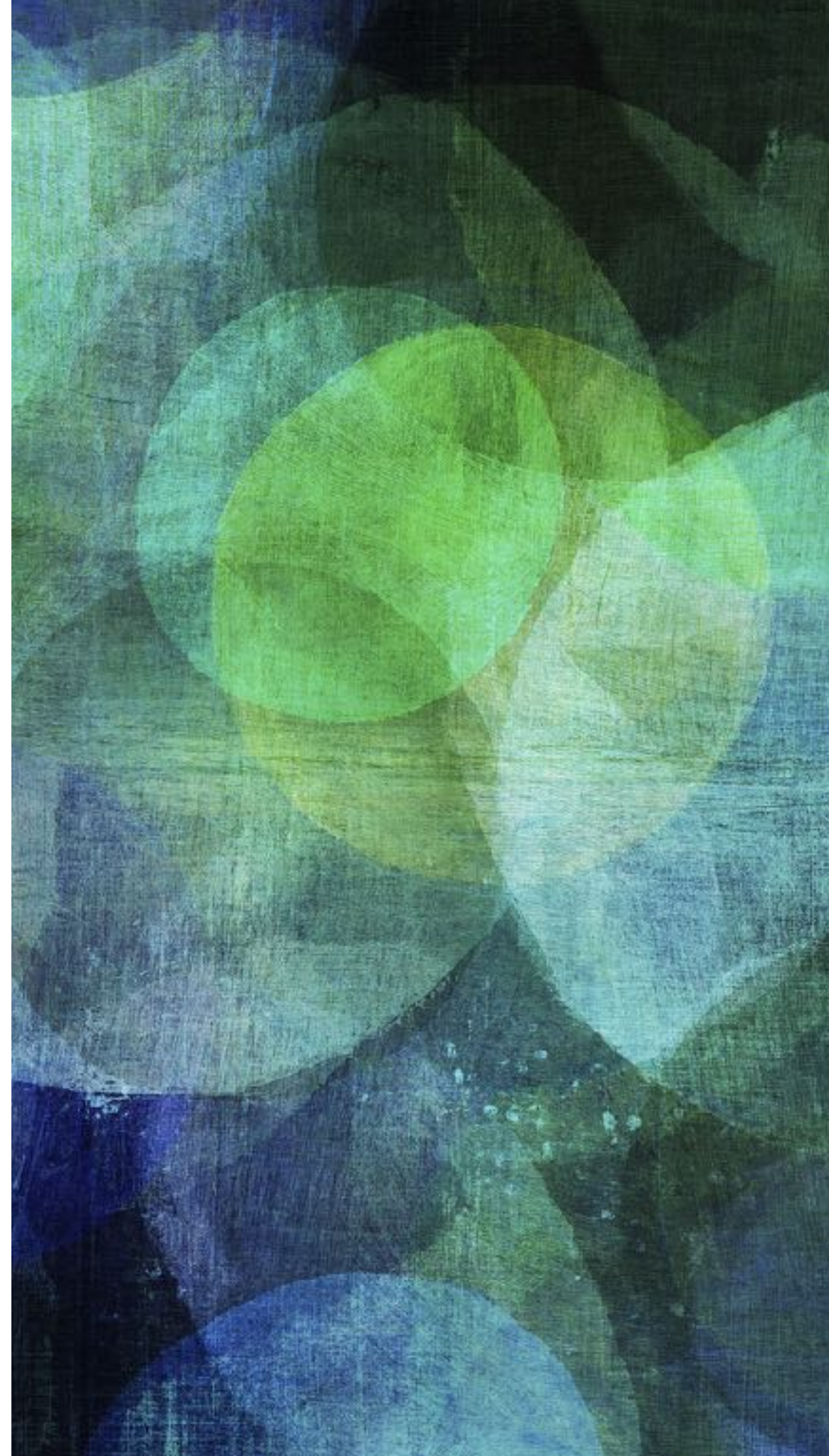
领域知识图谱的构建及应用

陈华钧　浙江大学计算机科学与技术学院 教授\博导

一、背景概述

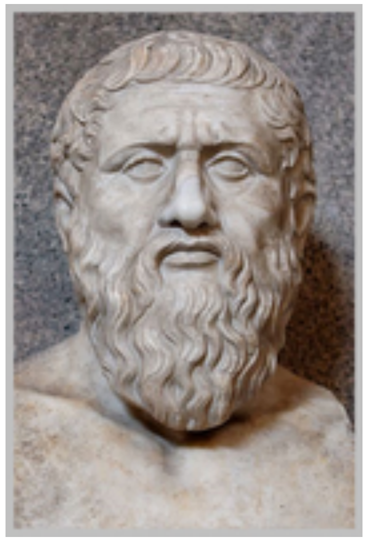


What is Knowledge

Knowledge is justified true belief.

