



知识图谱



童世炜

BDAA LAB

目录

contents

01

概述

什么是知识图谱

02

应用

如何利用知识图谱

03

构建

如何构建一个知识图谱

04

总结

总结、现状与推荐阅读

01

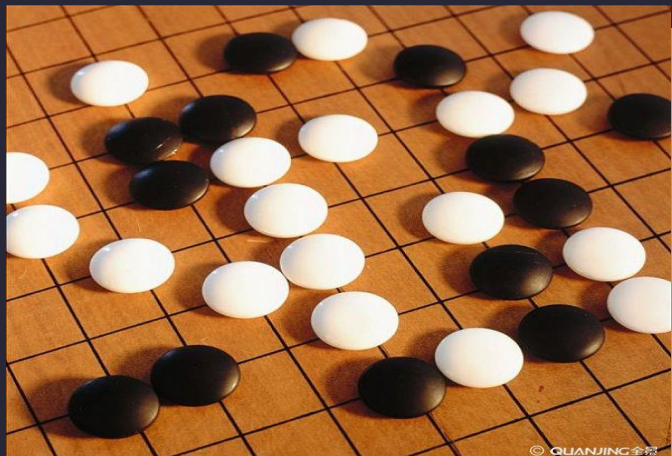
概述

Overview

什么是知识图谱



背景



Alpha Go

横扫围棋界



Siri

智能助理，帮助用户更加便捷地完成各项任务



Atlas

协助或取代人类繁重、危险的工作



人工智能



模拟脑的结构

连接派
神经网络
深度学习



模拟人的心智

符号派
知识表示
知识图谱

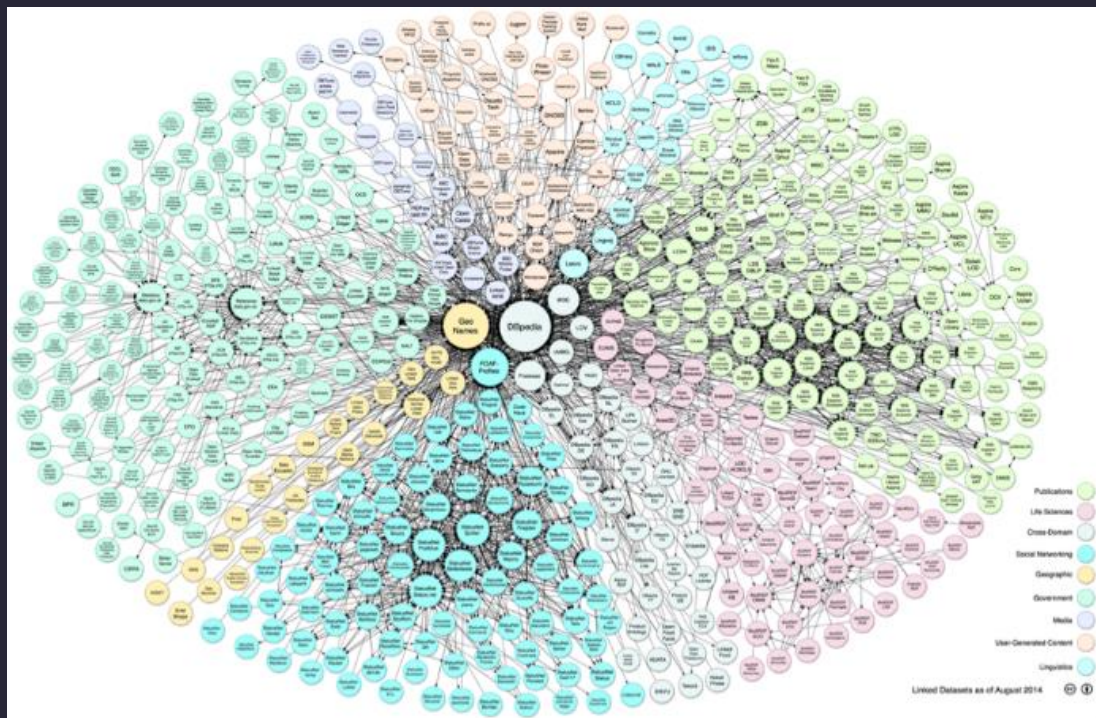


