

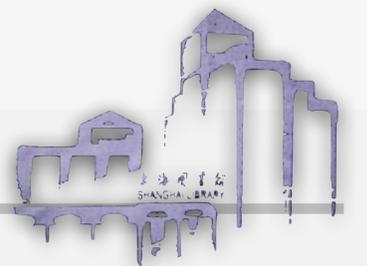
开放数据竞赛巡讲 · 上海大学站 · 210616



# 知识图谱与智慧图书馆服务

**KNOWLEDGE GRAPH AND SMART LIBRARY SERVICES**

刘炜 [wliu@libnet.sh.cn](mailto:wliu@libnet.sh.cn) 上海图书馆上海科学技术情报研究所



---

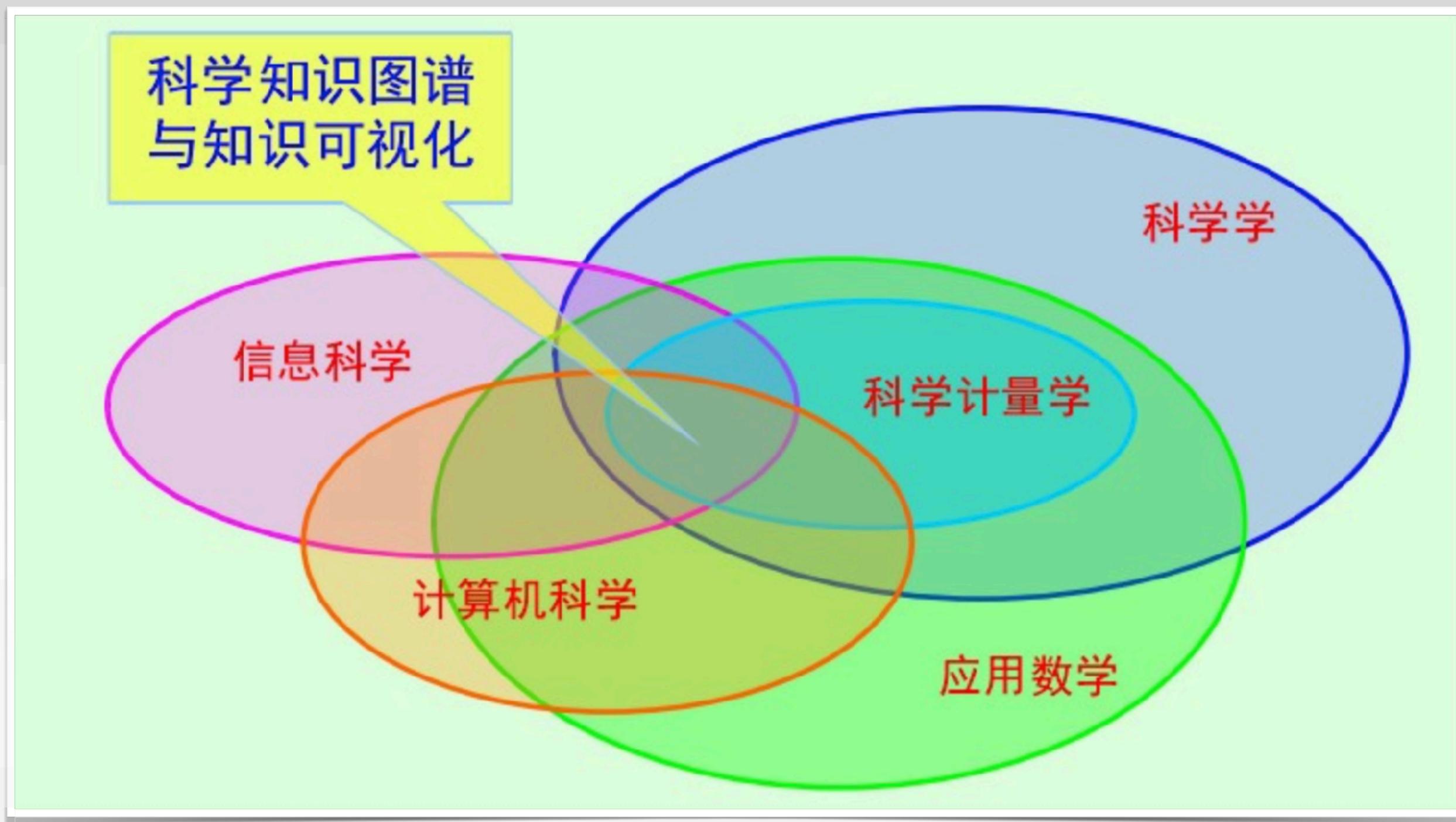
# 大纲

1. 揭秘知识图谱：三种知识图谱
  2. 知识图谱与图数据库的比较
  3. 知识图谱的能与不能
  4. 以知识图谱助力智慧服务
-





# 作为文献/知识/网络计量可视化的知识图谱

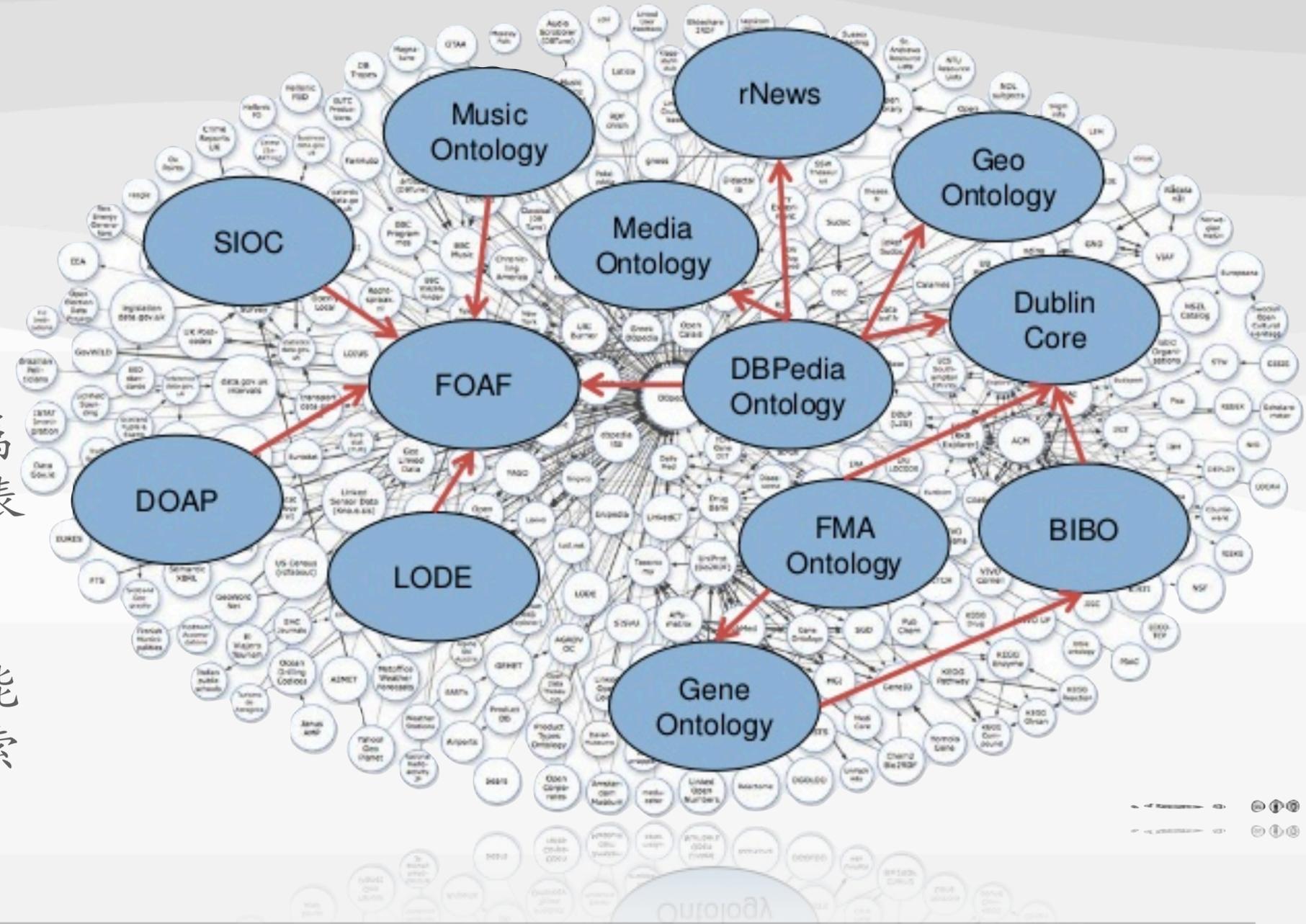




# KNOWLEDGE GRAPH

谷歌于2011/12年提出的一种在万维网上编码并关联碎片化知识单元的一种方案 (Schema.org), 该方案的实施目的和结构被谷歌称为“KG(知识图谱)”。

KG要求以RDF三元组模型表达实体-属性和属性值(statement), 推荐以规范的词表模式(即schema.org)描述各类事物(人、地、事件等), 以Microdata、RDFa、Jason-LD等方式进行三元组编码, 使相关语义信息能够包含于网页之中并相互关联, 并支持搜索引擎进行知识发现、索引以及可视化呈现。



# J.Crew

Specialty retailer company

J.Crew Group, Inc., is an American multi-brand, multi-channel, specialty retailer. The company offers an assortment of women's, men's and children's apparel and accessories, including swimwear, outerwear, ... [Wikipedia](#)

Customer service: 1 (800) 562-0258

Headquarters: New York City, NY

CEO: Mickey Drexler

Founder: Emily Scott

Founded: 1983



Logo: can be specified by using Organization Schema Mark-up

Company type: can be influenced by Wikidata or Wikipedia

Company details: can be influenced by Wikidata, Wikipedia, organization and local business schema mark-up

## Profiles



Facebook



Instagram



Twitter



LinkedIn



Google+

Social profiles: can be influenced by organization schema mark-up with social links specified

## Recent posts on Google+



J.Crew

1,488,424 followers • Shared publicly



Because THIS is the summer you finally learn how to surf. And THESE are the board shorts you'll be wearing when you catch the first wave.

