

# 法律声明

---

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院和讲者共同拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



# 第一讲 知识图谱概览

---

# 大纲

---

- 知识图谱与语义技术概述
- 典型知识库项目简介
- 知识图谱相关技术简介
  - 知识表示、知识抽取、知识存储、知识融合
  - 知识推理、知识众包、语义搜索、知识问答
- 典型应用案例：医药、电商、企业、金融、博物馆、媒体

---

# 第一部分: 知识图谱与语义技术概述

# MEMEX – 记忆机器

人的记忆偏重关联

"Wholly new forms of **encyclopedias** will appear, ready made with a **mesh of associative trails** running through them, ready to be dropped into the memex and there amplified"

As We May Think, The Atlantic, 1945

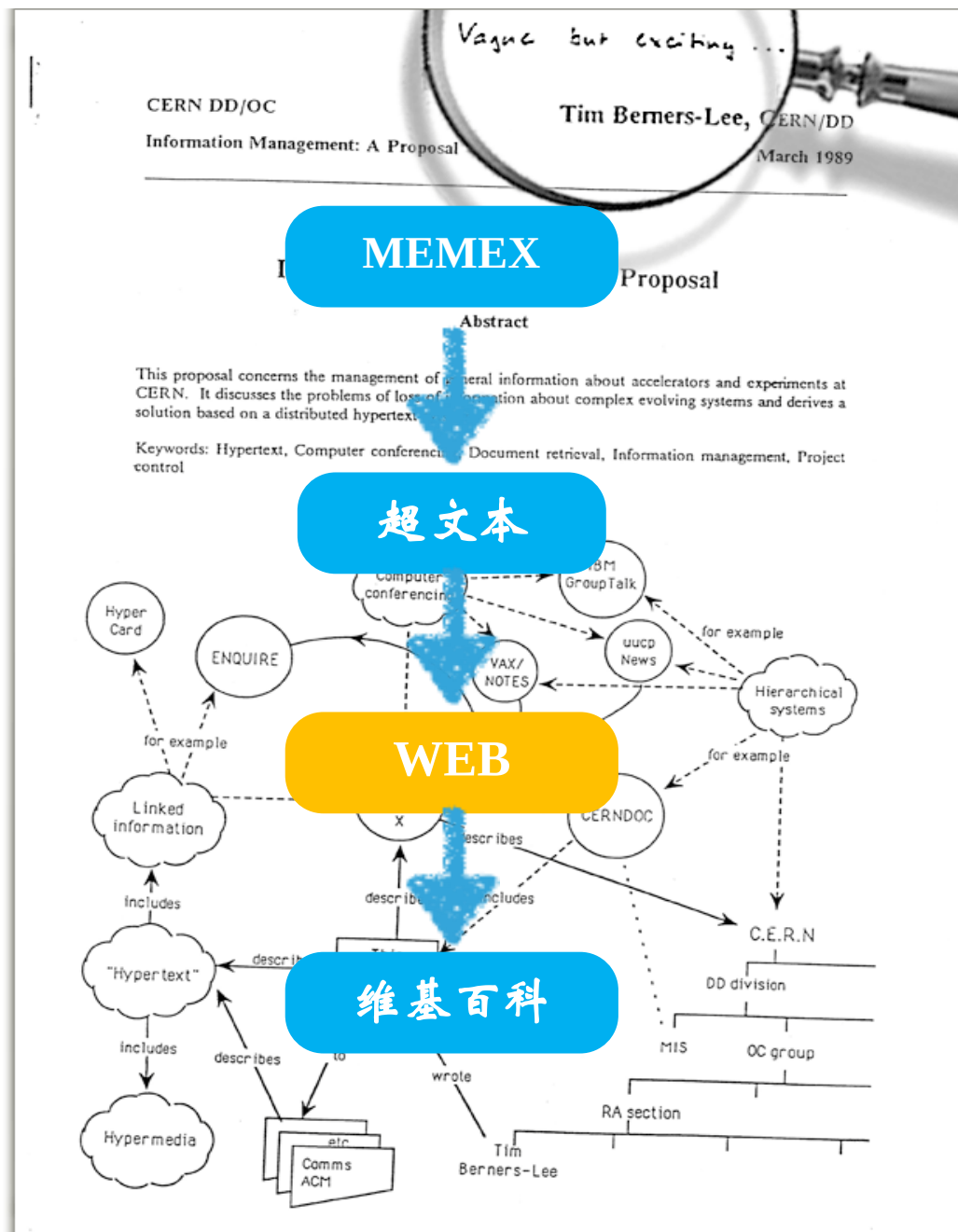


## Vannevar Bush

**National Science Foundation  
Manhattan Project  
First Presidential Science Advisor**

# Web: 以链接为中心的系统

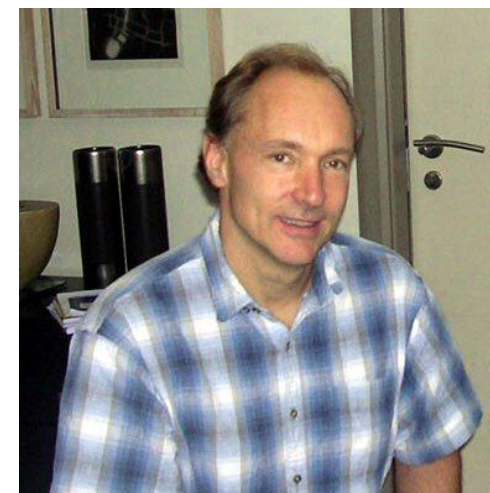
## Linked Information System



...This is why a "web" of notes with links between them is far more useful than a fixed hierarchical system. .... Circles and arrows leave one free to describe the interrelationships between things in a way that tables, for example, do not. The system we need is like a diagram of circles and arrows, where circles and arrows can stand for anything.

Information Management: A proposal 1989.

以“链接”为中心的系统，在开放的互联网环境里面更加容易生长和扩展。这一理念逐步被人们实现，并演化发展成为今天的万维网。



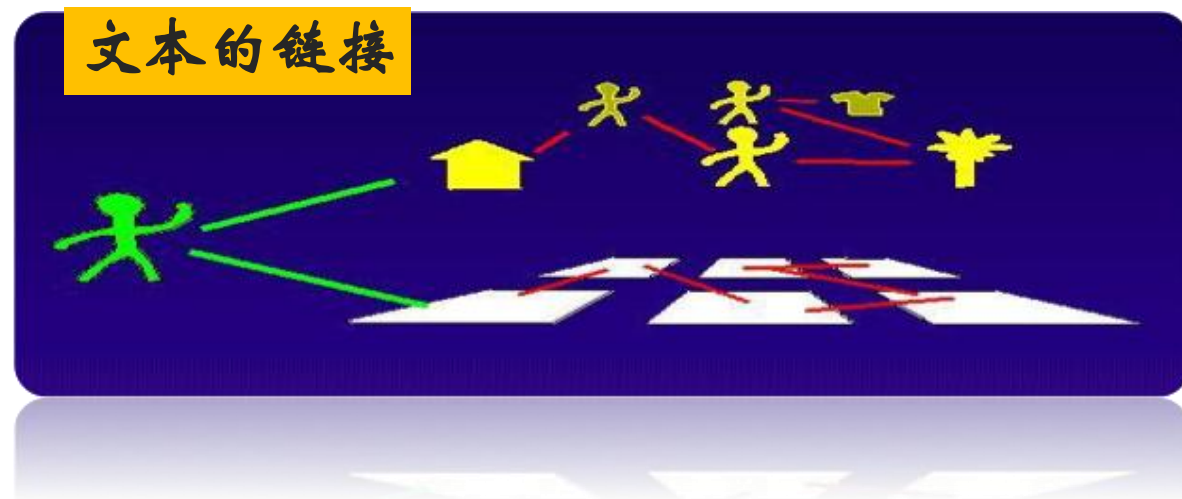
万维网创始人  
MIT教授  
图灵奖获得者

Sir Tim Berners-Lee

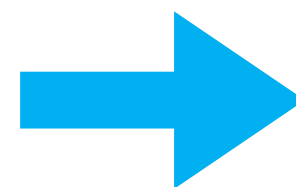
# 语义网：从链接文本到链接数据

"This is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them but we could not process them at all..."

Tim Berners-Lee , Inventor of the Web, @WWW Geneva, 1994



Web of Texts, Web of documents



数据的链接



Web of Objects, Web of Data, Web of Things

# Web中有多种类型的事物



/en/bank\_of\_china\_tower



/en/cn\_tower



/en/chrysler\_building



/en/central\_plaza\_hong\_kong



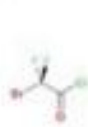
/en/eiffel\_tower



/en/empire\_state\_building



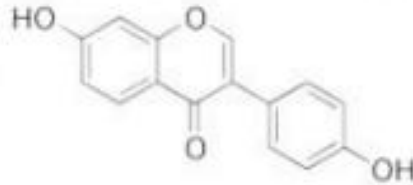
/en/hopewell\_centre\_hong\_kong



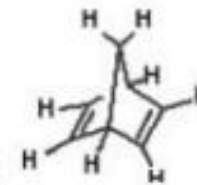
/en/bromodifluoroacetylchloride



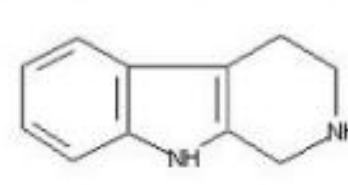
/en/neurine



/en/daidzein



/en/norbornadiene



/en/tryptoline



/en/chicomecoatl



/en/coatlicue



/en/coyolxauhqui



/en/ehecatl



/en/huehuecoyotl



/en/huehuetotl



/en/huitzilopochtli



/en/itzpapalotl



/en/xochipilli



/en/marco\_polo\_chocolate\_bar



/en/kvikk\_lunsj



/en/prince\_polo



/en/starbar



/en/big\_turk



/en/snack\_barz



/en/mona\_lisa



/en/the\_thinker



/en/guernica



/en/the\_scream



/en/the\_starry\_night

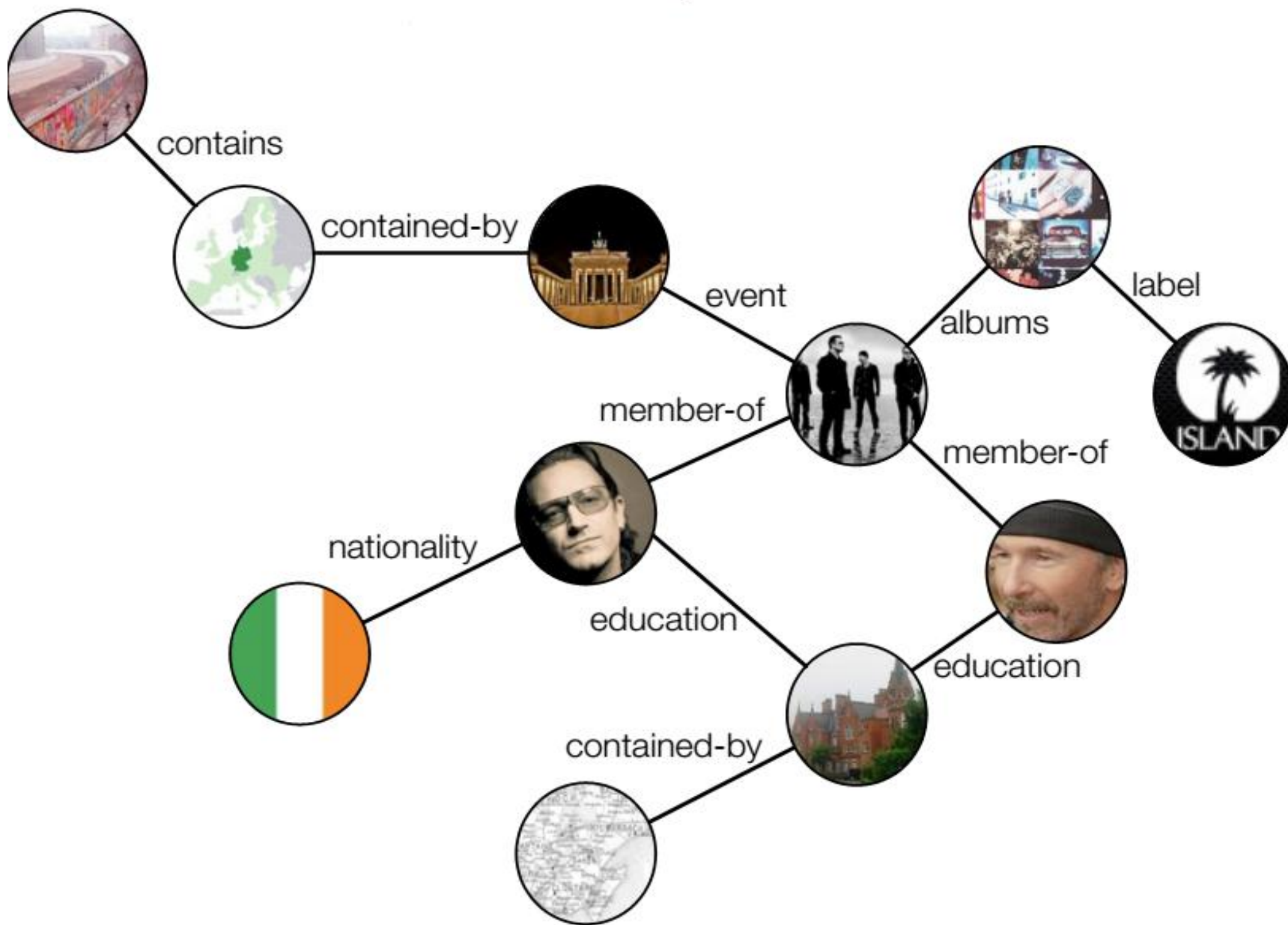


/en/self\_portrait\_with\_a\_friend



/en/the\_son\_of\_man

# 事物之间有多种类型的连接





**link**  
to the world

# 谷歌知识图谱: Things not strings

despicable me 2

Web Images Maps Shopping News More Search tools

About 163,000,000 results (0.29 seconds)

[Despicable Me 2 showtimes for San Francisco, CA](#)  
See showtimes for 3D

 1hr 38min - Rated PG - Animation  
In summer 2013, get ready for more Minion madness in Despicable Me 2. Chris Meledandri and his acclaimed filmmaking team ...  
[AMC Van Ness 14](#) - 1000 Van Ness Avenue, San Francisco, CA - [Map](#)  
11:25am - 2:05 - 4:55 - 7:40 - 10:30pm  
[Century San Francisco Centre 9 and XD](#) - 835 Market St., San Francisco, CA - [Map](#)  
7:00 - 9:25pm  
[+ Show more theaters](#)

[Despicable Me 2](#)  
[despicableme.com/](#)

**A short description of the movie, ratings, release date, directors, cast, etc.**

★★★★★ Rating: 7.8/10 - 51,274 votes  
Directed by Pierre Louis Padang Coffin, Chris Renaud. With Steve Carell, Kristen Wiig, Benjamin Bratt, Miranda Cosgrove. Gru is recruited by the Anti-Villain ...  
[Release Info](#) - [Full cast and crew](#) - [Videos](#) - [Version 3](#)

[Despicable Me 2 - Wikipedia, the free encyclopedia](#)  
[en.wikipedia.org/wiki/Despicable\\_Me\\_2](#)  
Despicable Me 2 is a 2013 American 3D computer-animated comedy film and the sequel to the 2010 animated film Despicable Me. Produced by Illumination ...  
[Minions \(film\)](#) - [Despicable Me \(franchise\)](#) - [Anecy International Animated ...](#)

[Despicable Me 2 - Official Trailer #3 \(HD\) Steve Carell - YouTube](#)  
[www.youtube.com/watch?v=HwXbtZxjbVE](#)  
Mar 19, 2013 - Uploaded by joblomovienetwork  
[http://www.joblo.com](#) - "Despicable Me 2 - Official Trailer #3"  
Universal Pictures and Illumination Entertainment ...

[Despicable Me 2 - Rotten Tomatoes](#)  
[www.rottentomatoes.com/m/despicable\\_me\\_2/](#)  
★★★★★ Rating: 75% - 162 reviews  
Review: It may not be as inspired as its predecessor, but Despicable Me 2 offers plenty of eye-popping visual inventiveness and a number of big...

[News for despicable me 2](#)  
[NBCUniversal CEO: 'Despicable Me 2' Will Be Most Profitable Film in Universal's History](#)

**Despicable Me 2**  
192,648 followers on Google+  
★★★★★ 7.8/10 - IMDb  
★★★★★ 75% - Rotten Tomatoes

Despicable Me 2 is a 2013 American 3D computer-animated comedy film and the sequel to the 2010 animated film Despicable Me.  
[Wikipedia](#)

**Release date:** July 3, 2013 (USA)  
**Directors:** [Pierre Coffin](#), [Chris Renaud](#)  
**Language:** English  
**Production company:** [Illumination Entertainment](#)  
**Music composed by:** [Pharrell Williams](#), [Heitor Pereira](#)

**Recent posts**

 Voting closes soon for the Evil Laugh Contest. Make sure you get your votes in or else... MUAHAHAHA! [http://www.evilloughlab.com/](#)  
Jul 24, 2013

**Cast**

 **Steve Carell**  
Gru

 **Kristen Wiig**  
Lucy Wilde

 **Miranda Cosgrove**  
Margo

 **Russell Brand**  
Dr. Nefario

 **Steve Coogan**  
Silas

**People also search for**

 **Despicable Me**  
2010

 **Monsters University**  
2013

 **The Lone Ranger**  
2013

 **Man of Steel**  
2013

 **The Smurfs 2**  
2013



# 搜索结果中体现关联

**Helsinki**  
Capital of Finland

Helsinki, Finland's southern capital, sits on a peninsula in the Gulf of Finland. Its central avenue, Mannerheimintie, is flanked by institutions including the National Museum, tracing Finnish history from the Stone Age to the present. Also on Mannerheimintie are the imposing Parliament House and Kiasma, a contemporary art museum. Ornate red-brick Uspenski Cathedral overlooks a harbor.

**Clash of Clans**  
Mobile game

4.6/5  
Google Play

4.1/5  
iTunes - Apple

Clash of Clans is a freemium mobile MMO strategy video game developed and published by Supercell. The game was released for iOS platforms on August 2, 2012, and on Google Play for Android on October 7, 2013. [Wikipedia](#)

Initial release date: August 2, 2012  
Developer: Supercell  
Publisher: Supercell  
Platforms: Android, iOS

**Supercell**  
Game developer

Supercell is a mobile game development company founded in June 2010 in Helsinki, Finland. The company's debut game was the browser game Gunshine.net, and after its release in 2011, Supercell started developing games for mobile devices. [Wikipedia](#)

Revenue: 2.11 billion EUR (2015)  
CEO: [Ilkka Paananen](#) (2010–)  
Founded: June 2010, Helsinki, Finland  
Owner: [Tencent](#) (84.3%)  
Founders: [Mikko Kodisoja](#), [Lassi Leppinen](#), [Niko Derome](#), [Visa Forsten](#), [Petri Styrman](#), [Ilkka Paananen](#)

Profiles  
[Facebook](#) [Twitter](#)

Video games  
[Clash of Clans](#) 2012  
[Clash Royale](#) 2016  
[Boom Beach](#) 2013  
[Hay Day](#) 2012

**Ilkka Paananen**  
Businessman

Born: 1978  
People also search for: [Mikko Kodisoja](#), [Niko Derome](#), [Lassi Leppinen](#), [Visa Forsten](#), [Petri Styrman](#)  
Organizations founded: [Supercell](#), [Sumea](#)

**Tencent**  
Internet company

Tencent Holdings Limited is a Chinese investment holding company whose subsidiaries provide media, entertainment, internet and mobile phone value-added services and operate online advertising services in China. [Wikipedia](#)

Stock price: 0700 (HKG) HK\$202.60 -1.80 (-0.88%)  
Nov 3, 2:30 PM GMT+8 - Disclaimer  
Headquarters: [Shenzhen, China](#)  
Revenue: 102.9 billion CNY (2015)  
Founded: November 11, 1998  
CFO: [John Lo](#)  
Subsidiaries: [Riot Games](#), [WeBank](#), [Epic Games](#), [Tencent Pictures](#)  
Founders: [Ma Huateng](#), [Zhang Zhidong](#)

创建地址  
CEO  
产品  
母公司

# Facebook 兴趣图谱

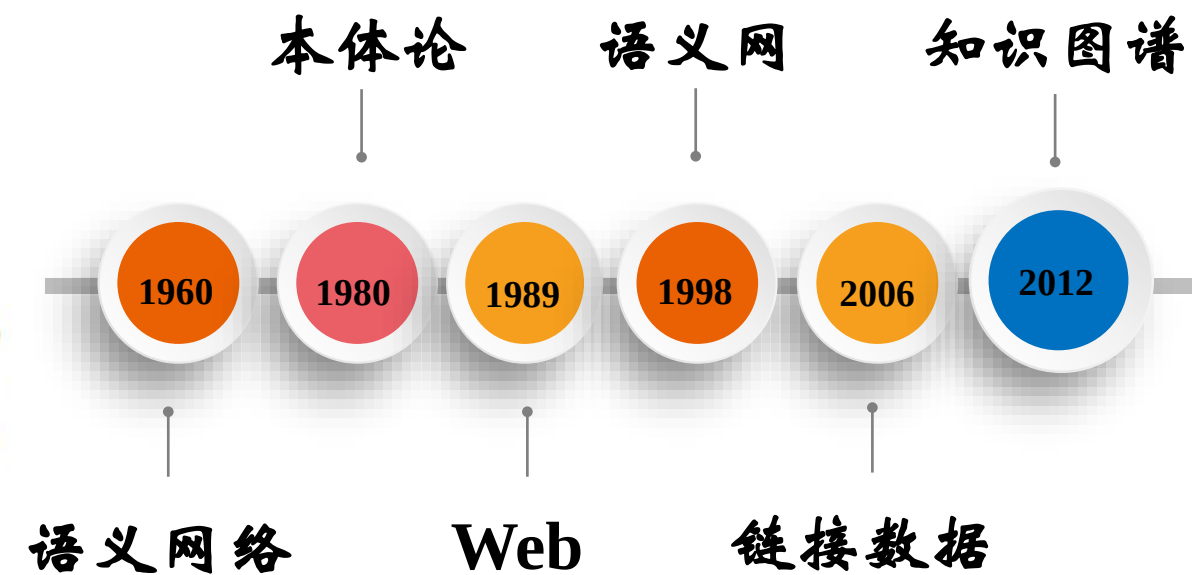
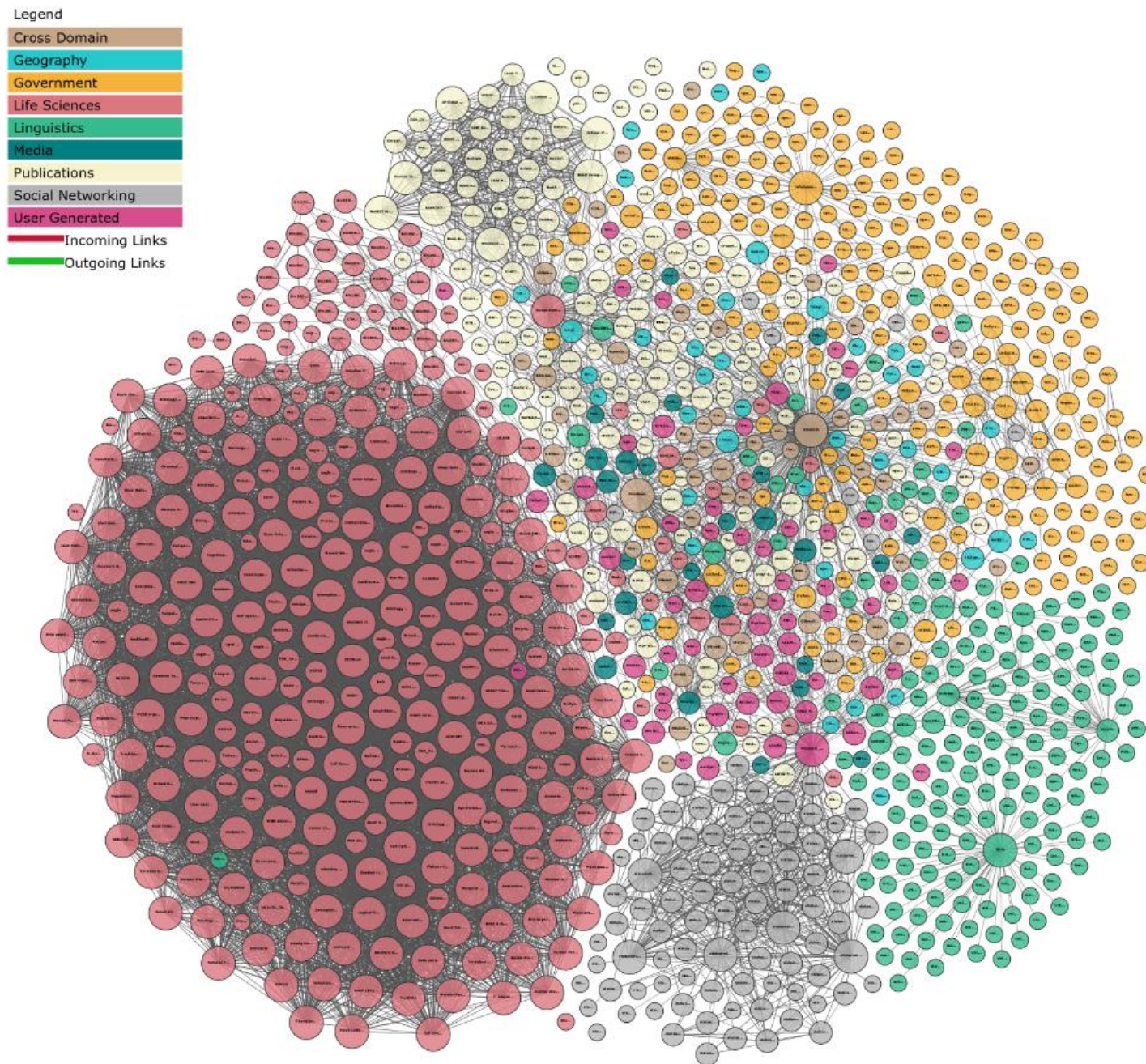


# Facebook兴趣图谱复杂示例

Structured search within the graph

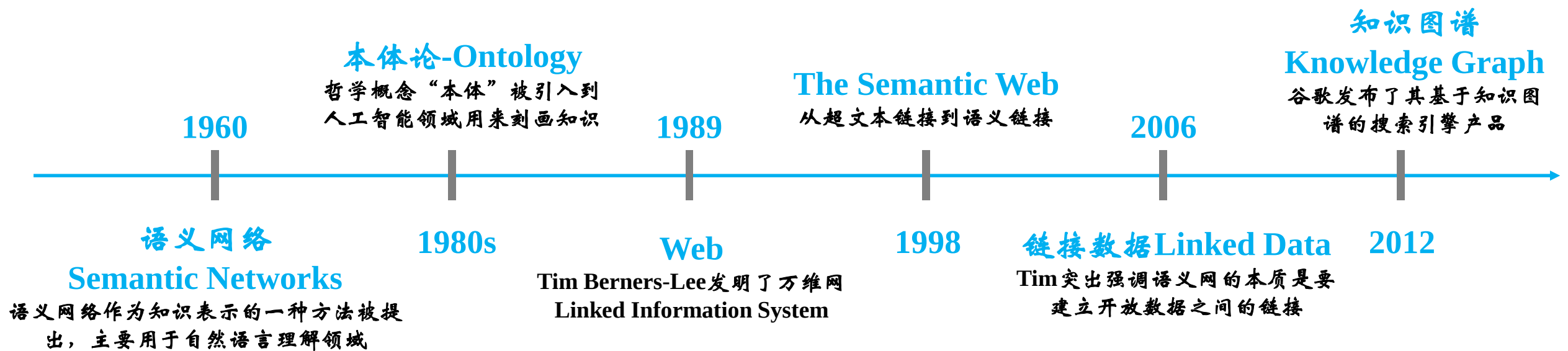


# 知识图谱 (Knowledge Graph, KG)



"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>"

# KG的概念演化



## 知识表示与知识库- Knowledge Representation / Knowledge Base

人工智能研究者陆续提出了大量知识表示的方法，如框架系统、产生式规则、描述逻辑等。

知识图谱得益于Web的发展（更多的是数据层面），有着来源于KR、NLP、Web、AI多个方面的基因。

# KG辅助搜索

Web的理想是万物的链接，搜索的理想是事物的搜索

网页搜索

语义搜索

Web of Docs

Web of Data



手工众包

格式转化

元组抽取

实体融合

链接预测

推理补全

语义嵌入

Freebase

社区协同构建

WIKIDATA

维基众包

schema.org

....the new SEO?

