知识图谱构建与应用最佳实践

关键词：知识图谱、图数据库、数据模型、MLOps、知识提取、图表征

2021.10.18
苏锦华

Presetation available at: https://github.com/SmartDataLab/KG-Share
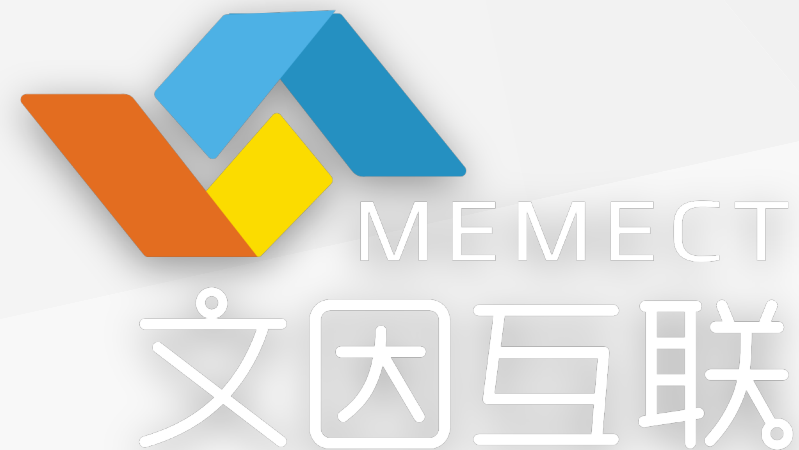
Written with Marp and Mermaid on Markdown

引子

" One hidden layer Perception with non-linear activation function -
- Logistic Regression

"