<http://jcrew.co/LVNQ1> ... 6 hours ago

Google + feed: can be influenced by Rel Publisher linking

## People also search for

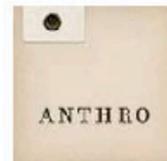
View 15+ more



Banana Republic



Nordstrom



Anthropol...



Ann Inc.



Vineyard Vines

Related companies/brands: cannot be influenced, entirely controlled by Google

# SCHEMA.ORG 举例

- Person
- Product
- Event
- Organization
- Movie
- Book
- Review

## [Pravir Gupta | LinkedIn](#)

[www.linkedin.com/pub/pravir-gupta/2/180/a70](http://www.linkedin.com/pub/pravir-gupta/2/180/a70)

San Francisco Bay Area - Software Engineer

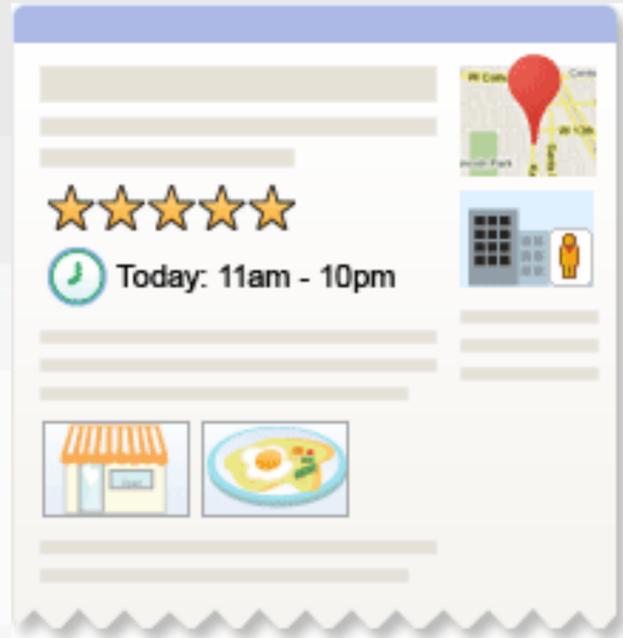
View **Pravir Gupta's** professional profile on LinkedIn. LinkedIn is the world's largest business network, helping professionals like **Pravir Gupta** discover inside ...

## [Office Star Professional Air Grid Deluxe Task Chair | Overst...](#)

[www.overstock.com](http://www.overstock.com) > ... > [Office Furniture](#) > [Office Chairs & Accessories](#) > [Ergonomic Chairs](#)  
- [Cached](#)

★★★★★ 181 reviews - \$189.99

The excerpt from the page will show up here. The reason we can't show text from your webpage is because the text depends on the query the user types.



## [San Francisco metro area, CA Events, Concerts, Festivals,](#)

[eventful.com/sanfrancisco/events](http://eventful.com/sanfrancisco/events) - [Cached](#)

The excerpt from the page will show up here. The reason we can't show text from your webpage is because the text depends on the query the user types.

<a href="#">Ethan Bortnick</a>	Sun, Aug 12	Palace of Fine Arts, San Francisco, CA, CA
<a href="#">Pitbull</a>	Sun, Aug 12	Bill Graham Civic Auditorium, San Francisco, CA, CA
<a href="#">Dead Can Dance</a>	Sun, Aug 12	Greek Theater, Berkeley, CA(San Francisco metro area), CA

## [The Princess Bride \(1987\) - IMDb](#)

[www.imdb.com/title/tt0093779/](http://www.imdb.com/title/tt0093779/)

Block all [www.imdb.com](http://www.imdb.com) results

★★★★★ Rating: 8.1/10 - 166718 votes

A classic fairy tale, with swordplay, giants, an evil prince, a beautiful **princess**, and yes, some kissing (as read by a kindly grandfather).

Directed by [Rob Reiner](#). Starring [Cary Elwes](#), [Mandy Patinkin](#).

[Full cast and crew](#) - [Quotes](#) - [Trivia](#) - [Cary Elwes](#)

## [J.K. Rowling \(Author of Harry Potter and the Sorcerer's ... - Goodreads](#)

[www.goodreads.com/author/show/1077326.J\\_K\\_Rowling](http://www.goodreads.com/author/show/1077326.J_K_Rowling)

**J.K. Rowling** (Joanne Jo Rowling) is the writer behind the best selling Harry ...

Thu, Sep 27 "The Casual Vacancy" Book Release - US

## [Dragon Age: Origins for PC - Dragon Age: Origins PC Game - Dragon ...](#)

★★★★★ Review by GameSpot - Nov 3, 2009

Wii, **Dragon Age**, and Tiger Woods in this GameSpot news update for ... I had a great hope for **Dragon Age** Origins, but, it didn't turn out quite what I think ...

[www.gamespot.com/pc/.../dragonage/index.html](http://www.gamespot.com/pc/.../dragonage/index.html) - 21 hours ago - [Cached](#) - [Similar](#)

# SCHEMA.ORG编辑器

## Schema Creator for Event

[schema.org/Event](http://schema.org/Event)

Type

Sports

Name

Four Hills Tournament in Innsbruck

URL

<http://www.bergiselspringen.at/>

Description

The third competition in the Four Hills Tournament traditionally takes place at Innsbruck's B

Start date

2013-11-27 11:30

End date

2014-01-04

Duration

Address

Bergisel Stadion Olympiaschanze

PO Box

City

Igls

## Four Hills Tournament in Innsbruck

The third competition in the Four Hills Tournament traditionally takes place at Innsbruck's Bergisel stadium, with its new ski jump built in 2002 and designed by London architect, Zaha Hadid. The futuristic-looking tower with the elegant curved approach run is a modern architectural symbol of the Tyrolean provincial capital, its great sporting significance is made clearly visible at the Bergisel ski jumping competition. On 4th January 2014, the world's top ski jumpers will meet at the Bergisel stadium and millions of television viewers will follow the jumping from the 134 metre high ski-jump live. You can check out the training and qualifying jumps for the competition on 3 January.

Starts: 11/27/2013 11:30AM

Ends: 2014-01-04:00.000

Bergisel Stadion Olympiaschanze

Igls , Tirol

A-6080

Austria

```
<div itemscope itemtype="http://schema.org/SportsEvent">
```

```
<a itemprop="url" href="http://www.bergiselspringen.at/"><div itemprop="name"><strong>Four Hills  
Tournament in Innsbruck</strong></div>
```

```
</a>
```

```
<div itemprop="description">The third competition in the Four Hills Tournament traditionally takes  
place at Innsbruck's Bergisel stadium, with its new ski jump built in 2002 and designed by London  
architect, Zaha Hadid. The futuristic-looking tower with the elegant curved approach run is a modern  
architectural symbol of the Tyrolean provincial capital, its great sporting significance is made clearly  
visible at the Bergisel ski jumping competition. On 4th January 2014, the world's top ski jumpers will  
meet at the Bergisel stadium and millions of television viewers will follow the jumping from the 134  
metre high ski-jump live. You can check out the training and qualifying jumps for the competition on 3  
January. </div>
```

```
<div><meta itemprop="startDate" content="2013-11-27T11:30">Starts: 11/27/2013 11:30AM</div>
```

```
<meta itemprop="endDate" content="2014-01-04:00.000">Ends: 2014-01-04:00.000
```

```
</div>
```

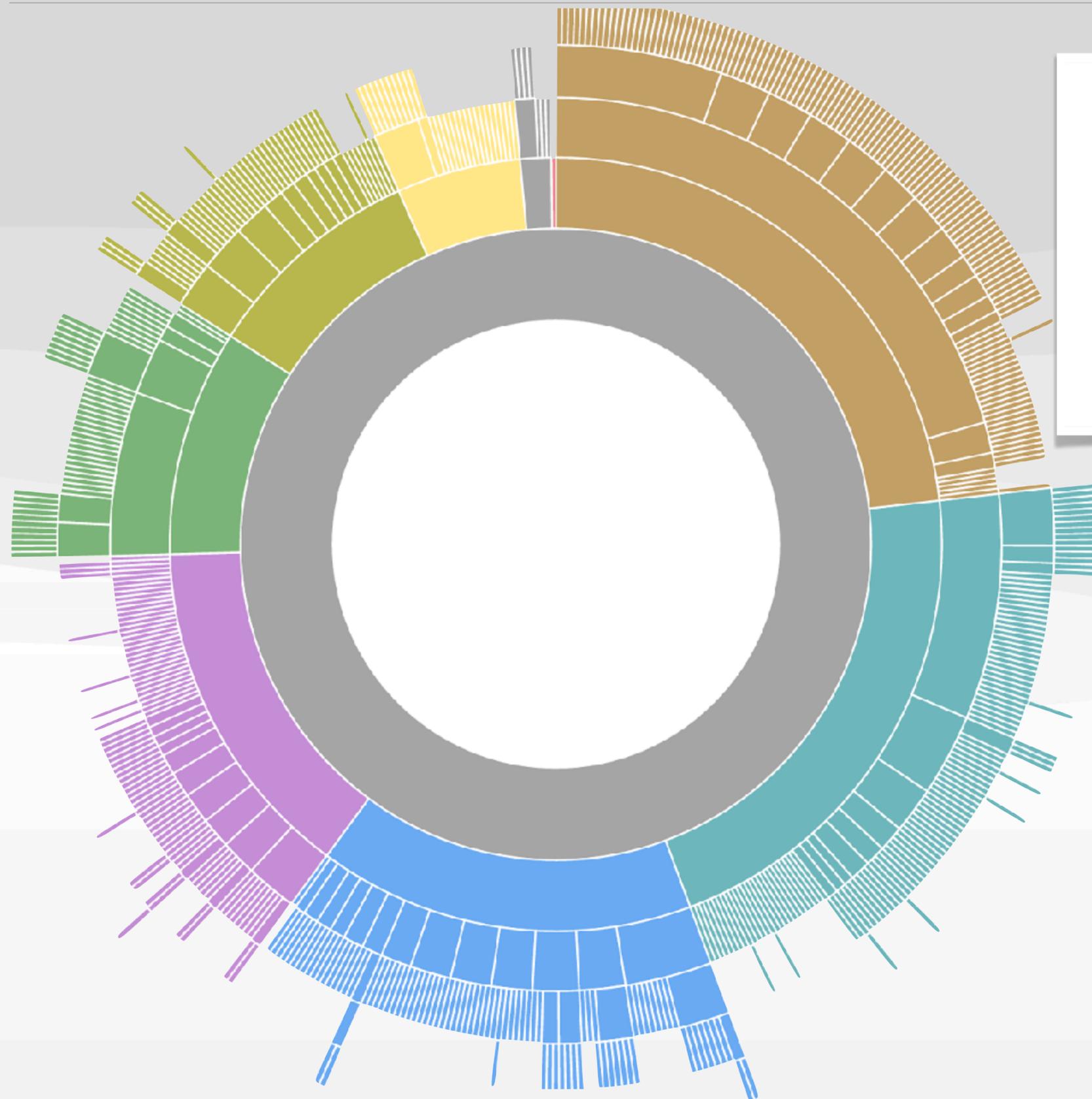
```
<div itemprop="location" itemscope itemtype="http://schema.org/PostalAddress">
```

```
<div itemprop="streetAddress">Bergisel Stadion Olympiaschanze</div>
```

```
<div><span itemprop="addressLocality">Igls </span>, <span  
itemprop="addressRegion">Tirol</span></div><div itemprop="postalCode">A-6080</div>
```

```
<div itemprop="addressCountry">Austria</div>
```

```
</div>
```