模拟人的行为

行为派
机器人
增强学习



知识图谱





知识图谱发展





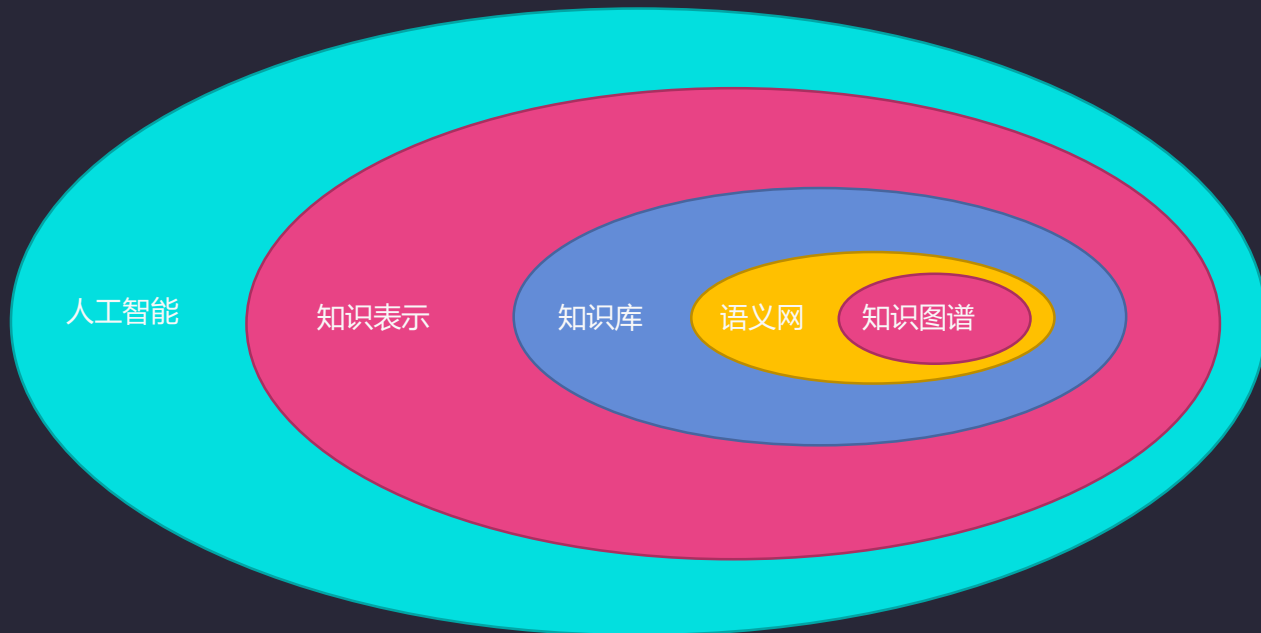
知识图谱

定义

知识图谱（knowledge graph）一词由谷歌于2012年提出。

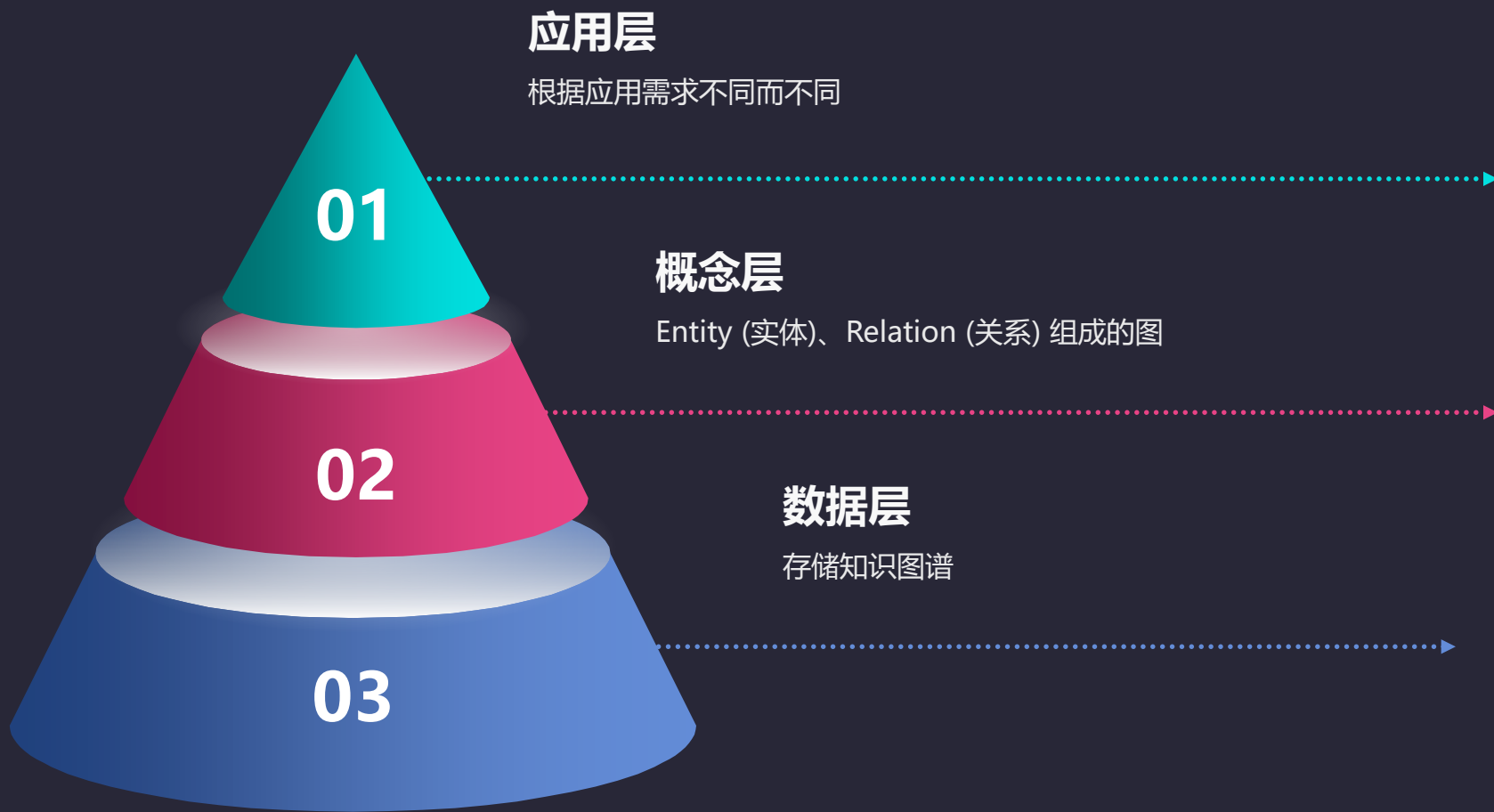
狭义上讲，知识图谱是一种用图结构来描述知识的方式，是一种语义网络（semantic network）。

广义上，采用任意结构描述知识的方式也可以被称为知识图谱（此时其含义和知识库knowledge base相同）。





知识图谱



02

应用

Applications

关系机器学习

基于知识图谱的推荐系统

基于知识图谱的搜索问题



关系机器学习



给定 N_d 个已观测的三元组 (点-边-点) 集合, 三元组特征记为 x , 标签记为 y

$$\mathcal{D} = \{(x^n, y^n) | n = 1, \dots, N_d\}$$

$$\begin{cases} (S, P, O)^n = x^n \\ label^n = y^n \end{cases}$$

独立假设

$$P(\underline{Y} | \mathcal{D}, \Theta) = \prod_{i=1}^{N_e} \prod_{j=1}^{N_e} \prod_{k=1}^{N_r} Ber(y_{ijk} | \sigma(f(x_{ijk}; \Theta)))$$