人类的自然语言、创作的绘画和音乐、数学语言、物理模型、化学公式等都是人类知识的表示形式和传承方式。具有获取、表示和处理知识的能力是人类心智区别于其它物种心智的重要特征。

AI的核心是研究怎样用计算机易于处理的方式表示、学习和处理各种各样的知识，广义的讲，神经网络也是一种知识表示形式。



柏拉图

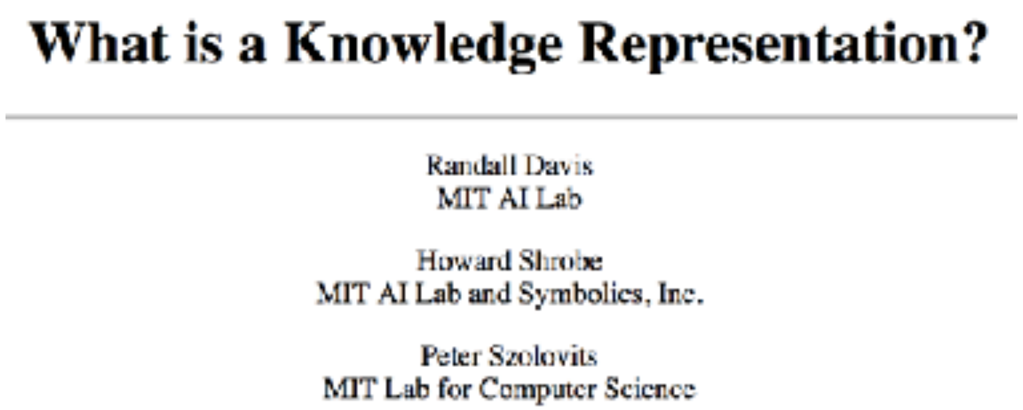
Knowledge is a familiarity, awareness, or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning.

KG is just one form of Knowledge, text is also knowledge.

What is Knowledge Representation

简单而言，知识表示（KR）就是用易于计算机处理的方式来描述人脑的知识。

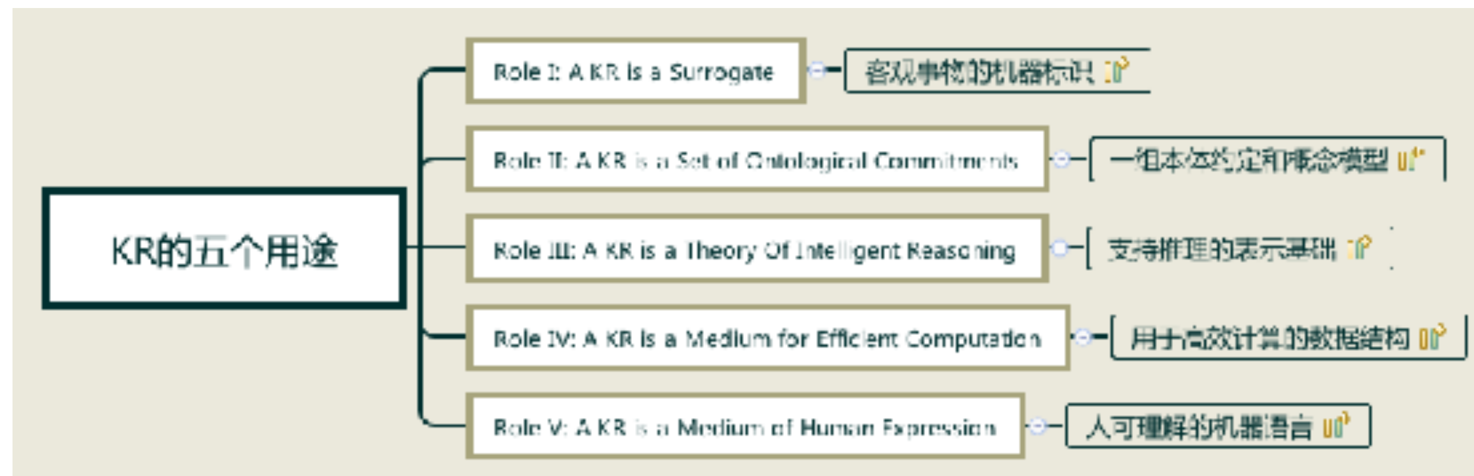
KR不是数据格式、不等同于数据结构、也不是编程语言，对于人工智能而言，数据与知识的区别在于KR支持推理。


What is a Knowledge Representation?

Randall Davis
MIT AI Lab

Howard Shrobe
MIT AI Lab and Symbolics, Inc.