知识图谱、图神经网络概念火热，国内有成熟的知识图谱产品吗？
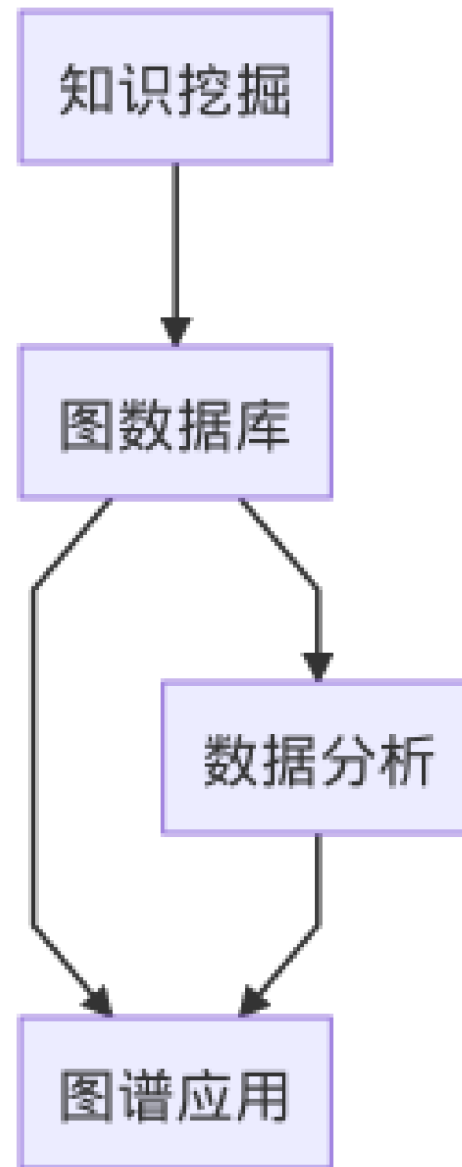
😆

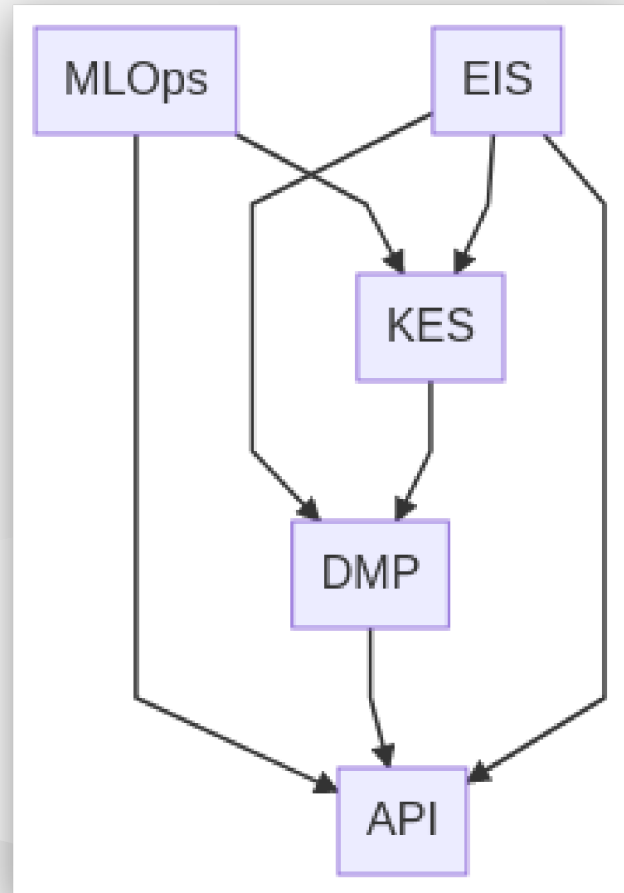知识图谱技术落地需要 **表子好看，里子实在**。

商业智能-行业数据库-知识挖掘

背后依托的技术

1. 知识挖掘：用户交互以及NLP
2. 数据存储：图数据库 😄
3. 数据分析：图算法及可视化技术 😄
4. 图谱应用：图表征以及图神经网络

```
graph TD;
    A[知识挖掘]-->B[图数据库];
    B-->C[数据分析];
    B-->D[图谱应用];
    C-->D;
```
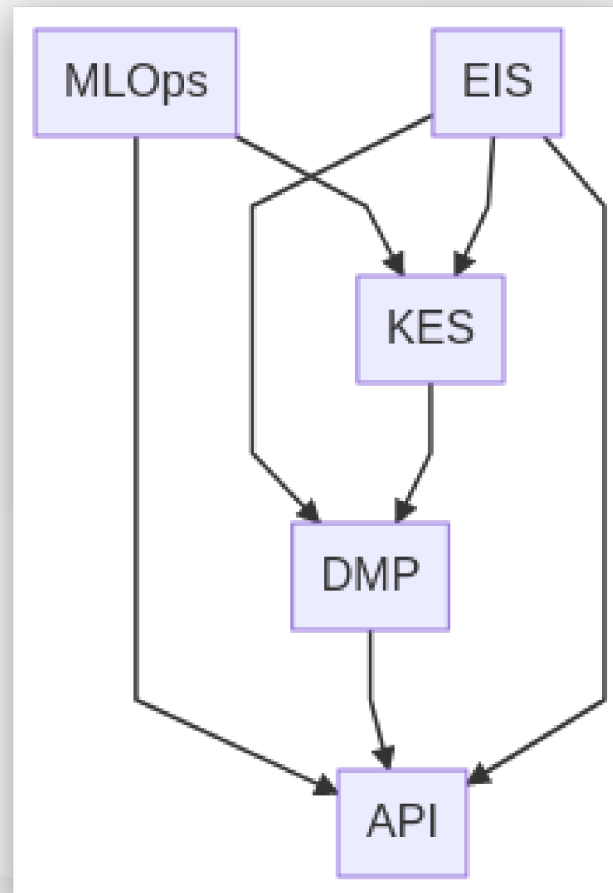
知识图谱产品架构

- MLOps System
  - Kubernetes
  - Jenkins X: MLOps [Quickstarts](#)
  - General ML Deployment: ONNX
- EIS: Expert Interact System
  - Human Labelling & Heuristic Algorithm
  - Monitoring & Back-testing
 - Quality: Bad Cases in KG

知识图谱产品架构 cont.

- KES: Knowledge Extraction System
  - Transfer from RDBS
  - Entity Recoginition
    - Node presentation
    - Breadth first priority search
  - Link Prediction
    - shortest path
    - NLP semantic distance
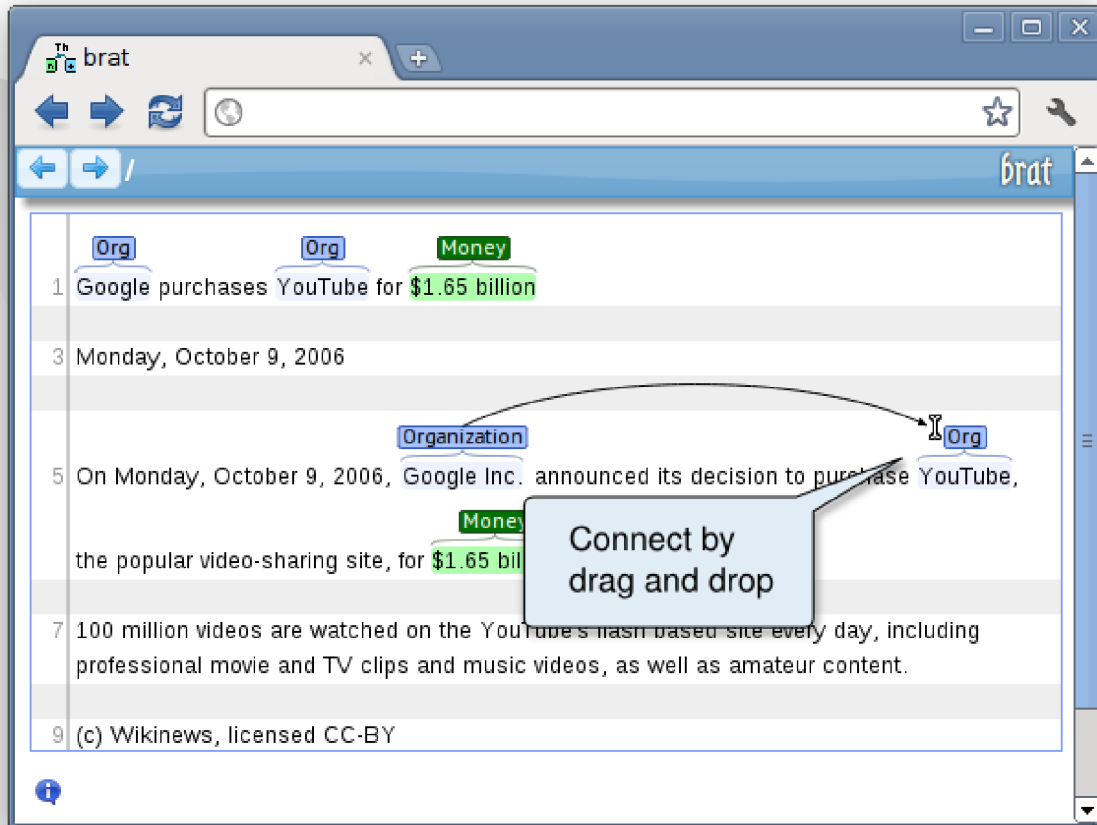 - Attribution Extration

知识图谱产品架构 cont.

