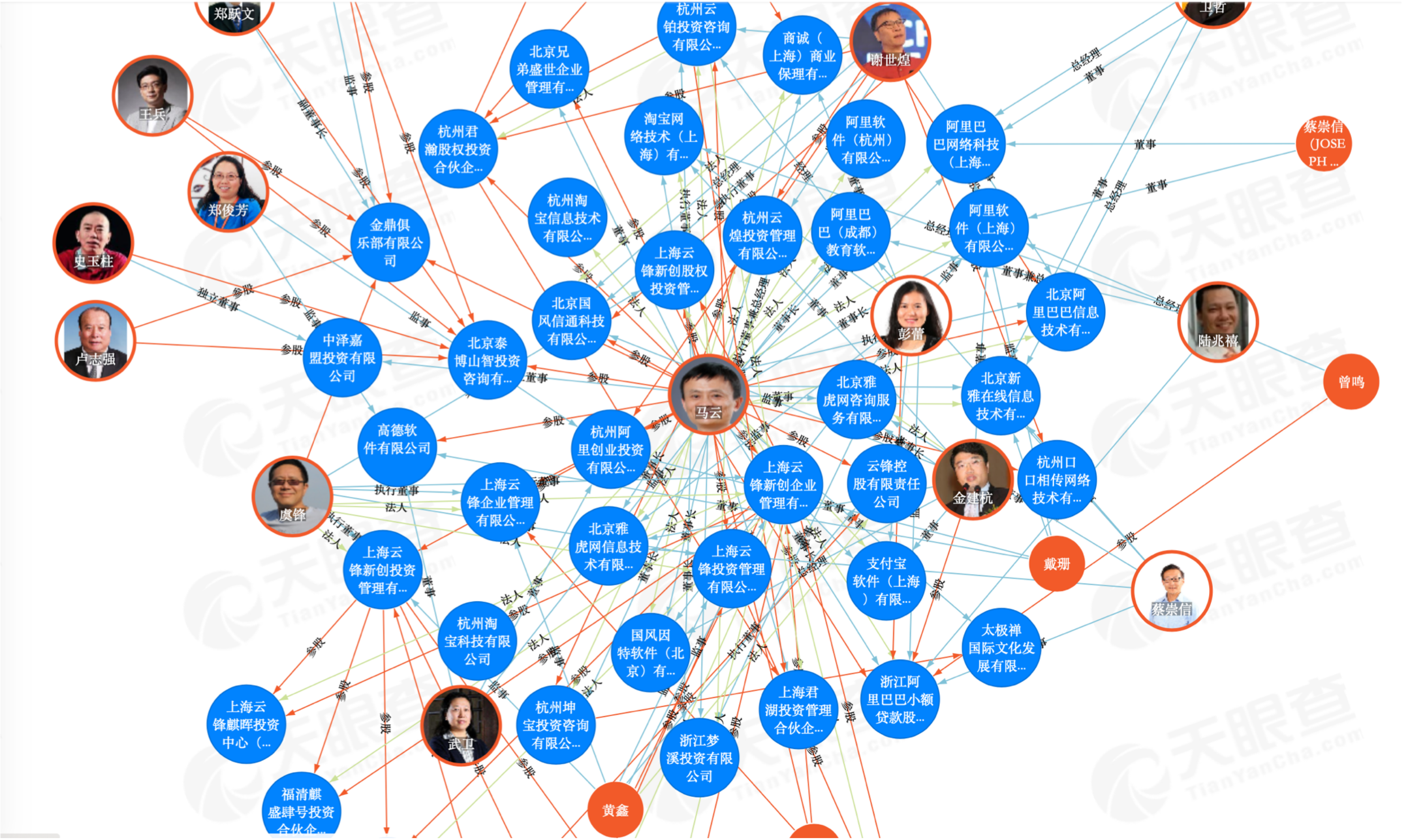
=

**Value Generation**
- **Reduce time** spent on understanding processes and following them
- **Reduce non-conformance cost** due to process violations
- **Increase employee satisfaction** by simplifying processes

需要联系谁进行批准?