概率密度函数

$$\sigma(u) = \frac{1}{1 + e^{-u}}$$

归一化函数

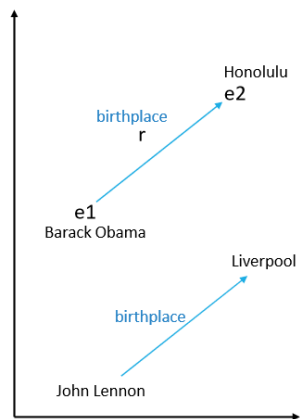
$$Ber(y|p) = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{if } y = 0 \end{cases}$$

伯努利分布函数



关系估分函数 $f(x_{ijk}; \Theta)$ 和 Pairwise Loss

Translation Embeddings



TransE

$$S(r(a, b)) = -\|e_a + \mathbf{R}_r - e_b\|_2^2$$

TransH

$$S(r(a, b)) = -\|e_a^\perp + \mathbf{R}_r - e_b^\perp\|_2^2$$

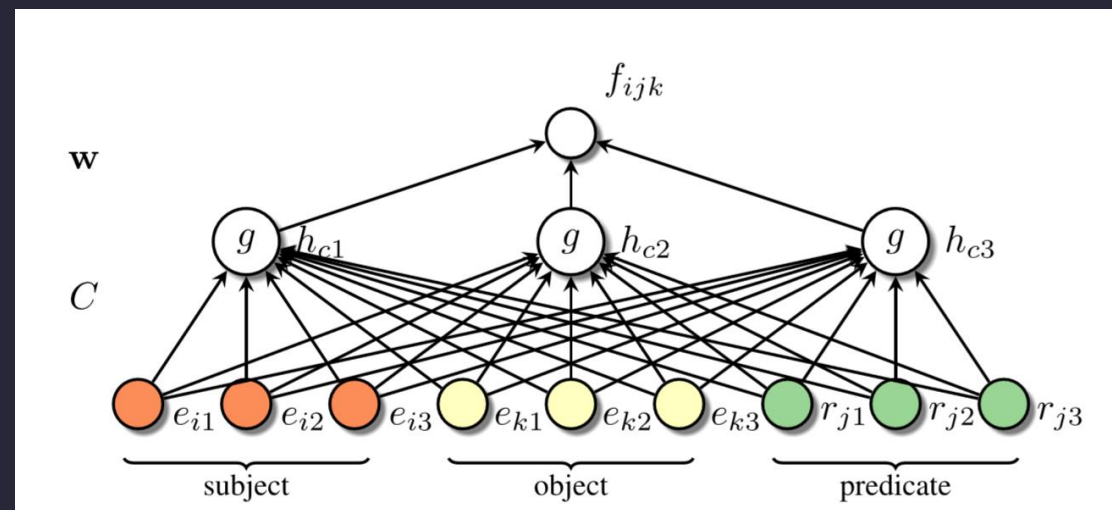
$$e_a^\perp = e_a - \mathbf{w}_r^T e_a \mathbf{w}_r$$

TransR

$$S(r(a, b)) = -\|e_a \mathbf{M}_r + \mathbf{R}_r - e_b \mathbf{M}_r\|_2^2$$

TransE: Bordes et al. XXX (2011), TransH: Bordes et al. XXX (2011), TransR: Bordes et al. XXX (2011)

23



$$\min_{\Theta} \sum_{x^+ \in \mathcal{D}^+} \sum_{x^- \in \mathcal{D}^-} \mathcal{L}(f(x^+; \Theta), f(x^-; \Theta)) + \lambda \text{reg}(\Theta)$$