- DMP: Data Manage Platform
  - DataHub to manage the dirty data
  - Graph Database
- API Platform
  - Graph Visualization
  - Node Representation
  - Prediction Tasks
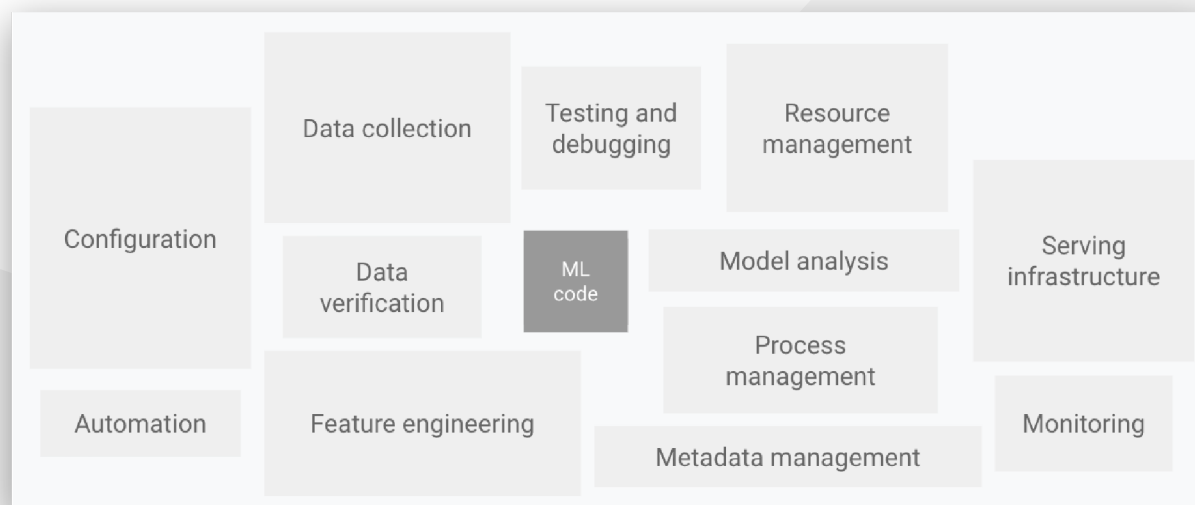   - intelligent search(google page rank)
    - KG-based Q&A (NLP)
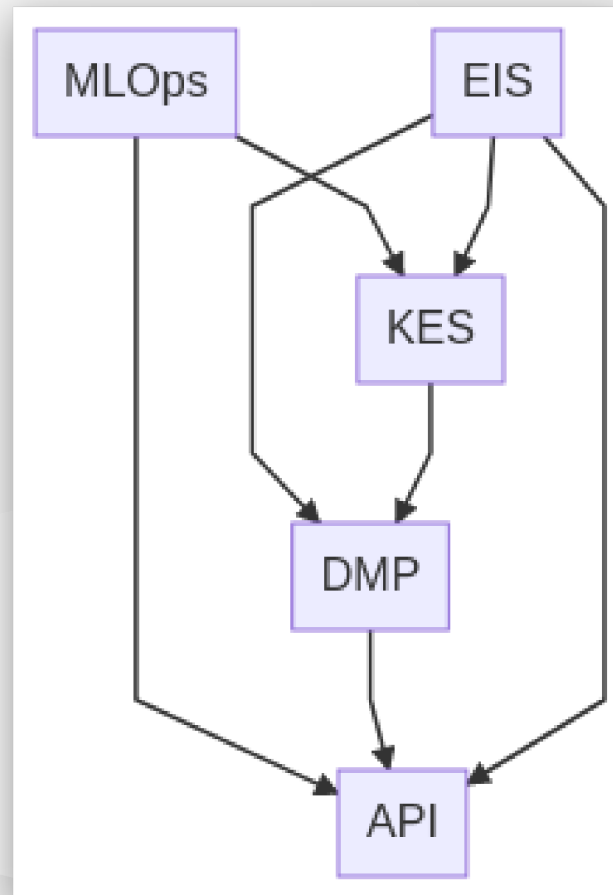
Expert Interact System
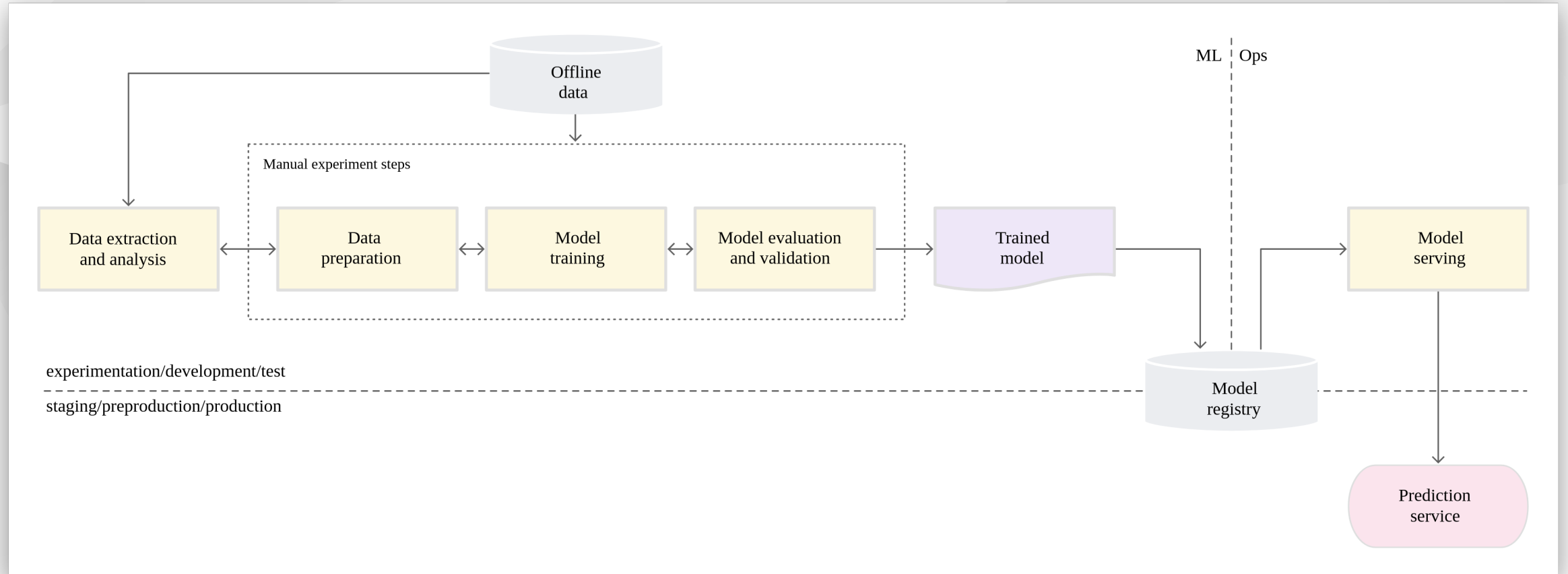
- Brat (labelling tool) [link](#)