网页嵌入语义数据

# KG辅助问答

机器人及IoT设备的智能化：给万物都挂接一个背景知识库

对话式的信息获取更加需要精准度和可靠度，知识图谱对于提升用户体验更加不可少。



DBpedia Yago  
Wordnet



True Knowledge/Evi



WolframAlpha  
DBpedia



zhishi.me



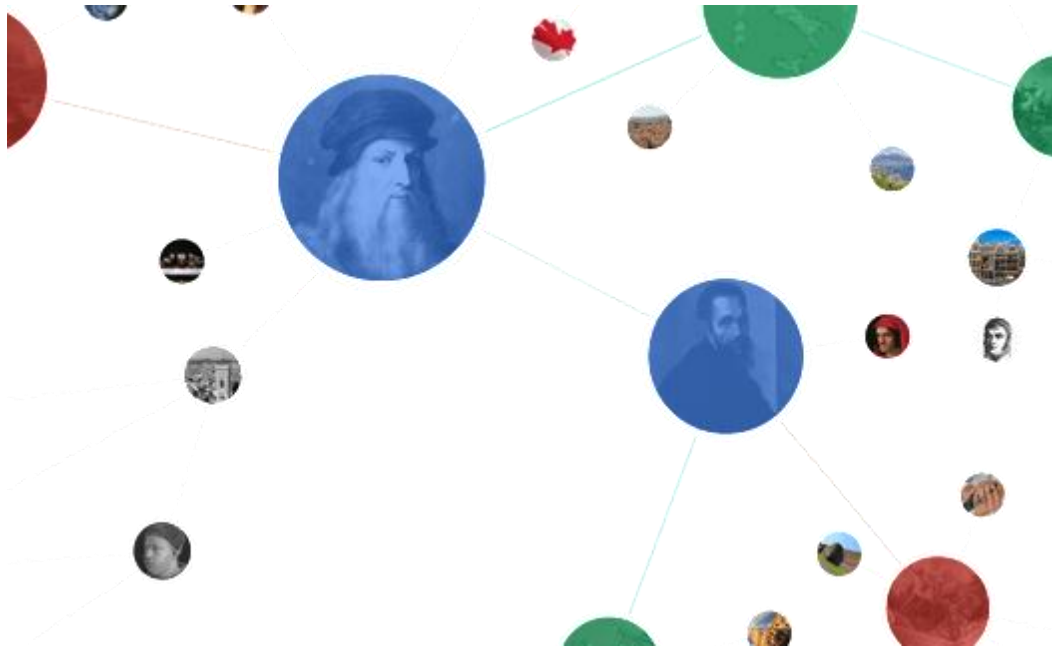
DUER OS

Conversational AI OS for everyone and everywhere



# KG辅助决策

**MORE MACHINE UNDERSTANDABLE**



建立数据链接

更加规范的数据表达

可计算的数据

预先抽取语义

粗糙数据



文本

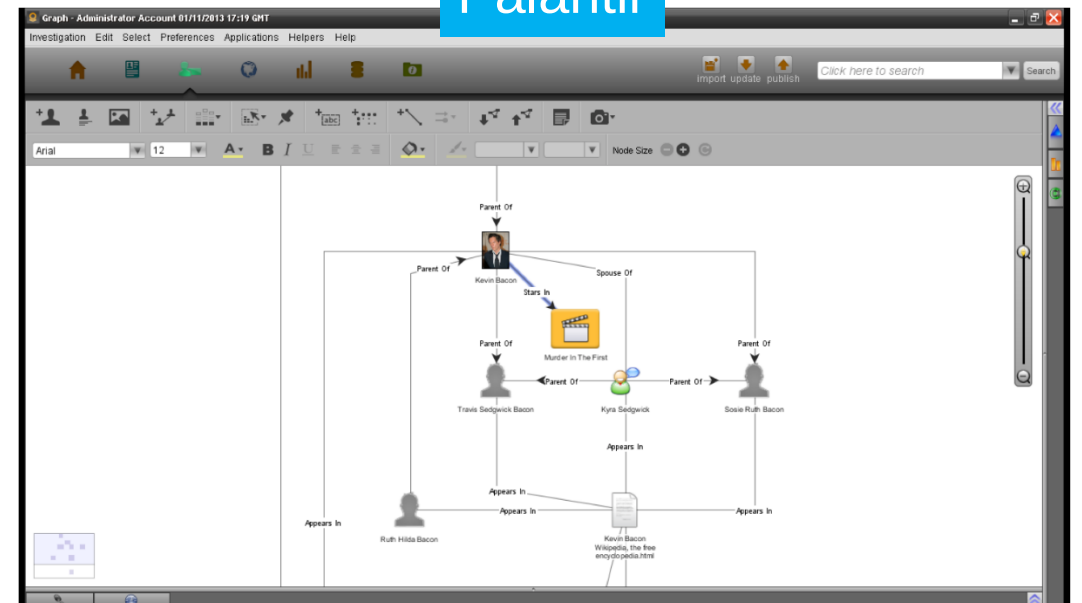


多媒体



传感器

Palantir



Kensho



# KG辅助AI: 常识推理

当一个人听到或看到一句话的时候，他使用自己所有的知识和智能去理解。这不仅包括语法，也包括他的词汇知识、上下文知识，更重要的，是对相关事务的理解

——Terry Winograd (自然语言理解系统SHRDLU作者)

## Winograd Schema Challenge

I. The trophy would not fit in the brown suitcase because **it** was too big (small). What was too big (small)?

Answer 0: the trophy

Answer 1: the suitcase

II. The town councilors refused to give the demonstrators a permit because **they** feared (advocated) violence. Who feared (advocated) violence?

Answer 0: the town councilors

Answer 1: the demonstrators

纯NLP: 50% → NLP+KB: 超过60% → 及格线: 90%

# KG的本质

---

**Web视角** 像建立文本之间的超链接一样，建立数据之间的语义链接，并支持语义搜索

**NLP视角** 怎样从文本中抽取语义和结构化数据

**KR视角** 怎样利用计算机符号来表示和处理知识

**AI视角** 怎样利用知识库来辅助理解人的语言

**DB视角** 用图的方式去存储知识

做好KG要兼容并蓄，综合利用好KR、NLP、Web、  
ML、DB等多方面的方法和技术

# 聪明的AI vs 有学识的AI

人的大脑依赖所学的知识进行思考、逻辑推理、理解语言

聪明的AI

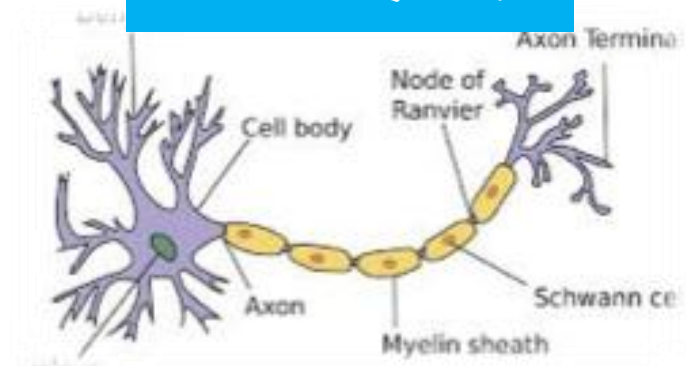
学习

感知

识别

判断

深度学习



推理

有学识的AI

思考

语言

推理

知识图谱



---

## 第二部分：典型知识库项目简介

# 各种知识图谱项目

 Freebase

 LINKING OPEN DATA  
W3C SWEO Community Project

**WordNet**  
A lexical database for English

 ZHISHI.me

PKUBASE

NELL

schema.org  
...the new SEO?

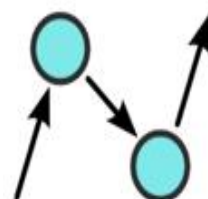
 DBpedia

 XLORE

 WIKIDATA

WEB CHILD

CN-DBpedia



**ConceptNet**  
An open, multilingual knowledge graph

**Herbnet**

 yago  
select knowledge

 LinkedGeoData.org

WEBKB

linked life data 

- ❑ Cyc是在1984年由Douglas Lenat开始创建。最初的目标是要建立人类最大的常识知识库。典型的常识知识如“Every tree is a plant”，“Plants die eventually”等
- ❑ Cyc知识库主要由术语Terms和断言Assertions组成。Terms包含概念、关系和实体的定义。Assertions用来建立Terms之间的关系，这既包括事实Fact描述，也包含规则Rule的描述
- ❑ 最新的Cyc知识库包含有50万条Terms和700万条Assertions。Cyc的主要特点是基于形式化的知识表示方法来刻画知识。形式化的优势是可以支持复杂的推理。但过于形式化也导致知识库的扩展性和应用的灵活性不够。Cyc提供开放版本OpenCyc

<http://www.cyc.com/>

# Wordnet

**WordNet**  
A lexical database for English

- WordNet是最著名的词典知识库，主要用于词义消歧。WordNet由普林斯顿大学认识科学实验室从1985年开始开发
- WordNet主要定义了名词、动词、形容词和副词之间的语义关系。例如名词之间的上下位关系（如：“猫科动物”是“猫”的上位词），动词之间的蕴含关系（如：“打鼾”蕴含着“睡眠”）等
- WordNet3.0已经包含超过15万个词和20万个语义关系

**WordNet: a lexical database for English**

GA Miller - Communications of the ACM, 1995 - [dl.acm.org](http://dl.acm.org)

Abstract Because meaningful sentences are composed of meaningful words, any system that hopes to process natural languages as people do must have information about words and their meanings. This information is traditionally provided through dictionaries, and

被引用次数: 9323 相关文章 所有 34 个版本 引用 保存 更多

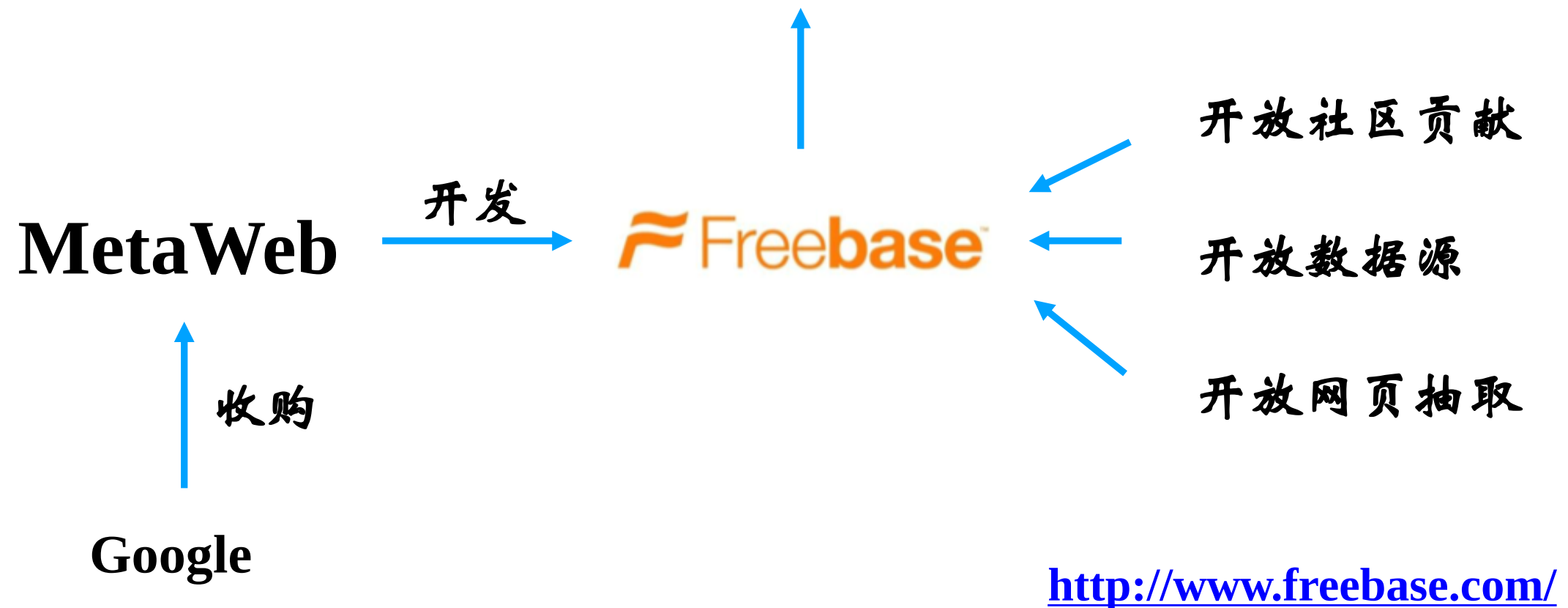
<http://wordnet.princeton.com/>

- ConceptNet是常识知识库。最早源于MIT媒体实验室的Open Mind Common Sense (OMCS)项目。OMCS项目是由著名人工智能专家Marvin Minsky于1999年建议创立
- ConceptNet主要依靠互联网众包、专家创建和游戏三种方法来构建。新版本导入大量开放的结构化数据，如DBPedia、Wikinary, Wordnet等
- ConceptNet知识库以三元组形式的关系型知识构成。ConceptNet5版本已经包含有2800万关系描述。与Cyc相比，ConceptNet采用了非形式化、更加接近自然语言的描述，而不是像Cyc那样采用形式化的谓词逻辑
- 与链接数据和谷歌知识图谱相比，ConceptNet比较侧重于词与词之间的关系。从这个角度看，ConceptNet更加接近于WordNet，但是又比WordNet包含的关系类型多。此外，ConceptNet完全免费开放，并支持多种语言

<http://www.conceptnet.io/>

# Freebase

完全免费并允许商业化的开放许可协议



通过开源免费吸引用户贡献数据，增值的应用及技术服务收费

# Wikidata

目标是构建全世界最大的免费知识库，采用CC0完全自由许可协议



维基基金会

导入维基数据



开放社区贡献



相对小众、专业、严格的众包社区

提供资金

导入Freebase数据

谷歌

提供资金

摩尔基金会

提供资金

Allen AI Institute

仍然面临知识缺失严重的问题

<http://www.wikidata.org/>

- DBPedia是早期的语义网项目。DBPedia意指数据库版本的Wikipedia，是从Wikipedia抽取出来的链接数据集。DBPedia采用了一个较为严格的本体，包含人、地点、音乐、电影、组织机构、物种、疾病等类定义
- 此外，DBPedia还与Freebase，OpenCYC、Bio2RDF等多个数据集建立了数据链接。DBPedia采用RDF语义数据模型，总共包含30亿RDF三元组

<http://dbpedia.org>

- YAGO是由德国马普研究所研制的链接数据库。YAGO主要集成了Wikipedia、WordNet和GeoNames三个来源的数据。YAGO将WordNet的词汇定义与Wikipedia的分类体系进行了融合集成，使得YAGO具有更加丰富的实体分类体系
- YAGO还考虑了时间和空间知识，为很多知识条目增加了时间和空间维度的属性描述。目前，YAGO包含1.2亿条三元组知识。YAGO是IBM Watson的后端知识库之一

<http://mpii.de/yago>

# Babelnet



- BabelNet是类似于WordNet的多语言词典知识库。BabelNet的目标是解决WordNet在非英语语种中数据缺乏的问题。BabelNet采用的方法是将WordNet词典与Wikipedia百科集成
- 首先建立WordNet中的词与Wikipedia的页面标题的映射，然后利用Wikipedia中的多语言链接，再辅以机器翻译技术，来给WordNet增加多种语言的词汇
- BabelNet3.7包含了271种语言，1400万同义词组，36.4万词语关系和3.8亿从Wikipedia中抽取的链接关系，总计超过19亿RDF三元组。BabelNet集成了WordNet在词语关系上的优势和Wikipedia在多语言语料方面的优势，构建成功了目前最大规模的多语言词典知识库。

<http://babelnet.org/>

浙江大学

CHINESE 翻译为...

搜索

设置

Chinese Arabic English French German Greek Hebrew Hindi + all preferred languages

bn:02763443n • NOUN • Named Entity • Categories: 浙江省重点建设大学, 211工程, 长三角高校合作联盟, 九校联盟...

浙江大学 • 浙大 • 国立浙江大学 • ZJU • 国立第三中山大学

浙江大学（英语：Zhejiang University，缩写为ZJU），简称浙大，位于中国浙江省杭州市，现有玉泉、西溪、华家池、之江、紫金港、舟山六个校区。 Wikipedia

IS A 大学 • 国立大学 • 公立大学  
HAS PART 浙江大学出版社  
CITY 杭州市  
COUNTRY 中华人民共和国  
LOCATED IN THE ADMINISTRATIVE TERRITORIAL ... 杭州市  
MEMBER OF 九校联盟

印度尼西亚 8K  
美国 8K  
北京 8K  
澳大利亚 8K  
香港 8K  
智利 8K  
上海市 8K  
台湾 8K  
新加坡 8K  
马来西亚 8K  
7.8K

道

（英语：Tsung-Dao Lee，1926年11月24日－）是一位美國華人物理學要知名於宇稱不守恒、李模型（Lee Model）、相對論重離子（RHIC）量子場論的非拓撲性孤立子和孤立子星以及破解粒子物理中的 $\theta$ - $\tau$ 之謎。

1050503n | Semantically related form

- NELL(Never-Ending Language Learner) 是卡内基梅隆大学开发的知识库。NELL主要采用互联网挖掘的方法从Web自动抽取三元组知识
- NELL的基本理念是：给定一个初始的本体（少量类和关系的定义）和少量样本，让机器能够通过自学习的方式不断的从Web学习和抽取新的知识。目前NELL已经抽取了400多万条高置信度三元组知识

<http://rtw.ml.cmu.edu/>

### Recently-Learned Facts [twitter](#)

Refresh

| Instance                                                                                         | Iteration | date learned | confidence |                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------|-----------|--------------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">n1 3 problems</a> is a <a href="#">song</a>                                          | 1075      | 04-sep-2017  | 98.5       |   |
| <a href="#">research programmes</a> is a <a href="#">research project</a>                        | 1075      | 04-sep-2017  | 100.0      |   |
| <a href="#">monetary damage</a> is an <a href="#">event outcome</a>                              | 1075      | 04-sep-2017  | 92.6       |   |
| <a href="#">metal coffee table</a> is a <a href="#">household item</a>                           | 1075      | 04-sep-2017  | 100.0      |   |
| <a href="#">n7 8 minutes</a> is a <a href="#">term used by physicists</a>                        | 1075      | 04-sep-2017  | 100.0      |   |
| <a href="#">group</a> is part of <a href="#">wheel</a>                                           | 1075      | 04-sep-2017  | 93.8       |   |
| <a href="#">ben gordon</a> is an athlete <a href="#">coached by</a> <a href="#">scott stiles</a> | 1079      | 28-sep-2017  | 100.0      |   |
| <a href="#">beds</a> is often found in <a href="#">story</a>                                     | 1078      | 23-sep-2017  | 93.8       |   |
| <a href="#">wwti</a> is a <a href="#">TV affiliate of</a> the network <a href="#">abc</a>        | 1079      | 28-sep-2017  | 93.8       |   |
| <a href="#">david selden</a> got <a href="#">married in</a> <a href="#">n1956</a>                | 1079      | 28-sep-2017  | 100.0      |   |

# 微软 Concept Graph

- Concept Graph是以概念层次体系为中心的知识图谱。与Freebase等知识图谱不同，Concept Graph以概念定义和概念之间的IsA关系为主
- 给定一个概念如“Microsoft”，Concept Graph返回一组与“微软”有IsA关系概念组，如：“Company”，“Software Company”，“Largest OS Vender”等。这被称为概念化“Conceptualization”
- Concept Graph可以用于短文本理解和语义消歧中。例如，给定一个短文本“the engineer is eating the apple”，可以利用Concept Graph来正确理解其中“apple”的含义是“吃的苹果”还是“苹果公司”。微软发布的第一个版本包含超过540万的概念，1255万的实体，和8760万的关系。Concept Graph主要通过从互联网和网络日志中挖掘来构建



<http://concept.research.microsoft.com>

# OpenKG: 中文知识图谱资源库



# Zhishi.me



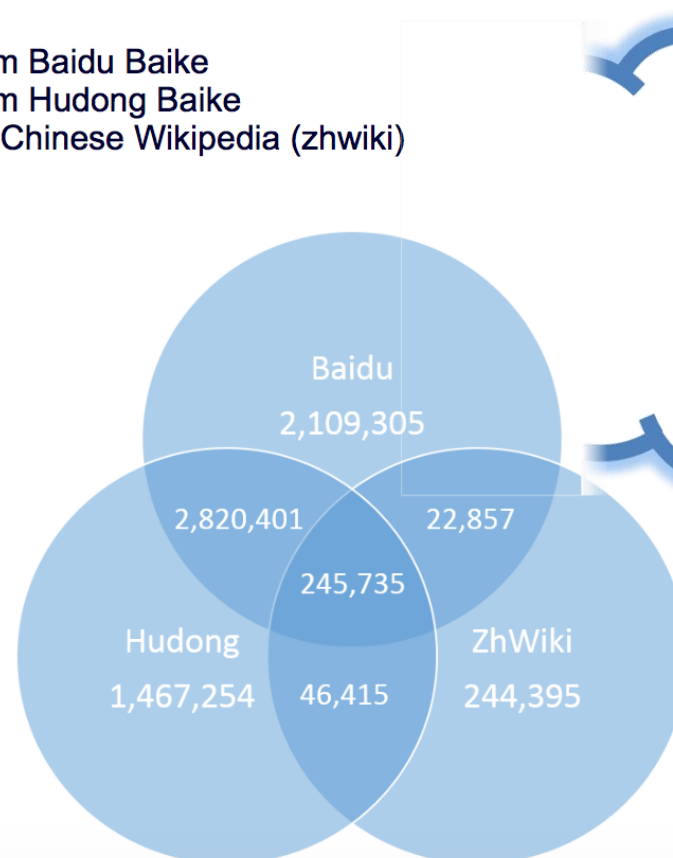
<http://zhishi.me>

About  
Lookup  
SPARQL  
Publications  
Credits  
API

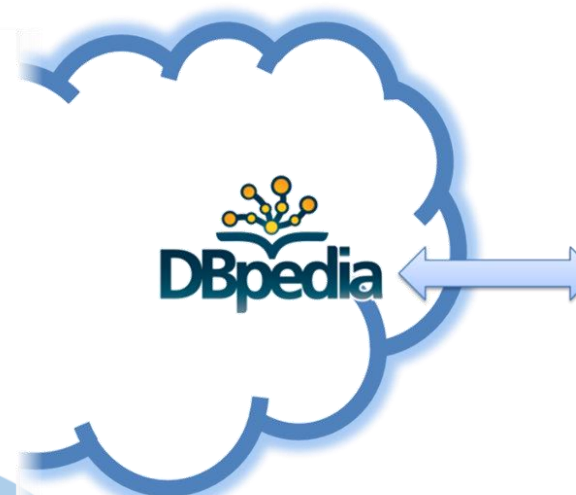
Zhishi.me is an effort to build Chinese Linking Open Data. Currently, it covers three largest Chinese encyclopedias: Baidu Baike, Hudong Baike and Chinese Wikipedia.