Peter Szolovits
MIT Lab for Computer Science



Entity ID | 概念模型 | 支持推理 | 易于计算 | 人可理解

R. Davis, H. Shrobe, and P. Szolovits. What is a Knowledge Representation? AI Magazine, 14(1):17-33, 1993.

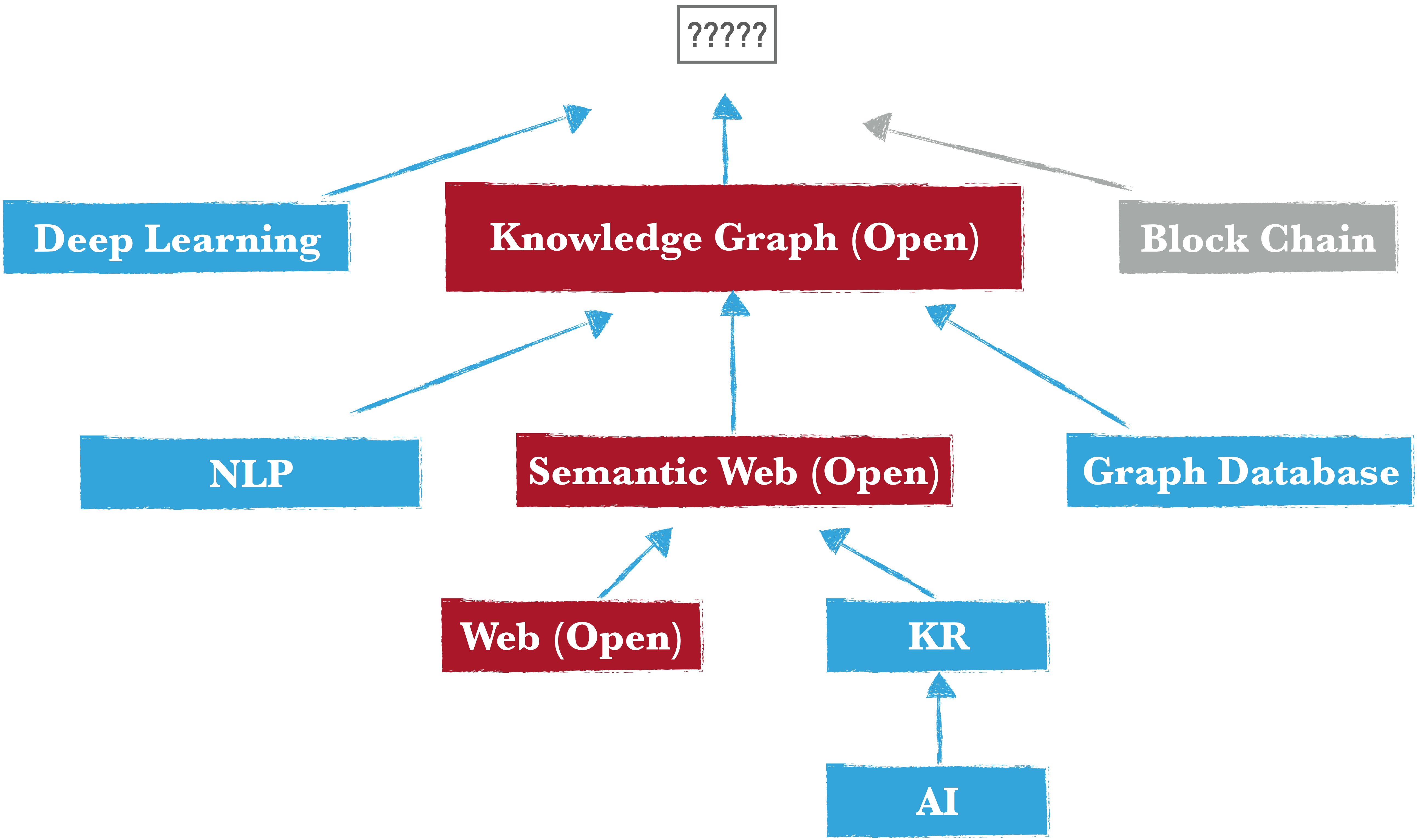
KR = Computational Model of Reality

知识图谱：事物关系的可计算模型

知识图谱旨在从数据中识别、发现和推断事物、概念之间的复杂关系，是**事物关系的可计算模型**，已经被广泛应用于搜索引擎、智能问答、语言理解、视觉场景理解、决策分析等领域。