- Creative works: [CreativeWork](#), [Book](#), [Movie](#), [MusicRecording](#), [Recipe](#), [TVSeries](#) ...
- Embedded non-text objects: [AudioObject](#), [ImageObject](#), [VideoObject](#)
- [Event](#)
- [Health and medical types](#): notes on the health and medical types under [MedicalEntity](#).
- [Organization](#)
- [Person](#)
- [Place](#), [LocalBusiness](#), [Restaurant](#) ...
- [Product](#), [Offer](#), [AggregateOffer](#)
- [Review](#), [AggregateRating](#)
- [Action](#)

- Person
- Organization
- Place
- CreativeWork
- Intangible
- Action
- MedicalEntity
- Event

## Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

## Thing

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for [data types](#).

Select vocabulary view:

Core vocabulary  Core plus all extensions

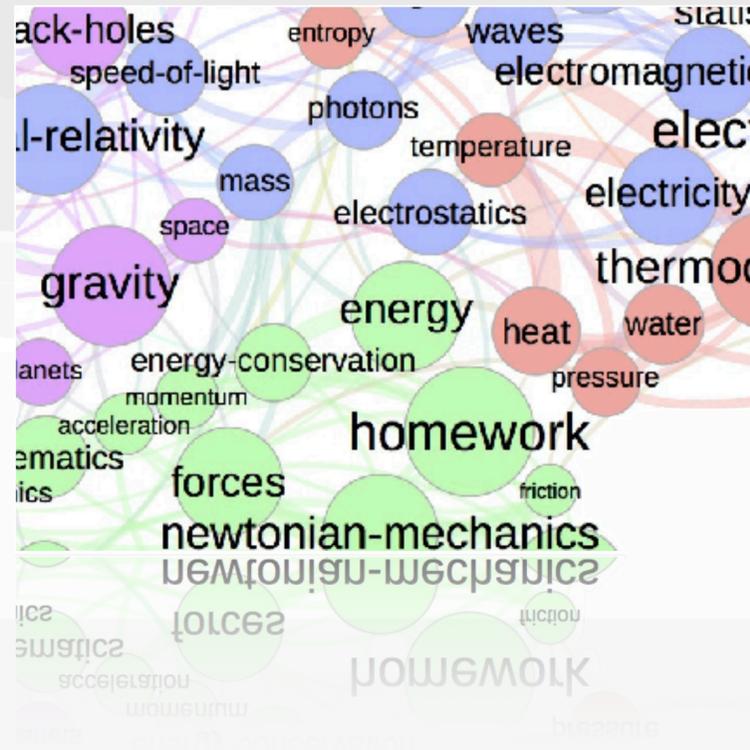
### Core plus extension vocabularies

- [Thing](#)
  - [Action](#)
    - [AchieveAction](#)
      - [LoseAction](#)
      - [TieAction](#)
      - [WinAction](#)
    - [AssessAction](#)
      - [ChooseAction](#)
        - [VoteAction](#)
      - [IgnoreAction](#)
      - [ReactAction](#)
        - [AgreeAction](#)
        - [DisagreeAction](#)
        - [DislikeAction](#)
        - [EndorseAction](#)
        - [LikeAction](#)
        - [WantAction](#)
      - [ReviewAction](#)
    - [ConsumeAction](#)
      - [DrinkAction](#)
      - [EatAction](#)
      - [InstallAction](#)
      - [ListenAction](#)
      - [ReadAction](#)
      - [UseAction](#)
        - [WearAction](#)
      - [ViewAction](#)
      - [WatchAction](#)
    - [ControlAction](#)
      - [ActivateAction](#)
      - [DeactivateAction](#)
      - [ResumeAction](#)
      - [SuspendAction](#)
    - [CreateAction](#)
      - [CookAction](#)
      - [DrawAction](#)
      - [FilmAction](#)

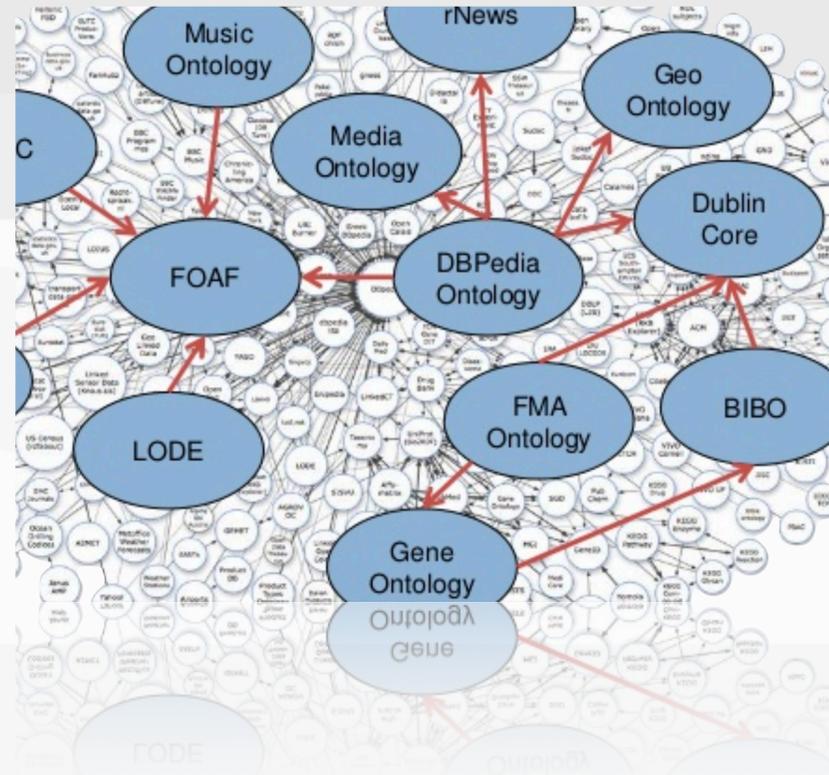
目前核心词表有597类, 875属性和114个取值词表  
<http://schema.org/docs/full.html>