# Application Example - Business Intelligence

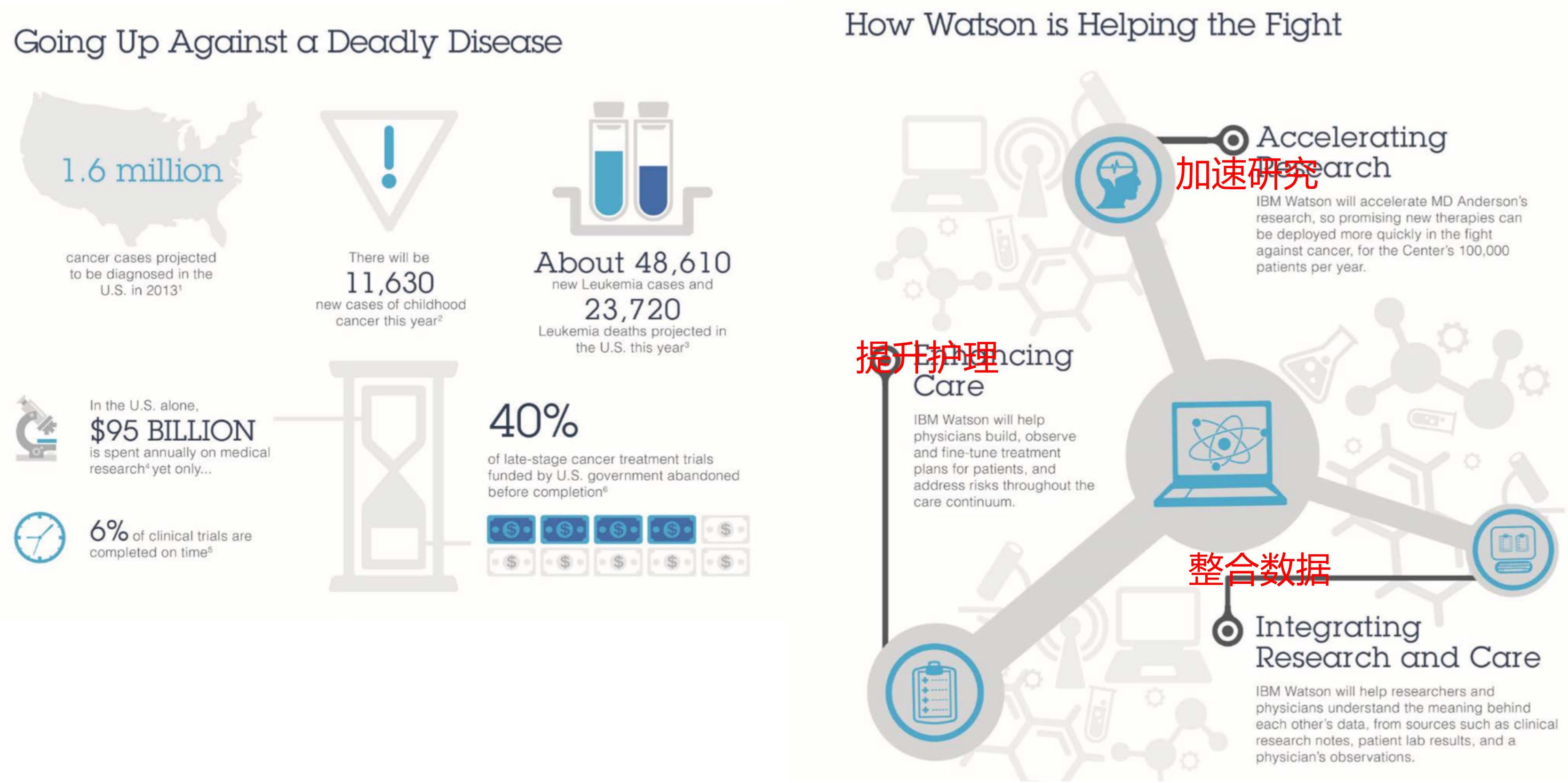




# Application example - medical diagnostics

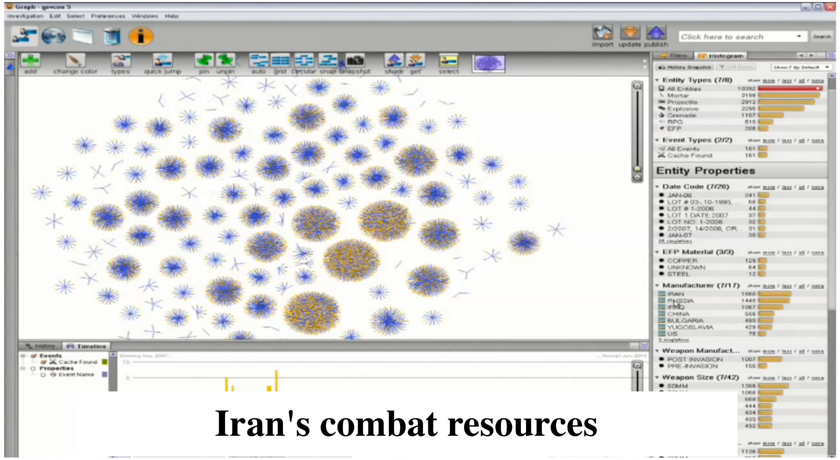


## Anderson Cancer Center teamed up with IBM Watson to end cancer

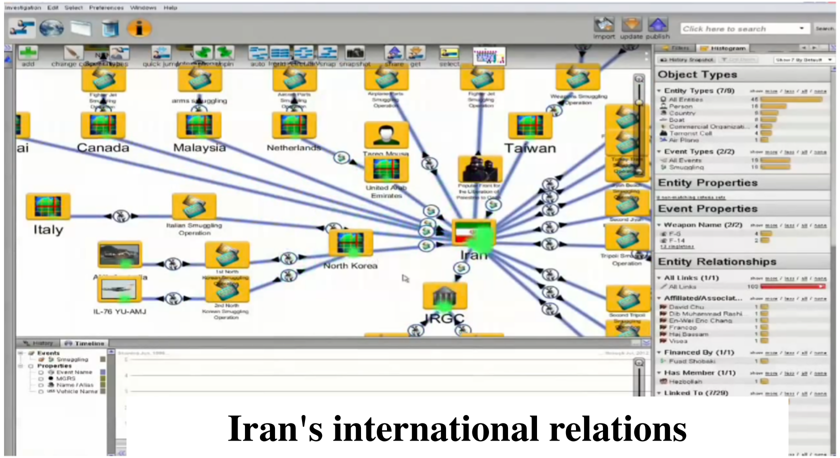


# Application example - Military/intelligence





Iran's combat resources



Iran's international relations

Iran's military facilities

Iran's chain of command

# The technical value of KG



- **Entity relations**
  - The entity relationship
  - Events, time and space
  - **Entity linking**
  - Build bridges between knowledge and entities
  - The intelligent foundation of understanding and interaction
  - **Integrate heterogeneous data**
  - by-product of knowledge graphs
  - KG has the advantage of flexible integration of heterogeneous data
  - **Large-scale knowledge reasoning**
  - Discover implicit, incorrect, inconsistent knowledge
  - Explainable intelligence foundation

# The academic frontier of AI



- **IJCAI2017 Award for Best Outstanding Paper**

**-Foundations of Declarative Data Analysis Using Limit Datalog Programs**

- **AAAI2017 Outstanding Paper Award**

**-Label-Free Supervision of Neural Networks with Physics and Domain Knowledge**

- **AAAI2018 Classic Paper Award**

**-PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment (AAAI 2000)**

- **IJCAI2017 Award for Excellence in Papers**

**-BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network**

**-YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia**

- **IJCAI2018 Award for Excellence in Papers**

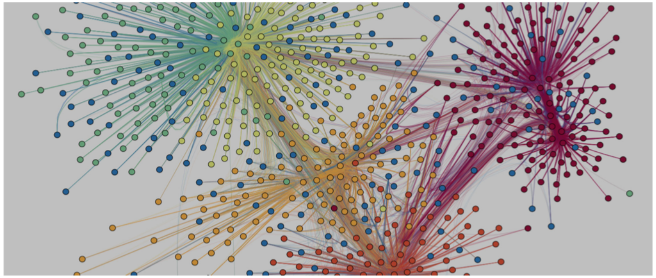
**-From Conjunctive Queries to Instance Queries in Ontology-Mediated Querying**

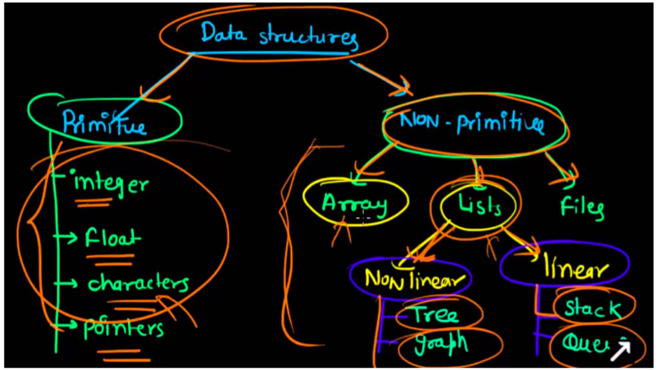
**-Commonsense Knowledge Aware Conversation Generation with Graph Attention**

# Multi-source heterogeneous data source

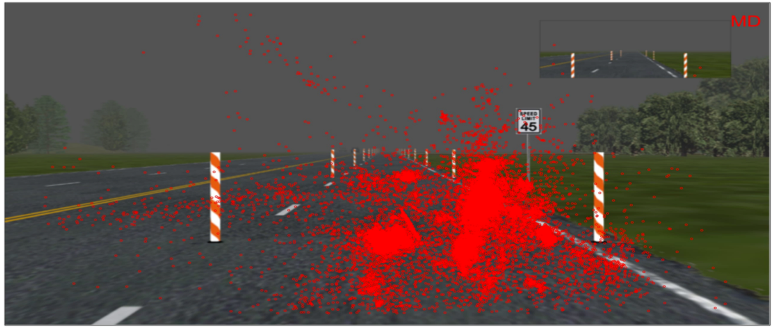


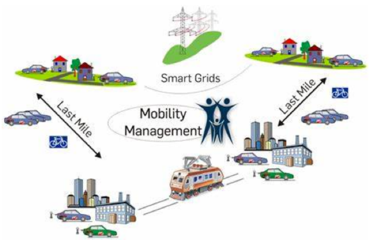
## Structured/semi-structured data

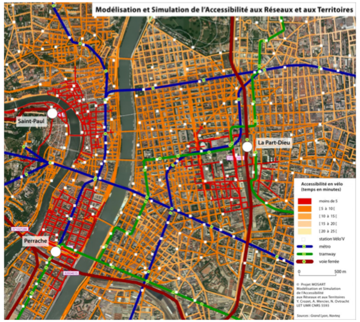





## Unstructured, multimodal, space-time, process data



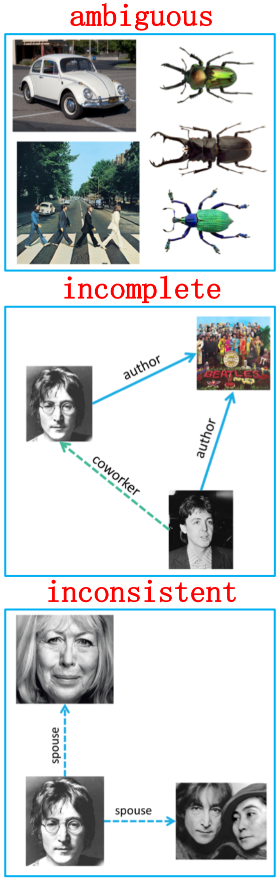



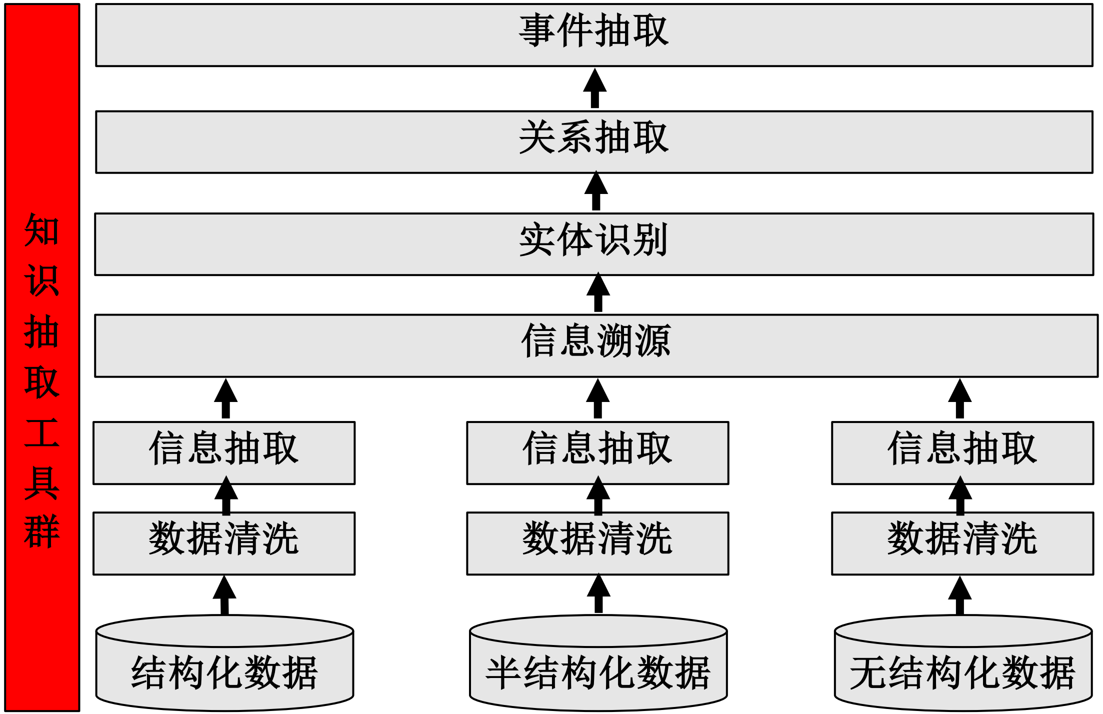




Multimodal (MM, code, signal, track)