Dataset overview:

- 5,198,298 entities from Baidu Baike
- 4,579,805 entities from Hudong Baike
- 559,402 entities from Chinese Wikipedia (zhwiki)



Linking Open Data



© 2006 The Authors  
Journal compilation © 2006 Blackwell Publishing Ltd

# C-Wikipedia



# Zhishi.me



## Label

■ **hudong:北京**

**foaf:page** <http://www.hudong.com/wiki/%E5%8C%97%E4%B8%A4%E5%8C%97%E4%BA%AC>  
**zhishi:resourceID** 757341

Index  
• **zhishi:abstract**  
• **infobox**  
• **dcterms:subject**  
• **zhishi:thumbnail**  
• **zhishi:relatedPage**  
• **zhishi:externalLink**  
• **zhishi:internalLink**

**owl:sameAs**

- ☒ **hudong:北京 (this)**
- ☒ **baidu:北京**
- ☒ **zhwiki:北京**

[MERGE PAGE](#)

## Abstract

**zhishi:abstract** ■

北京是中华人民共和国的首都，四个直辖市之一，是中国的政治、文化、科技、教育和国际交往中心，也是中国最大的航空交通枢纽和中国北方的经济中心。北京市位于华北平原北部，东南隅地区与天津市相连，其余为河北省所环绕，以市区人口数相比，北京为次于上海的中国第二大城市。北京是一座有三千年历史的文化名城，历史上共有五个朝代曾在此建都，北京荟萃了自元明清以来的中华文化，拥有诸多名胜古迹和人文景观。北京于2008年成功举办第29届夏季奥林匹克运动会（2008年8月8日至8月24日）及残疾人奥林匹克运动会（2008年9月6日至9月17日）。

[>>TOP](#)

## Infobox

**infobox**

**所属地区** ■ **中国**  
**GDP** ■ 11,865.9亿元人民币(2009年)  
**主要街道** ■ 长安街、平安大街  
**面积** ■ 16800平方公里  
**中文名** ■ 北京  
**下辖地区** ■ 海淀区 西城区 顺义区 大兴区  
**友好城市** ■ 纽约、东京、巴黎  
**知名企业** ■ 首钢集团、北京科技园置地有限公司、北京京宝公司、北京珠联公司  
**特产** ■ 烤鸭、豌豆黄、艾窝窝、豆汁  
**人口** ■ 1,755万人(2009年)  
**电话区码** ■ 010  
**英文名** ■ Municipality of Beijing

[>>TOP](#)

## Subjects

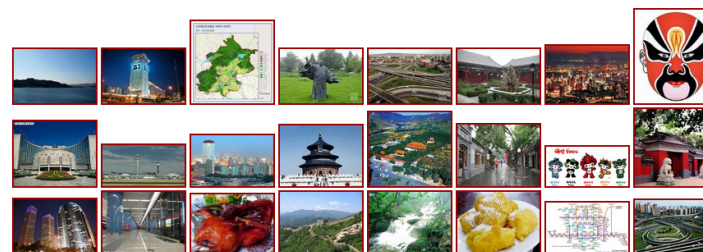
**dcterms:subject**

■ 水泮地名 ■ 文化 ■ 首都 ■ 各国首都 ■ 植物 ■ 历史 ■ 互联网 ■ 地理 ■ 中国 ■ 中国城市

[>>TOP](#)

## Thumbnails

**zhishi:thumbnail**



[>>TOP](#)

## Other properties

**zhishi:relatedPage**

■ **圆明园** [SHOW MORE \(10\)](#)

[>>TOP](#)

**zhishi:externalLink**

■ <http://www.cq6.cc/index.php%3Fdoc-innerlink-%25E5%258C%2597%25E4%25BA%25AC>  
[SHOW MORE \(14\)](#)

[>>TOP](#)

**zhishi:internalLink**

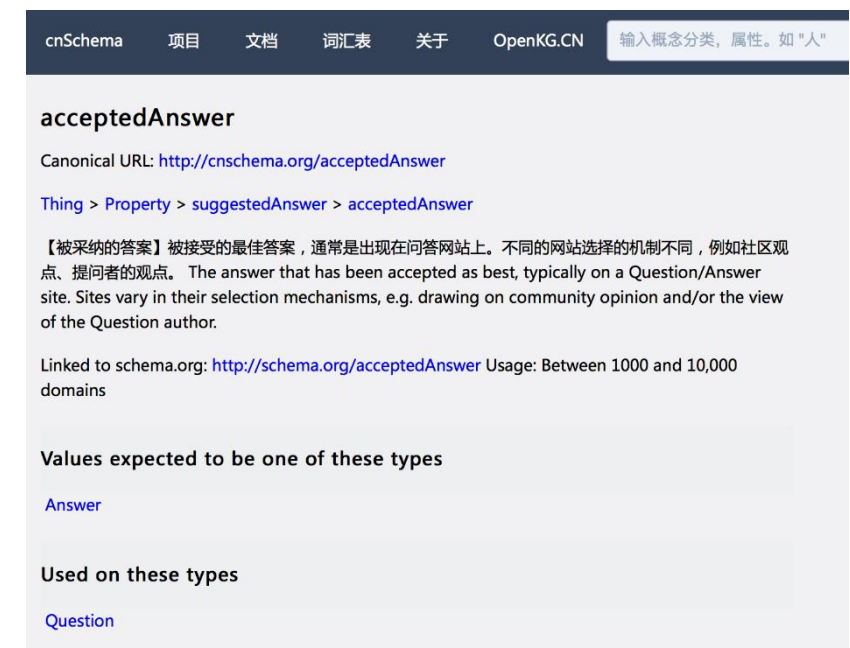
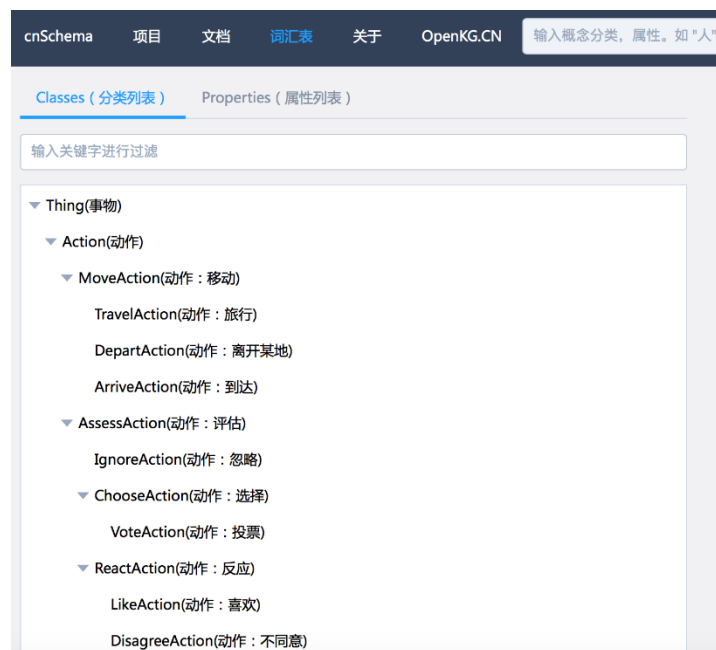
■ **福建省** [SHOW MORE \(798\)](#)

[>>TOP](#)

# cnSchema: 开放的中文知识图谱Schema

- cnSchema.org是一个基于社区维护的开放的知识图谱Schema标准。cnSchema分类(classes)、数据类型(data type)的词汇集包括了上千种概念(terms)、属性(properties)和关系(relations)等常用概念定义, 以支持知识图谱数据的通用性、复用性和流动性
- 结合中文的特点, 我们复用、连接并扩展了Schema.org, Wikidata, Wikipedia等已有的知识图谱Schema标准, 为中文领域的开放知识图谱、聊天机器人、搜索引擎优化等提供可供参考和扩展的数据描述和接口定义标准

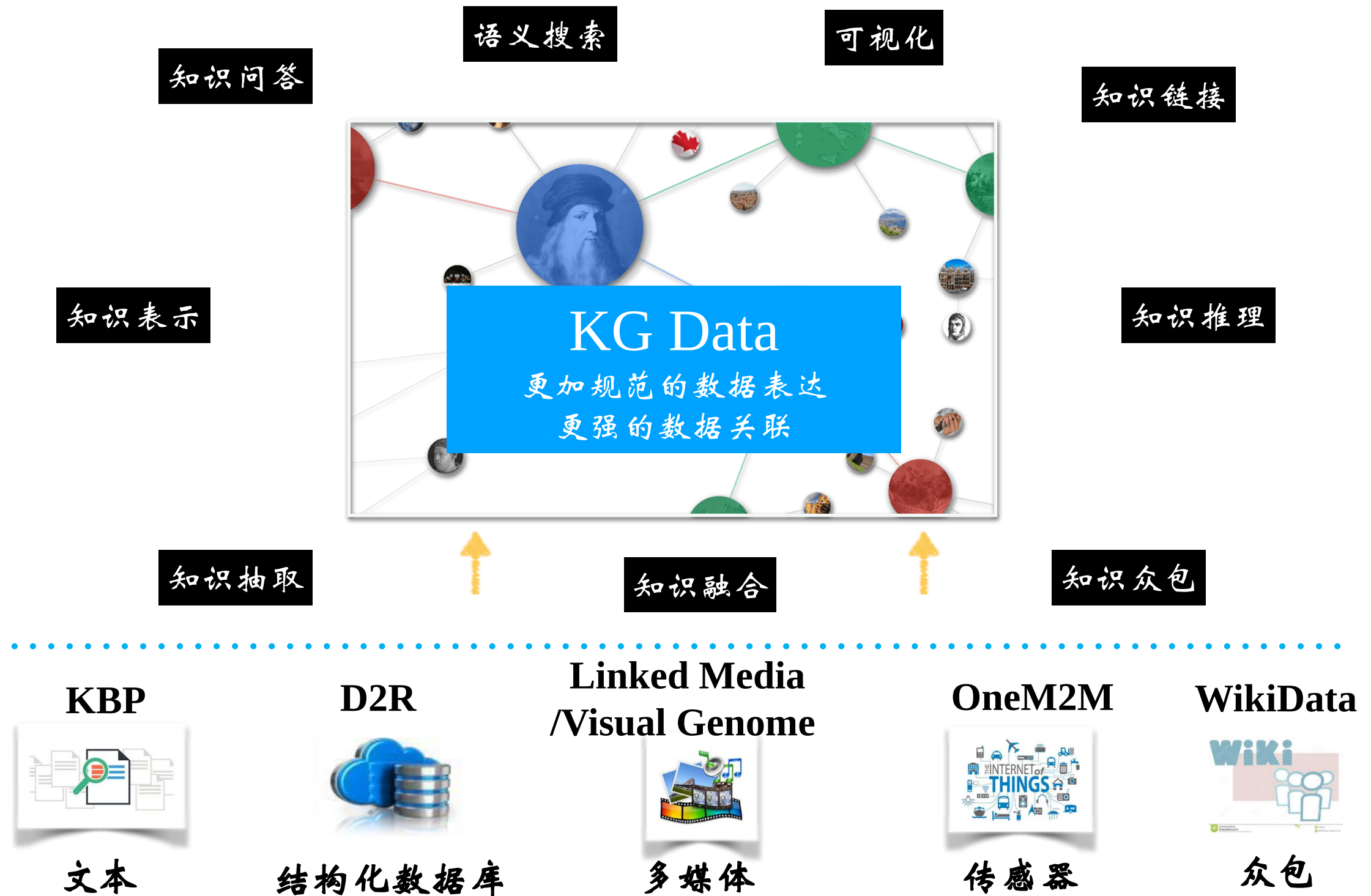
<http://cnschema.org>



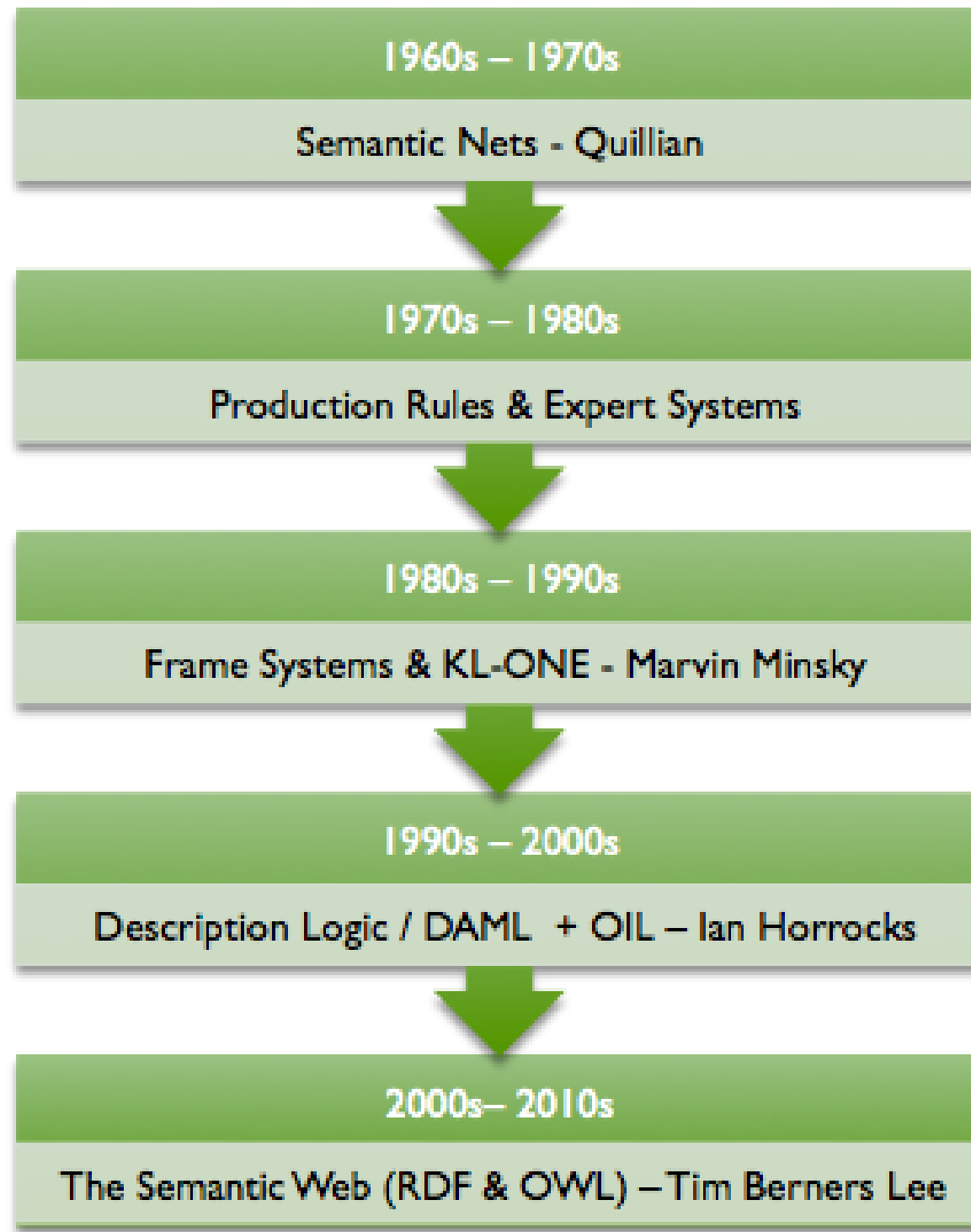
---

## 第三部分：知识图谱技术概览

# 知识图谱的技术体系



# 什么是知识表示



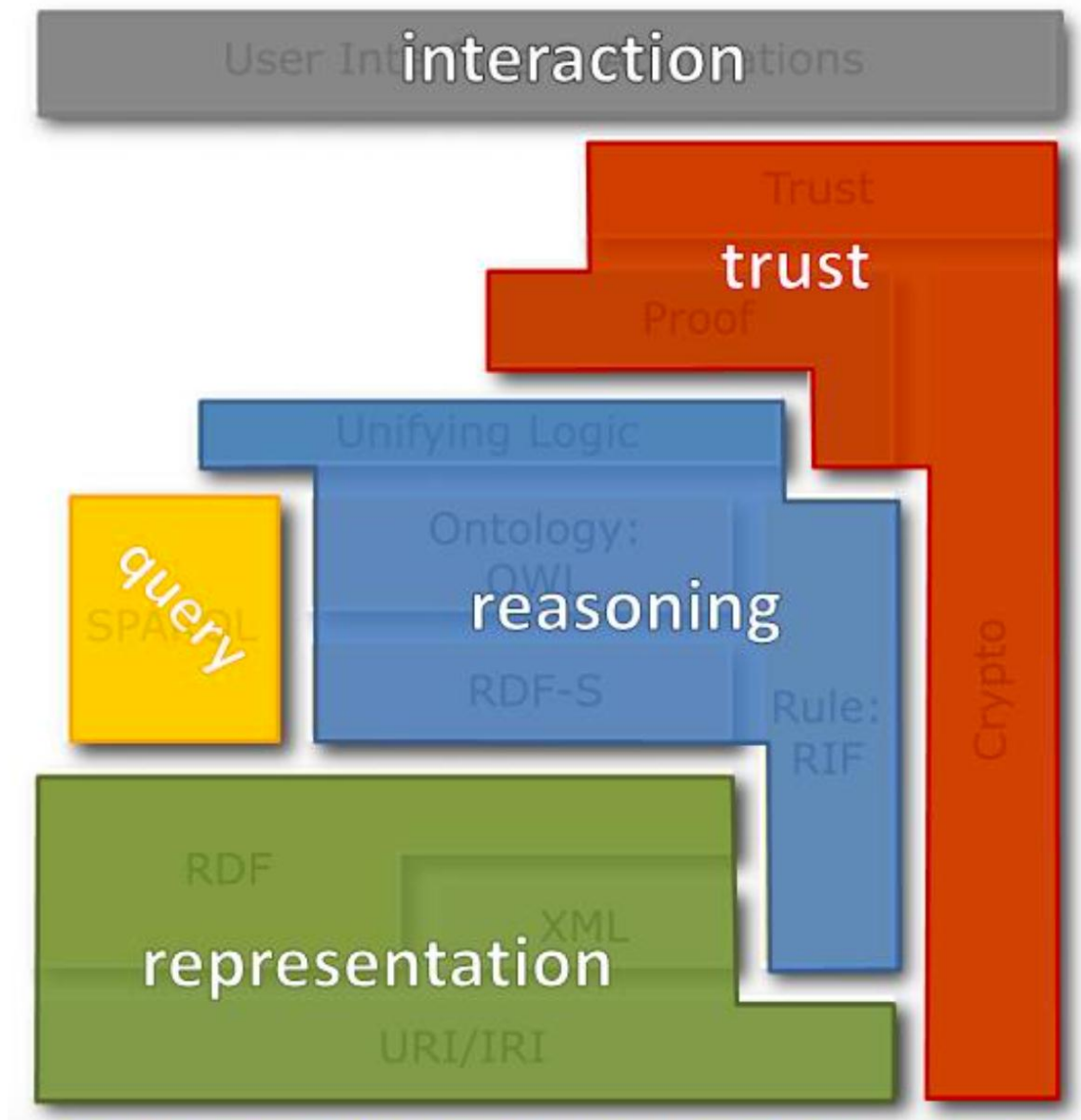
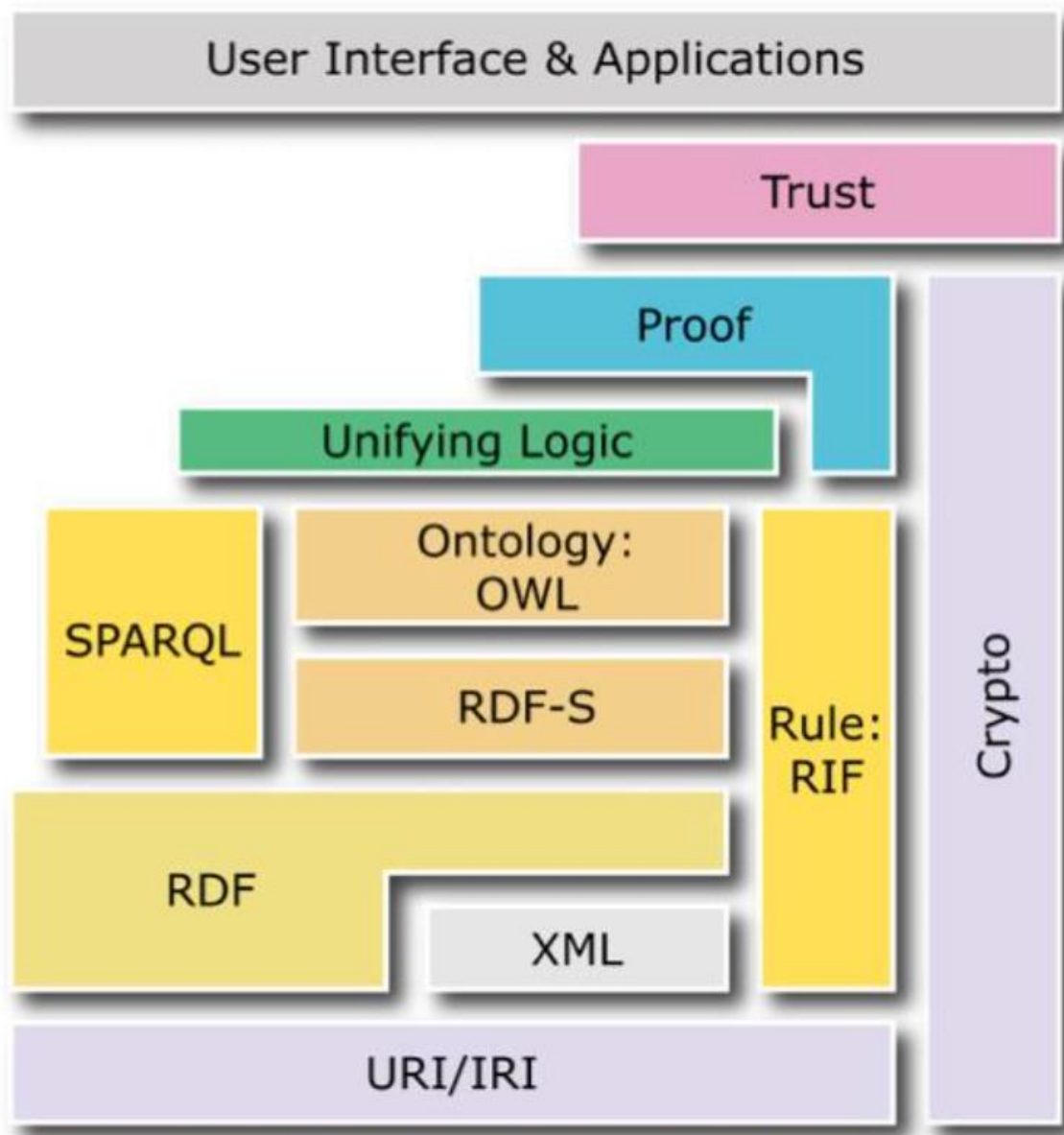
知识表示研究怎样用计算机符号来表示人脑中的知识，以及怎样通过符号之间的运算来模拟人脑的推理过程。