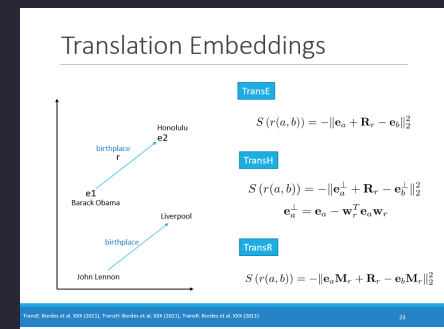
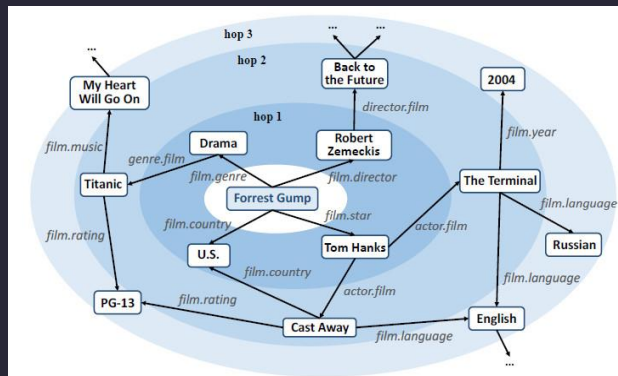
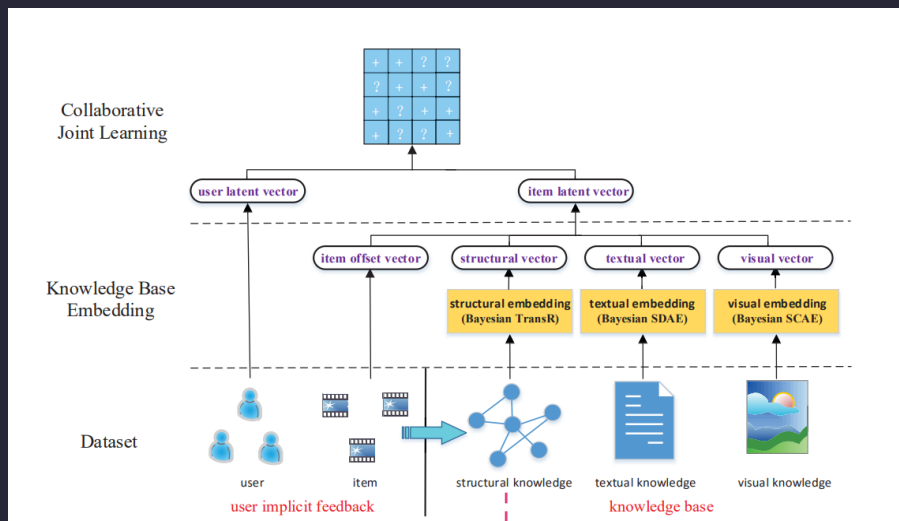


$$\mathcal{L}(f, f') = \max(1 + f' - f, 0).$$



基于知识图谱的推荐系统

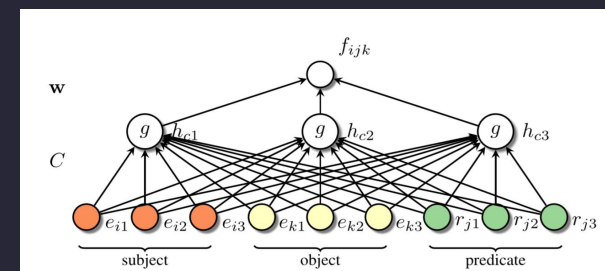
$$\begin{cases} (Item, R, Item)^n = x^n \\ label^n = y^n \end{cases}$$



物品之间的关系
(TransR)

图的嵌入式表达

电影、人物之间的关系
(latent vector)



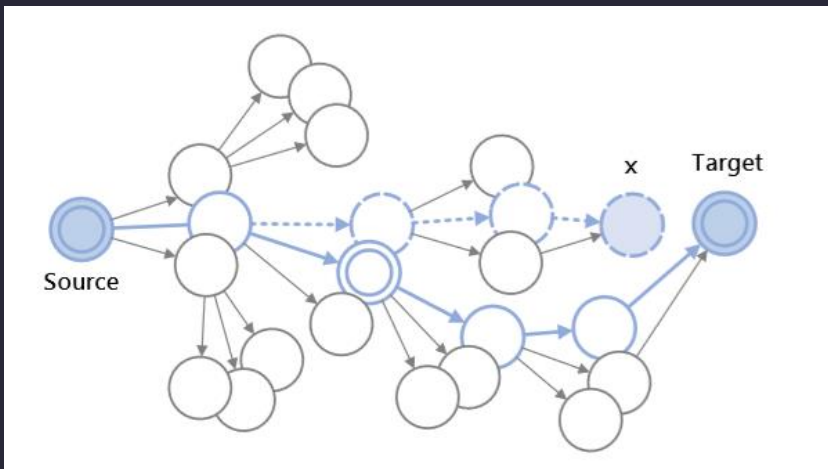
$$\begin{cases} (User, Like, Item)^n = x^n \\ label^n = y^n \end{cases}$$

Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016: 353-362.