# 三种知识图谱

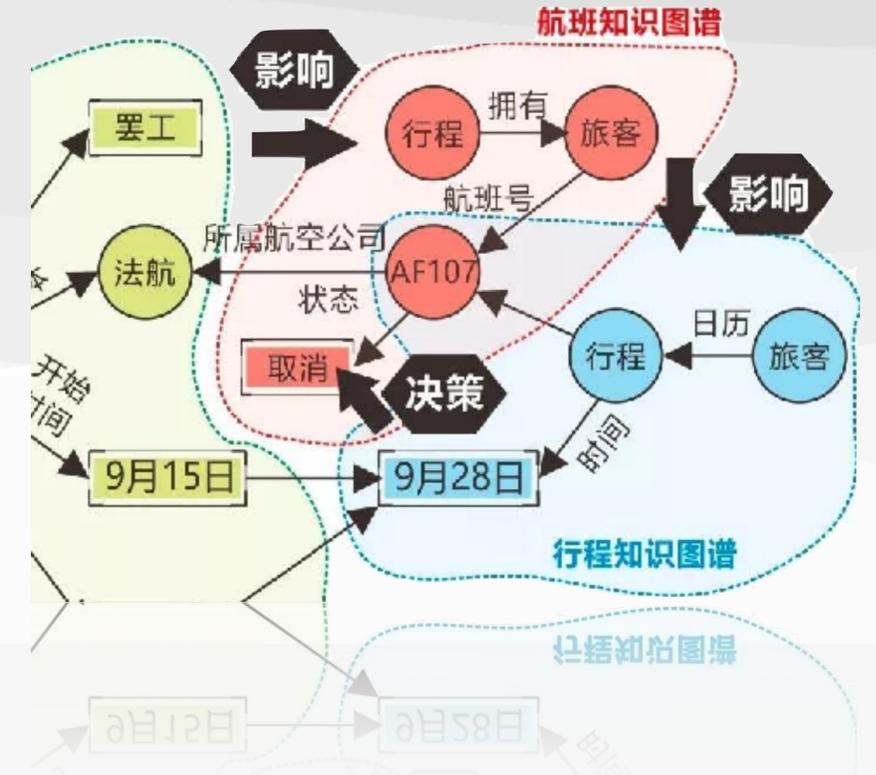
## Knowledge Map



## Knowledge Graph

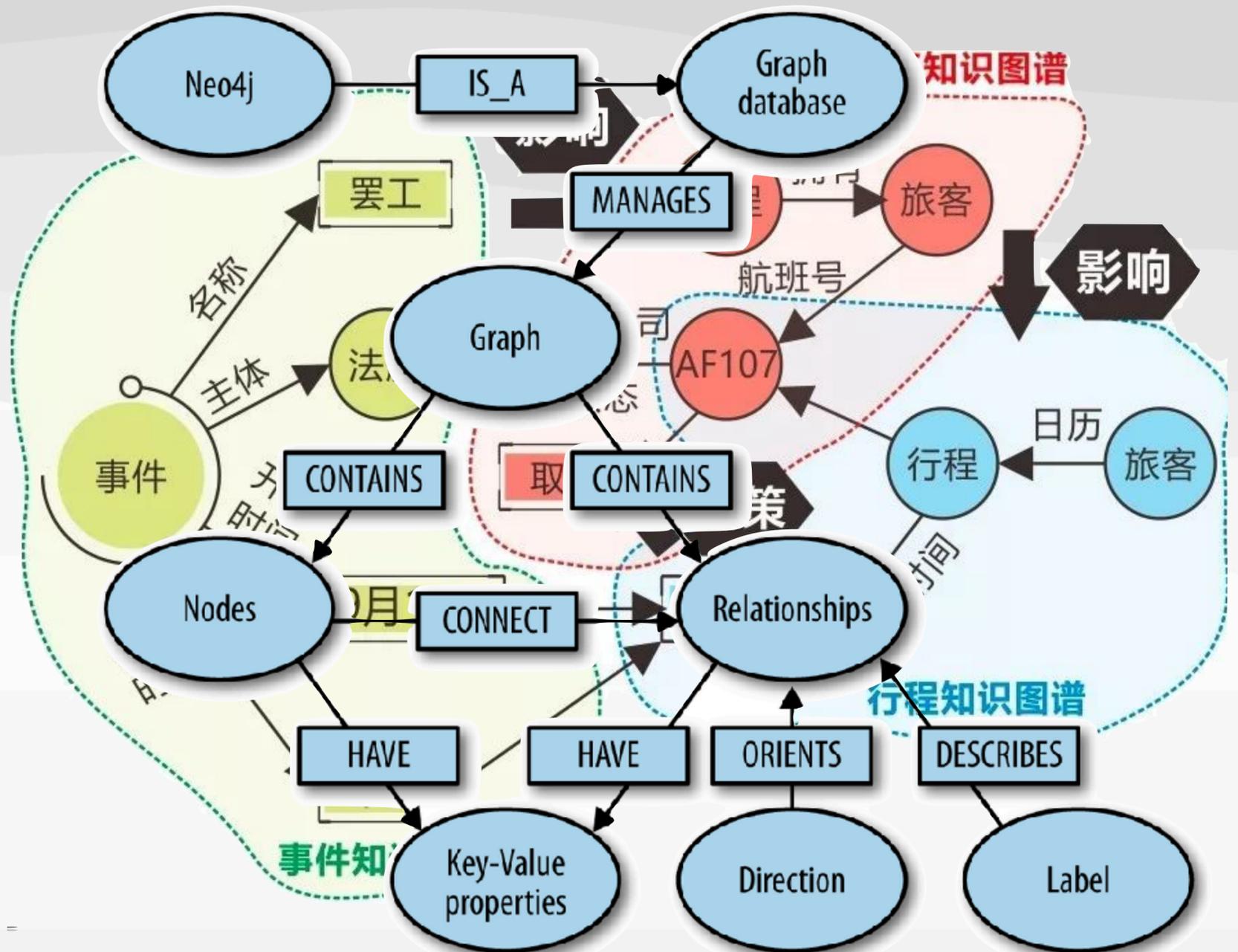


## Graph Database



# GRAPH DATABASE

即图数据库，以图形方式表示节点、属性和关系并进行存储和提供管理功能的数据库，如Neo4j、ArangoDB等，属于NoSQL的一种（其它还有键值对/大表/文档数据库），作为大数据的一种重要支撑技术，用来实现海量RDF数据的管理。