# Knowledge extraction

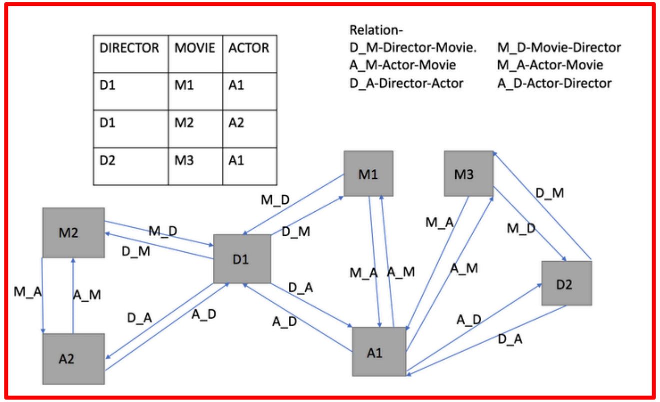




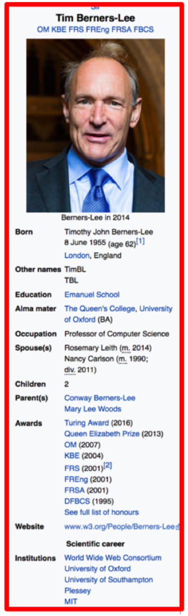



# Knowledge extraction for structured/semi-structured data

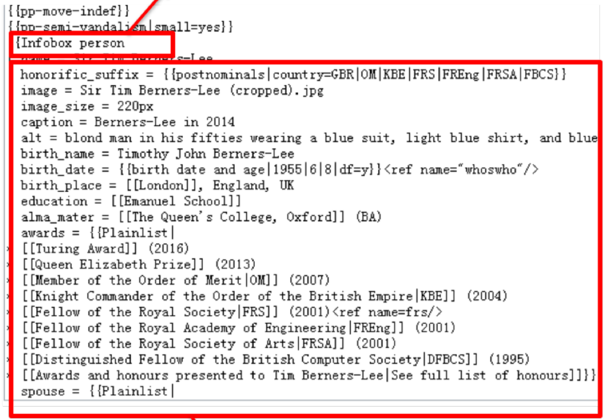





表(Table)—类(Class)
列(Column)—属性(Property)
行(Row)—资源/实例(Resource/Instance)
单元(Cell)—属性值(Property Value)
外键(Foreign Key)—指代(Reference)




*infobox template names=instance types*, (rdf: type)


```
{{pp-move-indef}}
{{pp-semi-vandalism|small=yes}}
{Infobox person
honorific_suffix = {{postnominals|country=GBR|OM|KBE|FRS|FREng|FRSA|FBCS}}
image = Sir Tim Berners-Lee (cropped).jpg
image_size = 220px
caption = Berners-Lee in 2014
alt = blond man in his fifties wearing a blue suit, light blue shirt, and blue
birth_name = Timothy John Berners-Lee
birth_date = {{birth date and age|1955|6|8|df=y}}<ref name="whoswho"/>
birth_place = [[London]], England, UK
education = [[Emanuel School]]
alma_mater = [[The Queen's College, Oxford]] (BA)
awards = {{Plainlist|
[[Turing Award]] (2016)
[[Queen Elizabeth Prize]] (2013)
[[Member of the Order of Merit|OM]] (2007)
[[Knight Commander of the Order of the British Empire|KBE]] (2004)
[[Fellow of the Royal Society|FRS]] (2001)<ref name=frs/>
[[Fellow of the Royal Academy of Engineering|FREng]] (2001)
[[Fellow of the Royal Society of Arts|FRSA]] (2001)
[[Distinguished Fellow of the British Computer Society|DFBCS]] (1995)
[[Awards and honours presented to Tim Berners-Lee|See full list of honours]]}}
spouse = {{Plainlist|
```

*infobox properties=instance properties*, (dbpedia:property/[propertyName])