基于数理逻辑的知识表示



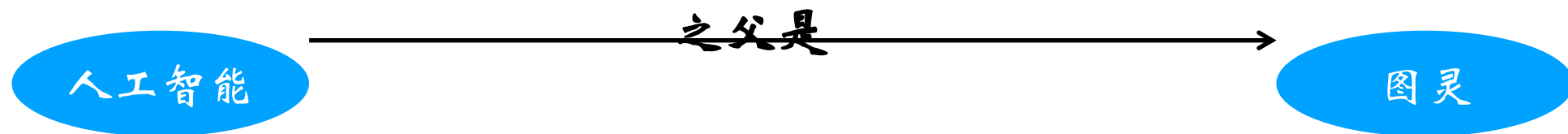
基于向量空间学习的分布式知识表示

# 语义网知识表示框架



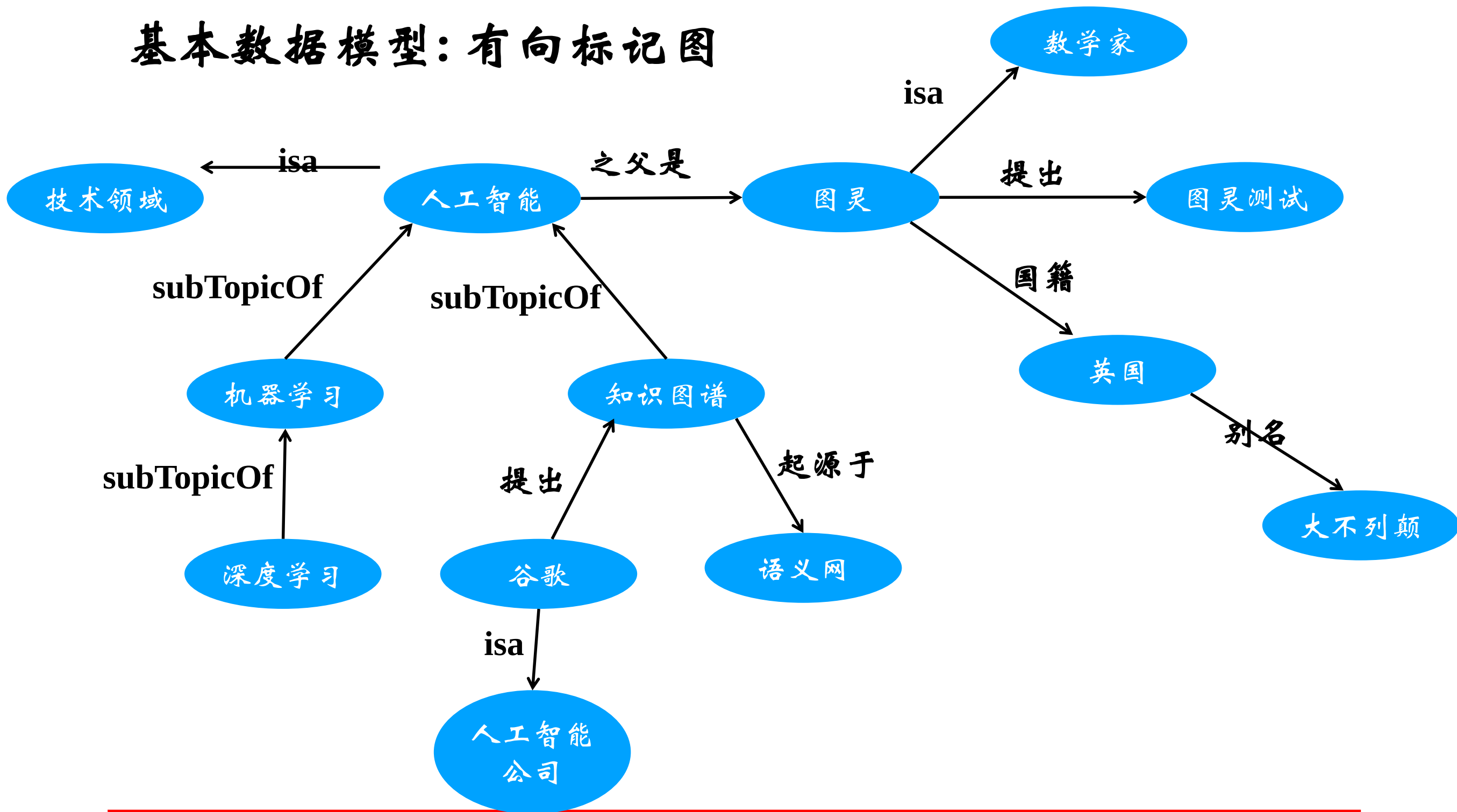
# RDF: Triple-based Assertion model

---



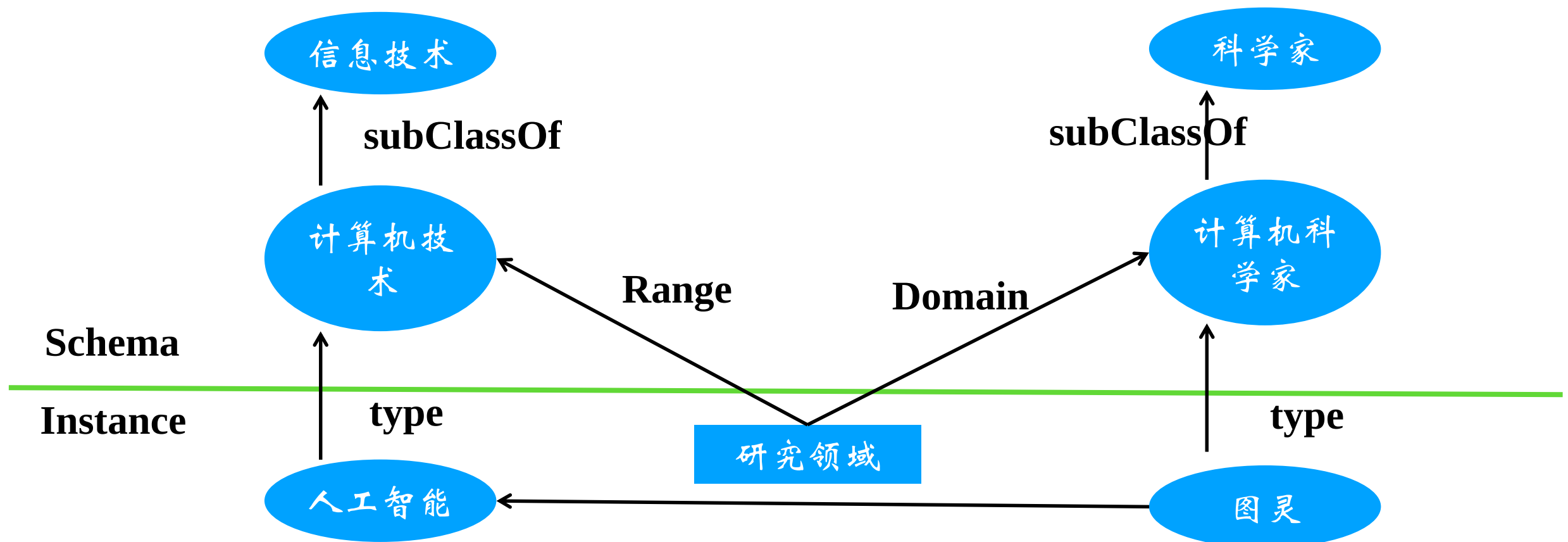
# RDF Graph: Directed Labeled Graph

## 基本数据模型：有向标记图



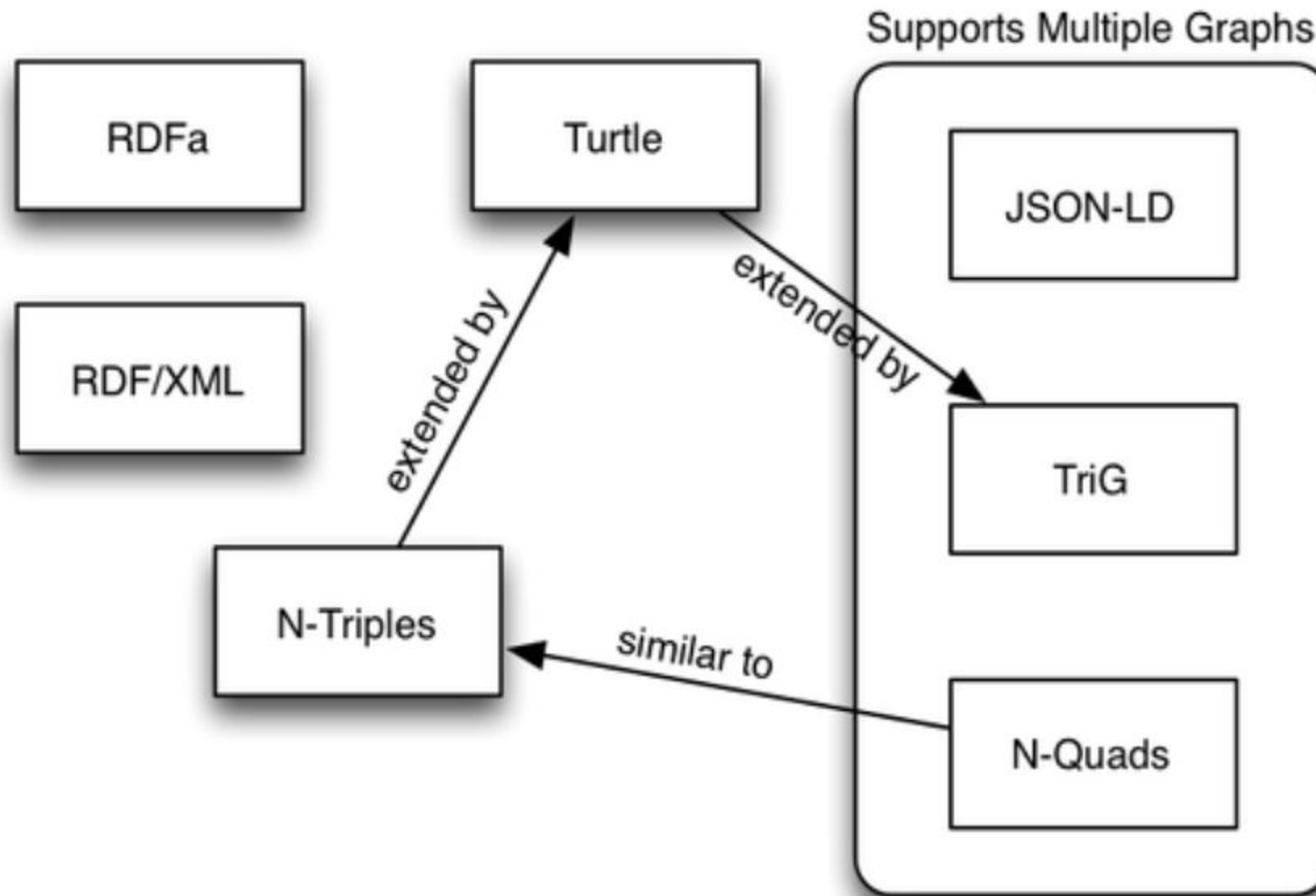
# RDFS: Simple Vocabulary and Schema

Defines small vocabulary for RDF:  
Class, subClassOf, type, Property, subPropertyOf, Domain, Range



# RDF的序列化格式

RDF 1.1



# OWL: Web Ontology Language

---

## Ontology in Philosophy

Ontology is the philosophical study of the nature of being, becoming, existence or reality, as well as the basic categories of being and their relations.

—— Merriam-Webster

## Ontology in Computer Science and Artificial Intelligence

An ontology is a description (like a formal specification of a program) of the concepts and relationships that can formally exist for an agent or a community of agents.

—— Tom Gruber, Founder of Siri

## Web Ontologies

Ontologies based on web standards such as RDFS/OWL. OWL is based on Description Logic, a very very long history of research in Artificial Intelligence.

# OWL extends RDF Schema

---

**Class, subclass, Property, subProperty.....**



**Complex Classes: intersection, union and complement**

**Property Restrictions: existential quantification, universal quantification, hasValue**

**Cardinality Restrictions: maxQualifiedCardinality, minQualifiedCardinality, qualifiedCardinality**

**Property Characteristics: inverseOf, SymmetricProperty, AsymmetricProperty, propertyDisjointWith, ReflexiveProperty, FunctionalProperty**

# OWL 示例

## 复杂类

```
:Mother owl:equivalentClass [  
  rdf:type owl:Class ;  
  owl:intersectionOf ( :Woman :Parent )  
] .
```

## 属性约束

```
:Parent owl:equivalentClass [  
  rdf:type owl:Restriction ;  
  owl:onProperty :hasChild ;  
  owl:someValuesFrom :Person  
] .
```

## 对称属性

```
:hasSpouse rdf:type owl:SymmetricProperty .
```

## 传递属性

```
:hasAncestor rdf:type owl:TransitiveProperty .
```

## 属性链

```
:hasGrandparent owl:propertyChainAxiom ( :hasParent :hasParent ) .
```

# SPARQL 简介

- RDF 的查询语言
  - 基于RDF数据模型
- 可以对不同的数据集撰写复杂的连接(joins)
- 由所有主流图数据库支持
- SPARQL Protocol and RDF Query Language

A SPARQL query comprises, in order:

- *Prefix declarations*, for abbreviating URIs
- *Dataset definition*, stating what RDF graph(s) are being queried
- A *result clause*, identifying what information to return from the query
- The *query pattern*, specifying what to query for in the underlying dataset
- *Query modifiers*, slicing, ordering, and otherwise rearranging query results

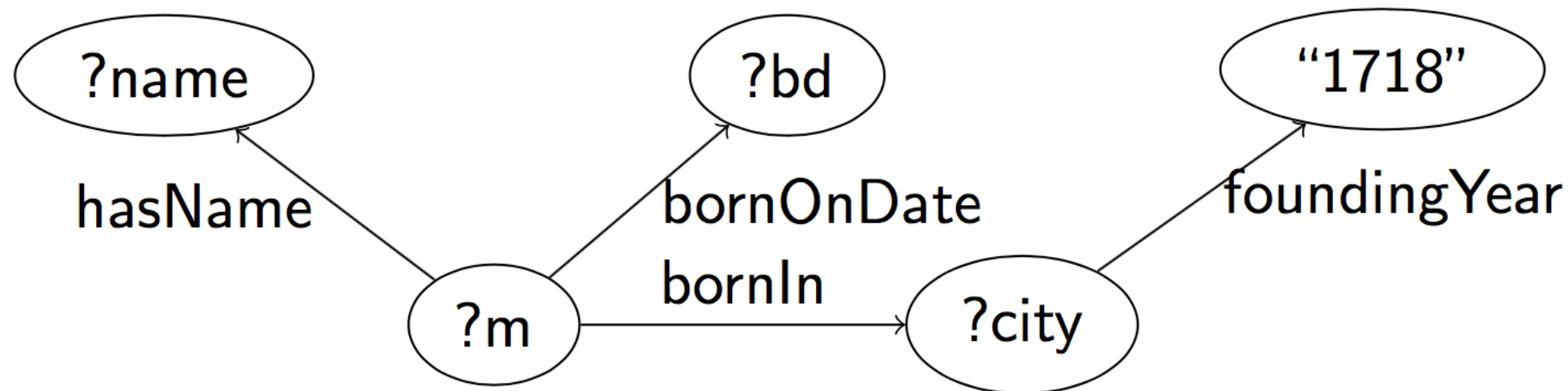
```
# prefix declarations
PREFIX foo: <http://example.com/resources/>
...
# dataset definition
FROM ...
# result clause
SELECT ...
# query pattern
WHERE {
    ...
}
# query modifiers
ORDER BY ...
```

详见: <http://www.w3.org/TR/rdf-sparql-query/>

# SPARQL 示例

```
SELECT ?name
WHERE {
  ?m <bornIn> ?city . ?m <hasName> ?name .
  ?m<bornOnDate> ?bd . ?city <foundingYear> '1718' .
  FILTER(regex(str(?bd), '1976'))
}
```

`FILTER(regex(str(?bd), "1976"))`



# JSON-LD: 数据交换格式

---

JSON for Linking Data: 适用于作为程序之间做数据交换，在网页中嵌入语义数据和Restful Web Service

```
{  
  "@context": "http://json-ld.org/contexts/person.jsonld",  
  "@id": "http://dbpedia.org/resource/John_Lennon",  
  "name": "John Lennon",  
  "born": "1940-10-09",  
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"  
}
```

<https://json-ld.org>

# RDFa, HTML5 MicroData: 在网页中嵌入语义数据

Semantic Markup: 在网页、邮件、应用程序中嵌入语义数据

```
<p>Big (1988) is a fantasy comedy movie starting Tom  
Hanks and directed by Penny Marshall.</p>
```

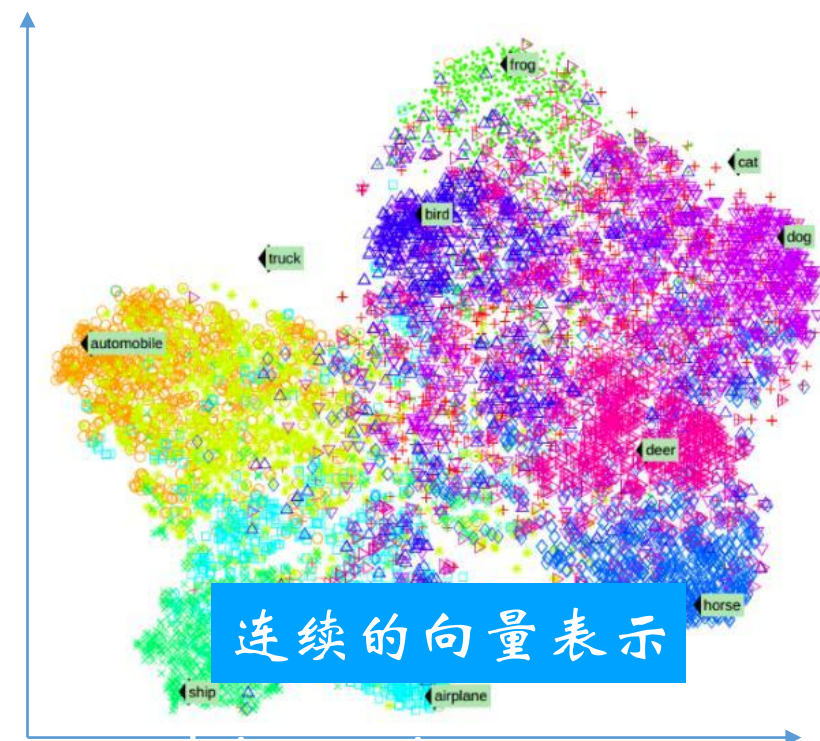
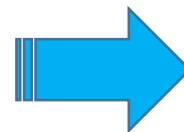
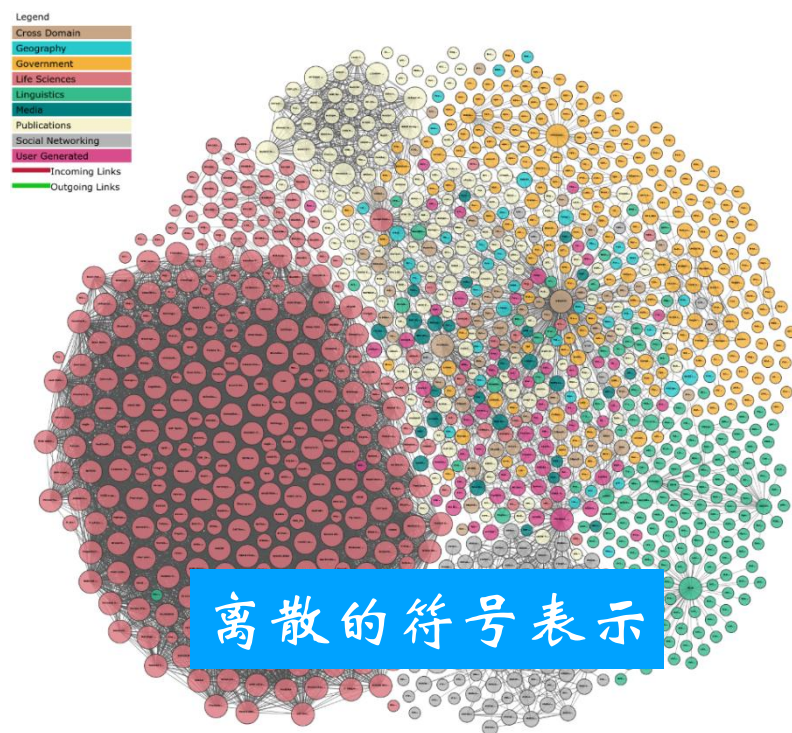
RDFa, JSON-LD, HTML5 Microdata

```
<div itemscope itemtype="http://schema.org/Movie"><p>  
  <span itemprop="name">Big </span> (1988) is a  
  <span itemprop="genre">fantasy comedy</span> movie starting  
  <span itemprop="actor" itemscope itemtype="http://schema.org/Person">  
  <span itemprop="name">Tom Hanks</span></span> and directed by  
  <span itemprop="director" itemscope itemtype="http://schema.org/person">  
    <span itemprop="name">Penny Marshall</span></span>.</p></div>
```

搜索引擎优化

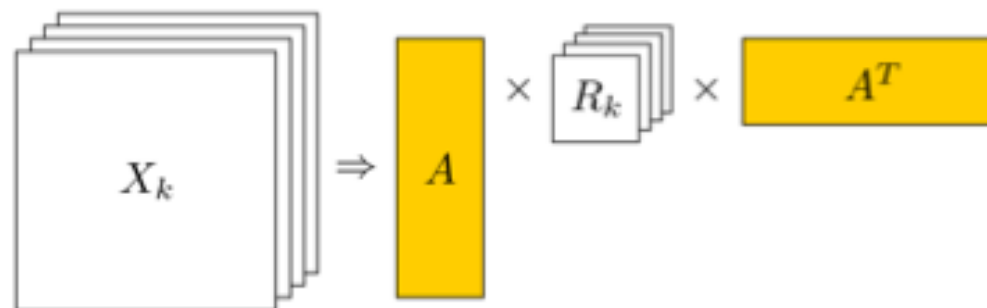
# 知识图谱的分布式表示 – KG Embedding

在保留语义的同时，将知识图谱中的实体和关系映射到连续的稠密的低维向量空间

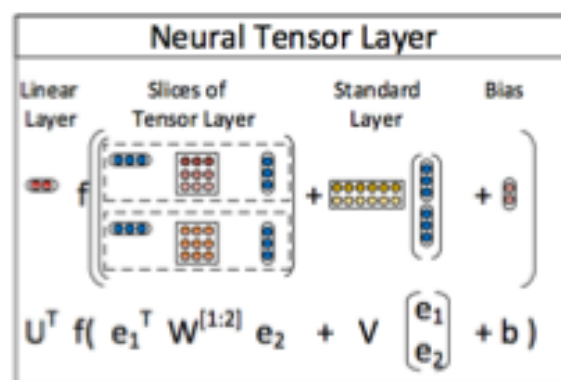


# 知识图谱的分布式表示

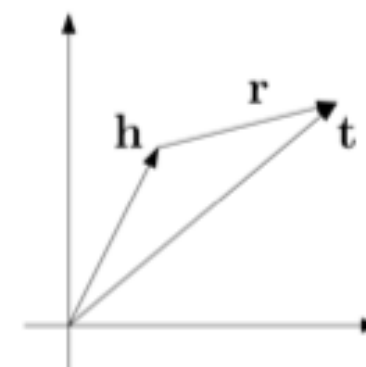
张量分解  
(e.g. RESCAL<sup>1</sup>)



神经网络  
(e.g. NTN<sup>2</sup>)



距离模型  
(e.g. TransE3)



M. Nickel, V. Tresp, and H. Kriegel, A three-way model for collective learning on multi- relational data. ICML2011  
Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. . Reasoning with neural tensor networks for knowledge base completion. NIPS2013  
A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, Translating em- beddings for modeling multi-relational data. NIPS2013

# 知识抽取: NLP + KR

KR: 三元组、多元关系、模态知识.....