---

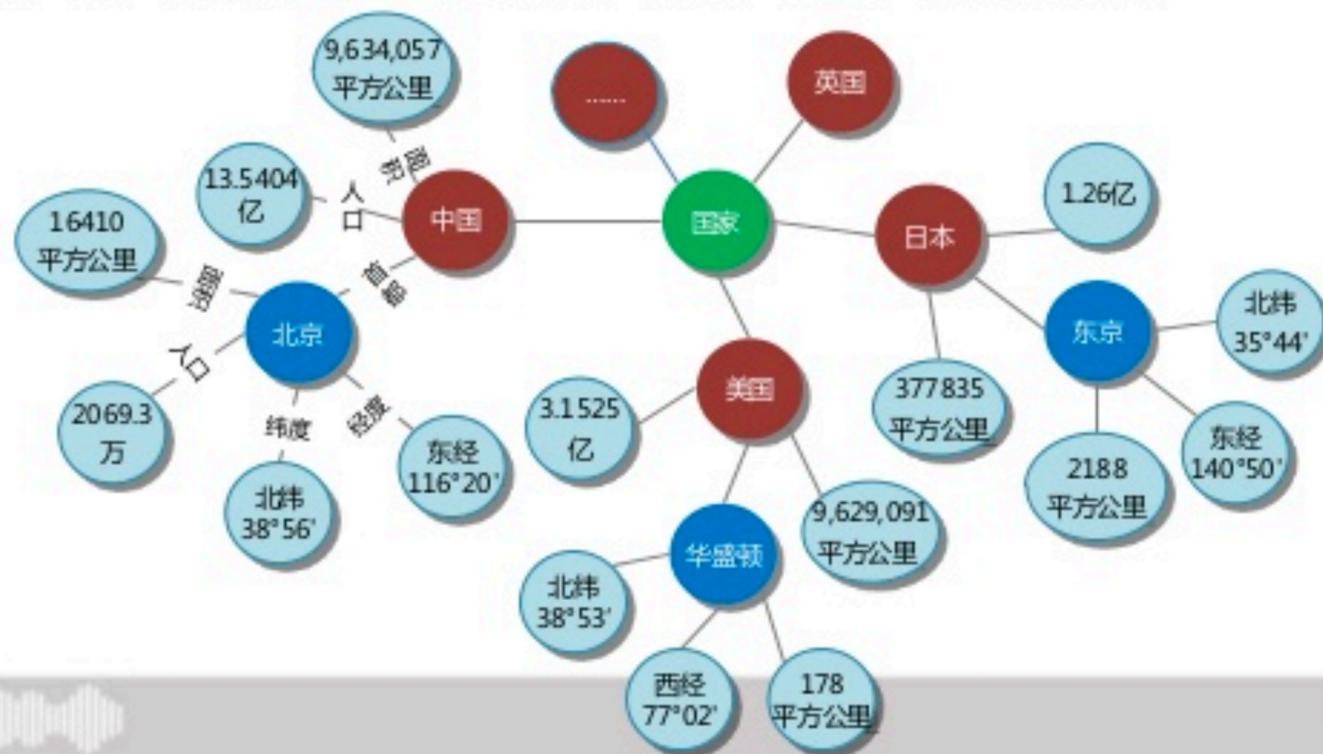
# 知识图谱与图数据库比较

---

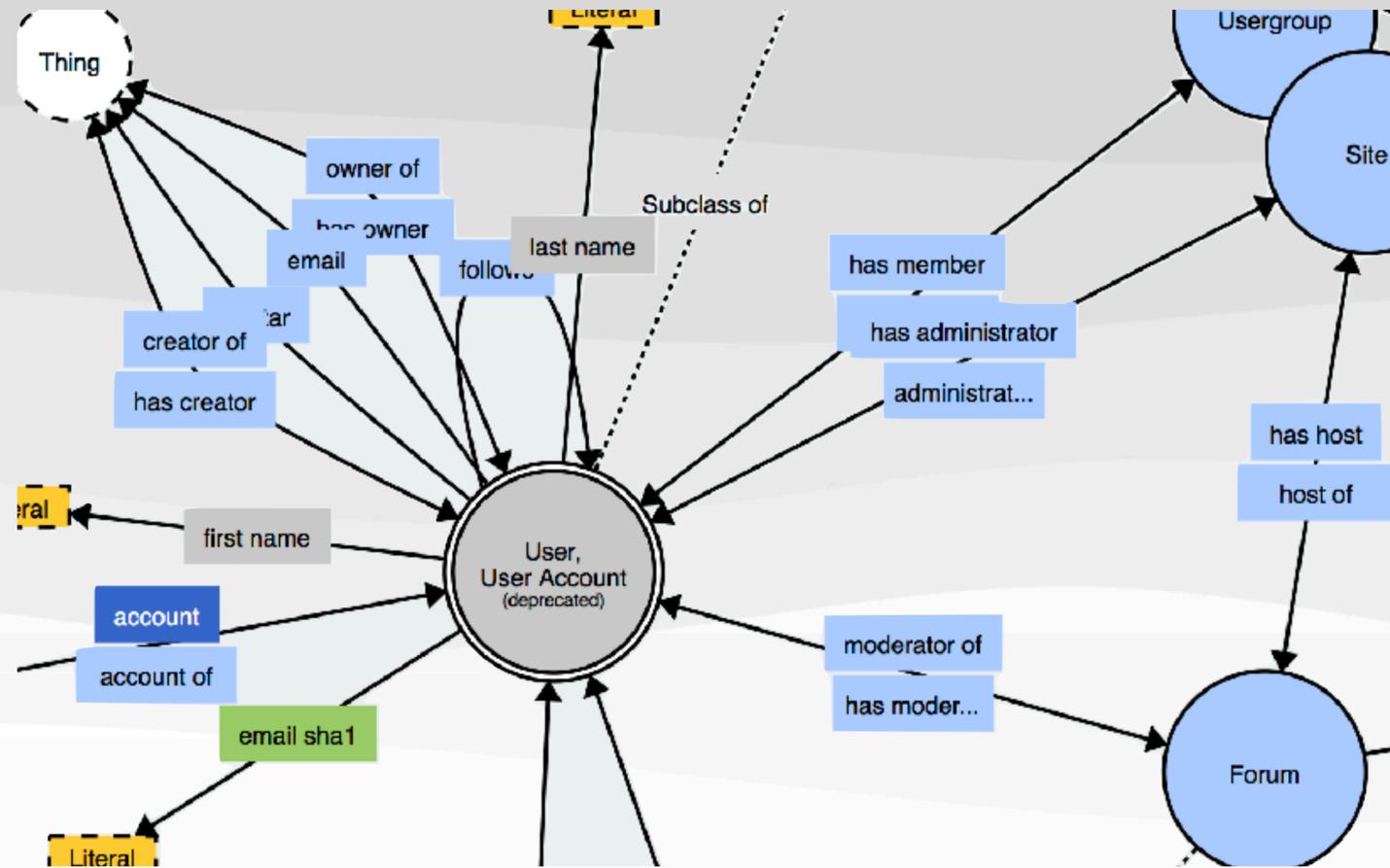
# 一种典型的混淆

## 什么是知识图谱(Knowledge Graph)?

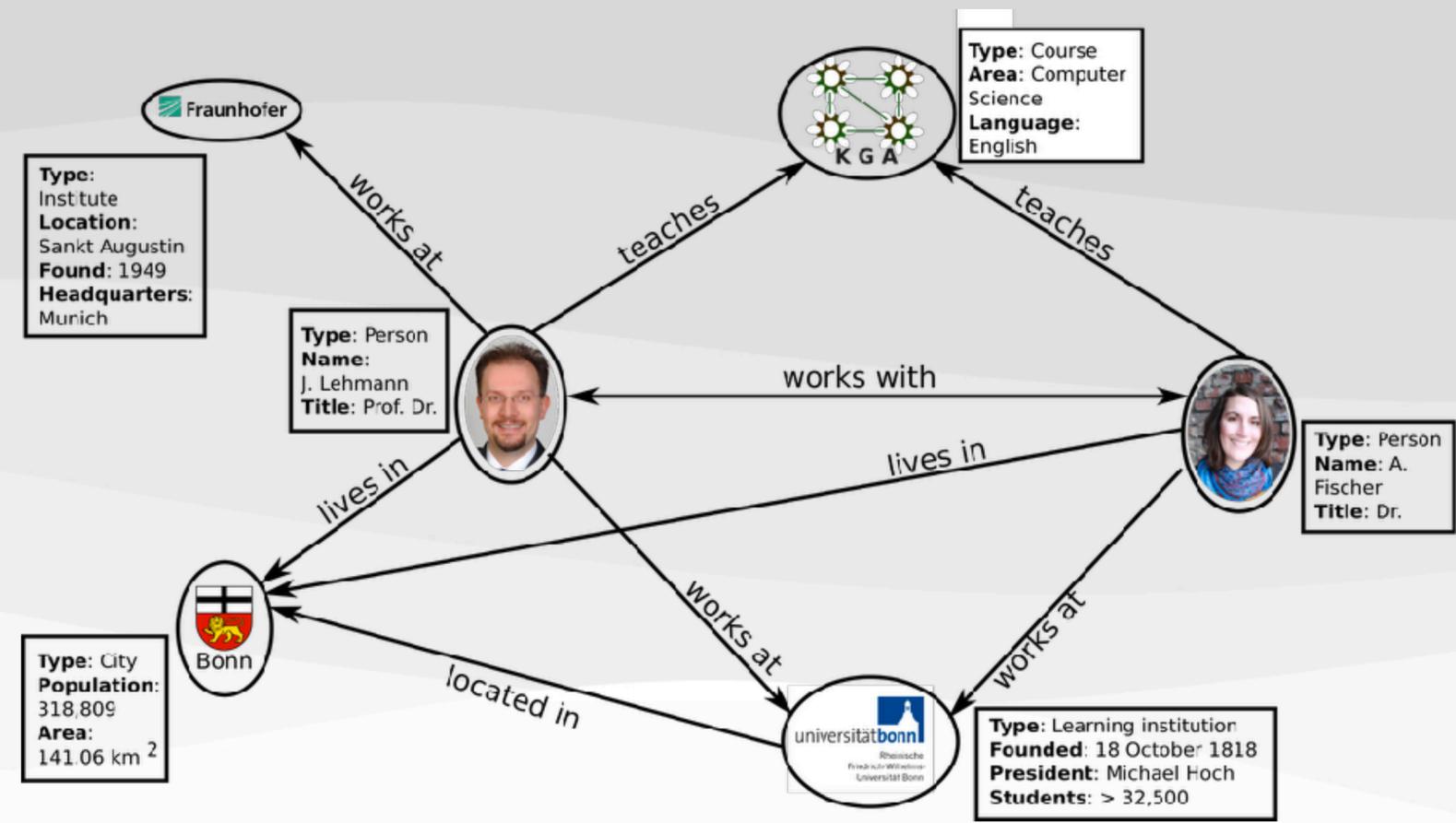
- 知识图谱旨在描述真实世界中存在的各种实体或概念及其关系，一般用三元组表示
- 知识图谱亦可被看作是一张巨大的图，节点表示实体或概念，边则由属性或关系构成



# KG与GD的比较



- 实体-关系(属性)-属性值 (即RDF三元组表示知识单元)
- 实体须以URI命名, 属性和值最好取自有URI的词表
- 本体即概念(词)及关系, 元数据即概念词表, 须显式说明
- 具有标准的SAPRQL查询语言, 可跨域查询
- 可以认为RDF Store是一种特殊的图数据库



- 包含节点和边, 节点上有属性 (键值对) 边也可以有属性
- 边有名字和方向, 并总是有一个开始节点和一个结束节点
- 节点可以有内部结构 (三元组), Neo4j查询用Cypher。
- 解决数据存储和索引问题, 不解决知识表示和全网域服务问题

# KG与GD的比较

## KG (RDF知识图谱)

- 优点：基于描述逻辑的数学基础，规范词表，标准化，工程实现
- 缺点：学习门槛高，高复杂性

## ■ GD (属性图谱)

- 优点：容易学习和实现，特别适用于社交网络，强计算性，效率高
- 缺点：缺乏标准，难以互通，应用中语义模糊

## #1 RDF vs Graph DBs. Apples to apples

*“RDF vs Graph DBs...”*

What are we comparing? Is it models or is it stores?

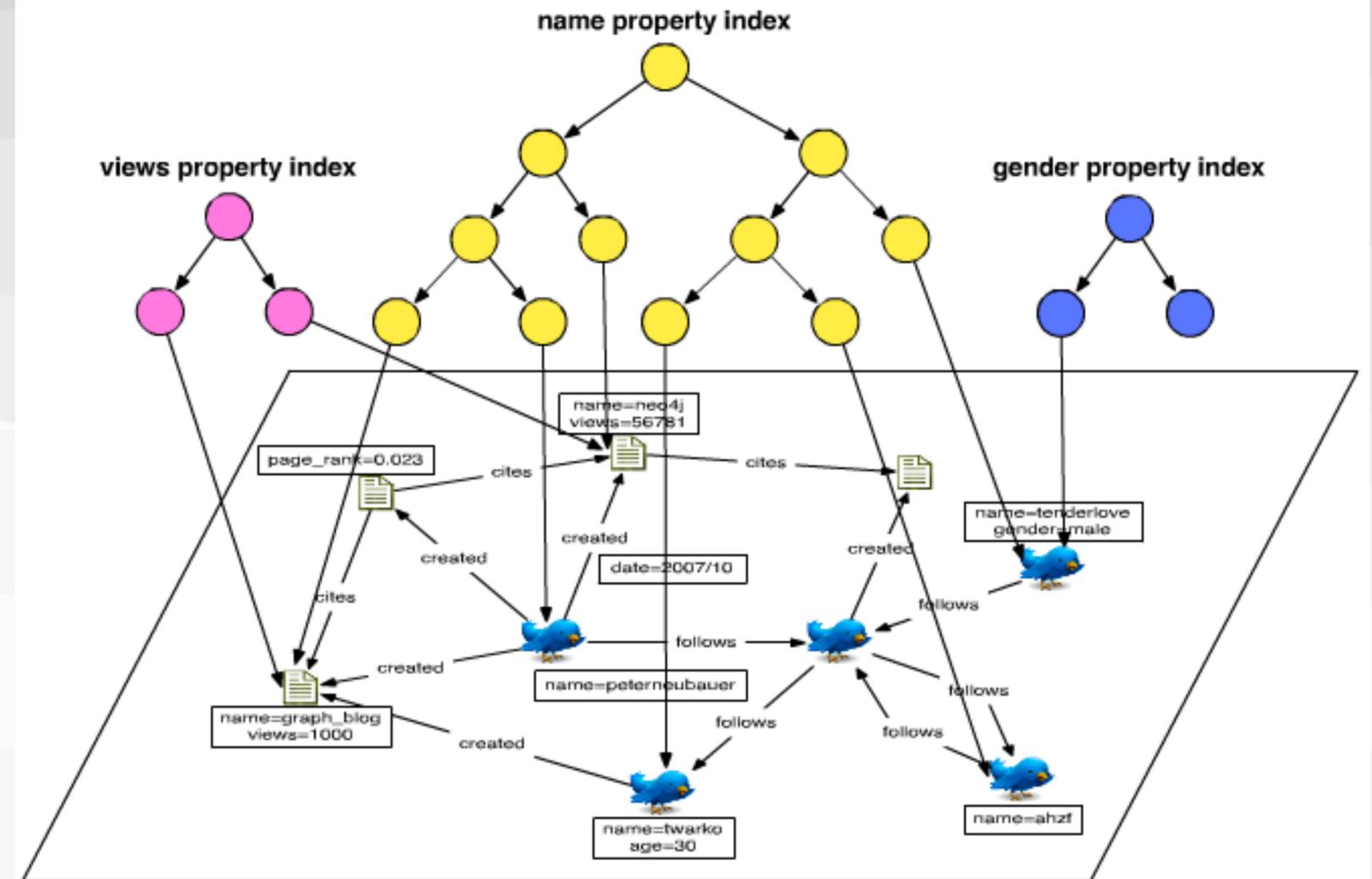
### RDF Stores

- Very strongly index based (some graph DBs are also strongly index based!! Graph as a feature)
- RDFsed on mostly additive, typically slow changing if not immutable data sets