**Structured data - knowledge**

**Semi-structured data - knowledge**

# Knowledge extraction - NER



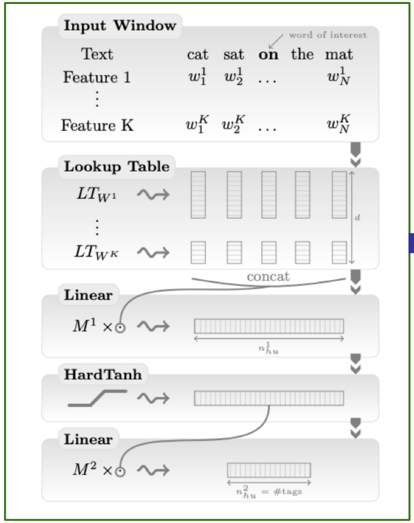
时间　　　　　　　　　　　　　　人名

北京时间3月23日0时50分许，美国总统特朗普在白宫正式签署对华贸易备忘录。特朗普当场宣布，将有可能对600亿美元的中国出口商品征收关税。

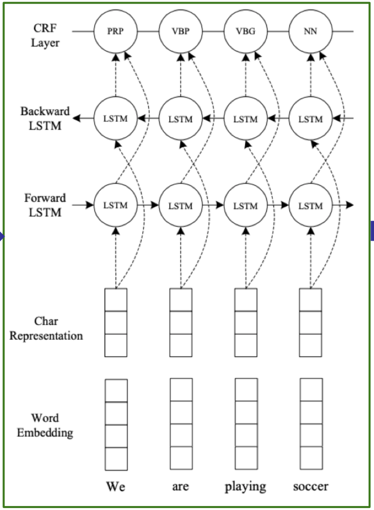
货币　　　　　地理位置

**NER methods: rules, dictionaries, machine learning models**

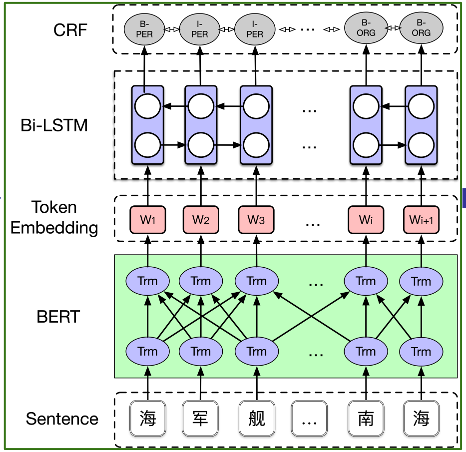
# Knowledge extraction - NER



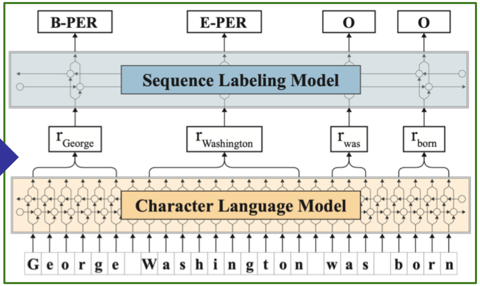

2011
NN/CNN + CRF
89%




2016
BiLSTM+CRF
91%



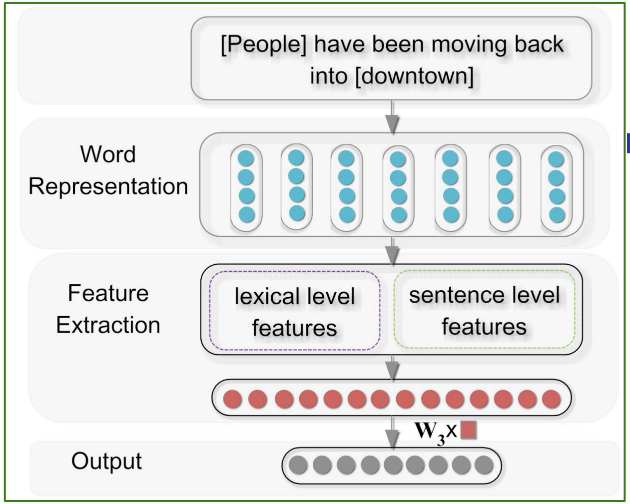

2018
BERT+BiLSTM+CRF
92.8%

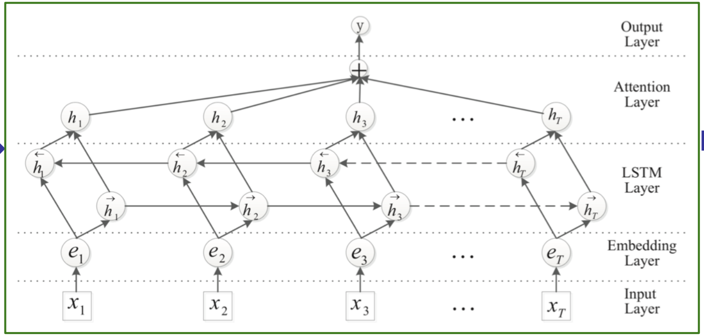


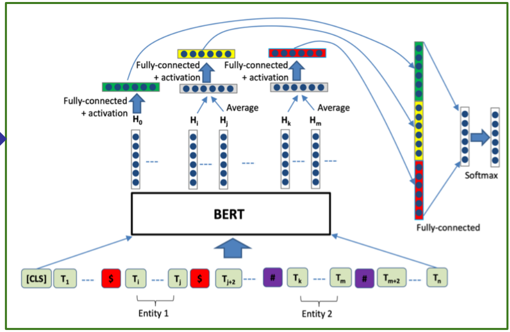

2019
Models+Flair/Fine tune
93.09%~93.5%

# Knowledge extraction - RE





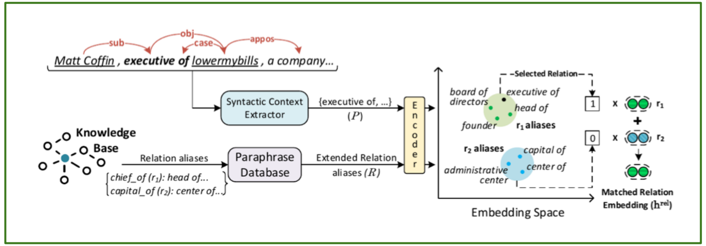




2012-2014
POS+CNN +RNN
82%

2016
Attention+BiLSTM
84%

2019
BERT-based Models
89.25%~89.5%




2018
Distantly-Supervised

# Knowledge Extraction - Event Extraction



Person Organization

Time Position

**Barry Diller on Wednesday quit as chief of Vivendi Universal Entertainment.**

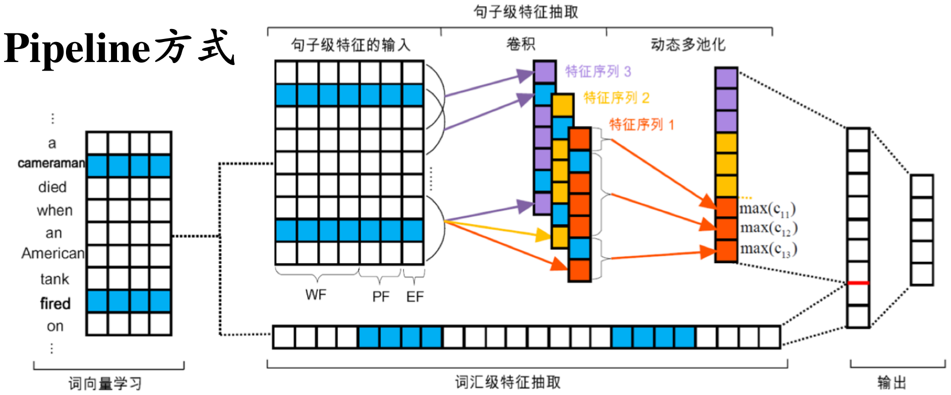
Trigger Word

| | | |
|---|---|---|
| **Trigger** | quite("Personnel/End-Position" event) | |
| **Argument** | **Role=Person** | Barry Diller |
| | **Role=Organization** | Vivendi Universal Entertainment |
| | **Role=Position** | Chief |
| | **Role=Time-within** | Wednesday(2003-03-04) |

# Knowledge Extraction - Event Extraction

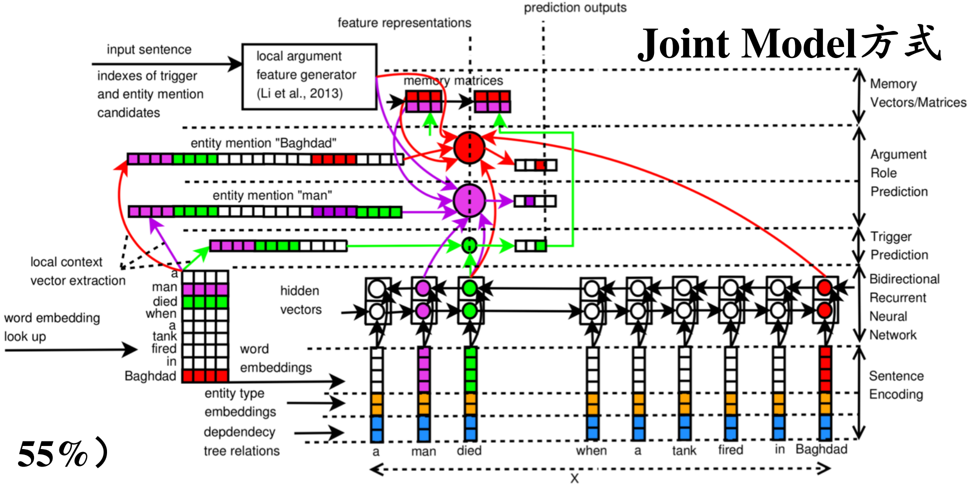


## Pipeline方式




**DMCNN（69%，53%）**

## Joint Model方式




**JRNN（69%，55%）**

# Knowledge Extraction – NER examples



When Adam had lived 130 years, he had a son in his own likeness, in his own image; and he named his son Seth. After Seth was born, Adam lived 800 years and had other sons and daughters. Altogether, Adam lived a total of 930 years

Select Example 提交

When God created mankind, he made them in the likeness of God. He created them male and female and blessed them . And he named them "Mankind" when they were created.

When Adam had lived 130 years, he had a son in his own likeness, in his own image; and he named his son Seth . After Seth was born, Adam lived 800 years and had other sons and daughters. Altogether, Adam lived a total of 930 years, and then he died.

When Seth had lived 105 years, he

Graph | Entities | Facts | Open Facts | Categories | Documentation

| Entity | Types | Salience | Sentiment |
|---|---|---|---|
| Seth | person, fictional entity | 1.00 | 0.66 |
| Methuselah | person, fictional entity | 1.00 | 0.78 |
| Jared | person, fictional entity | 1.00 | 0.44 |
| Enoch | person, fictional entity | 1.00 | 0.77 |
| Noah | person, fictional entity | 1.00 | 0.35 |
| Lamech | person, fictional entity | 1.00 | 0.75 |
| Adam | person, fictional entity | 1.00 | 0.68 |
| Kenan | person, fictional entity | 0.99 | 0.62 |
| Enos | person, fictional entity | 0.99 | 0.43 |

# Knowledge Extraction – RE examples



When Adam had lived 130 years, he had a son in his own likeness, in his own image; and he named his son Seth. After Seth was born, Adam lived 800 years and had other sons and daughters. Altogether, Adam lived a total of 930 years

Select Example

提交

When God created mankind, he made them in the likeness of God. He created them male and female and blessed them . And he named them "Mankind" when they were created.

When Adam had lived 130 years, he had a son in his own likeness, in his own image; and he named his son Seth . After Seth was born, Adam lived 800 years and had other sons and daughters. Altogether, Adam lived a total of 930 years, and then he died.