关系抽取 事件抽取

NER命名实体识别 Entity Linking 实体链接

分词 词性标注 语法解析 依存分析

文本预处理

各种非结构化文本数据



TAC 的 KBP Track

ACE

MUC-7

# 知识抽取：主要方法

---

## ☐ 知识工程

- 正则表达式

- 模板匹配

- ☐ BootStrap: 从种子数据中识别 Patterns, 用于抽取更多数据, 提取更多 Patterns

- 规则约束

- ☐ POS、NER约束、距离约束等

## ☐ 基于本体的抽取

- 知识挖掘 (推理)

- ☐ PRA 基于图的抽取

- ☐ TransE系列 基于Embedding的抽取

# 知识抽取：主要方法

---

## ☐ 基于模型的抽取

### ■ 模型

☐ SVM, Logistic Model

☐ 条件随机场 (CRF)

☐ LSTM等循环神经网络

### ■ 训练

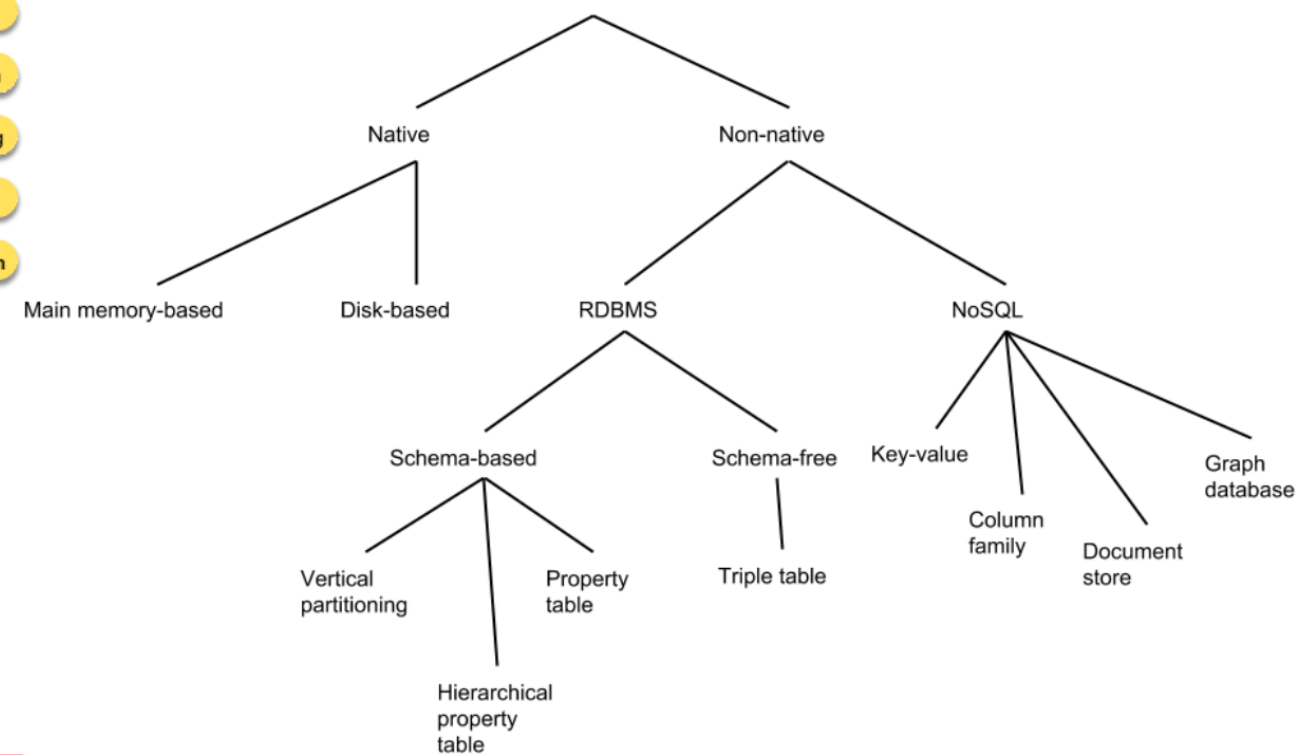
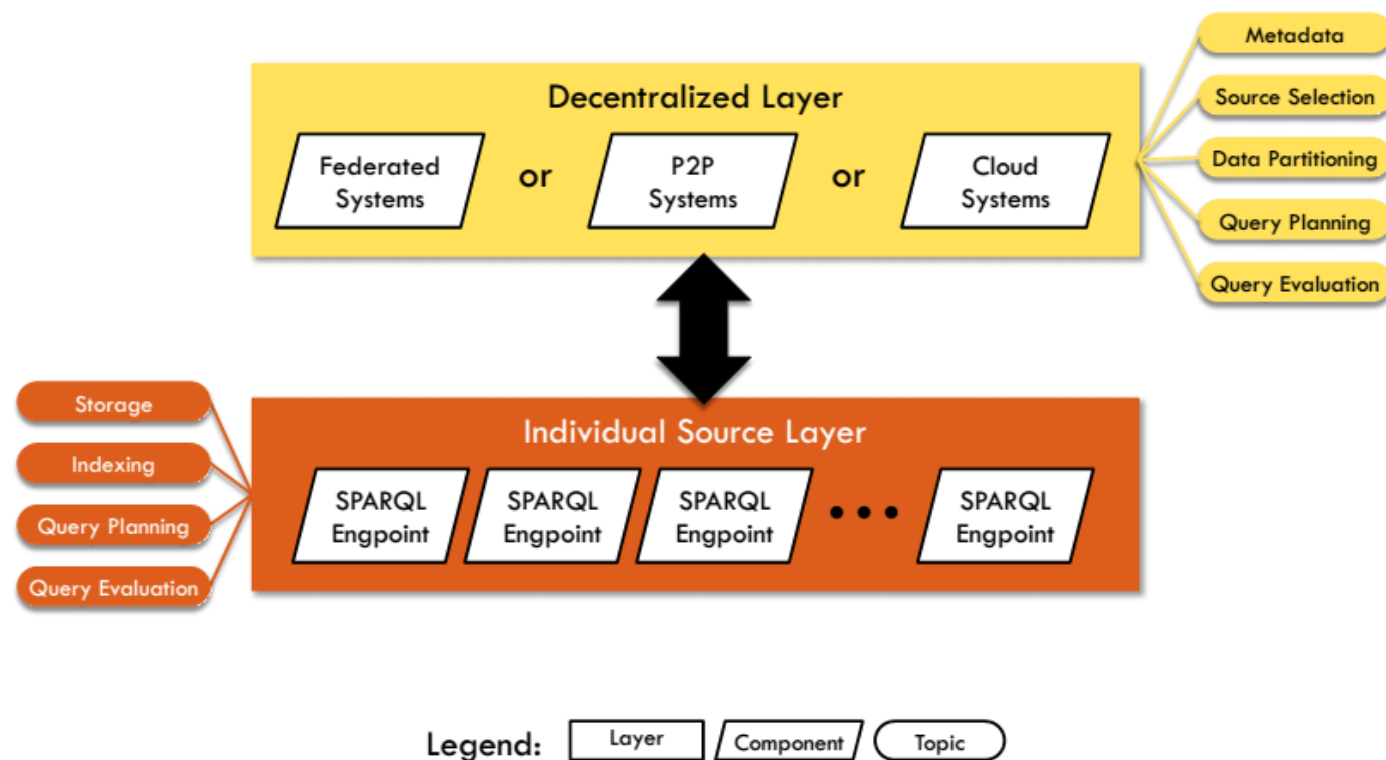
☐ 有监督学习

☐ 无监督聚类

☐ 远程监督 (Distant Supervision), 从已有知识库获取少量训练数据

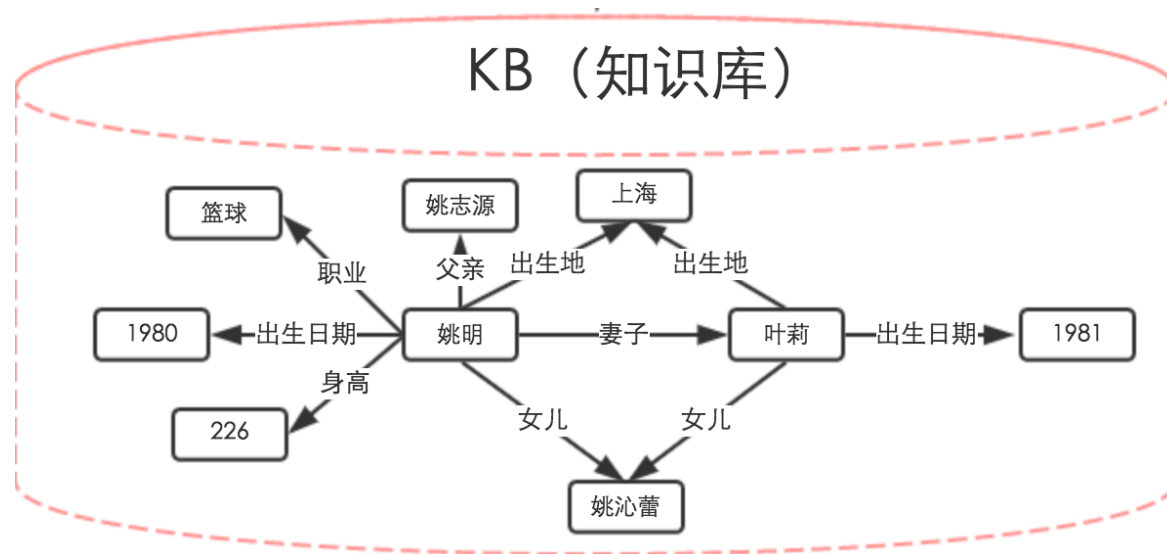
# 知识存储

- 知识比数据的结构更加复杂，知识的存储需要综合考虑图的特点、复杂的知识结构存储、索引和查询（支持推理）的优化等问题
- 典型的知识存储引擎分为基于关系数据库的存储和基于原生图的存储
- 在实践中，多为混合存储结构，图存储并非必须



# 知识问答

- KBQA (Knowledge-Based Question Answering, 基于知识库的问题回答)
- 以直接而准确的方式回答用户自然语言提问的自动问答系统将构成下一代搜索引擎的基本形态。



输入：自然语言的问句和知识库  
姚明的身高？

输出：问句对应的答案  
226



Prof. Oren Etzioni

Turing Center  
University of Washington

以直接而准确的方式  
回答用户自然语言提  
问的自动问答系统将  
构成下一代搜索引擎  
的基本形态

— 《Nature》2011.8

# KBQA例子

Baidu 姚明个子有多高 百度一下 百度首页 设置 登录

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约1,650,000个 搜索工具

姚明身高:  
226cm  
姚明, 1980年9月12日出生于上海市徐汇区, 祖籍江苏省苏州市吴江区震泽镇, 前中国职业篮球运动员, 司职中锋, 中联公司董事兼总经理。1998年4月, 姚明入选王非执教的国... 详情>>  
来自百度百科 | 报错

姚明个子有多高 百度知道  
1个回答 - 提问时间: 2016年12月01日  
最佳答案: 姚明, 1980年9月12日出生于上海市徐汇区, 祖籍江苏省苏州市吴江区震泽镇, 前中国职业篮球运动员, 司职中锋, 中联公司董事兼总经理[1]。1998年4月, ...  
zhidao.baidu.com/link?...  
姚明的身高是多少厘米? 4个回答 2014-04-28  
姚明的身高有多高? 1个回答 2014-04-30  
姚明的真正身高是多少? 6个回答 2014-04-30  
更多知道相关问题>>

姚明到底有多高? - 爱问知识人  
19个回答 - 最新回答: 2008年7月17日  
最佳答案: 姚明的身高到底是多少? 自从他进入NBA以来, 这个问题就一直存在着争论, 以前在中国CBA的秩序册上, 姚明的身高是2米23, 而到了美国后, 姚明的身高变成了2...  
iask.sina.com.cn/b/131... - 百度快照

相关人物 展开  
易建联 孙明明 迈克尔·乔丹 姚沁蕾  
叶莉 林书豪 穆铁柱 周琦  
科比·布莱恩特 沙奎尔·奥尼尔 蒙克·巴特尔 孙悦  
给百度提建议

Google who is the 35th president of the united states 语音 搜索

全部 新闻 图片 地图 视频 更多 设置 工具

找到约 3,490,000 条结果 (用时 0.61 秒)

美国总统 (35)

约翰·肯尼迪

更多有关“约翰·肯尼迪”的信息

John F. Kennedy - Wikipedia  
https://en.wikipedia.org/wiki/John\_F.\_Kennedy  
John Fitzgerald "Jack" Kennedy (May 29, 1917 – November 22, 1963), commonly referred to by his initials JFK, was an American politician who served as the 35th President of the United States from

反馈

WolframAlpha computational knowledge engine

Enter what you want to calculate or know about:

how big is China 星星 菜单

Examples Random

Assuming "how big" is international data | Use as referring to socioeconomic data or referring to species or referring to administrative divisions instead

Assuming total area | Use population instead

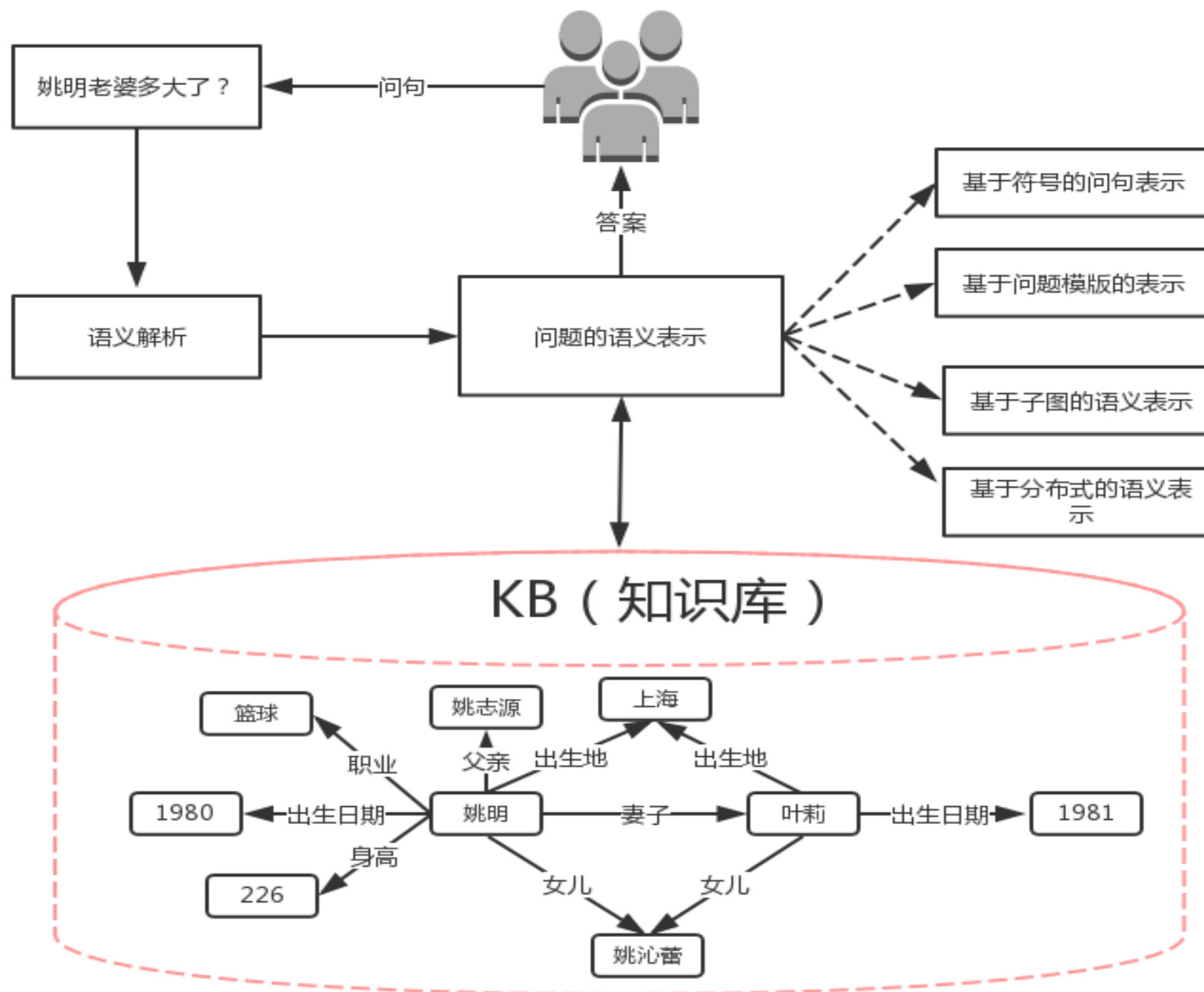
Input interpretation:  
China total area

Result: Show non-metric  
 $9.597 \times 10^6 \text{ km}^2$  (square kilometers) (world rank: 4<sup>th</sup>)

Unit conversions:  
 $9.597 \times 10^{12} \text{ m}^2$  (square meters)  
3.705 million mi<sup>2</sup> (square miles)  
 $1.033 \times 10^{14} \text{ ft}^2$  (square feet)

Comparisons as area:  
 $\approx 0.96 \times \text{total area of Canada}$  ( $9.98467 \times 10^6 \text{ km}^2$ )  
 $\approx 0.996 \times \text{total area of the United States}$  ( $9.63142 \times 10^6 \text{ km}^2$ )  
 $\approx \text{largest extent of the Roman Empire}$  ( $\approx 9 \text{ Mm}^2$ )

# KBQA基本实现流程



# 知识推理

简单而言，推理是指基于已知事实推出未知的事实的计算过程。

大部分搜索引擎目前都已经支持了一定的推理能力

百度输入：姚明的女儿的妈妈的老公

Baidu 百度

姚明的女儿的妈妈的老公

百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约1,910,000个

搜索工具

叶莉丈夫：  
**姚明**

姚明，1980年9月12日出生于上海市徐汇区，祖籍江苏省苏州市吴江区震泽镇，前中国职业篮球运动员，司职中锋，中职联公司董事长兼总经理。1998年4月，姚明入选王非执教的国... [详情>>](#)

来自百度百科图谱 | 报错

姚明的女儿的妈妈的老公\_百度图片



image.baidu.com - 查看全部8,274张图片

# 知识推理

---

按解决方法分类：

① 基于描述逻辑的推理

② 基于规则挖掘的推理

③ 基于概率逻辑的推理

④ 基于表示学习与神经网络的推理

按推理类型分类：

① 缺省推理 Default reasoning

② 连续变化推理 Continuous Change

③ 空间推理 Spatial reasoning

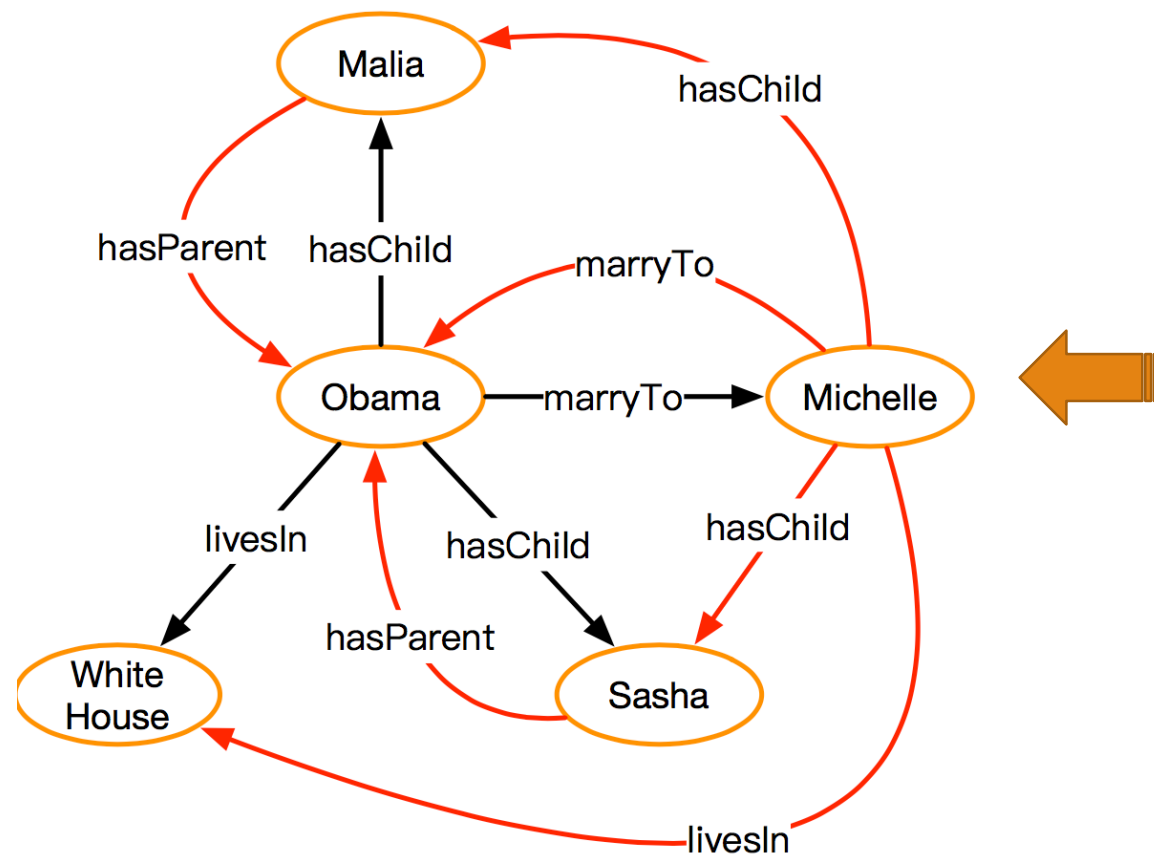
④ 因果关系推理 Causality reasoning

⑤ ...

# 基于描述逻辑的推理：本体推理

- 描述逻辑（description logic）是一种用于知识表示的逻辑语言和以其为对象的推理方法，主要用于描述概念分类及其概念之间的关系。描述逻辑是当前语义网发展中本体的理论基础。
- 主要方法：
  - 基于表运算（Tableaux）及改进的方法：FaCT++、Racer、Pellet Hermit等
  - 基于一阶查询重写的方法（Ontology based data access，基于本体的数据访问）
  - 基于产生式规则的算法（如rete）：Jena、Sesame、OWLIM等
  - 基于Datalog转换的方法如KAON、RDFox等。
  - 回答集程序 Answer set programming

# 基于统计规则挖掘的推理



$hasChild(A, B) \Rightarrow hasParent(B, A)$

$marryTo(A, B) \Rightarrow marryTo(B, A)$

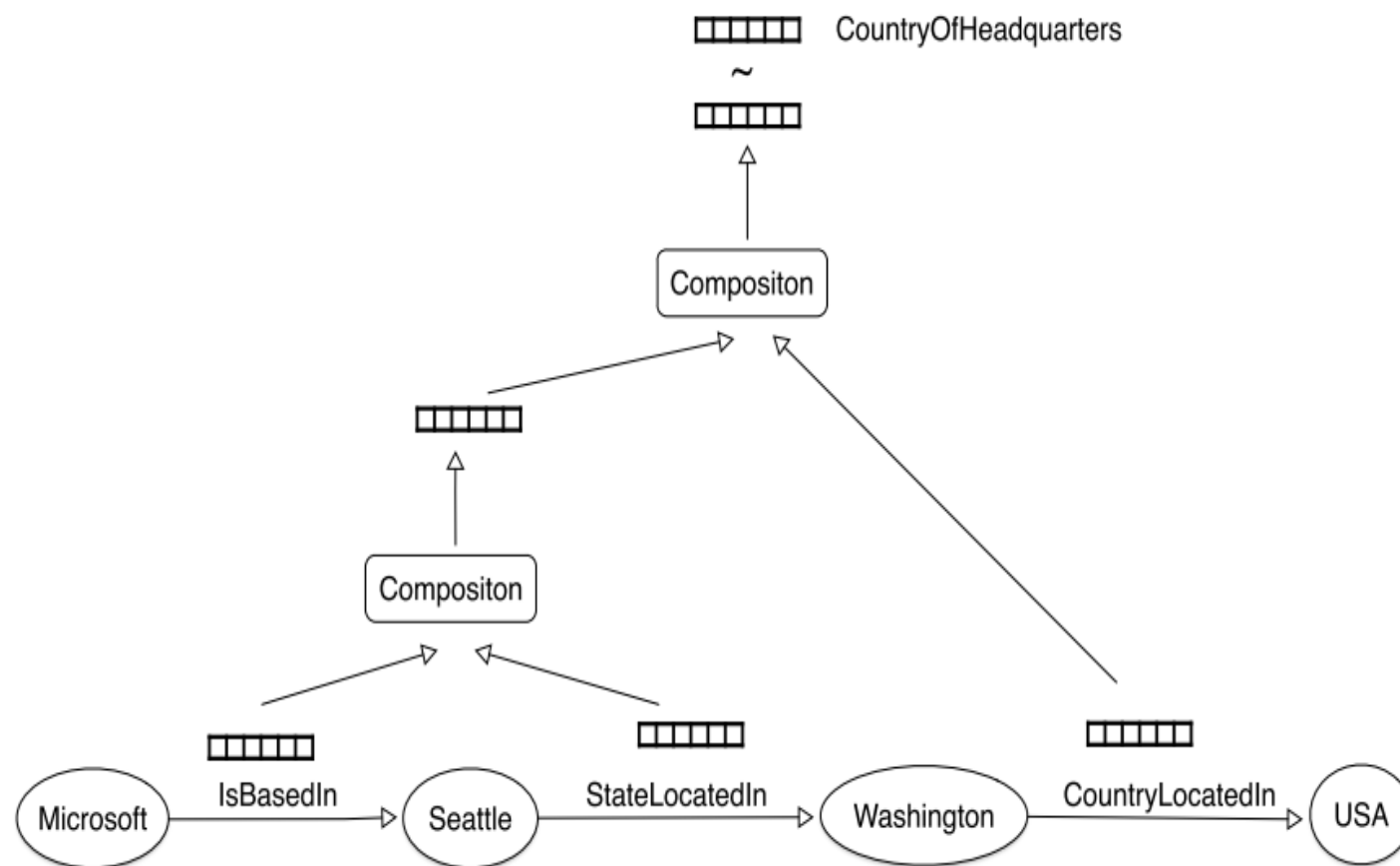
$marryTo(A, B) \wedge hasChild(A, C) \Rightarrow hasChild(B, C)$

$marryTo(A, B) \wedge livesIn(A, C) \Rightarrow livesIn(B, C)$

主要方法:

- 基于路径排序学习方法 (PRA, Path ranking Algorithm)
- 基于关联规则挖掘方法 (AMIE)