### Native Graph DB (Neo4j)

- Neo4j is Graph native
- Neo4j excels with highly dynamic datasets & transactional UC where data Integrity is key

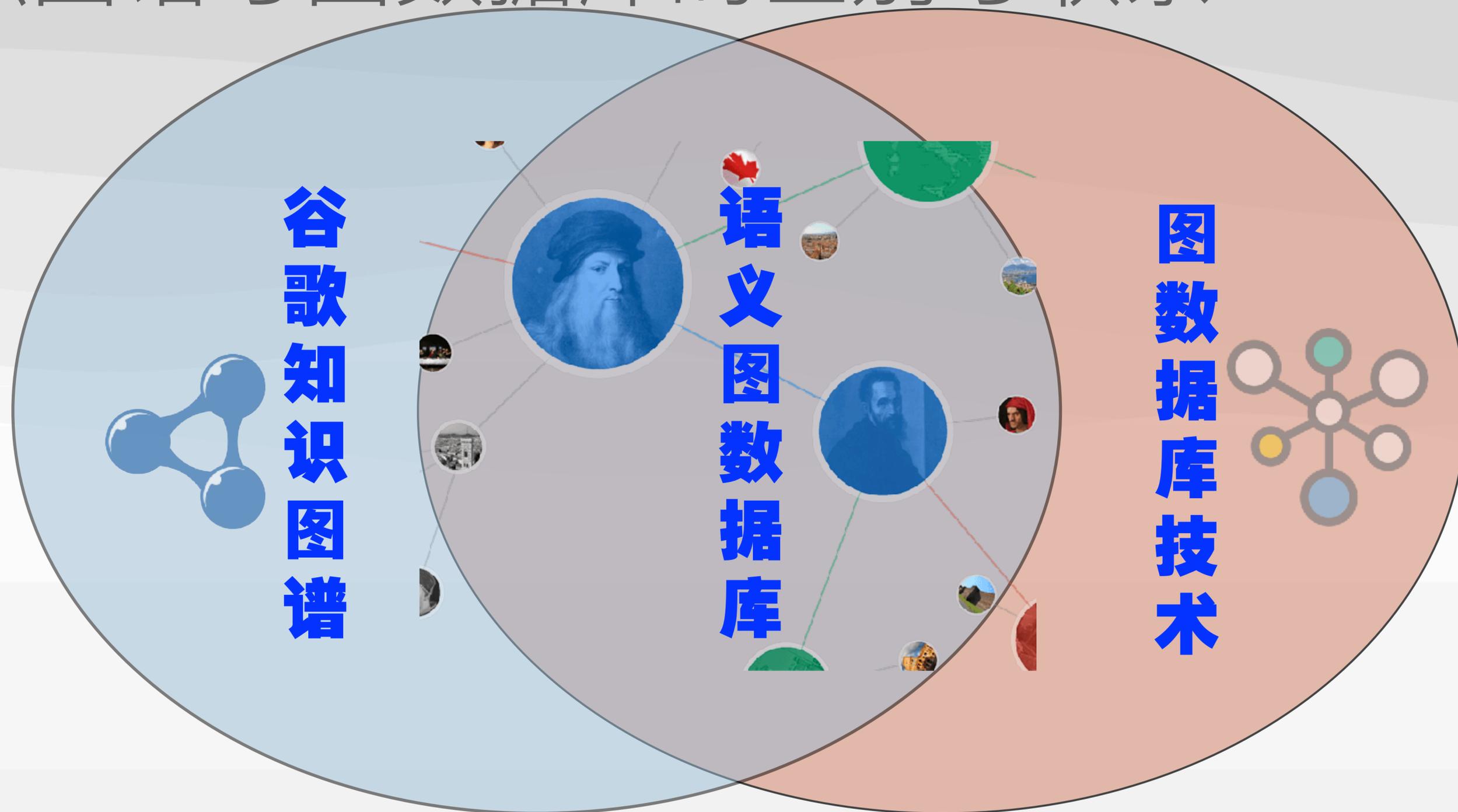
## Graph Databases and Endogenous Indices



# 知识图谱与图数据库的区别与联系

	主要功能	组成	特点	应用	能力	作用域	软件举例	应用举例
<b>知识图谱 KG</b>	编码, 关联	Schema	智慧数据	推理	认知	WWW	Virtuoso	谷歌KG
<b>图数据库 GD</b>	存储, 搜索	Pattern	大数据	挖掘	计算	Data Silo	Neo4j	脸书Social Graph

# 知识图谱与图数据库的区别与联系



---

# 知识图谱的能与不能

---

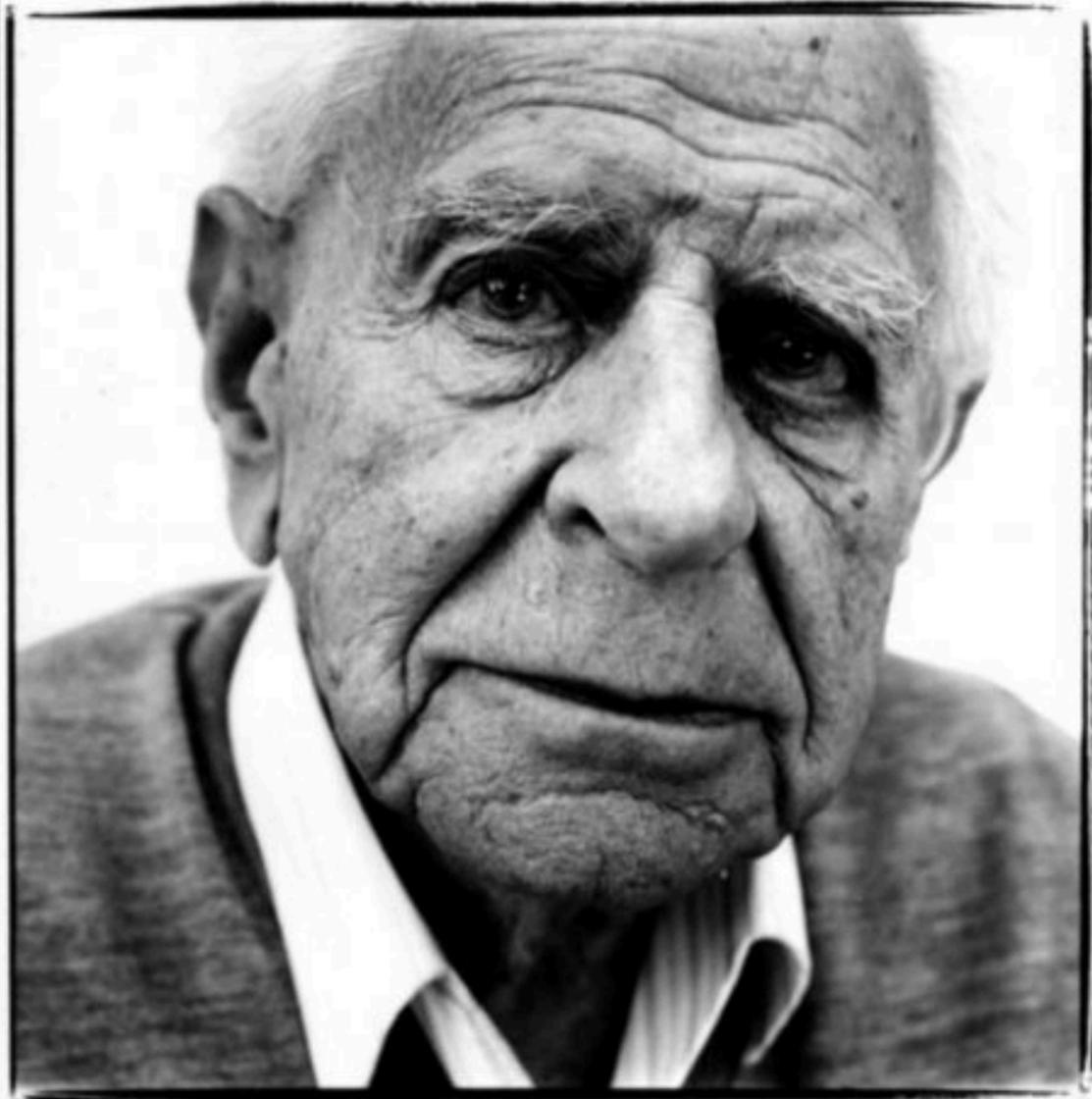
优酷



# 为什么要用知识图谱?

- 辨识网页中的“知识”，描述的细粒度化，帮助搜索引擎定位“知识”
- 更准确的检索，以提高用户的点击兴趣，从而提高CTR（点击率）、降低跳出率
- 促进大规模基于网页的信息融合。如即时信息（如天气、股票、演出等），使得用户直接在搜索页面中获得答案而不必进入信息页面（降低了点击率？但使得真正的用户提高了网站的滞留时间)

# 知识图谱背后的哲学



**If we call the world of “things” -- of physical objects -- the *first world*, and the world of subjective experiences, such as thought processes, the *second world*, we may call the world of statements themselves the *third world*.”**