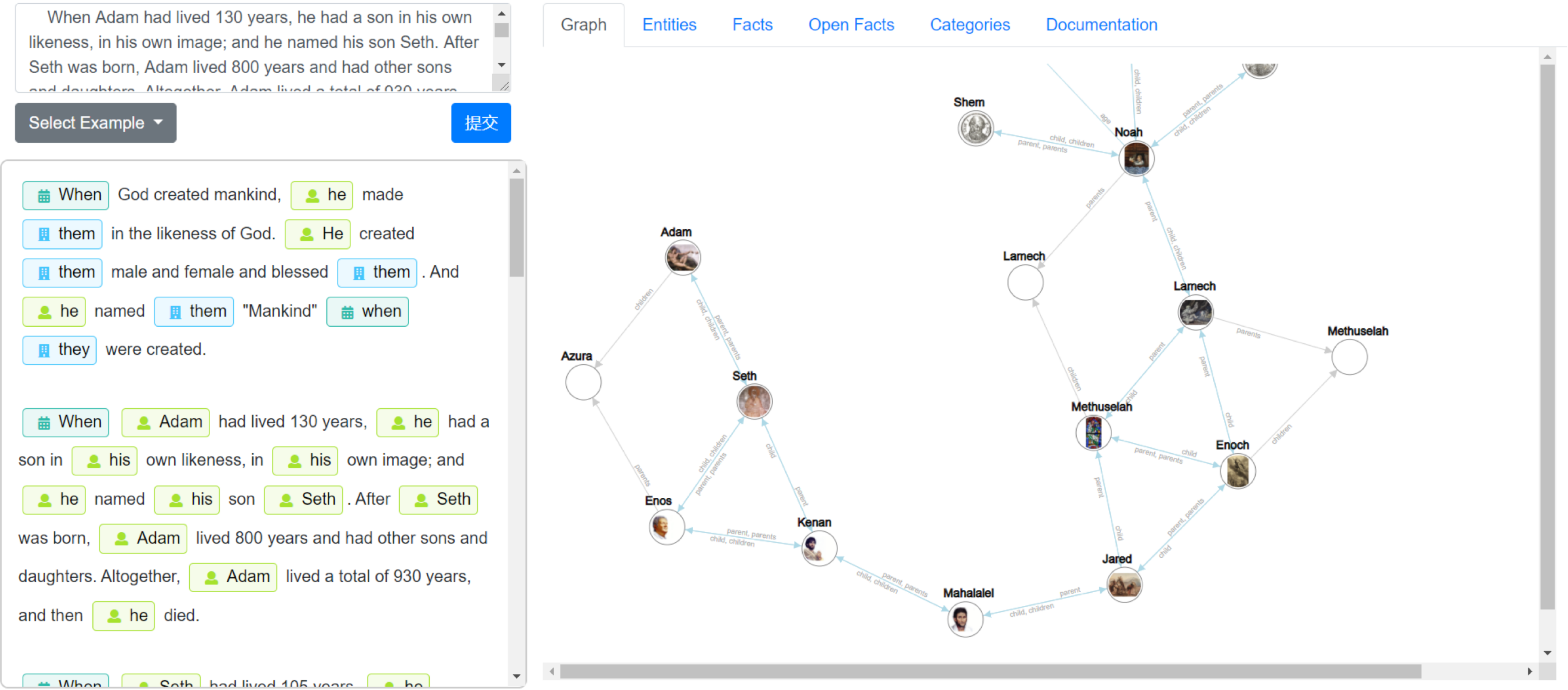
When Seth had lived 105 years he

Graph | Entities | Facts | Open Facts | Categories | Documentation

| Entity | Property | Value | Qualifiers |
|---|---|---|---|
| Seth | parent (1.00) | Adam | |
| Adam | child (1.00) | Seth | |
| Seth | child (1.00) | Enos | |
| Enos | parent (1.00) | Seth | |
| Enos | child (0.90) | Kenan | |
| Kenan | parent (0.90) | Enos | |
| Kenan | parent (1.00) | Seth | |
| Seth | child (1.00) | Kenan | |
| Mahalalel | parent (1.00) | Kenan | |