为什么要MLOps

Machine Learning: The High Interest Credit Card of Technical Debt
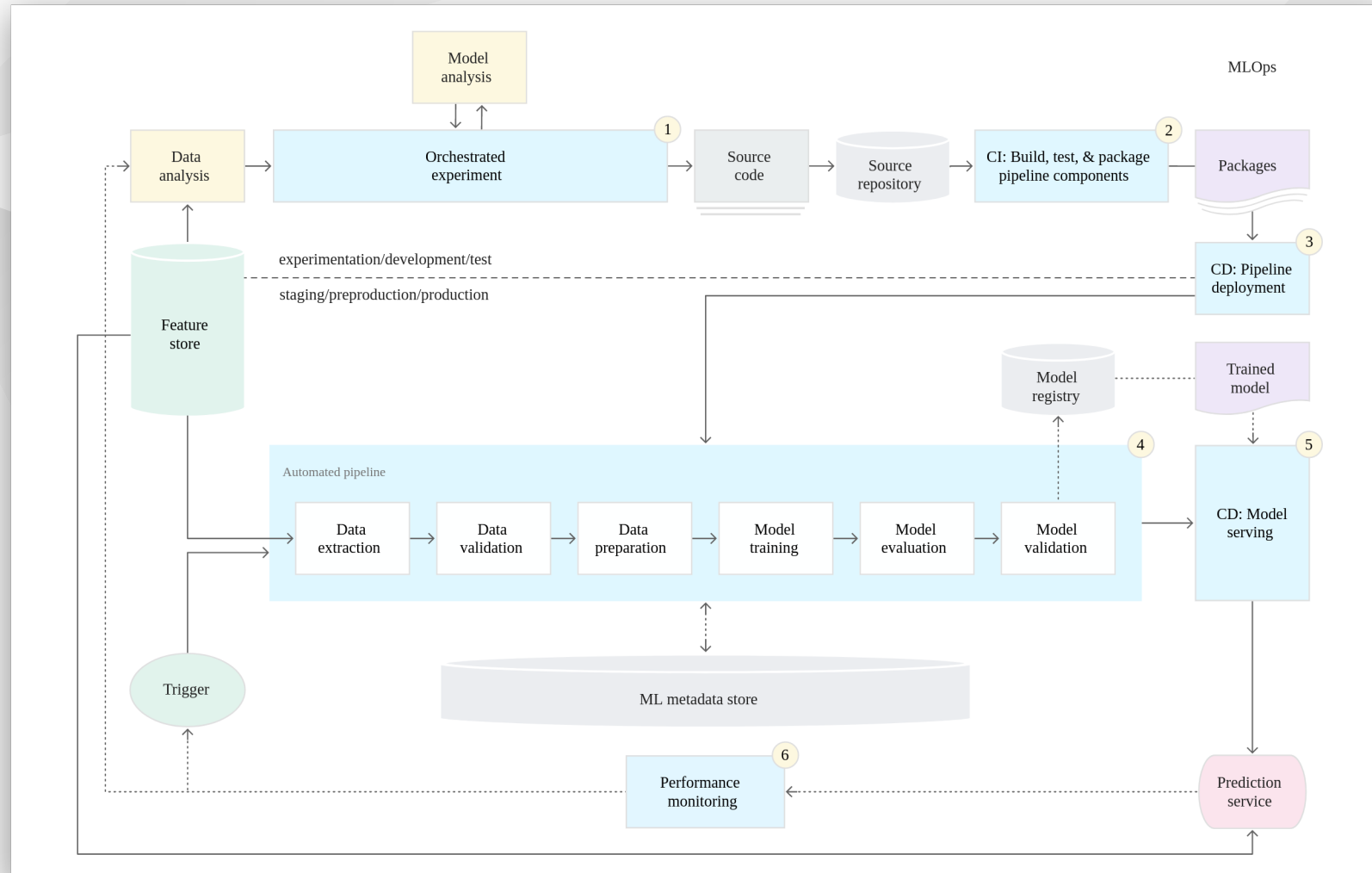A paper in NIPS 2014 workshop