**- Karl Popper (*Unended Quest: An Intellectual Biography* [p. 211])**









只有回归本源才是最终秩序

Nonlinearity

Decentralization

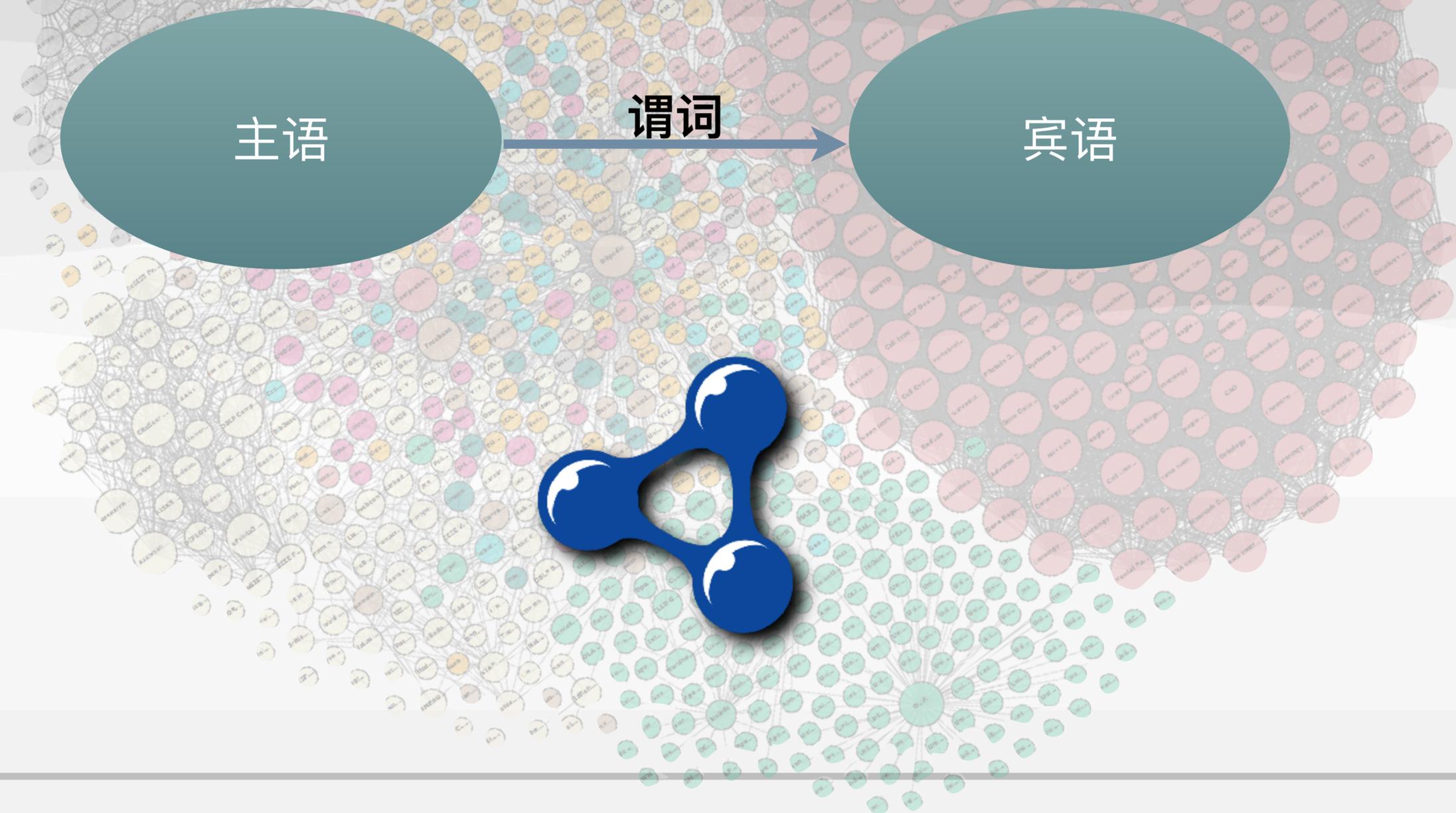
Interconnectedness

Interdependence

Multiplicity



# 只有一种联系是基本和普遍的



---

知识图谱是一种智慧数据

# 智慧数据

——有语义的、自解释的、可计算的、能够行动的信息单元。

## Smart Data

— —self-interpretable, machine understandable and actionable  
formatted information unit with semantic annotations

---

# 知识图谱不是什么？

- 知识图谱不是神经网络，但可以促进机器学习
  - “知识图谱”不总是知识，KG的语义离开本体无法解释（机器理解）
  - 知识图谱不保证正确，可以容错
  - 知识图谱不会创新知识，除非机器学习
-

---

以知识图谱助力智慧服务

---

# 利用知识图谱发布图书馆数据

- 提高图书馆资源的搜索引擎可见度
  - 以W3C标准全网域呈现图书馆各类数据
  - 通过整合融合和与其它数据互操作，使书目数据及其相关模型更有用
    - The Schema Bib Extend W3C Community group
    - OCLC's experiments with SchemaBibEx and BIBFRAME
    - 关注新模型如Linked Archive Model
-

---

# 如何使用知识图谱

- 分析网站发布的内容，选择合适的类型（人地时事），以[schema.org](http://schema.org)规范格式发布
  - 利用诸如[schema-creator.org](http://schema-creator.org)之类的工具
  - 在Wikipedia、Wikidata、Sinopedia、KBpedia等网站上发布内容
  - 在<http://www.google.com/webmasters/tools/richsnippets>上测试
-

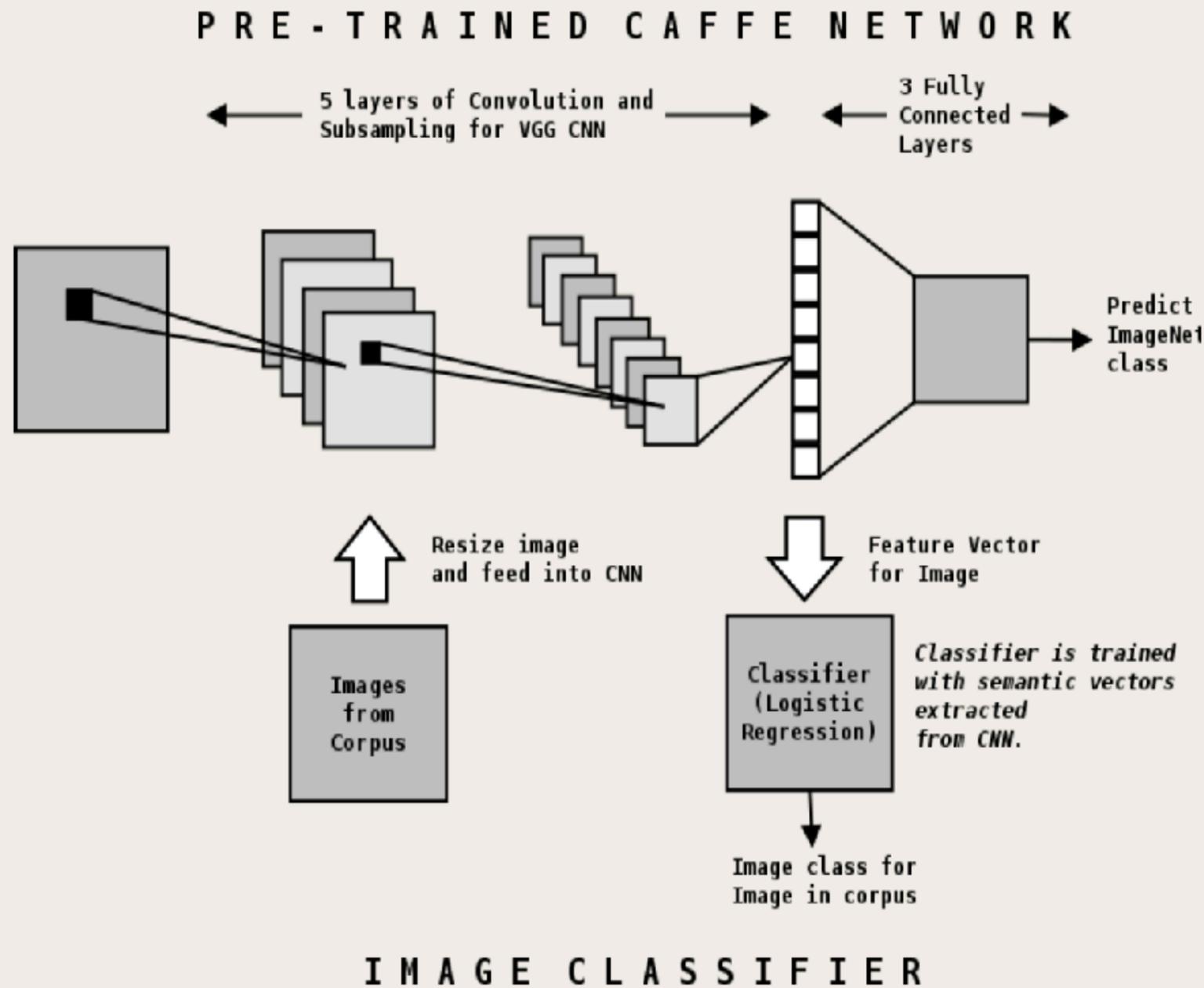
# 新编目：利用知识图谱描述、揭示、组织和计算知识



这些过程，其实就是计算机技术所追求的终极目标。

其中每个环节，没有人工智能无法应用的  
目前相关技术还不够成熟和普及，并且还不具备足够的基础设施条件

# ELSEVIER: DEEP LEARNING FOR IMAGE ANNOTATION



ClinicalKey

English CME Login Register

All Types hypertension Books Journals More

Filter By: Source Type Study Type Specialties Date

34991 results Rate Results

Relevance Subscribed Content

**Hypertension**

Definition

Normal blood pressure (BP) in adults can be defined as systolic BP <120 mm Hg and diastolic BP <80 mm Hg. Prehypertension is defined as systolic BP between 120 and 139 mm Hg or diastolic between 80 and 89 mm Hg. Hypertension can be divided into stage 1: systolic BP from 140 to 159 mmHg and diastolic BP from 90 to 99 mm Hg and stage 2: systolic BP ≥160 mm Hg or diastolic BP ≥100 mm Hg. Measurements should be appropriately taken as described below.