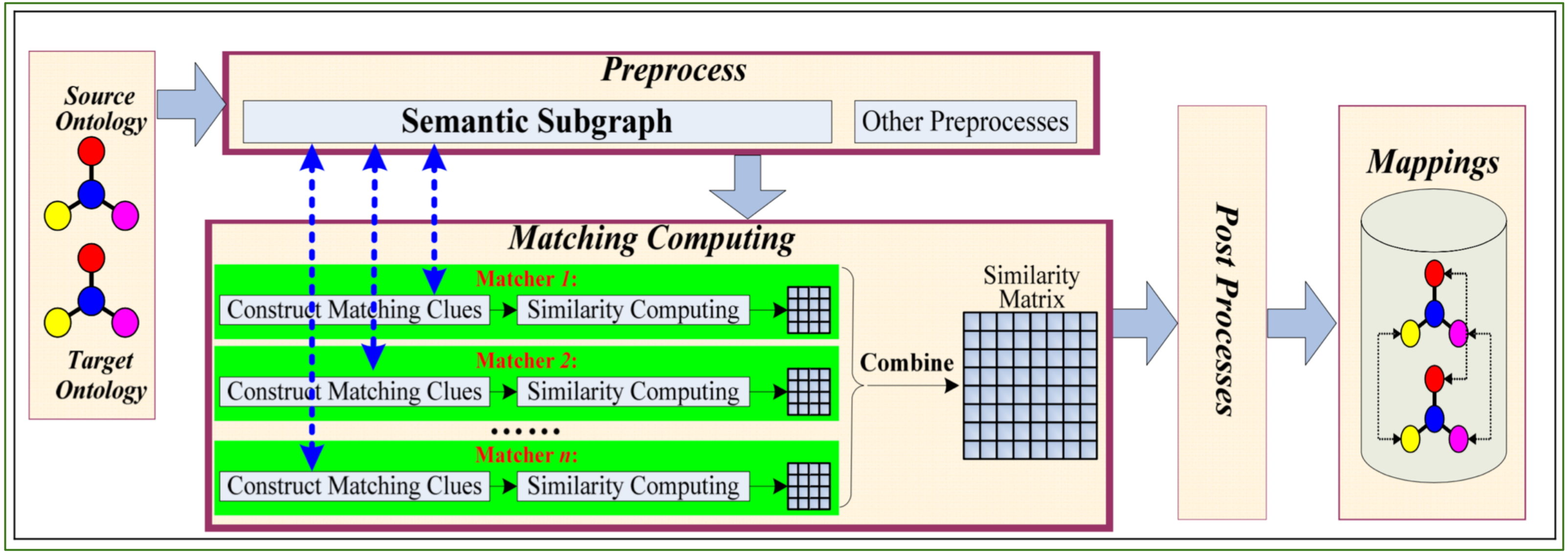
# Knowledge Extraction – KG examples





# Knowledge Fusion - Ontology Matching





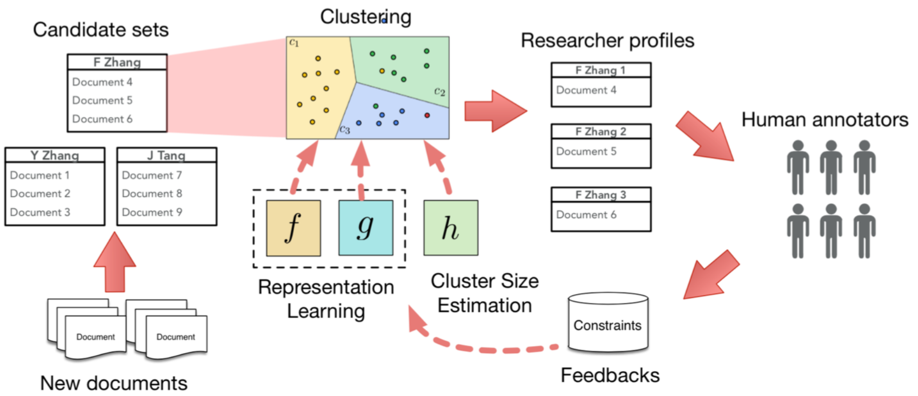

预处理：解析、数据清洗、构造基础数据
匹配计算：构造匹配线索、相似度计算
后处理：匹配结果抽取、匹配调谐

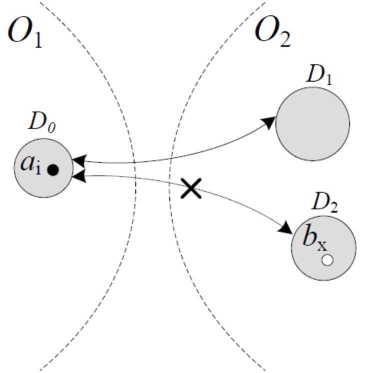
- 本体异构示例：华东野战军-三野 二炮-火箭军 胖妞-运20

# Knowledge Fusion - Ontology Matching

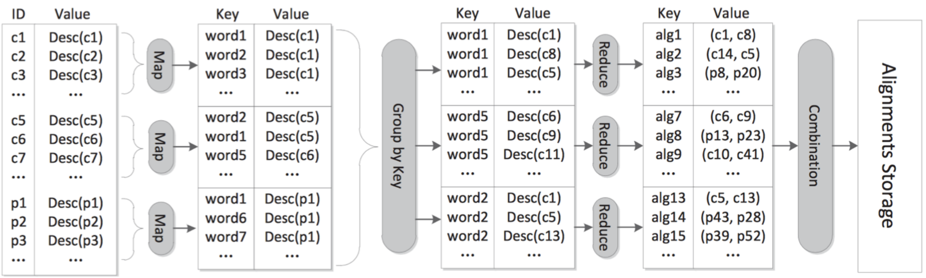


分治策略



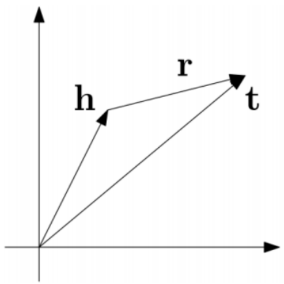




并行提速



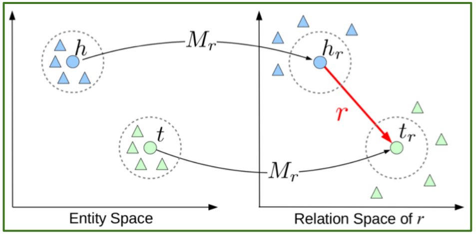

- 实例异构示例：（官员）郎平 <> （女排）郎平

# KG representation learning



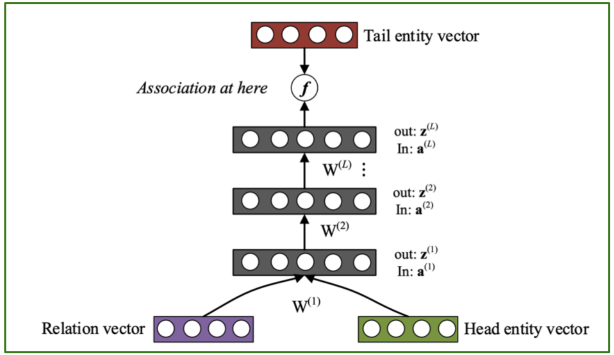

$$f(h,r,t) = \| h + r - t \|_{l1/l2}$$

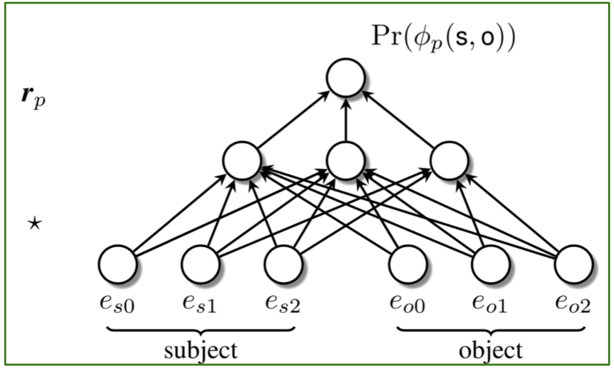
$$L = \sum_{\xi \in T} \sum_{\xi' \in T'} [\gamma + f(\xi) - f(\xi')]_{+}$$