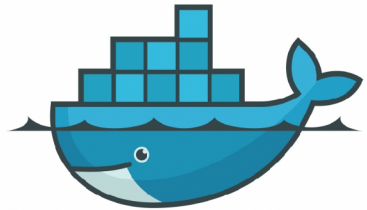Rules of Machine Learning: Best Practices for ML Engineering link

training-serving skew 😄
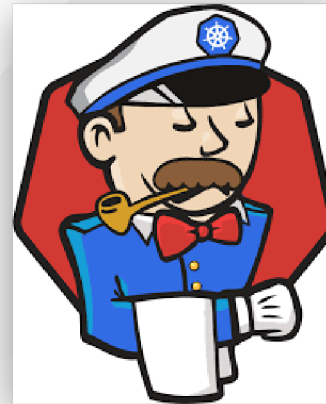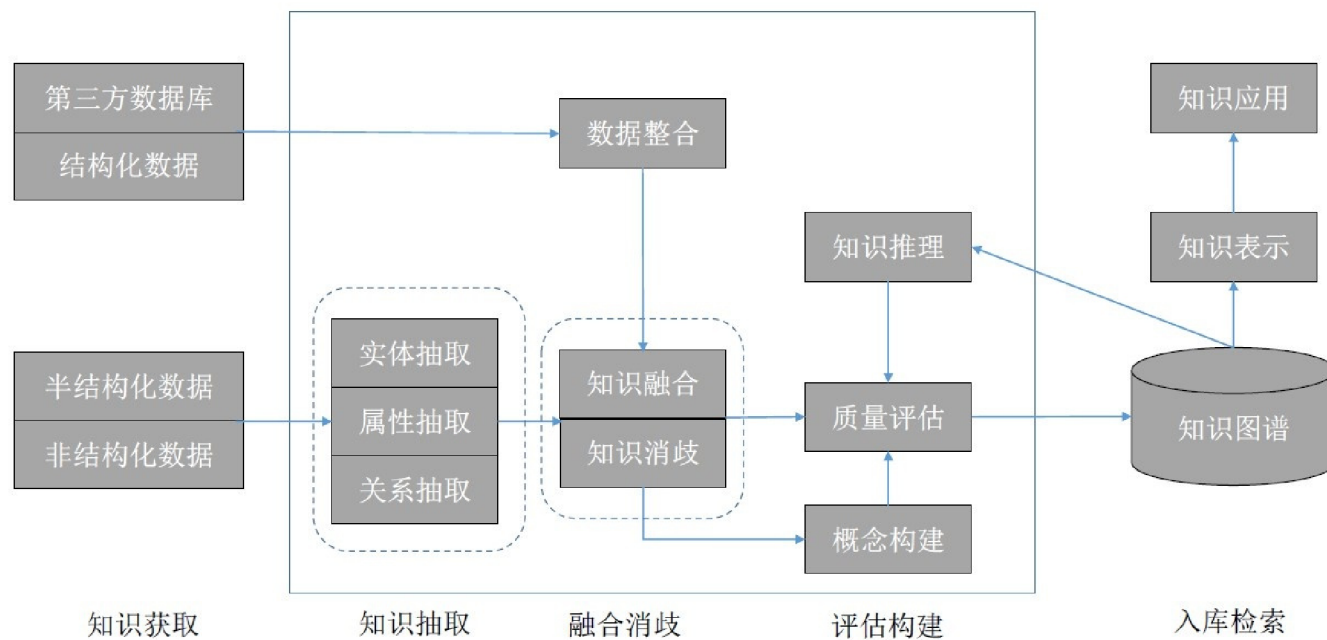
MLOps(low level)

MLOps(high level)

MLOps Tools(support gpu)