**Selected Content (6)**

- Comparison of Hypertension Guidelines  
Mayo Clinic Proceedings • February 2015
- Standardised effects of a 10 mm Hg reduction in systolic blood pressure stratified by blood pressure  
The Lancet • March 2016
- Evidence-based dosing of antihypertensive medications  
Medical Clinics of North America • July 2015

Was this helpful? Yes or No

# ELSEVIER: MACHINE LEARNING FOR LINK DISCOVERY

diseases 2791370 glaucoma  
 diseases 2791370 glaucoma

have been documented to cause  
 is assessed through  
 progresses more rapidly than  
 recommend  
 supports the assumption that  
 is the death of

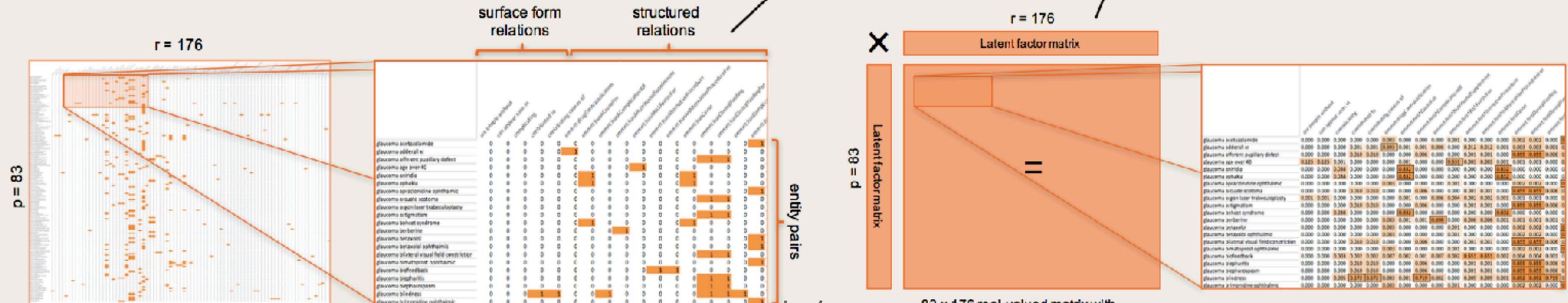
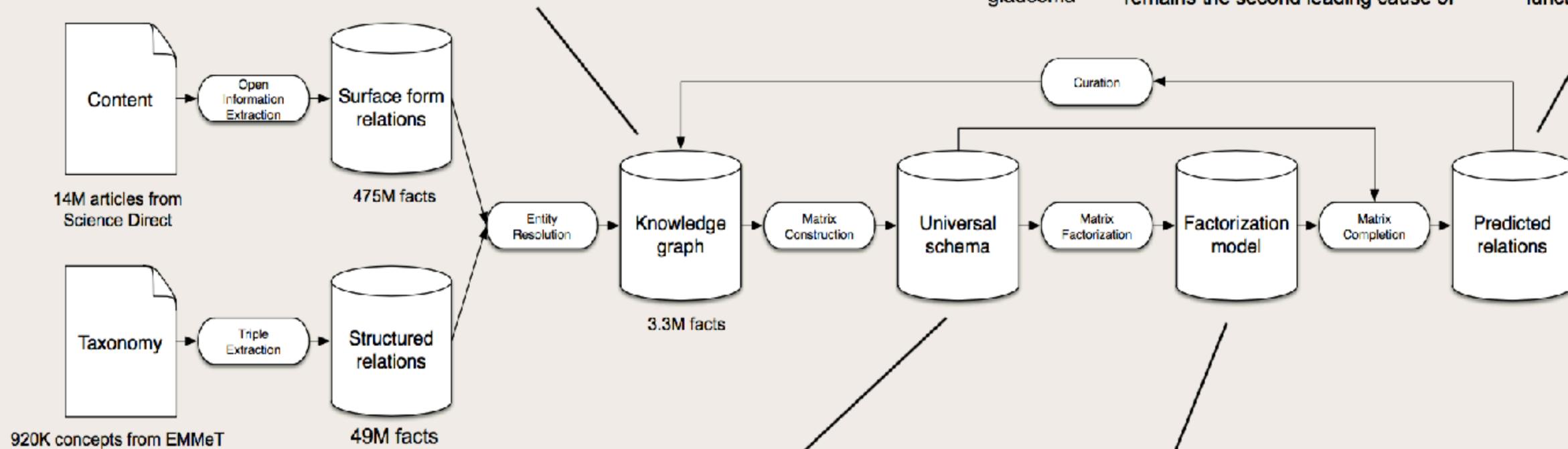
contact dermatitis 3815093  
 evaluation 5415395  
 primary open-angle glaucoma 8247149  
 treatment 5216597  
 oxidative stress 8184588  
 retinal ganglion cells 8002088

diseases  
 qualifier  
 diseases  
 procedures  
 diseases  
 anatomy

glaucoma  
 glaucoma  
 glaucoma  
 glaucoma  
 glaucoma  
 glaucoma  
 glaucoma

developed many years after  
 develop following  
 can appear soon in  
 can appear soon in  
 the risk of  
 contributing causes of  
 contributed to  
 is considered the second leading cause of  
 remains the second leading cause of

chronic inflammation of uveal tract  
 chronic inflammation of uveal tract  
 family history of glaucoma  
 age over 40  
 functional visual field loss  
 functional visual field loss  
 functional visual field loss  
 functional visual field loss  
 functional visual field loss



83 x 176 sparse binary-valued matrix with 366 entries

83 x 176 real-valued matrix with 14608 entries

来自: Paul Allen's Keynotes during DC-2016, Oct. 13, 2016

Back

# Networked advocacy

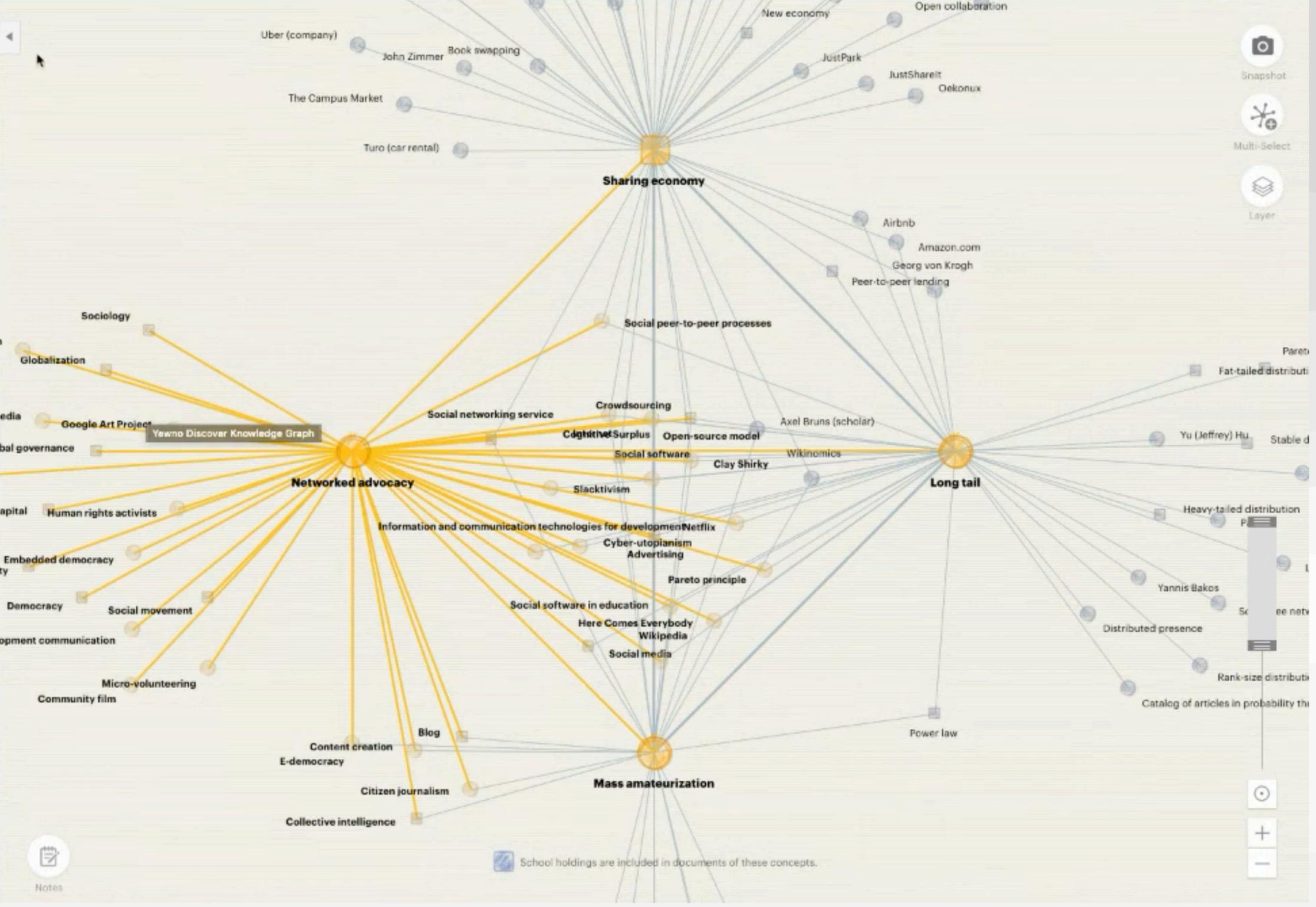
Political Science / Political History & Theory  
Web & Social Media / Social Aspects of the Internet

Overview Concepts Documents



### DEFINITIONS

Networked advocacy or net-centric advocacy refers to a specific type of advocacy. While networked advocacy has existed for centuries, it has become significantly more efficacious in recent years due in large part to the widespread availability of the internet,



Yewno Unearth



Settings

Notes

# 知识图谱的自动实现和应用技术： 认知计算

- 自然语言处理、理解和服务：NER名称实体，主题提取，文本标注，自动摘要，自动问答
- 图像处理：OCR、对象标注
- 音视频处理：文本化，映射
- 社会关系分析：实体属性标注
- 关联数据：服务整合、语义互操作

谢谢!



刘炜 [wliu@libnet.sh.cn](mailto:wliu@libnet.sh.cn)

上海图书馆上海科学技术情报研究所