翻译模型

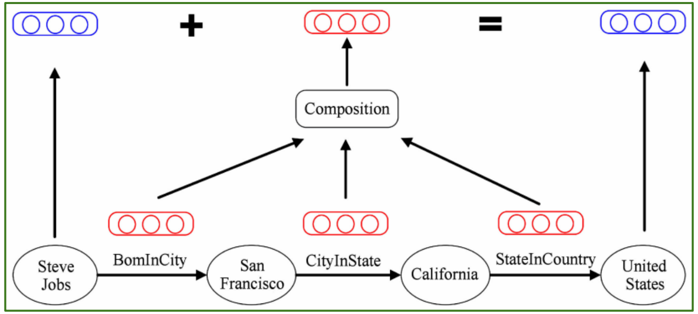
**TransE TransH TransR 等**

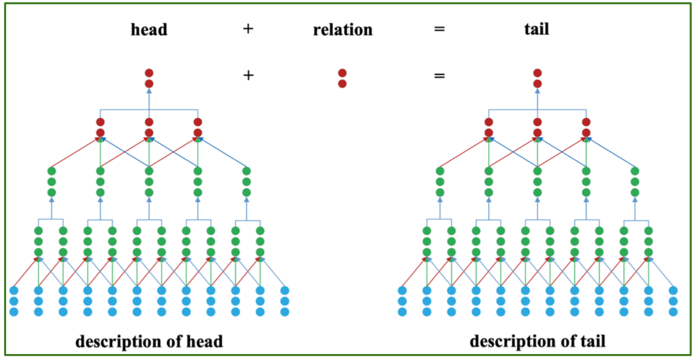



语义匹配模型

**RESCAL MLP NAM等**

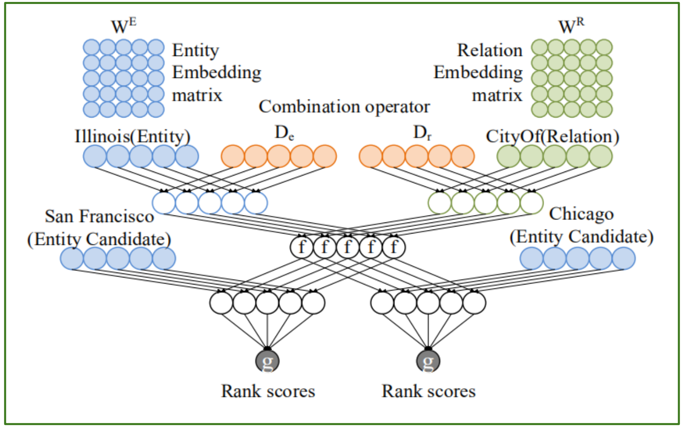


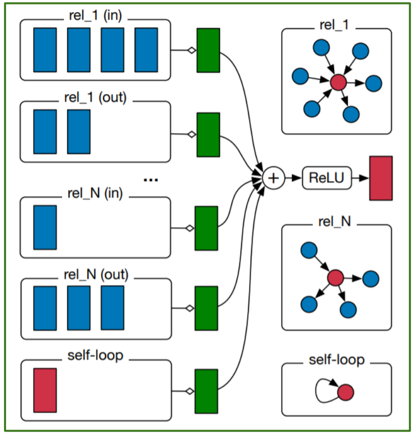

融合模型

**PTransE TKRL DKRL 等**

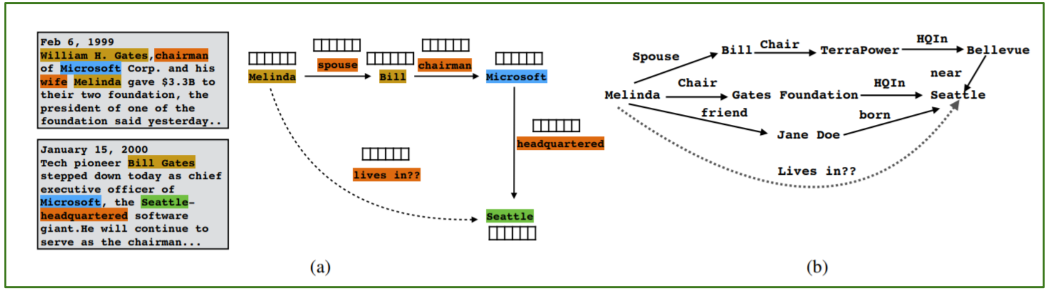
# KG reasoning




NN+基于语义的推理




NN+基于结构的推理

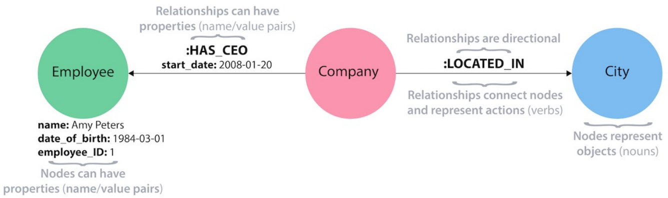



NN+基于路径的推理

# KG storage

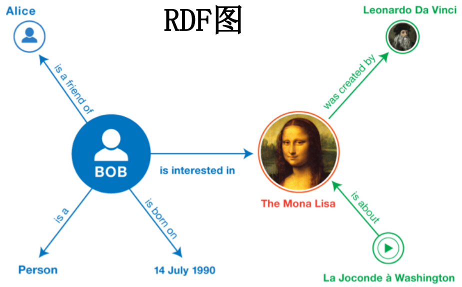


## 属性图

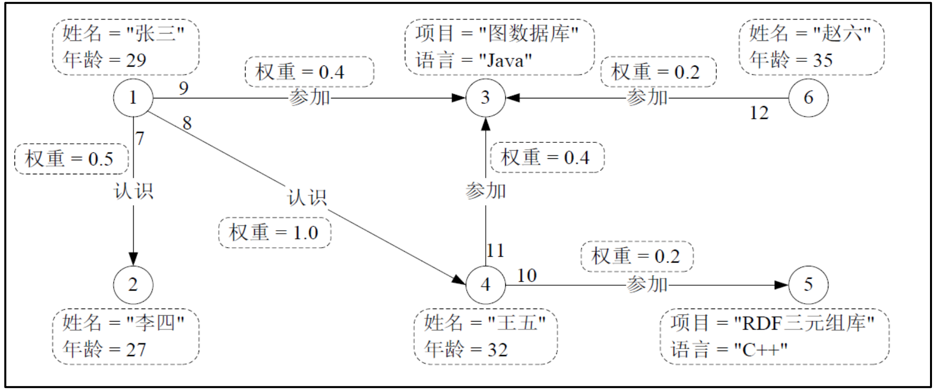


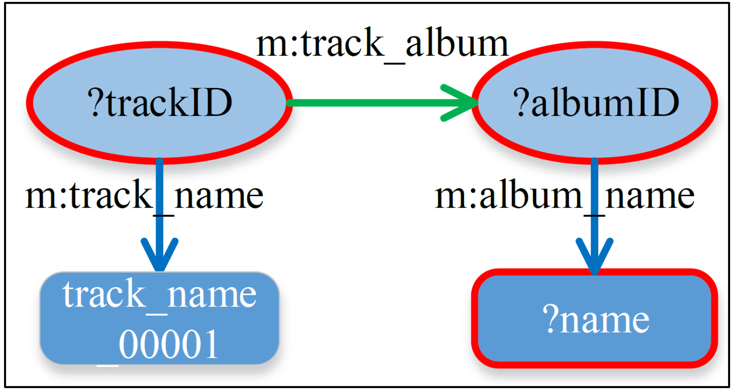

Neo4j, JanusGraph, ArangoDB, TigerGraph

## RDF图




gStore、Virtuoso、AllegroGraph、RDF4J




Gremlin: *g.v(1).out(‘认识’).filter{it.年龄>30}.out(‘参加‘).项目*

Cypher: *MATCH (n)-[:认识]->(p), (p)-[:参加]->(pr)*
*WHERE n.id==1, p.年龄 > 30*
*RETURN pr.项目*

SPARQL

```
PREFIX m: <http://kg.course/music/>
SELECT ?name
WHERE {
        ?trackID m:track_name "track_name_00001"
        ?trackID m:track_album ?albumID
        ?albumID m:album_name ?name
}
```

# KG in SEU

## 公开课程　https://github.com/npubird/KnowledgeGraphCourse

Unwatch 78　Unstar 1.7k　Fork 608

访问量已超过35万次

**A systematic course about knowledge graph for graduate students, interested researchers and engineers.**

东南大学《知识图谱》研究生课程
时间：2019年春季（2月下旬~5月中旬）
每周五下午2:00~4:30
地点：东南大学九龙湖校区，纪忠楼Y205
答疑/讨论/建议：请致信 pwang AT seu.edu.cn

### 课程内容

#### 第1讲 知识图谱概论 （2019-3-1,2019-3-8）

1.1 知识图谱起源和发展

爱可可-爱生活
8-31 来自Mac客户端

'东南大学《知识图谱》研究生课程(资料)' by Peng Wang GitHub: 网页链接

课程内容

第1讲 知识图谱概论 （2019-3-1,2019-3-8）

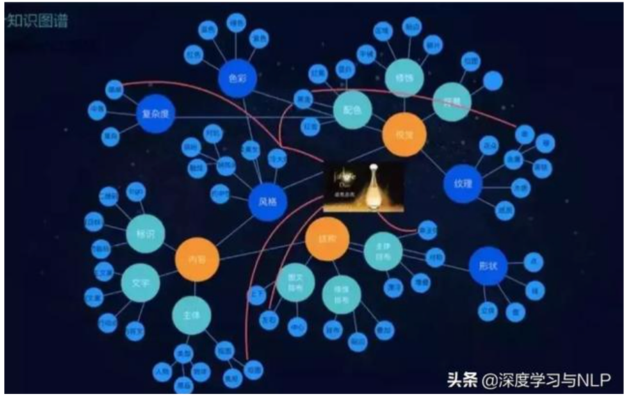
1.1 知识图谱起源和发展
1.2 知识图谱 VS 深度学习

今日头条　首页 / 教育 / 正文

24　转发　微博　Qzone　微信

**东南大学2019年最新研究生精品课程《知识图谱》资源分享**

原创 深度学习与NLP 2019-09-15 09:50:50




国内访问最多的知识图谱课程之一

# research interests

东南大学 SOUTHEAST UNIVERSITY

**智能**

- 智能问答：KBQA、NL2SQL、多轮问答
- 搜索推荐：语义索引、语义排序、精准推荐
- 智能决策：语义关联, 信息汇聚、可解释AI、推演复盘
- 文本生成：古诗/笑话生成、看图写作、辅助科学写作

**认知**

- 知识图谱：大规模知识图谱构建理论与方法、多模态知识图谱、大规模知识图谱存储与查询、知识可视化、面向复杂场景的多策略知识抽取技术、知识融合、知识质量评估、知识推理、知识图谱表示学习

**工具**

- 自然语言处理：文本分类、情感分析、实体识别、关系抽取、事件抽取

**社会**

- 社交网络：网络演化、链接预测、网络表示学习、网络挖掘与分析

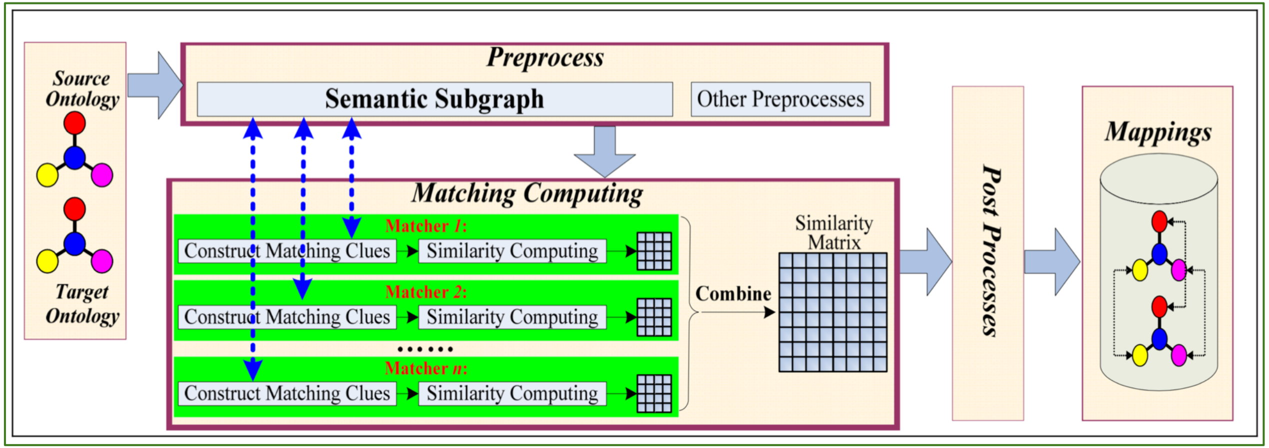
**技术**

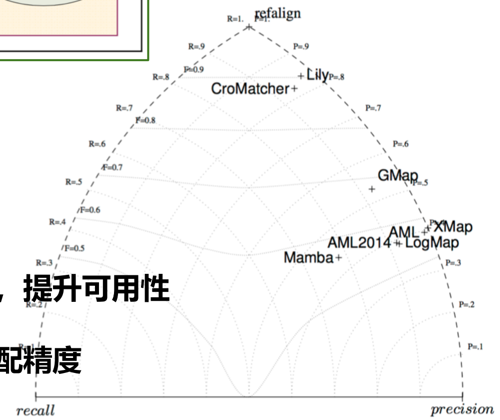
- 支撑技术：深度学习、机器学习、知识工程、博弈论、信息检索、数据库、算法、编码、……

# Knowledge Fusion/Integration



## Lily




- ✓ 基于语义子图的通用本体匹配框架（2006-2010）
  特色：精确描述语义、高精确度、高召回率
- ✓ 大规模本体匹配（2008-2013）
  特色：巧妙利用层次结构和匹配区域性，无需划分本体
- ✓ 弱信息本体匹配（2008-2013）
  特色：利用语义子图和相似度传播
- ✓ 本体匹配调试 (2008-2009)
  特色：首次提出并给出启发式解决之一
- ✓ 本体匹配调谐 (2012-至今)
  特色：基于机器学习实现匹配系统自动调谐，提升可用性
- ✓ 实例匹配 (2013-至今)
  特色：高效处理大规模实例匹配，并保证匹配精度
- ✓ 本体匹配系统Lily (2006-至今)

# Challenges and the future



- **Challenges**
  - 成本瓶颈：缺少高效、重用性好、成本低的相对通用知识图谱平台
    - ✓ 本体构建（非常难用）
    - ✓ 知识抽取（非常复杂）数据处理复杂、模型训练复杂
    - ✓ 知识存储
    - ✓ 知识查询（非常耗时）慢且受制于NLU
    - ✓ 知识可视（非常难看）
    - ✓ 知识管理
    - ✓ 领域知识（非常难积累）
    - ✓ API服务