Wang H, Zhang F, Wang J, et al. Ripple Network: Propagating User Preferences on the Knowledge Graph for Recommender Systems[J]. arXiv preprint arXiv:1803.03467, 2018.



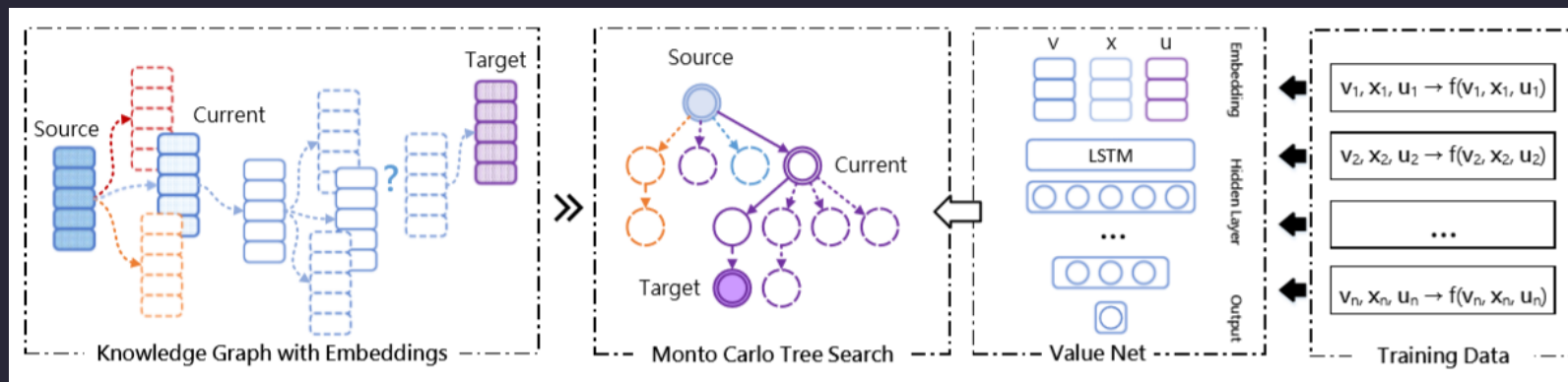
基于知识图谱的搜索问题



图结构限制



构建基于图的嵌入式表达的启发式函数





知识图谱



03

构建

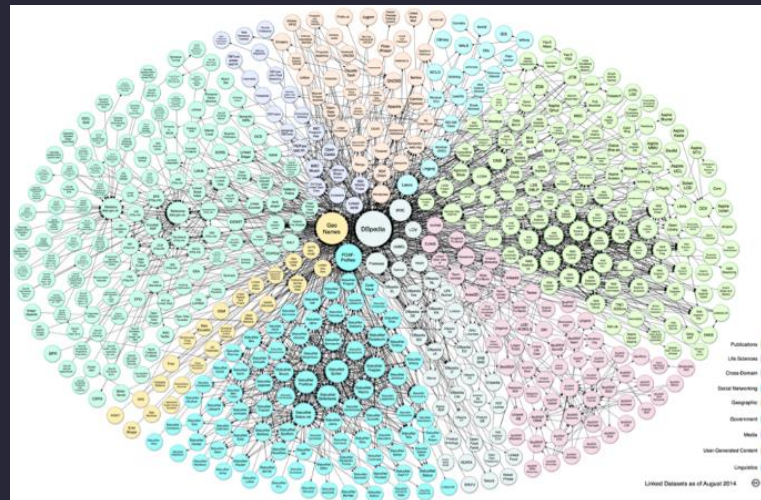
Construction

Surface → Thing → Graph → Data



知识图谱的本质

- 统一的数据管理层
 - 不同格式的数据转换为统一的数据格式
 - 消除不同数据之间的冲突
 - 同名异指
 - 张伟、刘洋
 - 他、她、它
 - 异名同指
 - 飞人-乔丹、外星人-罗纳尔多
 - 他-那个人
- 高效的数据管理层
 - 概念化的缓存 Cache