# 基于表示学习的推理-Representational Learning

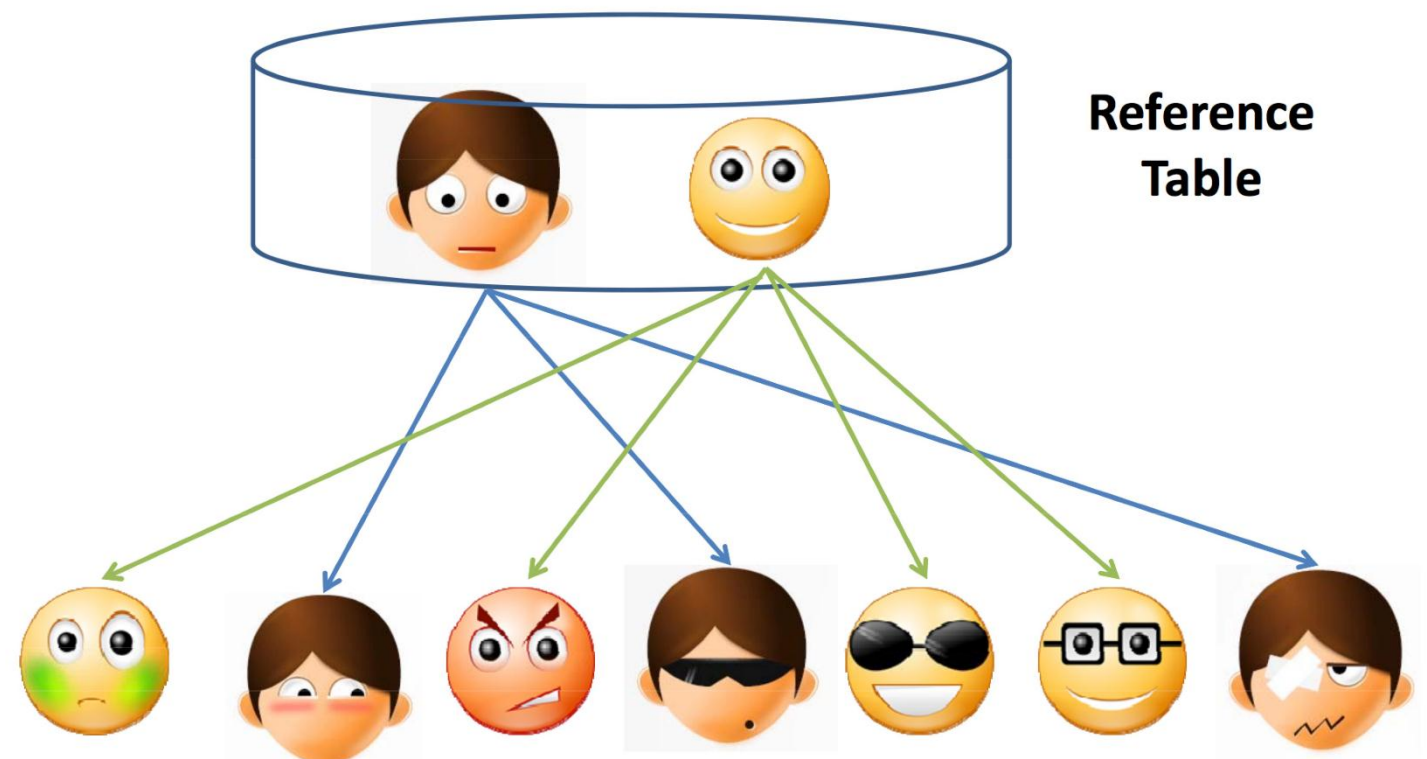


## Representation Learning

- 将实体和关系都表示为向量
- 通过向量之间的计算代替图的遍历和搜索来预测三元组的存在，由于向量的表示已经包含了实体原有的语义信息，计算含有一定的推理能力。
- 可应用于链接预测，基于路径的多度查询等。

# 知识融合

- 其它相关叫法: Record Linkage, Entity Resolution, Data Linking, Knowledge Fusion, Entity Alignment.....
- 都是指在不同数据集中找出同一个实体的描述记录, 主要目的是对不同数据源中的实体信息进行整合, 形成更加全面的实体信息



引自丁力  
《实体融合、实体链接技术》

# 知识融合: 典型工具-Dedupe

---

- ❑ Dedupe, 基于Python的工具包, 实现了包括fuzzy matching, deduplication, entity resolution在内的常见任务。主要处理流程是先对所有records通过Clustering/Blocking的方法进行分组, 然后在组内部通过计算相似度特征和机器学习分类模型对任一一对records进行预测是否为同一实体。
- ❑ 适合于两个数据集有相似的结构, 如果两个数据集的实体属性个数差异很大, 不建议采用。

具体方法参考: [http://vldb.org/pvldb/vol5/p1878\\_larskolb\\_vldb2012.pdf](http://vldb.org/pvldb/vol5/p1878_larskolb_vldb2012.pdf)

<http://openkg.cn/tool/dedupe>

# 知识融合: 典型工具-LIMES

- ❑ Limes: A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data. IJCAI2013
- ❑ 专门针对链接数据Linked Data设计的链接框架，不要求两个数据集的实体具有相似的结构
- ❑ 可灵活配置匹配规则，自定义距离计算模型，也支持基于 Active Learning的机器学习方法

<http://openkg.cn/tool/limes>

The screenshot shows the LIMES web interface. At the top, there are links for 'Examples (toggle)', 'Drugbank', 'Diseases', 'SanCities', and 'Download: Manual | Distribution'. The main configuration area is divided into 'Source' and 'Target' sections. The 'Source' section has fields for 'Endpoint' (http://mesh.bio2rdf.org/sparql), 'Var' (?y), 'Pagesize' (1000), 'Restriction' (?y rdf:type mesh:Concept), and 'Property' (dc:title). The 'Target' section has fields for 'Endpoint' (http://data.linkedct.org/sparql), 'Var' (?x), 'Pagesize' (1000), 'Restriction' (?x rdf:type linkedct:condition), and 'Property' (linkedct:condition\_name). Below these, there is a 'Metric' field (levenshtein(y.dc.title, x.linkedct:condition\_name)), an 'Exemplars' field, and 'Acceptance' and 'Review' sections with 'Threshold' (0.98 and 0.95) and 'Relation' (owl:sameAs) fields. A 'Detected prefixes' section lists various URIs. At the bottom, there is a 'Start Linking' button and a visual graph showing the linking process between source and target properties, with a central 'and' node and an 'output' node.

# 知识众包: Wikibase



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

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

[English](#) [Not logged in](#) [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Item [Discussion](#)

Read [View history](#)

## Erin Moran (Q1281080)

American actress

[edit](#)

Erin Moran Fleischmann | Erin Fleischmann | Erin Marie Moran | Erin Marie Moran Fleischmann | Erin Marie Fleischmann

[In more languages](#) [Configure](#)

| Language   | Label            | Description            | Also known as                                                                                                            |
|------------|------------------|------------------------|--------------------------------------------------------------------------------------------------------------------------|
| English    | Erin Moran       | American actress       | Erin Moran Fleischmann<br>Erin Fleischmann<br>Erin Marie Moran<br>Erin Marie Moran Fleischmann<br>Erin Marie Fleischmann |
| Chinese    | No label defined | No description defined |                                                                                                                          |
| Wu Chinese | No label defined | No description defined |                                                                                                                          |
| Cantonese  | No label defined | No description defined |                                                                                                                          |

[All entered languages](#)

## Statements

instance of



human

[edit](#)

[1 reference](#)

# 知识众包: Schema.ORG

## Semantic Markup:

在网页、邮件、应用程序中嵌入语义数据

```
<p>Big (1988) is a fantasy comedy movie starting Tom Hanks and directed by Penny Marshall.</p>
```

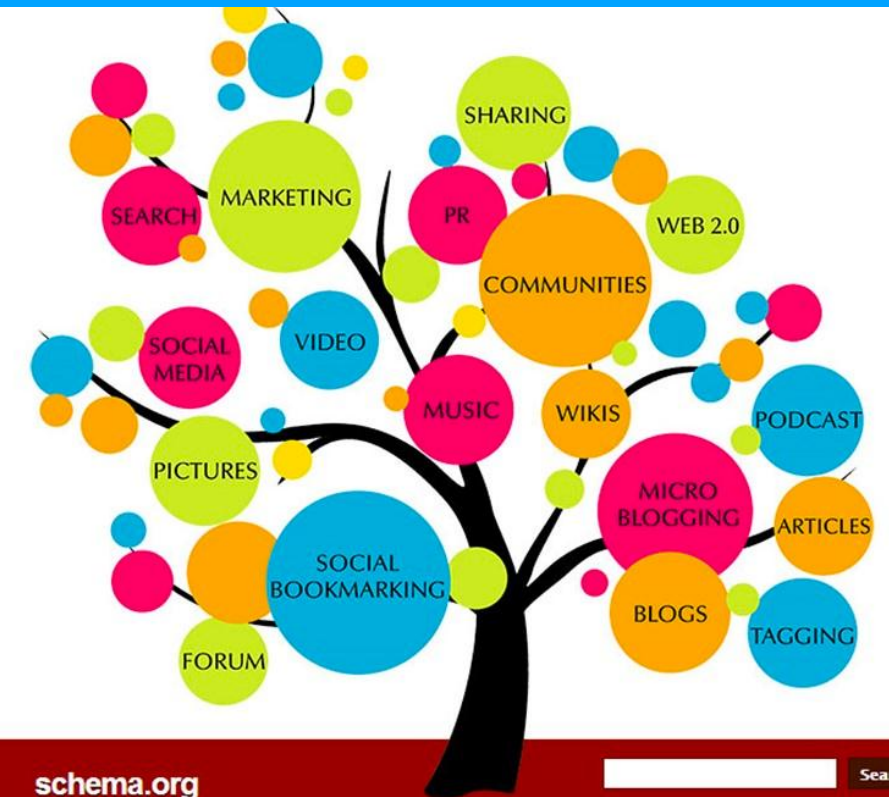
RDFa, JSON-LD, HTML5 Microdata

```
<div itemscope itemtype="http://schema.org/Movie"><p>  
  <span itemprop="name">Big </span> (1988) is a  
  <span itemprop="genre">fantasy comedy</span> movie starting  
  <span itemprop="actor" itemscope itemtype="http://schema.org/Person">  
    <span itemprop="name">Tom Hanks</span></span> and directed by  
  <span itemprop="director" itemscope itemtype="http://schema.org/person">  
    <span itemprop="name">Penny Marshall</span></span>.</p></div>
```

搜索引擎优化

只针对开放域

已经包含642个类, 992个关系

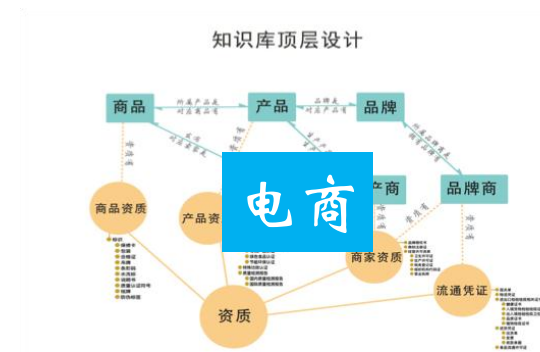
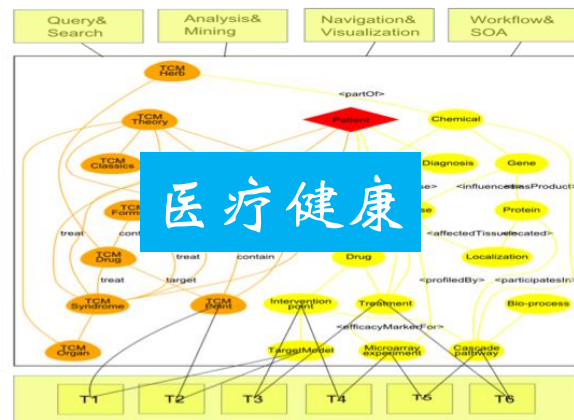


截止目前, 全球约有1.2亿的网站, 超过30%的网页已经嵌入有Schema.org的语义数据。谷歌2015年12月推出的Customizable Knowledge Graph, 允许各网站基于Schema.org, 以RDFa, JSON-LD, HTML Microdata等方式在网页、邮件等数据源中嵌入语义化数据, 支持个人和企业定制自己的知识图谱信息。

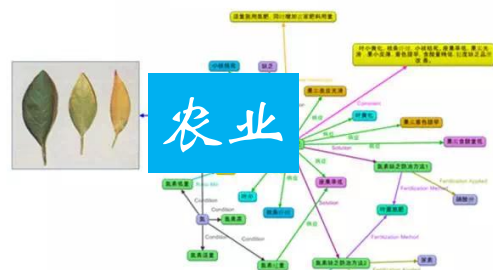
---

## 第四部分：典型案例简介

# 蓬勃发展的知识图谱应用



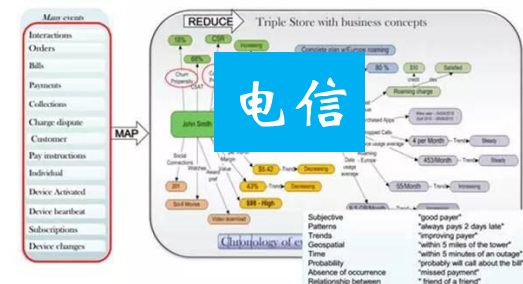
Complete Ontology (knowledge) for Lacking Nitrogen



Palantir government



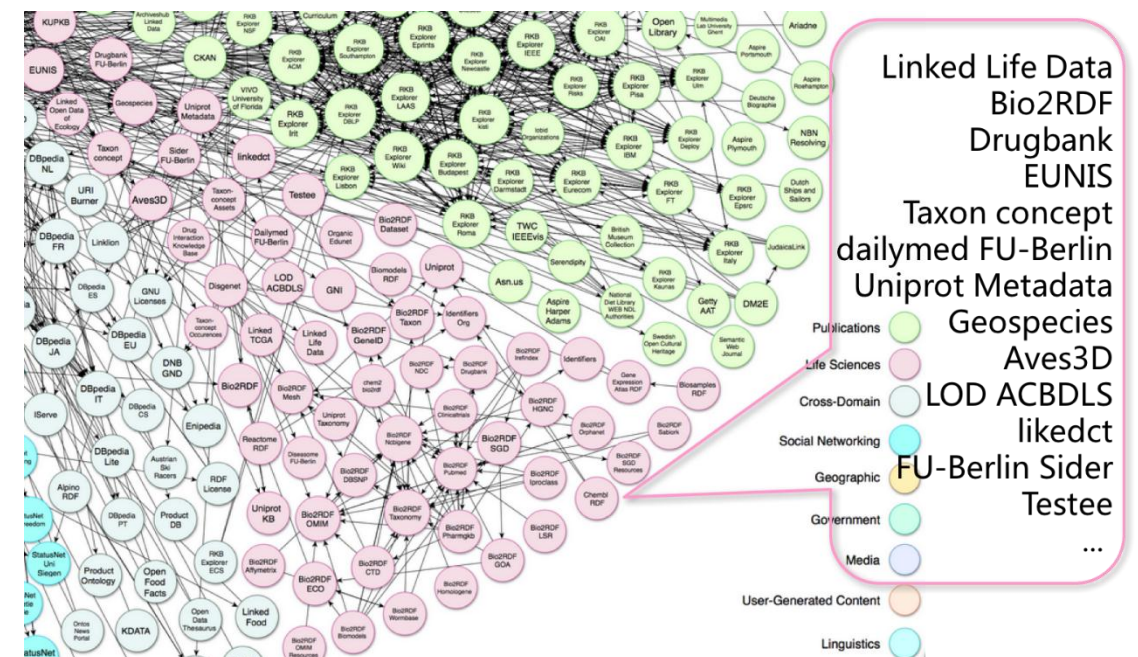
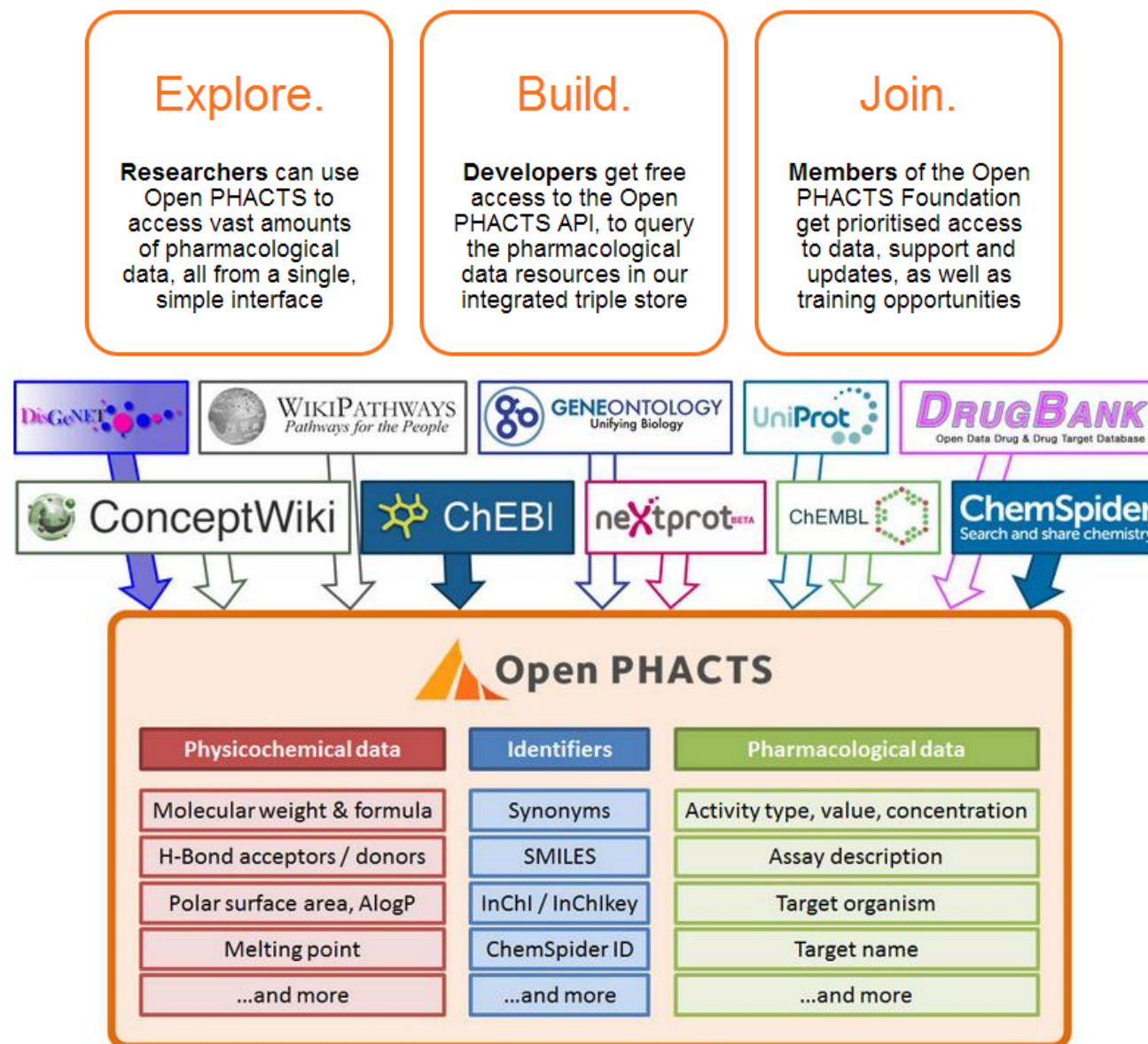
Turn Massive Raw Data Into Business Concepts



# Open PHACTS



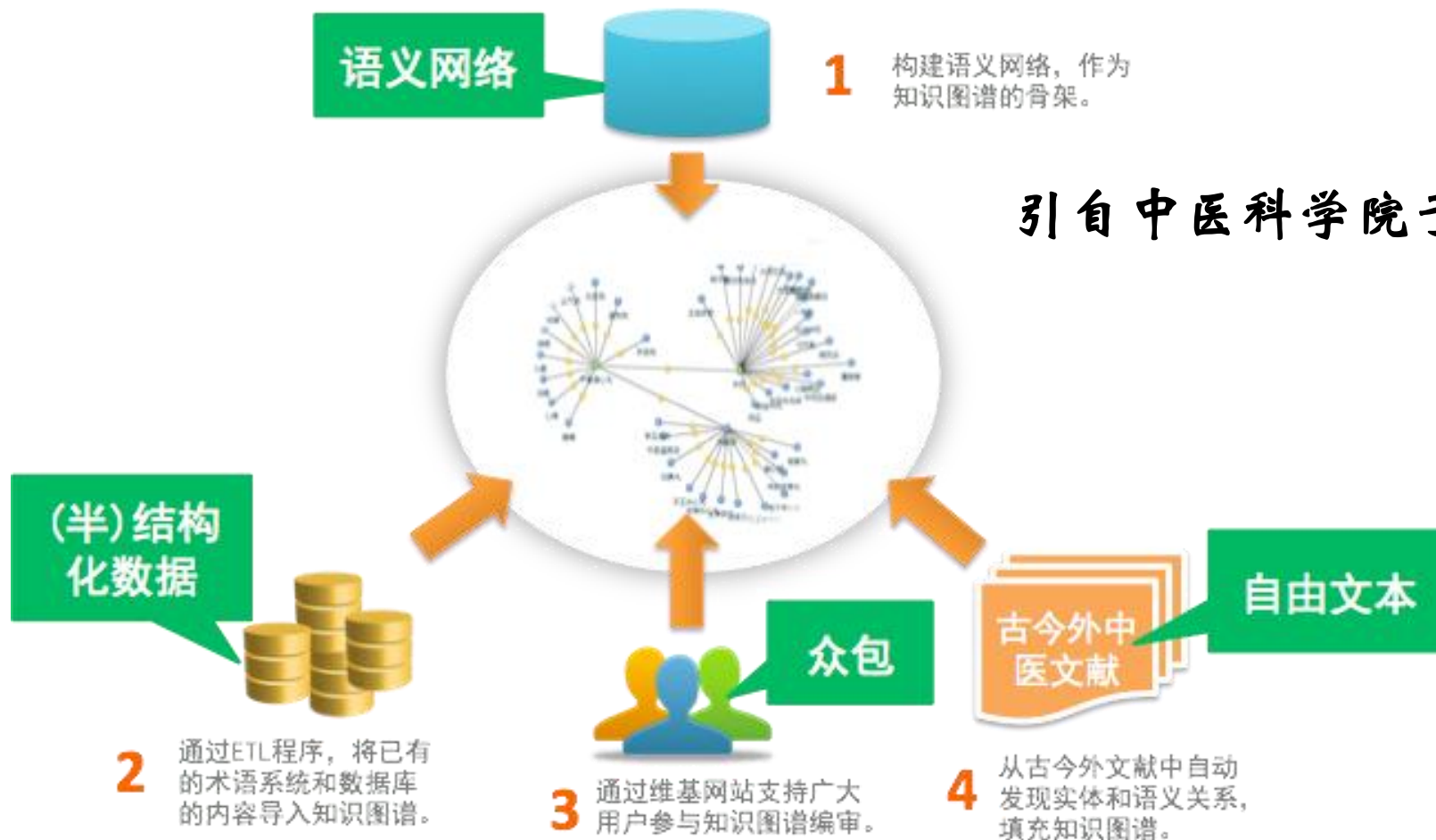
Bringing together **pharmacological data resources** in an integrated, interoperable infrastructure



## 欧盟重大联合攻关项目

面向药物研发的开放数据访问平台开发，其核心技术就是采用语义技术为有关研究人员提供高效的数据访问技术环境的支持

# 中医药知识平台



引自中医科学院于彤博士《中医药知识服务平台研发》



## 主要目标

- 对中医药知识体系进行系统梳理、建模和展示
- 知识可视化以图形方式凸显核心概念之间的关系
- 辅助中医专家厘清学术发展脉络，浏览中医知识，发现知识点之间的联系。与阅读文献等手段相比，可节约知识检索获取时间