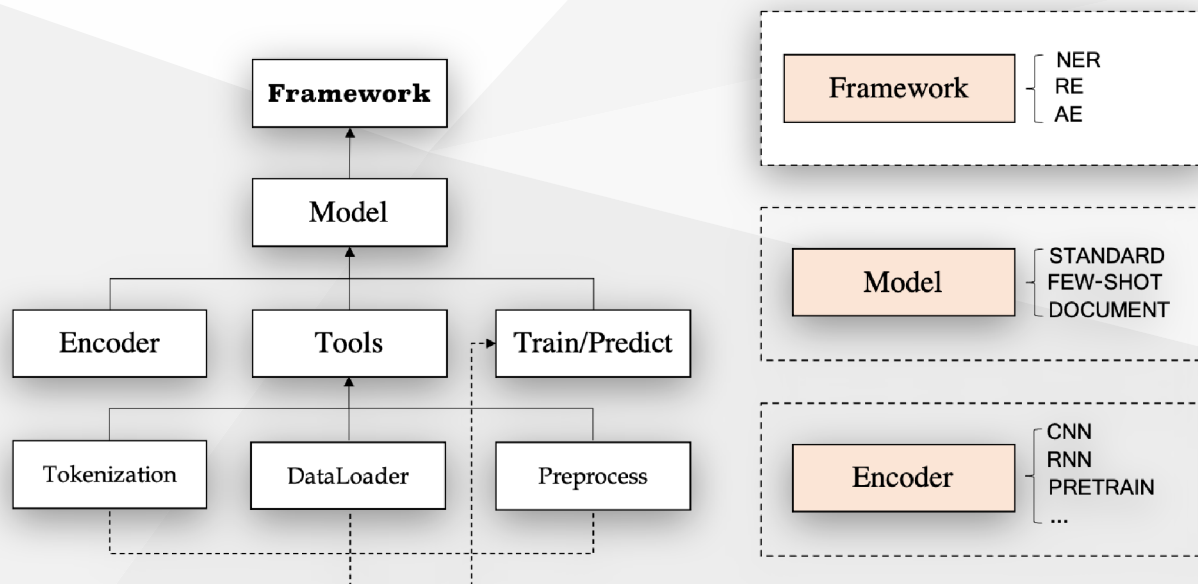# Knowledge Extraction System

A structure from CSDN(author: coder_oyang)