数据格式

• 非结构化

- 文本
- 图片
- 视频
- 知识

• 半结构化

- 百科知识库
- 表格、网页

• 结构化

- 知识库
- 知识图谱

一组封闭的专家输入

WordNet
Freebase



自动半结构化

Wikipedia infoboxes
YAGO
DBpedia

一组开放的志愿者输入

WIKIPEDIA
WikiData



自动非结构化

NELL
Knowledge Vault





知识图谱构建流程

Surface → Thing

- 基底对齐
- 实体抽取
- 关系抽取

01



02



Thing → Graph

- 知识与知识的联系
- 知识与知识可能存在冲突和重复

Graph → Data

- 用什么方式来存储数据
- 如何高效地管理数据

03



概念层

- 统一的概念化的表示

- SPO三元组



- 点代表实体/属性
- 边代表了实体-实体或实体-属性的关系

- NER

- 命名实体识别

- Resolution

- 消解



Entities with Same Name

Same type of entities share names

Kevin Smith, John Smith,
Springfield, ...

Things named after each other

Clinton, Washington, Paris,
Amazon, Princeton, Kingston, ...

Partial Reference

First names of people, Location
instead of team name, Nick names

Different Names for Entities

Nick Names

Bam Bam, Drumpf, ...

Typos/Misspellings

Baarak, Barak, Barrack, ...

Inconsistent References

MSFT, APPL, GOOG...



数据层

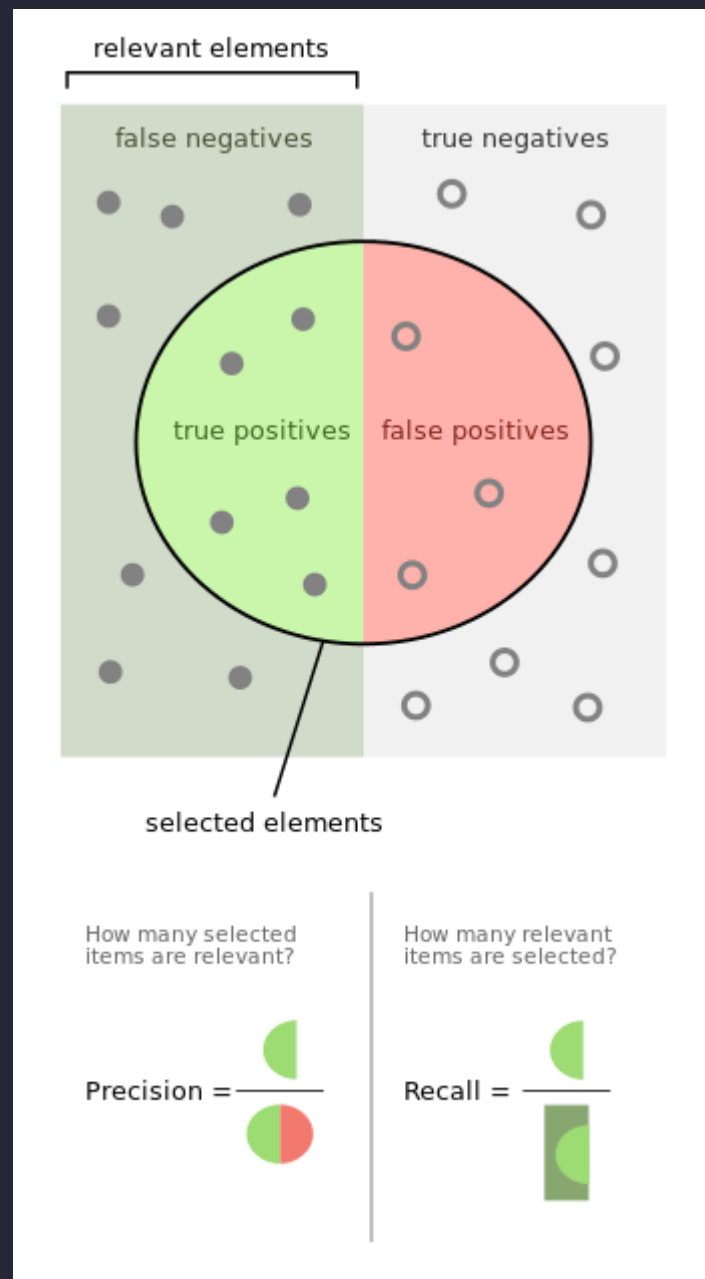
- 概念化
 - SPO三元组
- Data组织
 - RDF (SPO三元组)
 - RDFs, RDFa
 - OWL
 - N-Triples
 - JSON-LD
- 底层支撑
 - 关系型数据库
 - 图数据库