# Challenges and the future



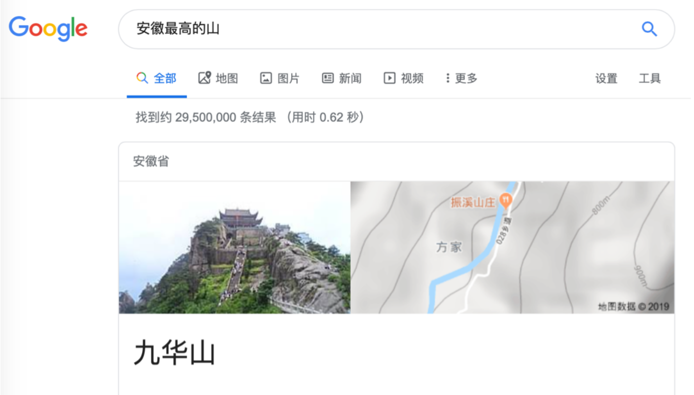
## Challenges

- 知识构建瓶颈
  - ✓ 知识密集型领域知识构建需要领域专家深度参与
    - 西医、中医、生物医学、化工、军事、……
  - ✓ 工业领域知识构建与传统场景差别很大
    - 大量ERP, CRM信息系统，数据异构严重，缺乏训练模型的语料
  - ✓ 大量的多模态场景需要有效的知识构建技术
    - 大量非结构化文本中的高价值知识亟待抽取、整理和总结
    - 大量图像、视频、音频中存在高价值知识
  - ✓ 大量长尾和离散知识的构建
    - 大量知识的分布具有长尾特点，训练数据少，抽取困难且代价高
    - 大量离散的/非记录的/机密的/高商业价值的知识需要持续总结和构建
      - 老工人的生产经验，老中医的诊疗经验，警察破案经验，供应链知识，……
  - ✓ 长尾实体和关系难以有效抽取
    - 超过80%的实体和关系都是长尾分布的
  - ✓ 没有通用意义上的知识构建技术

# Challenges and the future

## Challenges

- 知识质量瓶颈
  - ✓ 前期忽视了知识的质量问题
    - 知识不全难以解决
    - 知识错误和冲突
  - ✓ 知识补全/推理侧重于“边”，忽略了“点”
  - ✓ 缺乏公认的直接知识质量评估体系和标准
  - ✓ 通过任务间接评估知识质量的可信度不够
  - ✓ 大规模知识图谱评估中如何有效抽样


【#亚马逊智能音箱劝主人自杀#：活着会给地球造成负担😲】据英国媒体每日邮报12月19日消息，英国的Danni询问智能助手Alexa心动周期问题时，Alexa突然说“心跳是件坏事，人活着会给地球造成负担”，还建议主人“将刀插入心脏”。😲Danni被吓坏了。亚马逊称，它出故障念了百科。 全球视频大魔王的微博视频 收起全文 ^

# Challenges and the future



- **Challenges**
  - 大规模知识存储的瓶颈
    - ✓ 缺乏开源、易用、高性能、易扩展的图数据库
    - ✓ 属性图 VS RDF图
    - ✓ CYPHER VS GREMLIN VS SPARQL
    - ✓ 多节点 VS 分布式
  - 大规模知识推理的瓶颈
    - ✓ 图是一种天然稀疏的数据
    - ✓ 知识图谱中语义信息不足，传统逻辑推理效果不好
    - ✓ 概率推理在知识图谱中难解释
    - ✓ 知识不全、错误和冲突普遍存在，阻碍推理的进行
    - ✓ 链接预测和实体分类无法满足很多应用场景
    - ✓ 大规模知识推理任重道远

# Challenges and the future



## Challenges

- 知识图谱表示学习的瓶颈
  - ✓ 大规模知识图谱表示学习性能瓶颈；空间向量不等于原始语义。
- 知识融合的瓶颈
  - ✓ 大规模实例匹配、属性值匹配、跨语言匹配、跨模态匹配的性能。
- 基于知识智能问答的瓶颈
  - ✓ 意图识别受制于NLP和NLU的瓶颈
  - ✓ 能精准回答已有知识，但无法有效回答相关的知识
- 处理数据的弱点
  - ✓ 不适合处理表格数据、时序数据、大规模统计数据、过程、计算等。
- 表达能力的弱点
  - ✓ 三元组表达能力有限，不适合表示一切知识，自然语言中很多信息难以被三元组表示。

# Challenges and the future



## Challenges

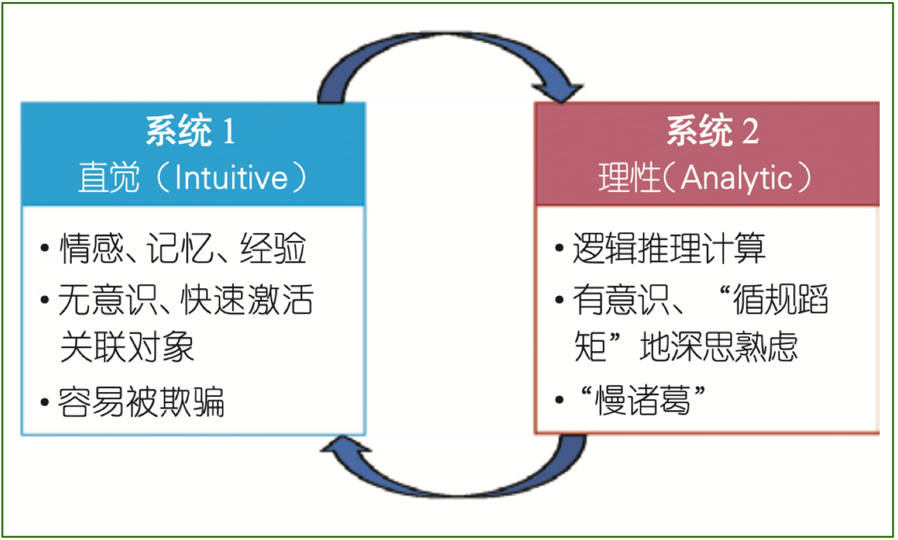
- 应用瓶颈（希望、期望和失望）
  - ✓ 成功的语义搜索：挖掘高价值信息，改善传统信息检索(IR)的精度问题
  - ✓ 人机交互的曙光：产生可用的智能问答，自然人机交互出现曙光
  - ✓ 失望：
    - 大量应用集中在浅层，进行可视化、查询、问答等辅助智能工作
    - 无法在大量的行业中明显提升生产效率，解放大量人力，展示智能潜力
    - 解决之前技术无法突破的问题

# Challenges and the future

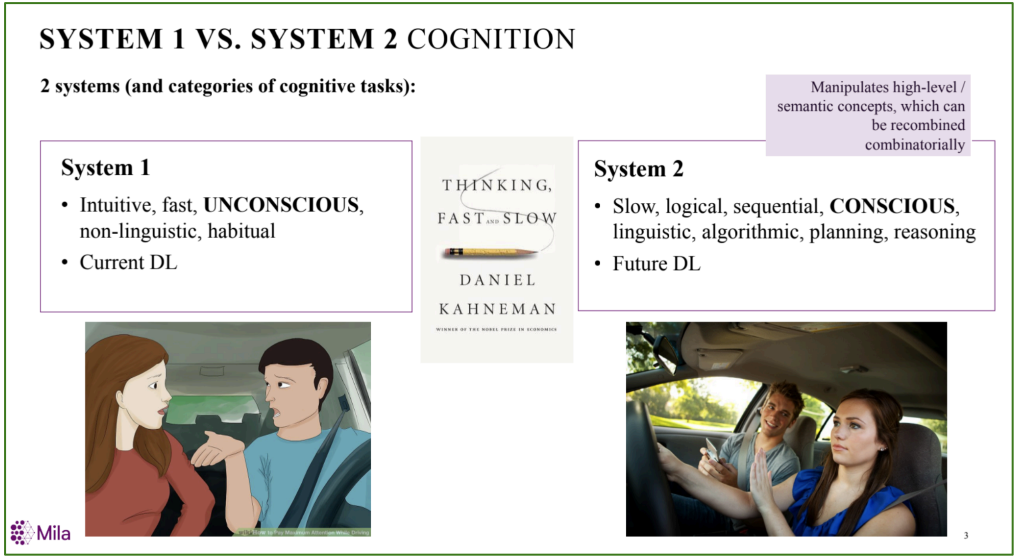


## ● Cognitive Dual Systems: (System1+System2)

- System1:直觉、快、无意识、非语言、习惯性
- System2:理性、慢、有意识、有语言、有逻辑、连续、算法、规划、推理




Kahneman, Thinking, Fast and Slow, 2002




Yoshua Bengio，NeurIPS 2019

धन्यवाद
Hindi
Hindi

多謝
Traditional Chinese

ขอบคุณ
Thai

Спасибо
Russian

Thank You
English

Gracias
Spanish

شكراً
Arabic

Obrigado
Brazilian Portuguese

Danke
German

Grazie
Italian

多谢
Simplified Chinese

Merci
French

நன்றி
Tamil
Tamil

ありがとうございました
Japanese

감사합니다
Korean