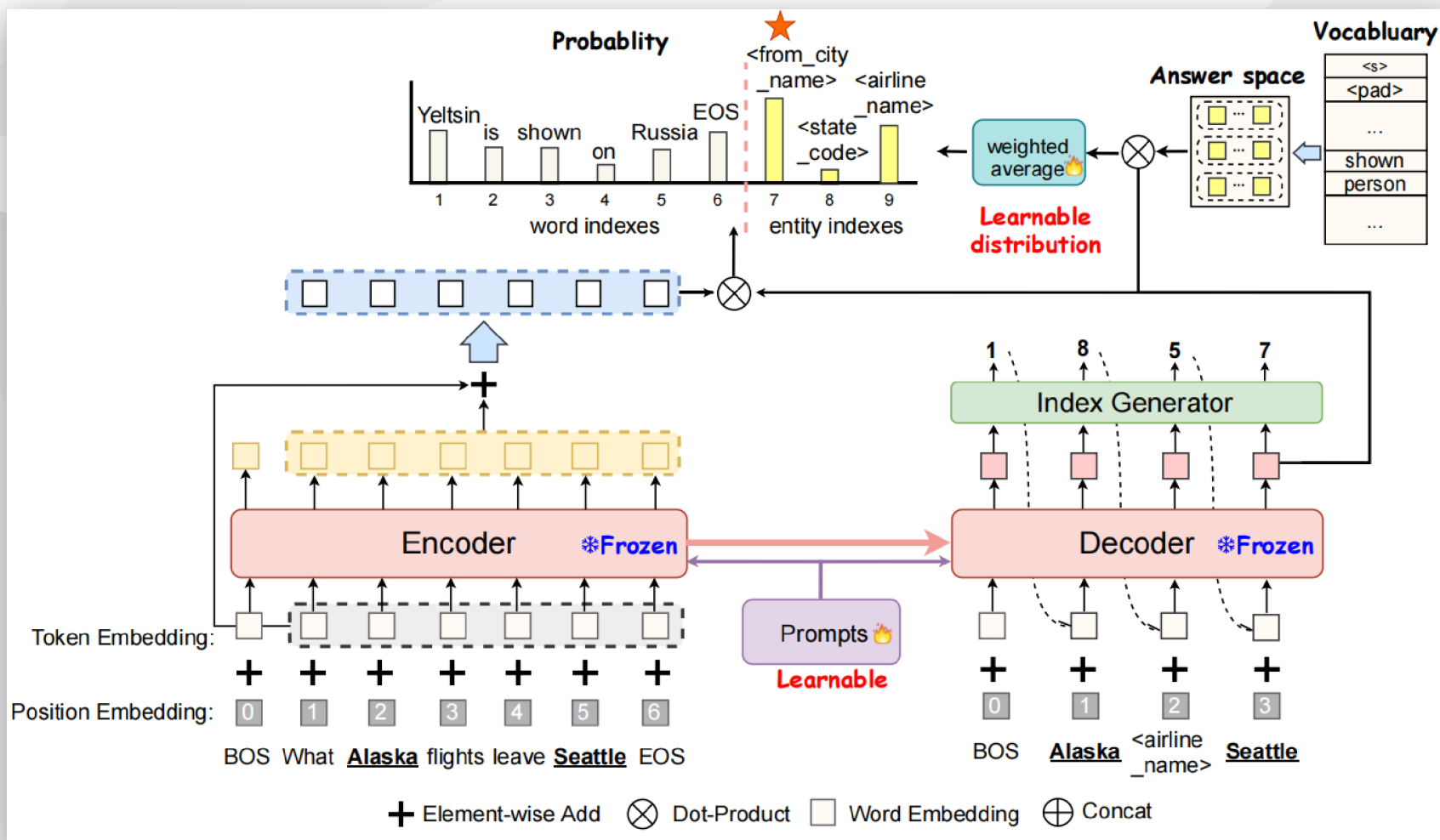# 构建工具

推荐OpenKG(浙大),网站整理了大量工具和相关技术文章

- 中文NLP工具:(哈工大LTP，车万翔)
- 知识提取工具:DeepKE(浙大，结合prompt，适合few-shot场景)

LightNER arXiv

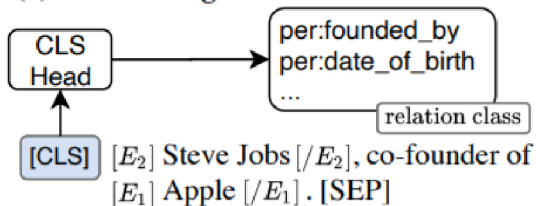# LightRE github
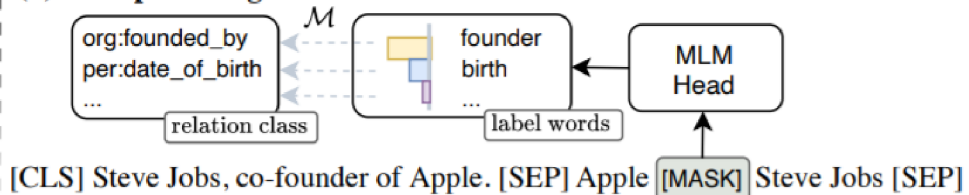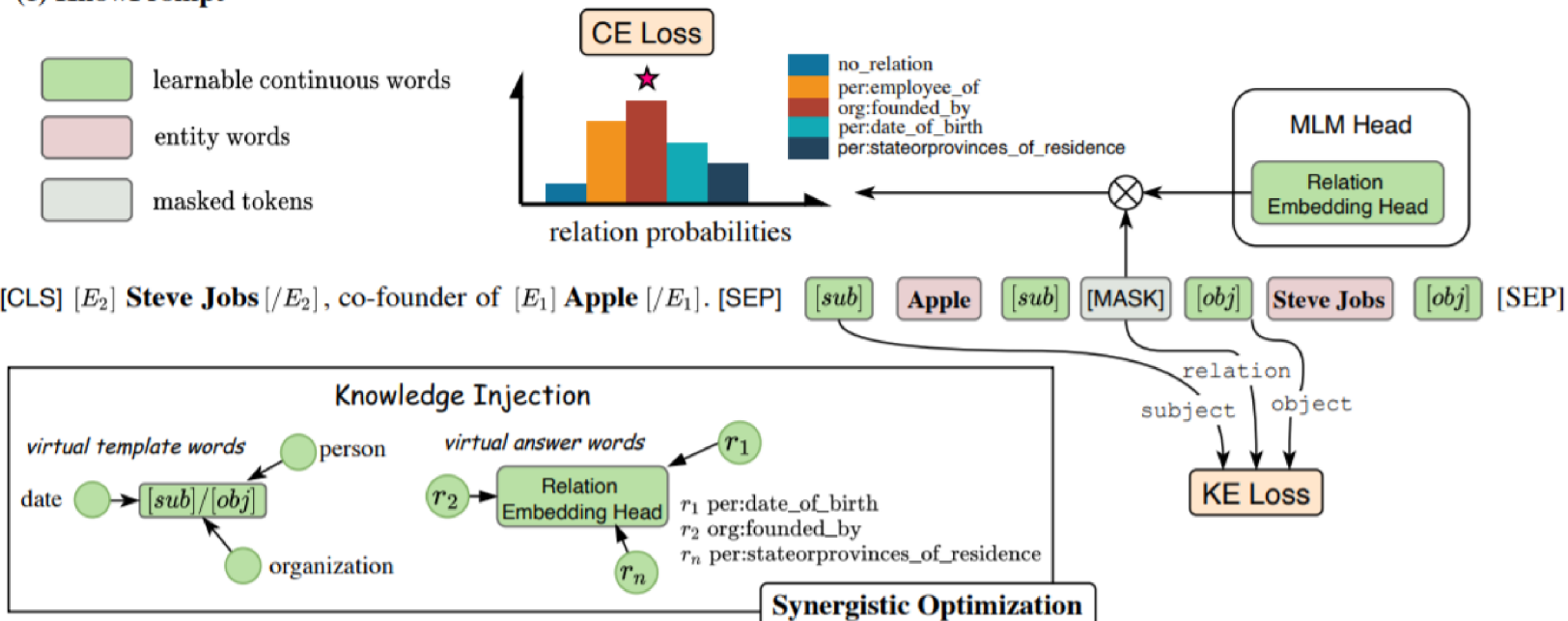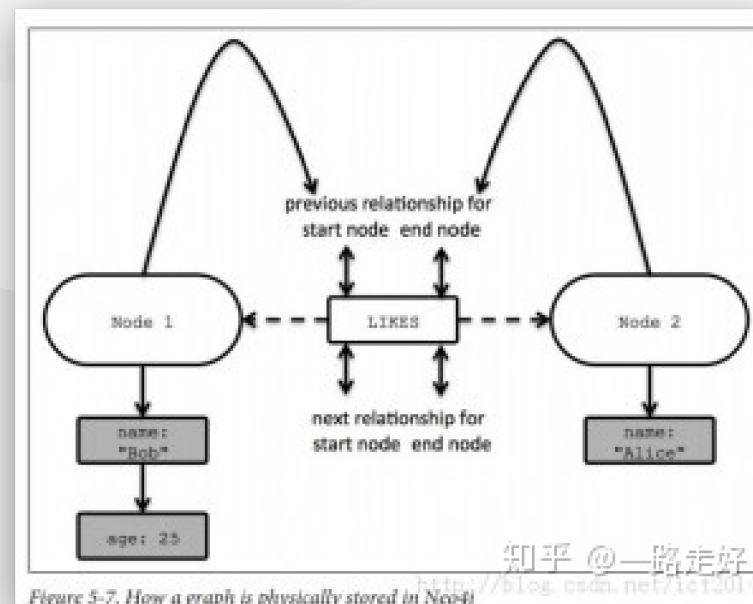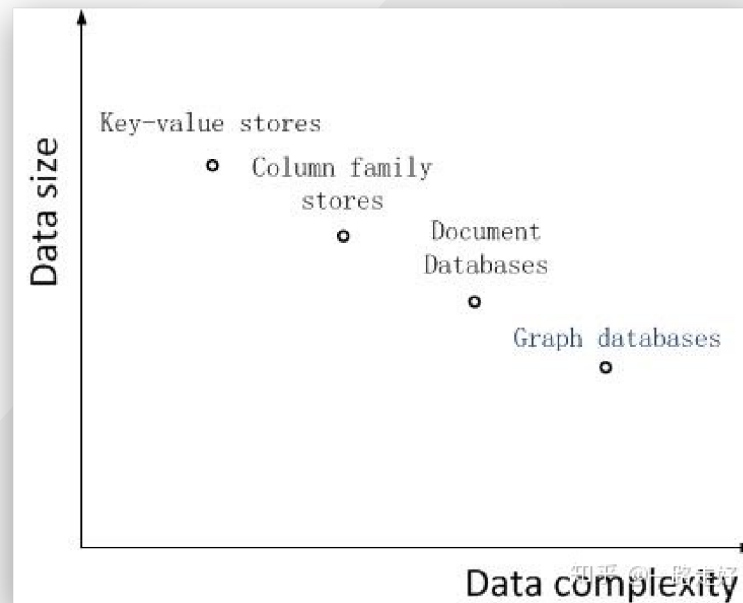
Data Manage Platform