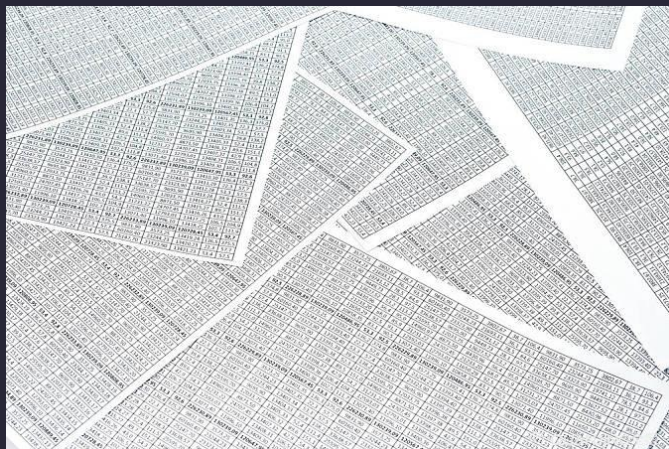
质量评估

- Precision
 - 知识准确性
- Recall
 - 知识丰富度
- Time
 - 查询时间





知识图谱



Surface → Graph



应用层

根据应用需求不同而不同
根据应用确定需要使用的表示方式

概念层

Entity (实体)、Relation (关系) 组成的图

数据层

存储知识图谱

图表示学习

- Latent Vector
- Random Walk
- Markov Random Field



```

31100110100101010011100111010100101
101101010011111110000101101101010
111110110110001101101010101010101
101001101101101010101010101010101
100001100000000000000000000000000
100000001111011111100001010101010
100101100111011110101010101010101
100101100111100011110100010110000
11001110011001010100011100000000000
  
```

04

总结

Summary

总结

现状

参考文献



研究现状

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

<https://nlp.stanford.edu/software/openie.html>

<http://knowledgeworks.cn/>

<http://openkg.cn/>

<https://www.jiqizhixin.com/articles/2017-03-20>

<https://developers.google.com/knowledge-graph/>

<https://ai.baidu.com/tech/kg/schema>

<https://www.zhihu.com/question/62164852>

| 机构 | 产品/奖项 | 备注 |
|----------|---|---------------|
| CMU | NELL | |
| Stanford | OpenIE | |
| 复旦 | 知识工场 | 肖仰华 |
| 浙大等 | OpenKG | 刘知远、漆桂林、王昊奋 |
| 微软 | Satori, Probase/Microsoft Concept Graph | |
| 谷歌 | Knowledge Graph, Knowledge Vault[2] | |
| 百度 | Schema | |
| | DBpedia | 维基百科 |
| | Yago | 维基百科与 WordNet |
| 维基媒体基金会 | Wikidata | |



推荐阅读

- Nickel M, Murphy K, Tresp V, et al. A review of relational machine learning for knowledge graphs[J]. Proceedings of the IEEE, 2016, 104(1): 11-33.

Q&A