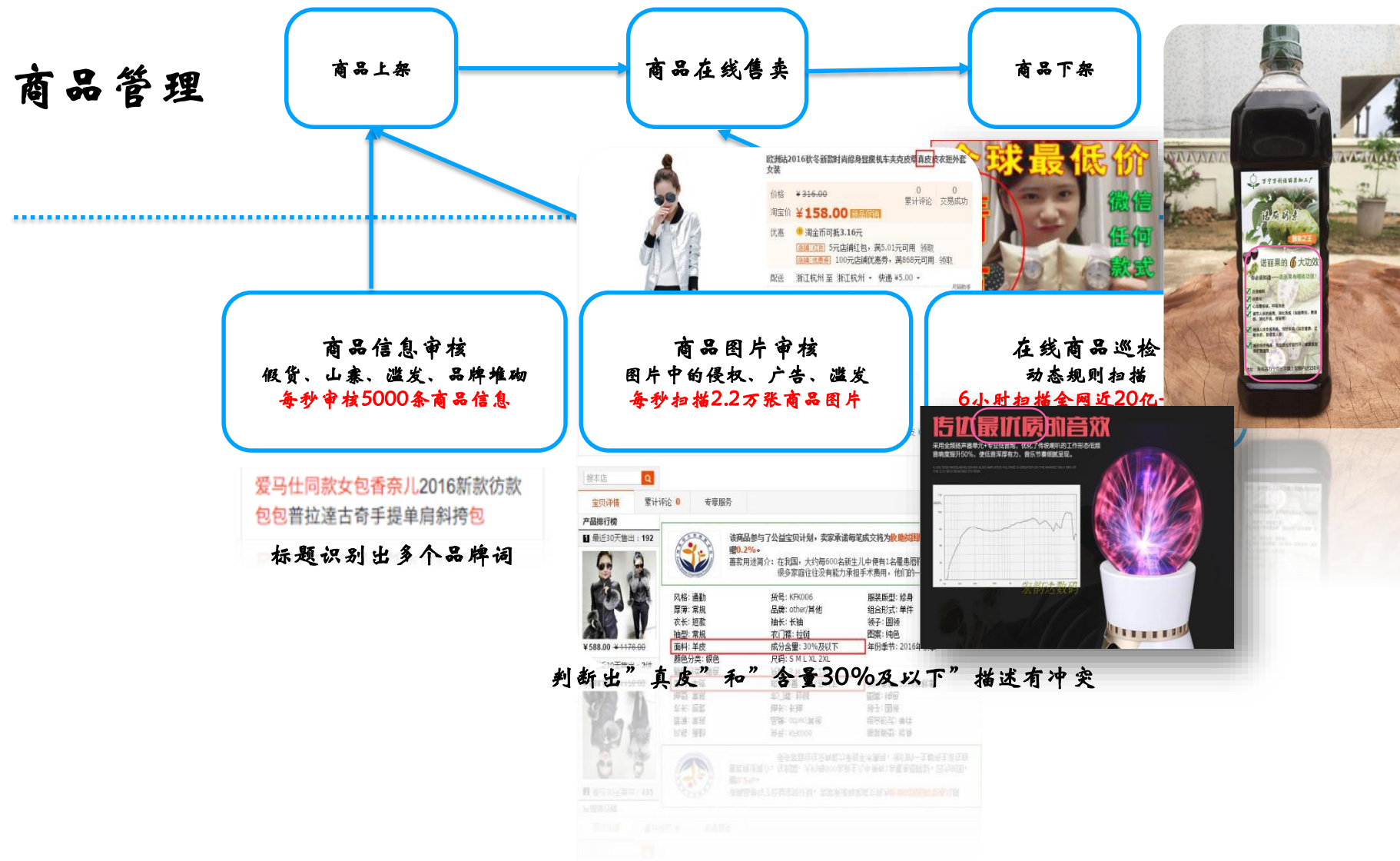
# 电商知识图谱

- 百亿条信息实体、百亿条信息关联(边)



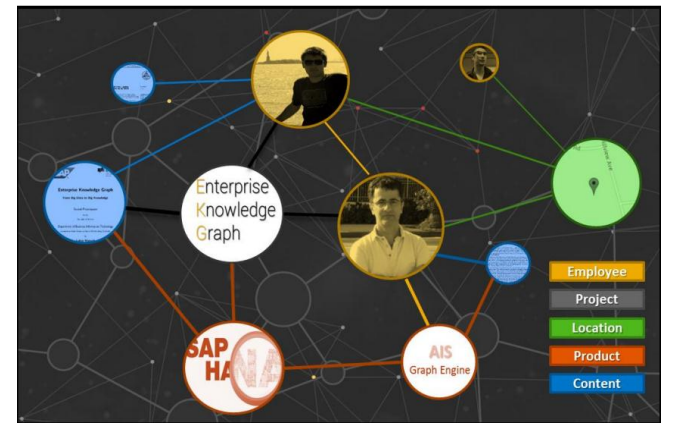
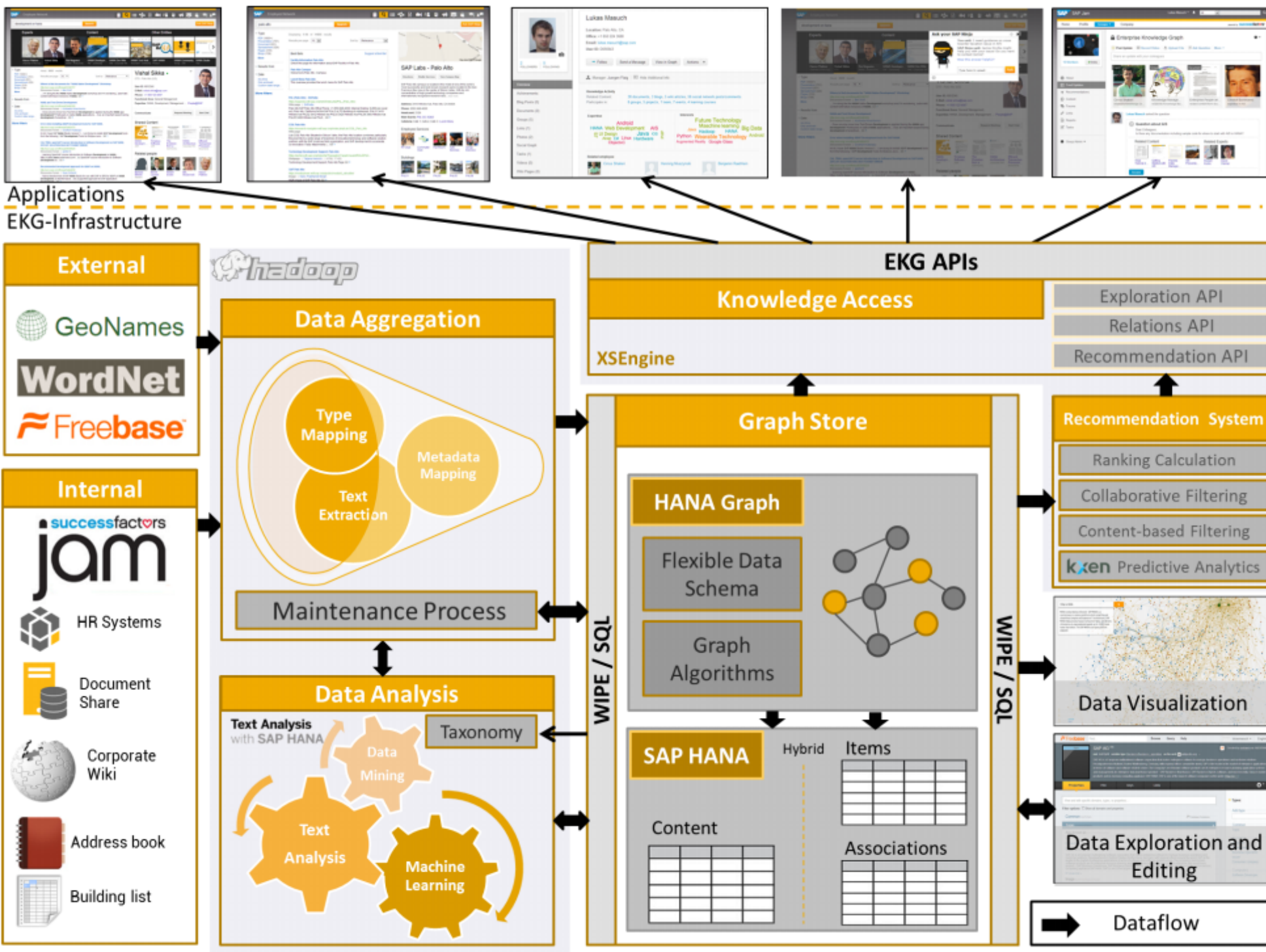
引自：张伟（览图）——“阿里巴巴知识图谱”

# 电商在管控中的应用 让判断更智能

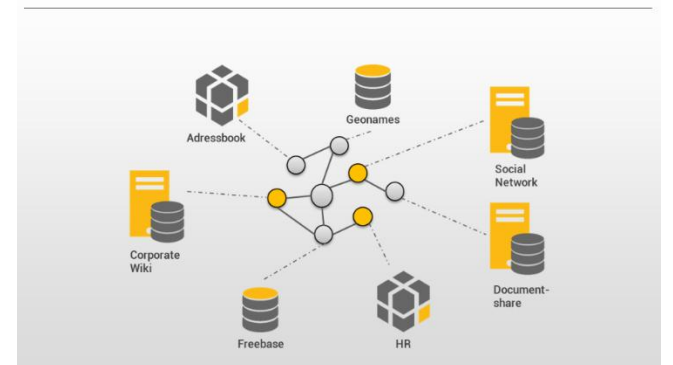


引自：张伟（览图）——“阿里巴巴知识图谱”

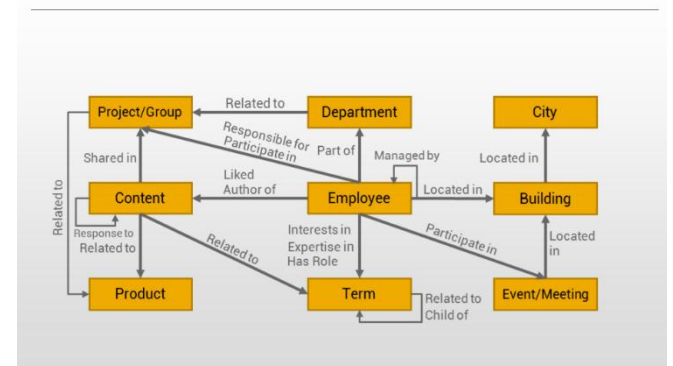
# 企业知识图谱应用 (SAP)

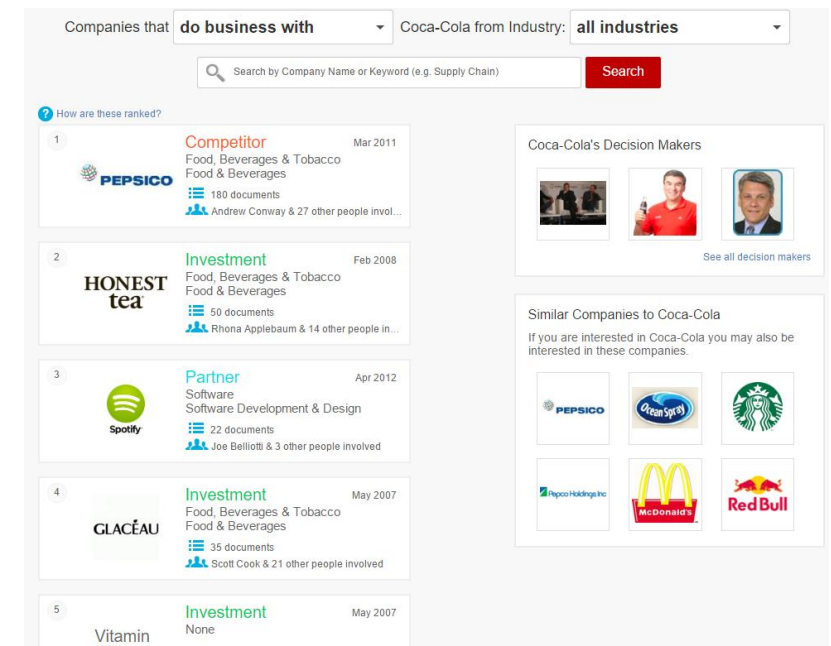
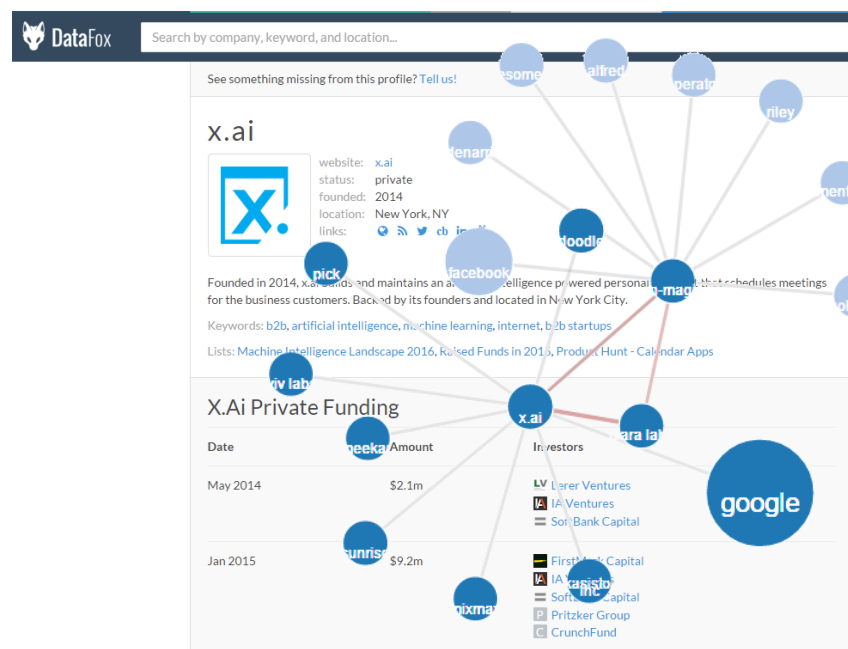
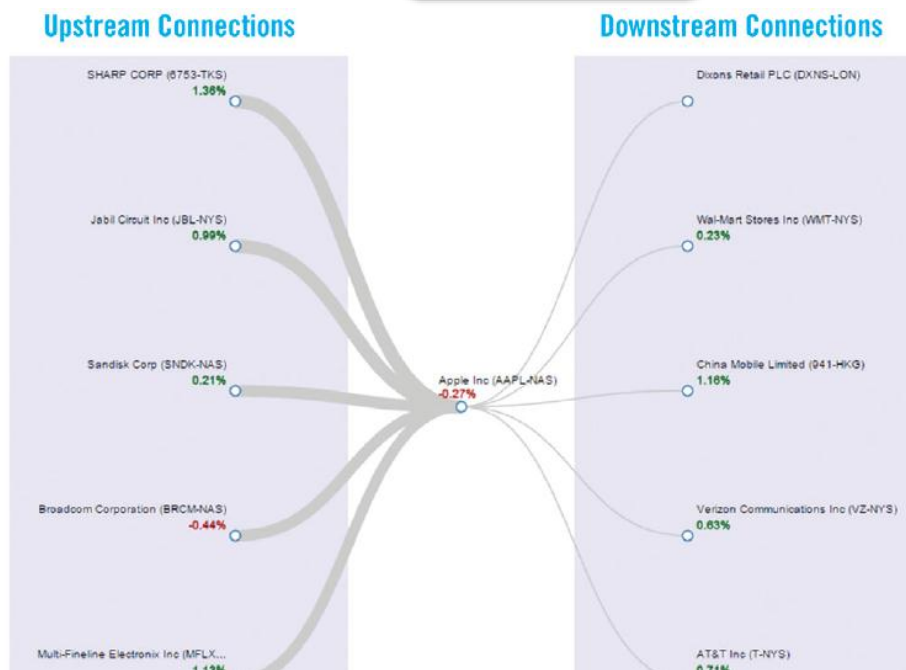
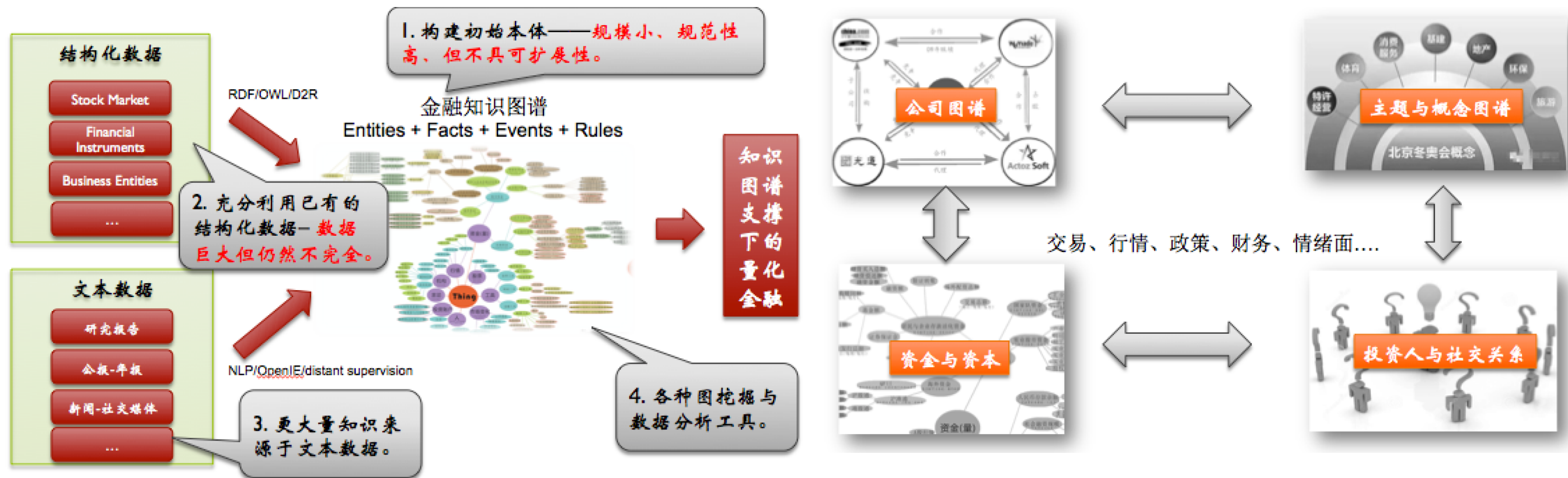


Data Aggregation



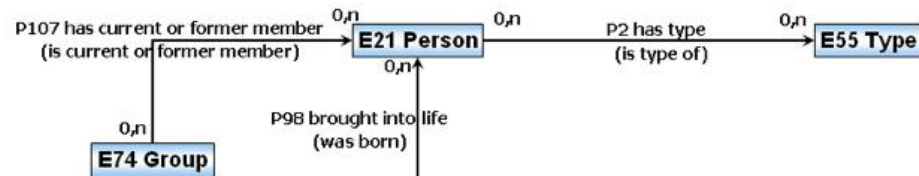
Enterprise Ontology Model (simplified)



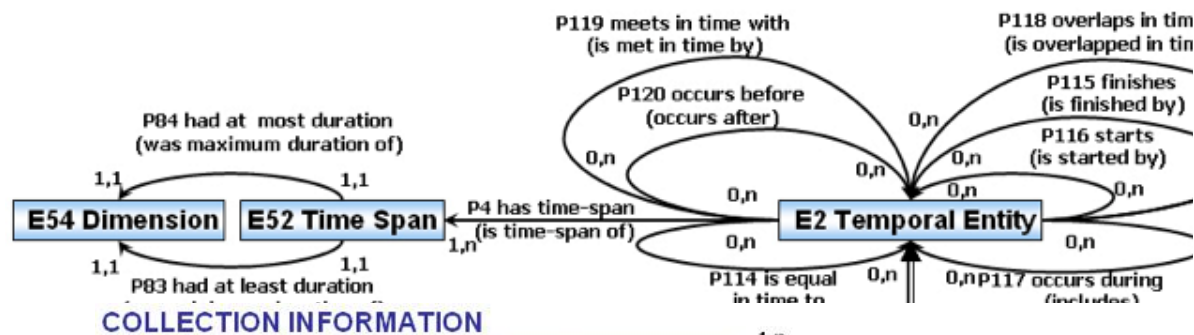
[illegible]

# 大英博物馆：知识图谱与本体设计

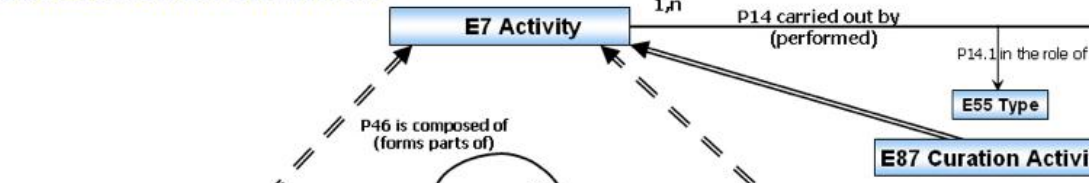
## PERSON NATIONALITY INFORMATION



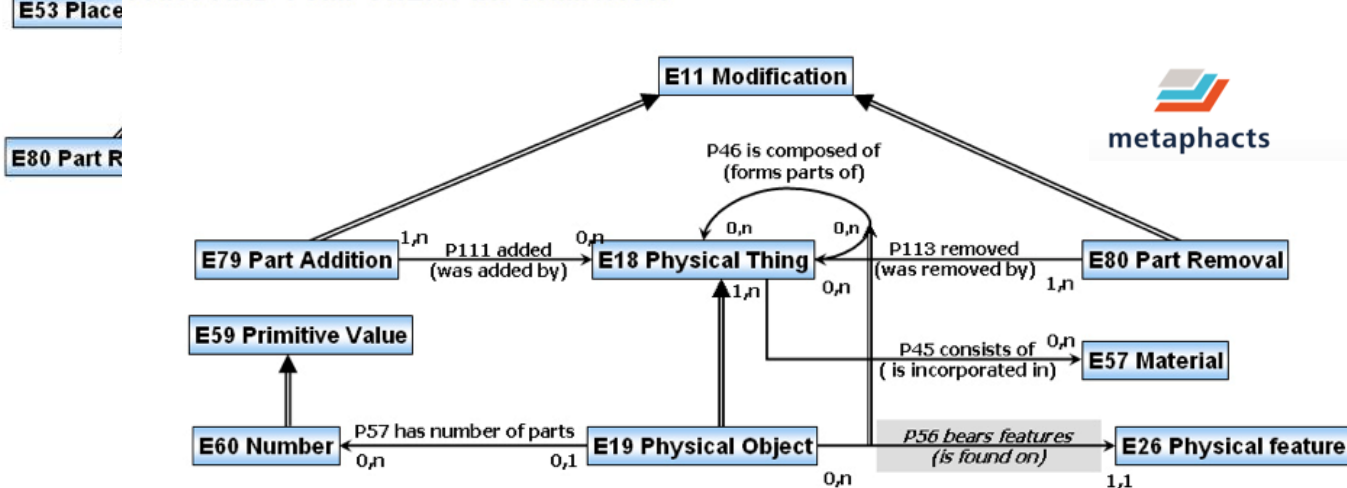
## Spatial – Temporal Relationships



## COLLECTION INFORMATION



## PART AND COMPONENT INFORMATION



- Resources are typed into classes in rich poly-hierarchies:



object/YCA62958

[crm:E18\\_Physical\\_Thing](#)

[crm:E71\\_Man-Made\\_Thing](#)

[crm:E19\\_Physical\\_Object](#)

[crm:E24\\_Physical\\_Man-Made\\_Thing](#)

[crm:E22\\_Man-Made\\_Object](#)

$\xrightarrow{\text{rdf:type}}$

$\xrightarrow{\text{rdfs:subClassOf}}$

metaphacts

RESEARCH  
SPACE

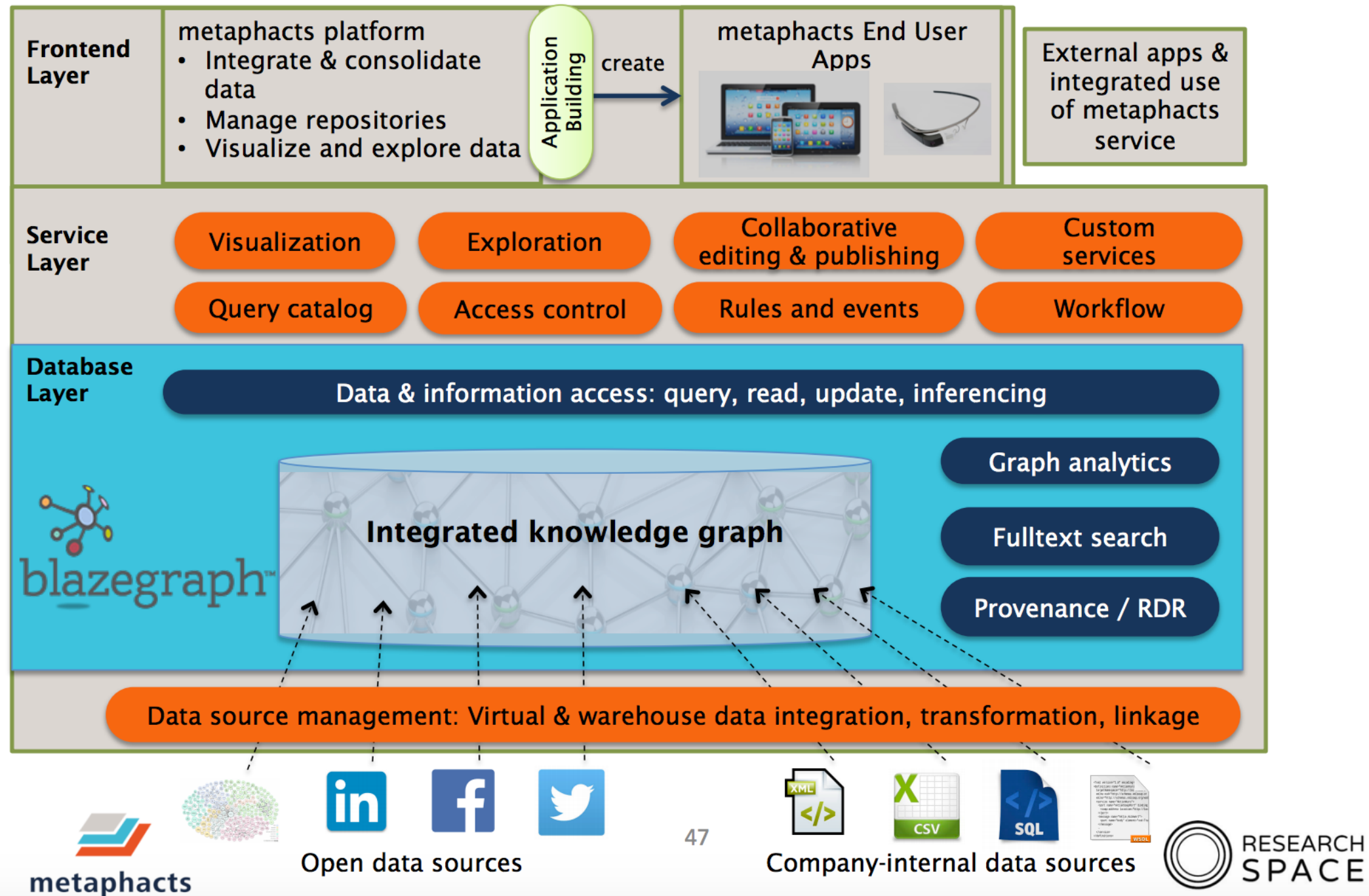
# 大英博物馆：语义搜索

The image displays the ResearchSpace beta interface, which is a platform for semantic search and data exploration. The interface is divided into several sections:

- Header:** Includes the ResearchSpace logo, a 'beta' badge, and navigation links for 'Example Records', 'Sparql', and 'Account'.
- Home Section:** Features a 'Clipboard' on the left with a search bar and a list of folders (Sloane, edo, my actors, france, Lowestoft, Fuji). Below the folders are two portrait images of Sir Joseph ... and Hendrik Go....
- Search Interface:** A central area for finding places. It includes a 'Find a Place' dropdown menu with a search bar and a 'cancel' button. The dropdown shows a hierarchical list of locations: 'france' (selected), 'Ile de France Mauritius - archaic', 'Americas', 'North America', 'Canada', 'Ontario', 'Fort Frances', 'Europe', 'Gaul', 'France', 'Ile-de-France region', 'Paris France', and 'Seine-et-Marne'. 'Cancel' and 'Select' buttons are at the bottom.
- Results Section:** Displays search results for 'Find: Things REFERS TO Vesuvius, Mount'. It shows a list of 625 matches. The results are presented in a grid of image thumbnails, each with a caption. The captions include: 'Cup and Saucer, gold (Fran...', 'Eruption of the Soufrière o...', 'Vue du mont Vézouze tel qu...', 'A Napoli, paper (1849, 103...', 'Prospetto del Vesuvio nel F...', 'Montis Vesuvii juxta Neapo...', 'Montis Vesuvii Saucos et vi...', 'The bay of Naples with a m...', 'Lazzaroni and androm Group...', 'An eruption of Mount Vesu...', 'Bay of Naples with Mount ...', 'House of Pansa, paper (18...', and 'Street Scene in the City of ...'. There are also two more thumbnails at the bottom: 'Tombs of the Calventus Q...' and 'View up the Street of Tombs...'. The interface includes a 'Chart Table Grid' toggle and a 'Download CSV' button.
- Footer:** Contains the British Museum logo, copyright information, and a note that ResearchSpace is funded by the Andrew W. Mellon Foundation. It also mentions 'Platform powered by Metaphacts'.