思考🤔关系型数据库(RDBMS)与NoSQL的发展和知识图谱有什么联系

" Web2.0进入了"可读可写"模式，交互内容的急剧增加，RDBMS在超大规模数据的高并发处理力不从心,NoSQL应运而生。 "

Figure 5-7. How a graph is physically stored in Neo4j

图数据库的选择

include secondary database models				33 systems in ranking, December 2019	

Rank			DBMS	Database Model	Score		
Dec 2019	Nov 2019	Dec 2018			Dec 2019	Nov 2019	Dec 2018
1.	1.	1.	Neo4j ➕	Graph	50.56	+0.03	+5.00
2.	2.	2.	Microsoft Azure Cosmos DB ➕	Multi-model ℹ	31.43	-0.54	+8.06
3.	3.	3.	OrientDB	Multi-model ℹ	4.93	-0.46	-1.15
4.	4.	4.	ArangoDB ➕	Multi-model ℹ	4.87	-0.14	+0.60
5.	5.	5.	Virtuoso ➕	Multi-model ℹ	2.64	+0.00	+0.16
6.	6.	6.	JanusGraph	Graph	1.75	-0.05	+0.49
7.	7.	7.	Amazon Neptune	Multi-model ℹ	1.57	-0.03	+0.35
8.	8.	⬆10.	GraphDB ➕	Multi-model ℹ	1.15	+0.01	+0.43
9.	⬆11.	⬇8.	Giraph	Graph	1.04	+0.03	+0.04
10.	10.	⬆14.	TigerGraph ➕	Graph	0.96	-0.05	+0.49

Neo4j生态

Transfer from RDBMS

Schema: Model Design

图谱应用工具

- 知识表示：清华大学OpenKE
  - RESCAL
  - DistMult, ComplEx, Analogy
  - TransE, TransH, TransR, TransD
  - SimplE
  - RotatE
- 可视化工具
  - Boom(Neo4j Lab)
  - Echarts(Baidu)

Thanks