# 大英博物馆：后端平台

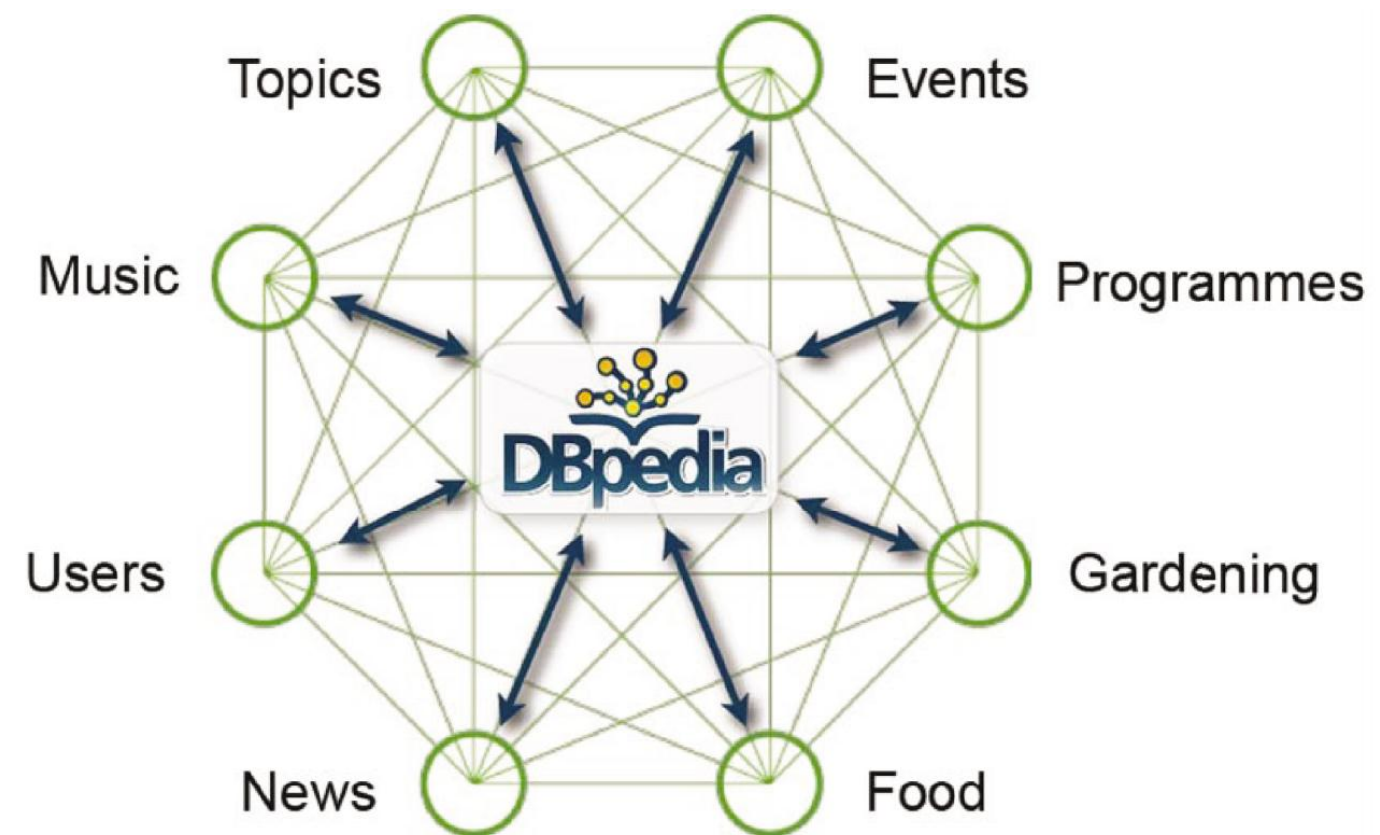
## metaphacts platform



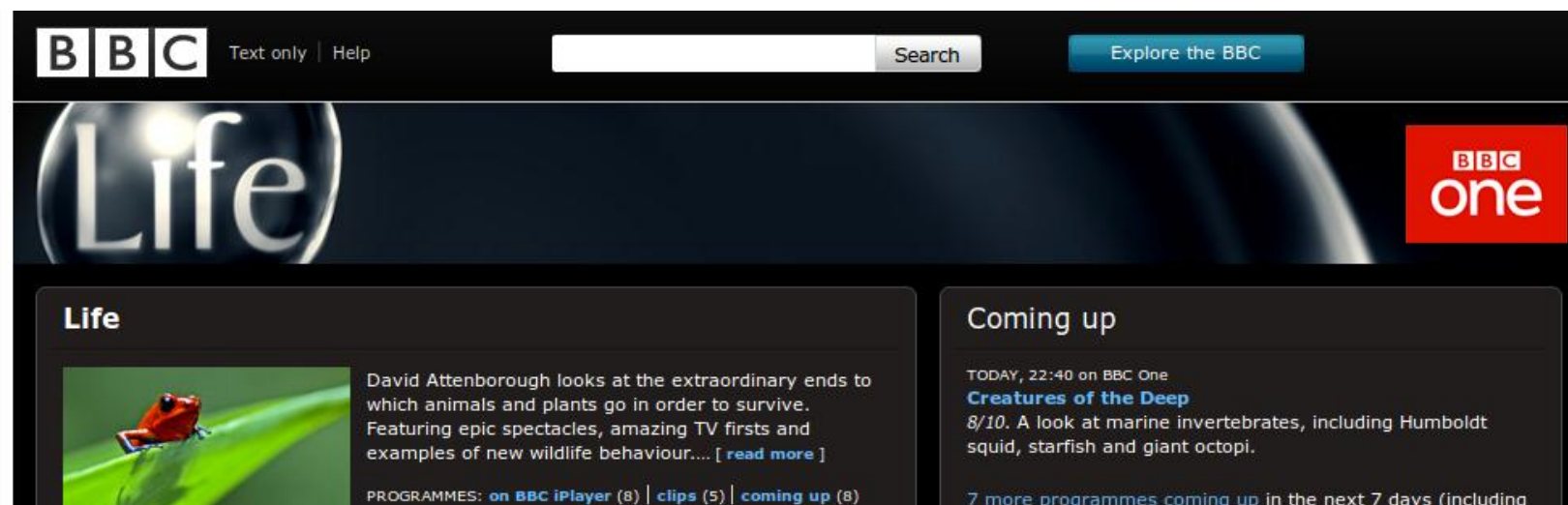
# BBC: 打通BBC全领域

## BBC ONTOLOGIES

- Programmes Ontology
- Wildlife Ontology
- Sport Ontology
- Curriculum Ontology
- Corenews Ontology
- Storyline Ontology



# BBC: Semantic Programme



## THE WEB SITE IS THE API

```
-<rdf:RDF>
- <rdf:Description rdf:about="/programmes/b00vk2lp.rdf">
  <rdfs:label>Description of the brand "Luther"</rdfs:label>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-10-08T14:41:01+01:00</dcterms:created>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-07-11T11:50:55+01:00</dcterms:modified>
  <foaf:primaryTopic rdf:resource="/programmes/b00vk2lp#programme"/>
</rdf:Description>
- <po:Brand rdf:about="/programmes/b00vk2lp#programme">
  <po:pid>b00vk2lp</po:pid>
  <dc:title>Luther</dc:title>
- <po:short_synopsis>
  Crime drama series starring Idris Elba as DCI John Luther.
</po:short_synopsis>
- <po:medium_synopsis>
  Crime drama series starring Idris Elba as Luther, a near-genius murder detective whose brilliant mind can't always save him from the dangerous violence of his passions
</po:medium_synopsis>
<foaf:depiction rdf:resource="http://www.bbc.co.uk/iplayer/images/progbrand/b00vk2lp_512_288.jpg"/>
<po:genre rdf:resource="/programmes/genres/drama/crime#genre"/>
<po:genre rdf:resource="/programmes/genres/drama/thriller#genre"/>
<po:masterbrand rdf:resource="/bbcone#service"/>
<po:series rdf:resource="/programmes/b00s8jk0#programme"/>
<po:series rdf:resource="/programmes/b00vk2mq#programme"/>
<po:series rdf:resource="/programmes/p01b2b2g#programme"/>
<po:clip rdf:resource="/programmes/p01bqtts#programme"/>
<po:clip rdf:resource="/programmes/p01cgbsj#programme"/>
<po:clip rdf:resource="/programmes/p01c6fvk#programme"/>
<po:clip rdf:resource="/programmes/p01bzlgm#programme"/>
<po:clip rdf:resource="/programmes/p01bqg6y#programme"/>
<po:clip rdf:resource="/programmes/p011wxy8#programme"/>
<po:clip rdf:resource="/programmes/p0109njf#programme"/>
</po:Brand>
</rdf:RDF>
```

BBC Programmes is to provide a web identifier, with associated HTML pages and machine-readable feeds (RDF/XML, JSON and XML), for every programme the BBC broadcasts—allowing other teams within the BBC to incorporate those pages into new and existing programme support sites, TV Channel and Radio Station sites.

# 知识图谱在媒体出版行业应用

## Real Proof: The BBC's World Cup 2010 Web Site



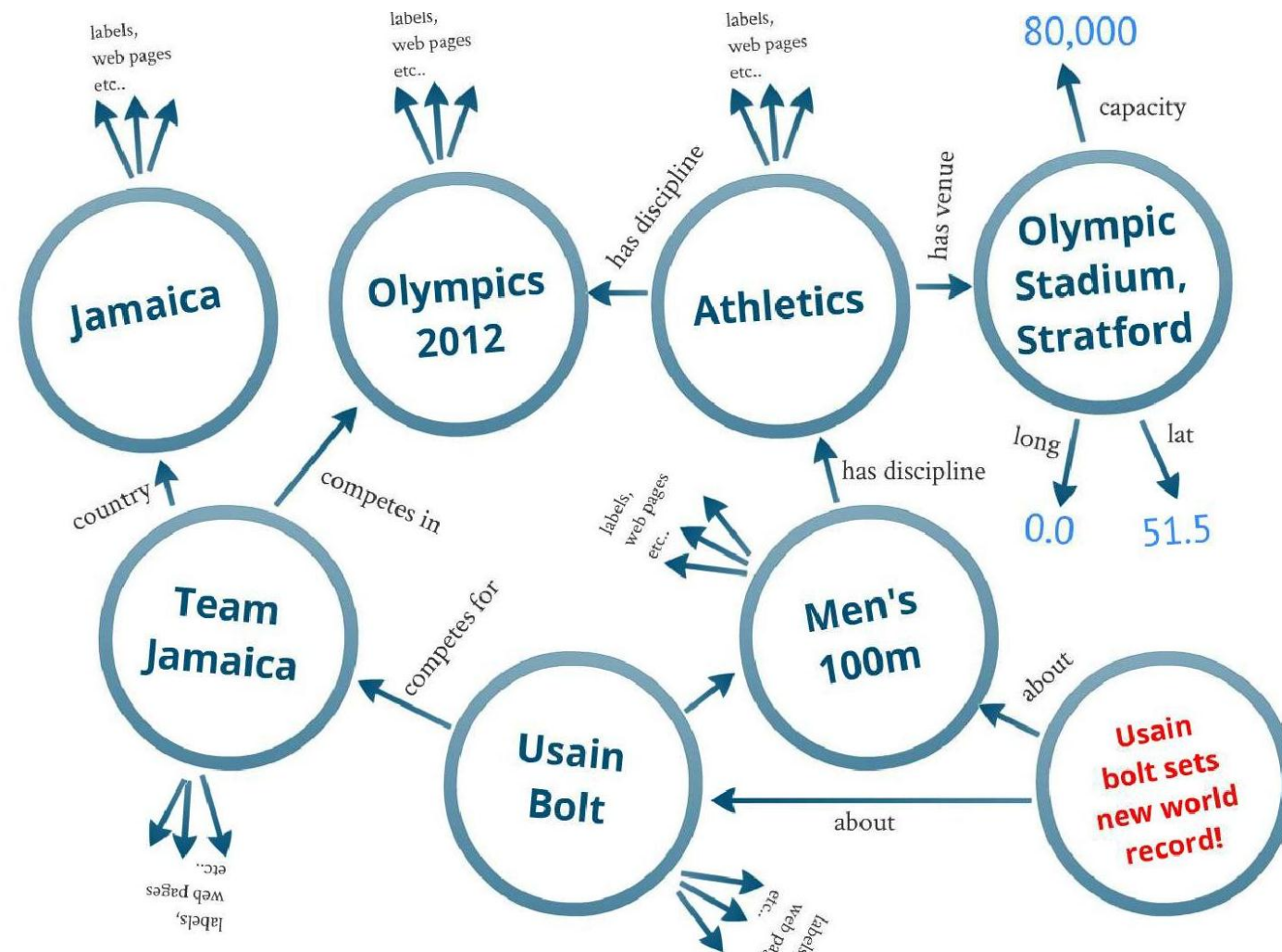
BBC's 2010 World Cup website was the first project running on its Dynamic Semantic Publishing architecture and was deemed a success by both BBC and visitors.

Some statistics about BBC's World Cup web site:

- 800+ Dynamic aggregations/pages (Player, Team, Group, etc.), generated through SPARQL queries
- Average unique page requests/day: 2 million
- Average SPARQL queries/day: 1 million
- 100s repository updates/inserts per minute with OWL 2 RL reasoning
- Multi data center fully resilient, clustered 6 node triplestore

With the DSP Framework, the BBC achieved the desired efficiencies in content creation and curation. As a result they announced that the framework would be expanded initially to the rest BBC Sports section, including their coverage of the 2012 Olympics.

# 知识图谱在伦敦奥运会报道中应用



**BBC** Sign in | Preferences News | Sport | Weather | iPlayer | TV | Radio | More Search the BBC

**SPORT OLYMPICS** The Olympic Broadcaster London 2012

Home | Football | Formula 1 | Cricket | Rugby Union | Olympics | The Ashes | More sports

Countries | Athlete | Britta Steffen | Schedule & Results | Medals | More Olympic sports

### Britta Steffen

**Biography**

**SWIMMING**

Britta Steffen (born 16 November 1983 in Schwedt, Germany) is a German swimmer who specializes in freestyle, who is the current holder of the world record in women's 50 and 100 metre freestyle.

Full biography

**Key facts**

- Country: Germany
- Competes in: Women's 50m Freestyle, Women's 100m Freestyle
- Team: Germany Women's 4 x 100m Freestyle Relay
- Date of birth: 03/03/83
- Height: 1.88m

**WOMEN'S 100M FREESTYLE**

**Britta Steffen sets new record 100m**

Jamaican sprint superstar Usain Bolt claims yet another Olympic record with a stunning victory in the 100m heats.

**Olympic schedule**

**414 days to go**

Olympic Games 27 July-12 Aug 2012

View full olympic schedule

**Olympic venues**

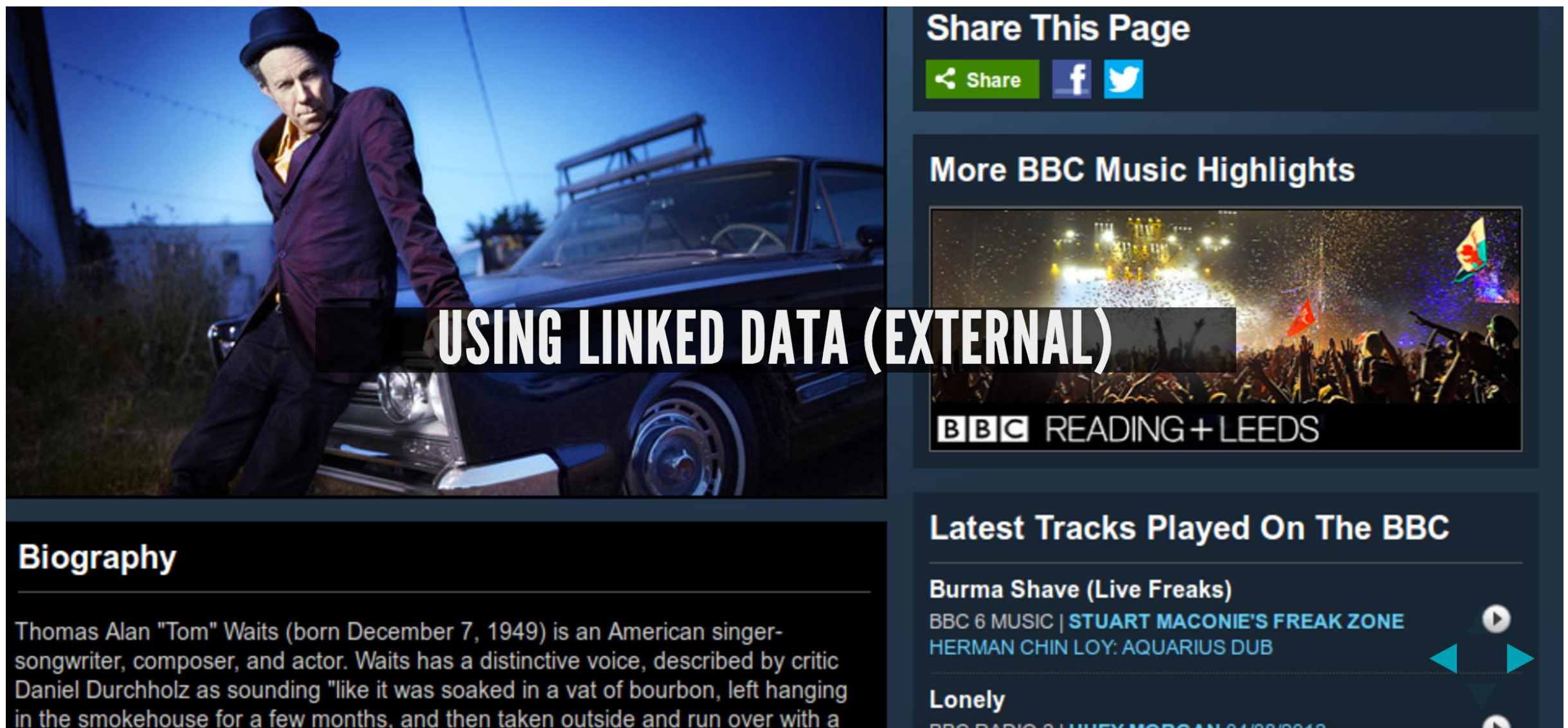
From the 27th of July 2012, 17,000 athletes will compete in 31 venues in the Olympic Park, across London and at out of London venues

View all London 2012 venues

**Olympic 2012 programmes**

# BBC: BBC Music-Using Linked Data (External)

BBC Music is underpinned by the Musicbrainz music database and Wikipedia, thereby linking out into the Web as well as improving links within the BBC site. BBC Music takes the approach that the Web itself is its content management system. Our editors directly contribute to Musicbrainz and Wikipedia, and BBC Music will show an aggregated view of this information, put in a BBC context.



The screenshot displays the BBC Music interface for Tom Waits. A large image of Tom Waits in a purple suit and bowler hat is on the left. A semi-transparent black box with the text "USING LINKED DATA (EXTERNAL)" is overlaid on the image. To the right, there are sections for "Share This Page" with social media icons, "More BBC Music Highlights" featuring a concert image and the text "BBC READING + LEEDS", and "Latest Tracks Played On The BBC" listing "Burma Shave (Live Freaks)" and "Lonely".

## Biography

Thomas Alan "Tom" Waits (born December 7, 1949) is an American singer-songwriter, composer, and actor. Waits has a distinctive voice, described by critic Daniel Durchholz as sounding "like it was soaked in a vat of bourbon, left hanging in the smokehouse for a few months, and then taken outside and run over with a

## Share This Page

Share

## More BBC Music Highlights

BBC READING + LEEDS

## Latest Tracks Played On The BBC

Burma Shave (Live Freaks)  
BBC 6 MUSIC | [STUART MACONIE'S FREAK ZONE](#)  
[HERMAN CHIN LOY: AQUARIUS DUB](#)

Lonely  
BBC RADIO 2 | [HUEY MORGAN 04/08/2012](#)

# BBC: Schema.Org

## [BBC One - EastEnders](#)

[www.bbc.co.uk/programmes/b006m86d](http://www.bbc.co.uk/programmes/b006m86d)

Following the residents of Albert Square as their lives intertwine with drama and tears.

[EastEnders episodes ...](#) - [BBC EastEnders blog](#) - [Characters](#) - [Weekly Quiz](#)

## [EastEnders](#) - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/EastEnders](http://en.wikipedia.org/wiki/EastEnders)

**EastEnders** is a British television soap opera, first broadcast in the United Kingdom on BBC One on 19 February 1985. **EastEnders** storylines examine the ...

## [EastEnders](#) spoilers and news - Digital Spy

[www.digitalspy.co.uk/soaps/s2/eastenders/](http://www.digitalspy.co.uk/soaps/s2/eastenders/)

1 hour ago – The latest episode spoilers, news and pictures from **EastEnders** on Digital Spy.

[Ben and Lola sleep together](#) - [Ben's shocking betrayal is ...](#) - [Alfie is left devastated](#)

## [EastEnders](#) (TV Series 1985) - IMDb

[www.imdb.com/title/tt0088512/](http://www.imdb.com/title/tt0088512/)

★★★★★ Rating: 5.2/10 - 1239 votes

The everyday lives of the inhabitants of Albert Square in the East-End of London. Directed by [Dermot Boyd](#), [Tom Kingdon](#). Starring [Steve McFadden](#).

## [News for eastenders](#)



[Shakey hits the Square: British Superbike star visits the set of Eastenders](#)

[Mirror.co.uk](#) - by [Simon Head](#) - 1 day ago

## EastEnders



[thedrum.co.uk](http://thedrum.co.uk)

EastEnders is a British television soap opera, first broadcast in the United Kingdom on BBC One on 19 February 1985. [Wikipedia](#)

**First episode:** February 19, 1985

**Adaptations:** [Eastenders](#)

**Writers:** [Simon Ashdown](#), [Tony Jordan](#)

**Program creators:** [Julia Smith](#), [Tony Holland](#)

**Genres:** [Soap opera](#), [Drama](#)

SCHEMA.ORG

## Cast



[Jessie Wallace](#)  
Kat Slater



[Lacey Turner](#)  
Stacey Slater



[Shane Richie](#)



[Pam St. Clement](#)



[Letitia Dean](#)

## People also search for



---

# 谢谢大家！

本课程课件由OpenKG提供支持

# 疑问

---

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回答问题

# 联系我们

---

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop