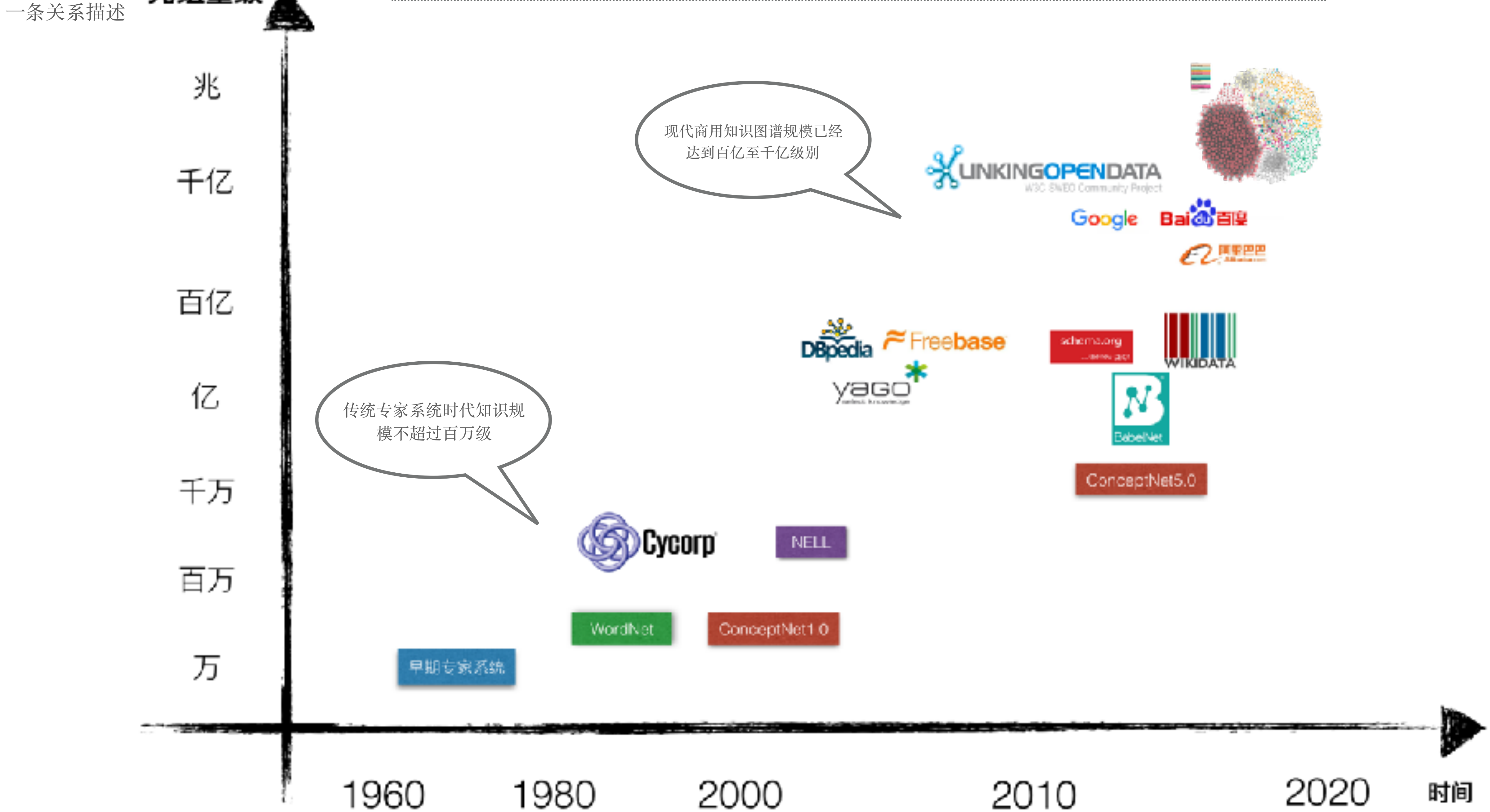
KG = Computational Model of Relations

知识图谱：事物关系的可计算模型

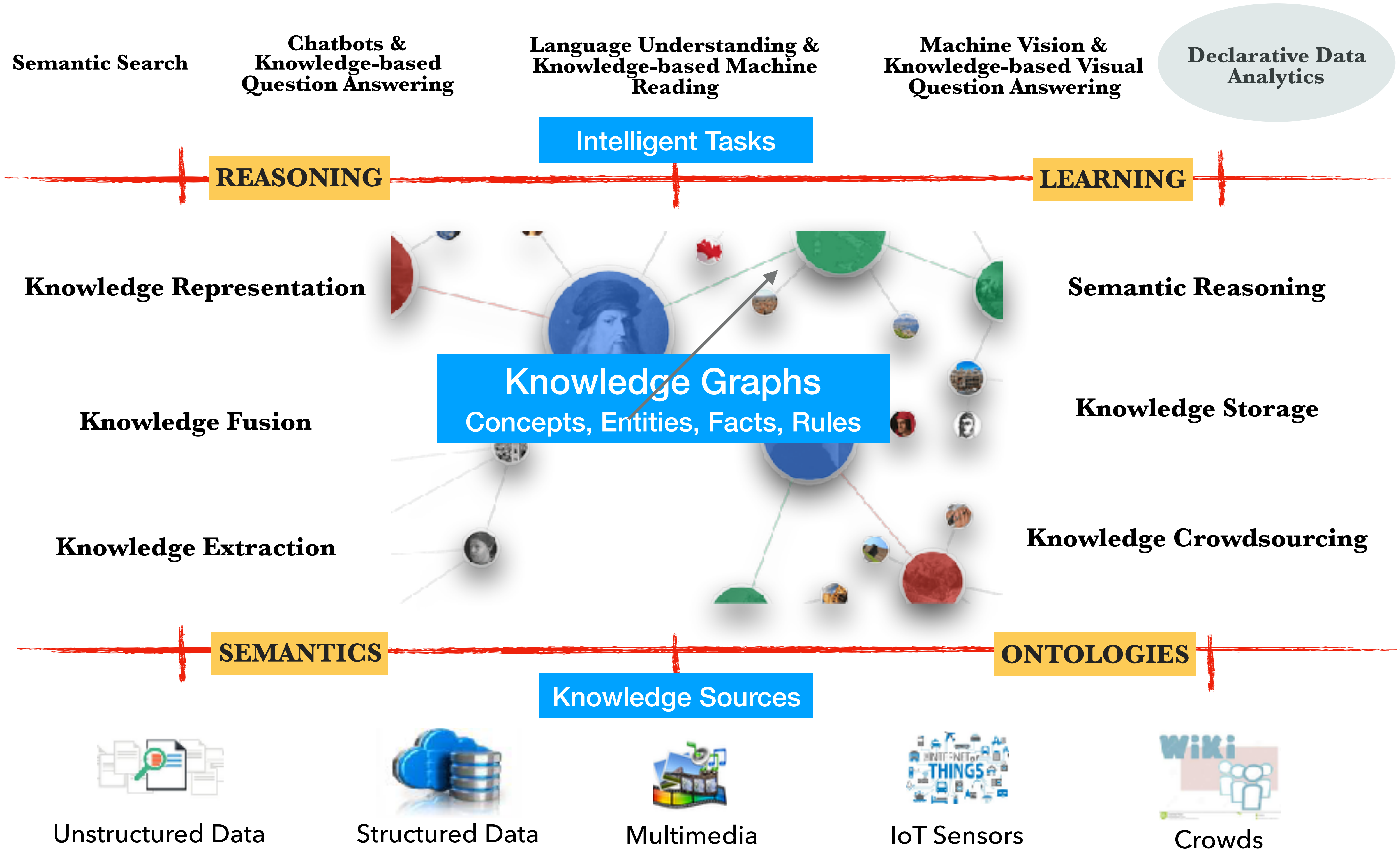


知识图谱 ≠ 专家系统

冯诺依曼曾估计单个个体的大脑中的全量知识需要 $2.4*10^{20}$ 字节存储，知识工程的根本性科学问题是知识完备性问题，即规模化知识获取与处理能力。



知识图谱 ≠ 专家系统



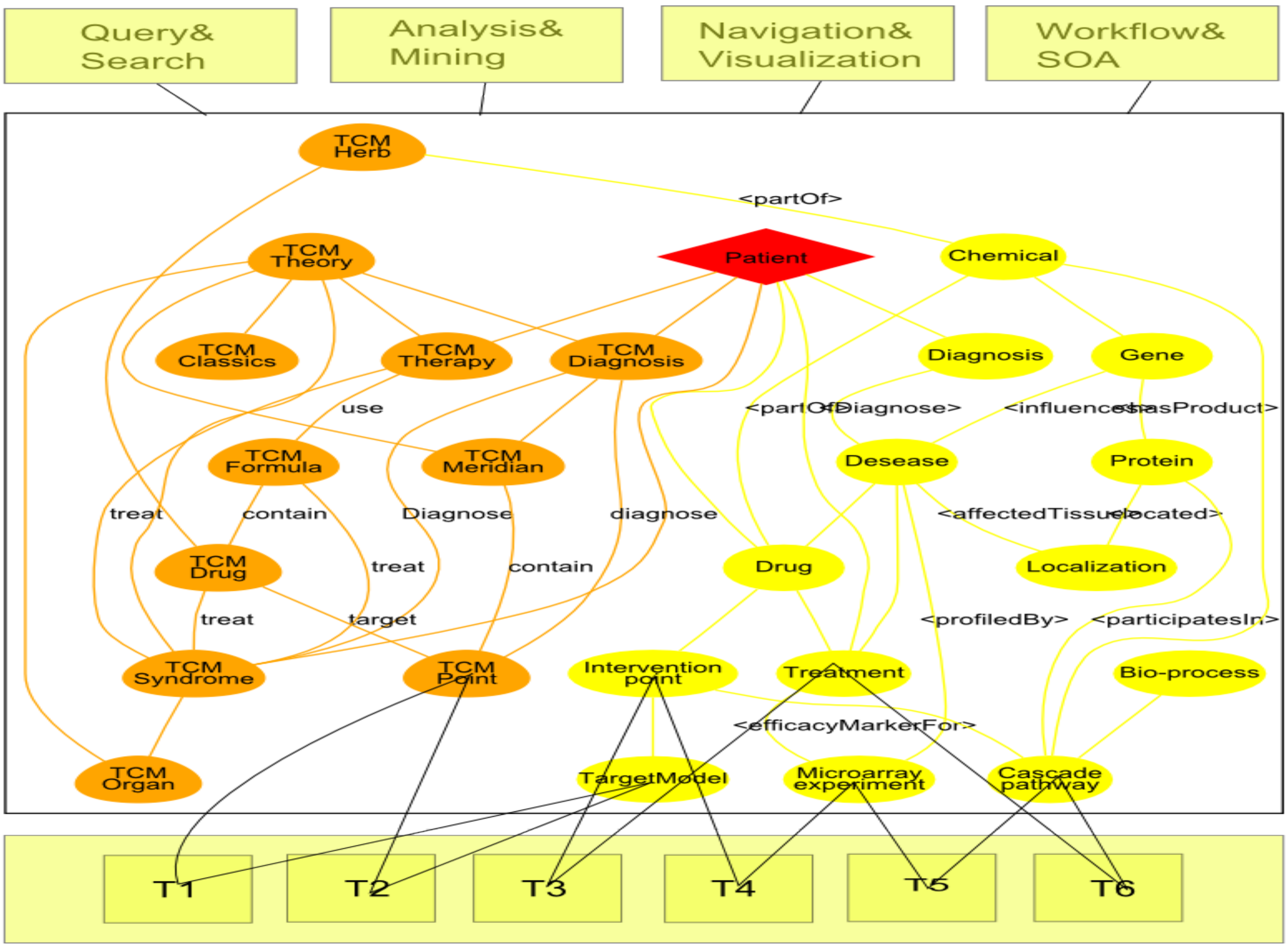
Semantic Search
Chatbots & Knowledge-based Question Answering
Language Understanding & Knowledge-based Machine Reading
Machine Vision & Knowledge-based Visual Question Answering
Declarative Data Analytics
Intelligent Tasks
REASONING
LEARNING
Knowledge Representation
Knowledge Fusion
Knowledge Extraction
Knowledge Graphs
Concepts, Entities, Facts, Rules
Semantic Reasoning
Knowledge Storage
Knowledge Crowdsourcing
SEMANTICS
ONTOLOGIES
Knowledge Sources
Unstructured Data
Structured Data
Multimedia
IoT Sensors
Crowds

领域知识图谱 vs 通用知识图谱

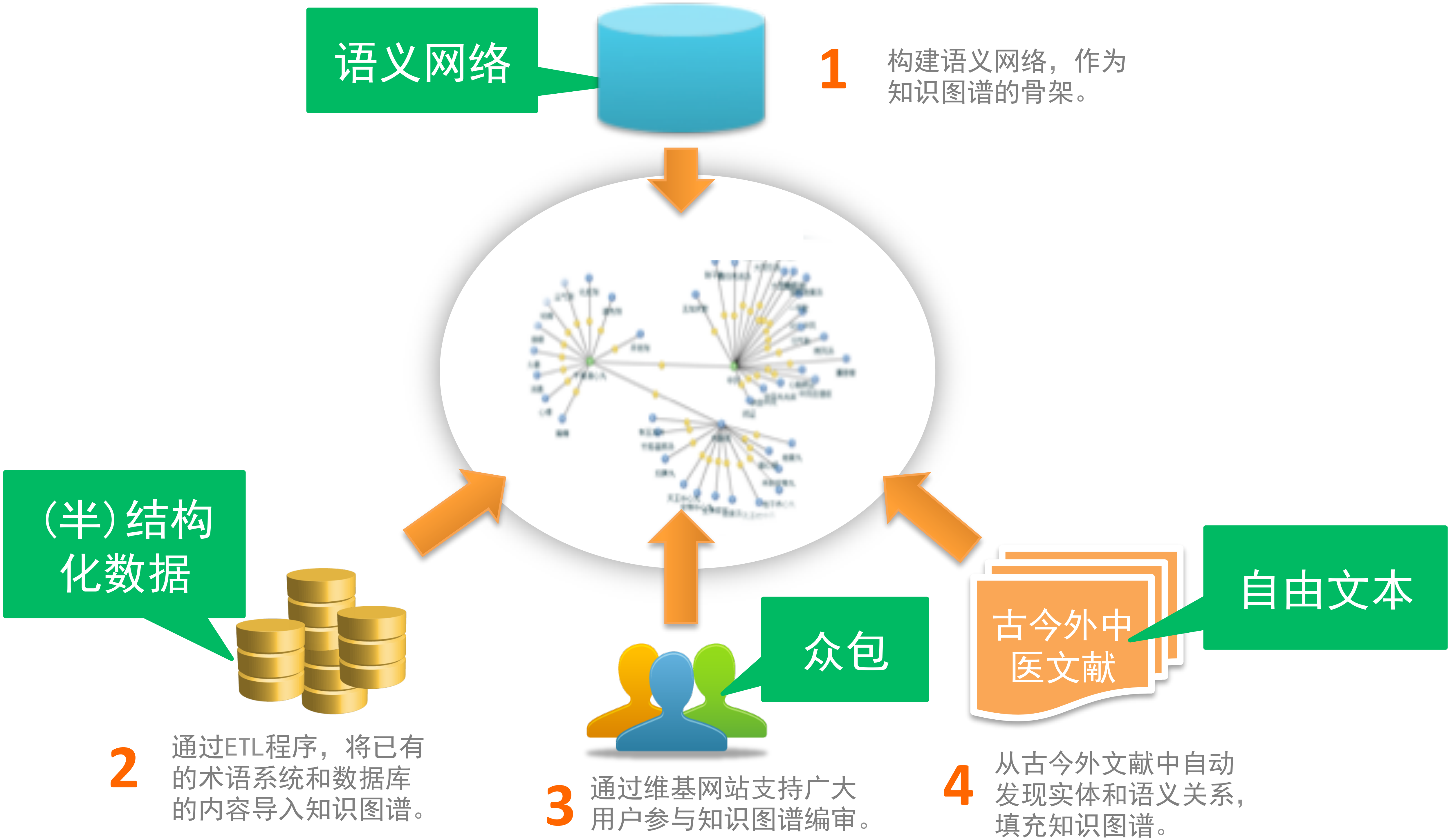
	通用知识图谱	领域知识图谱
知识来源及规模化	以互联网开放数据，如维基百科，或社区众包为主要来源，逐步扩大规模	以领域或企业内部的数据为主要来源，通常要求快速的扩大规模
对知识表示的要求	主要以三元组事实型知识为主	知识结构更加复杂，通常包含较为复杂的本体工程和规则型知识
对知识质量的要求	较多的采用面向开放域的 Web 抽取，对知识抽取质量有一定容忍度	知识抽取的质量要求更高，较多的依靠从企业内部的结构化、非结构化数据进行联合抽取，并依靠人工进行审核校验，保障质量
对知识融合的要求	融合主要起到提升质量的作用	融合多来源的领域数据是扩大构建规模的有效手段
知识的应用形式	一般主要以搜索和问答为主要应用形式，对推理要求较低	应用形式更加全面，除搜索问答外，还通常包括决策分析、业务管理等，并对推理的要求更高，并有较强的可解释性要求。
举例	DBPedia、Yago、百度、谷歌等	电商、医疗、金融、农业、安全等

二、主要实践

早期实践：中医药语义网络



中医药：知识图谱的构建



引自中医科学院于彤博士《中医药知识服务平台研发》

中医药语义搜索（2004-2006）

中医药知识平台

http://www.tcmkb.cn

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。、

引自中医科学院于彤博士《中医药知识服务平台研发》

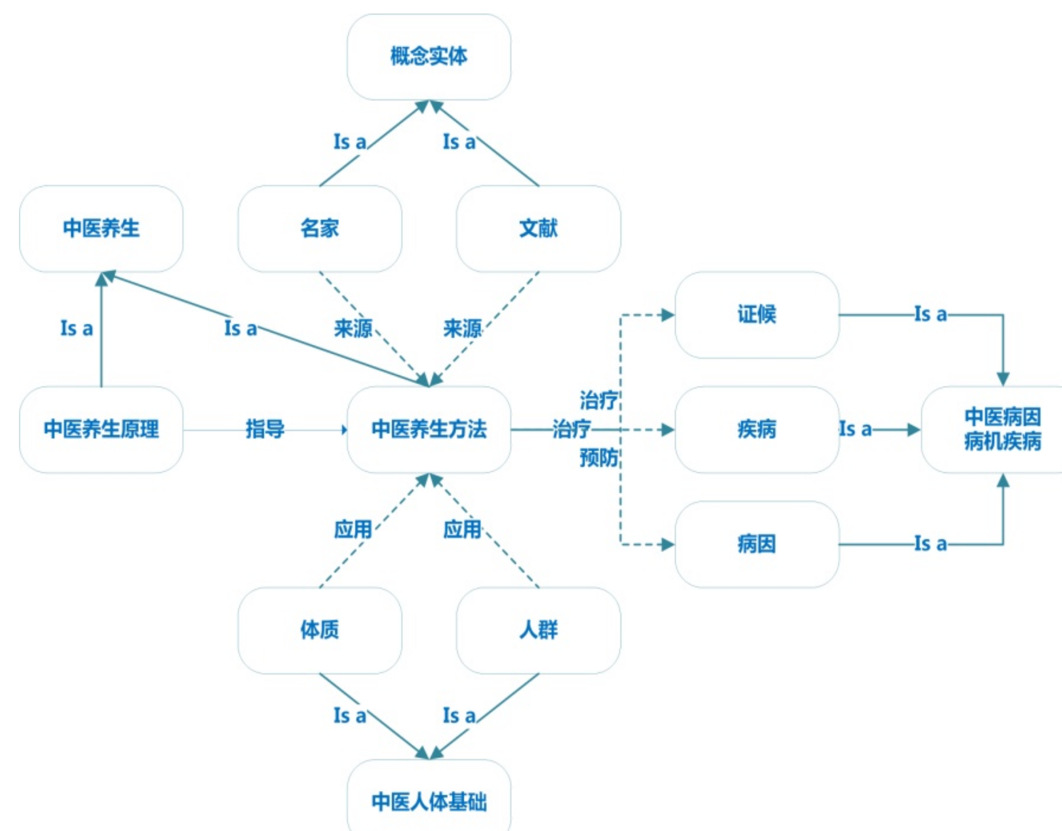
中医药：知识资源

中医临床知识库

整合临床研究、指南、医案、诊疗技术等临床知识资源，辅助临床决策和临床研究。

中医养生知识库

收集中医养生文献，总结名医养生经验，整理中医养生思想、原则、方法和应用，辅助养生知识体系梳理和百姓养生保健。



引自中医科学院于彤博士《中医药知识服务平台研发》

中医药：知识资源

中药知识库

集成中药的化学成分、药理作用、功效、毒性等信息，以及中药研究相关数据（如中药实验），辅助中药研究。

中医方剂知识库

收集整理中医方剂知识，支持中医药工作者对方剂知识进行分析和挖掘，揭示方剂配伍规律。

引自中医科学院于彤博士《中医药知识服务平台研发》

中医药：知识资源

中医医案知识库

全面收集古代、现代、名医医案，建立医案共享机制，辅助医案大数据分析和临床研究。

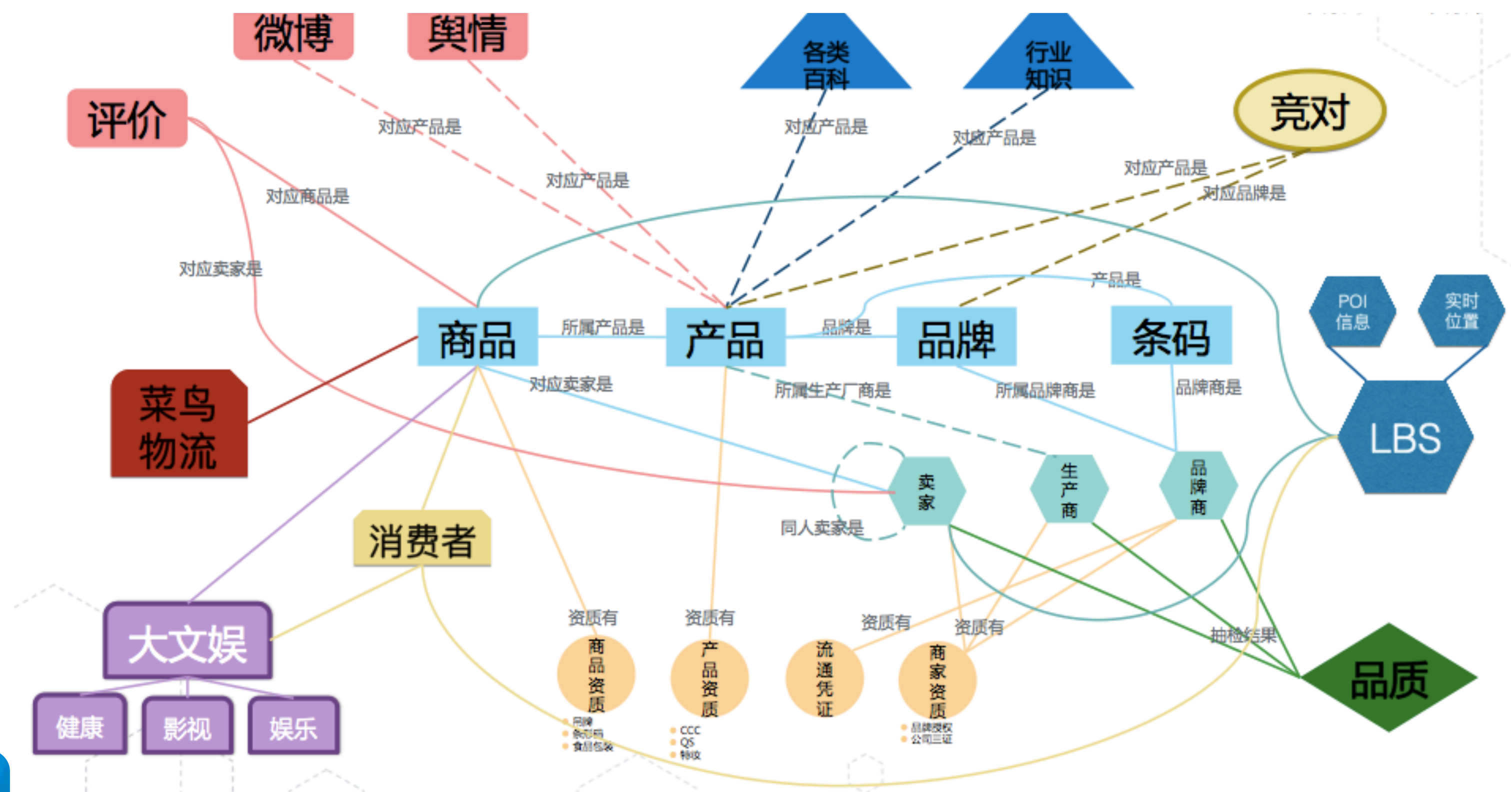
中医药文献库

构建中医药领域的结构化文献库，收集期刊、古籍、外文130余万篇（本），支持文献检索和文本挖掘。

引自中医科学院于彤博士《中医药知识服务平台研发》

近期工作：阿里电商知识图谱

不同于搜索引擎，更强的知识表示需求，搜索、导购、问答、决策多种场景的综合。



百亿条信息实体、百亿条信息关联(边)、海量的规则知识库

在线服务：毫秒级响应

分级存储：图数据库，在线关系数据库，搜索引擎，缓存，全量离线关系数据库

阿里商品知识图谱

商品标准化（商品通）：统一规范的ID体系、类、属性、关系、以及概念层次体系

智能导购　商品画像　智能搜索　平台治理与打假

智能问答　人工智能产品（阿里小蜜、天猫精灵）　热点挖掘与追踪

商品跨市场互通协作：如果商品各市场商品体系标准不统一，商品无论是在国内跨市场、线上线下打通成本都很高；

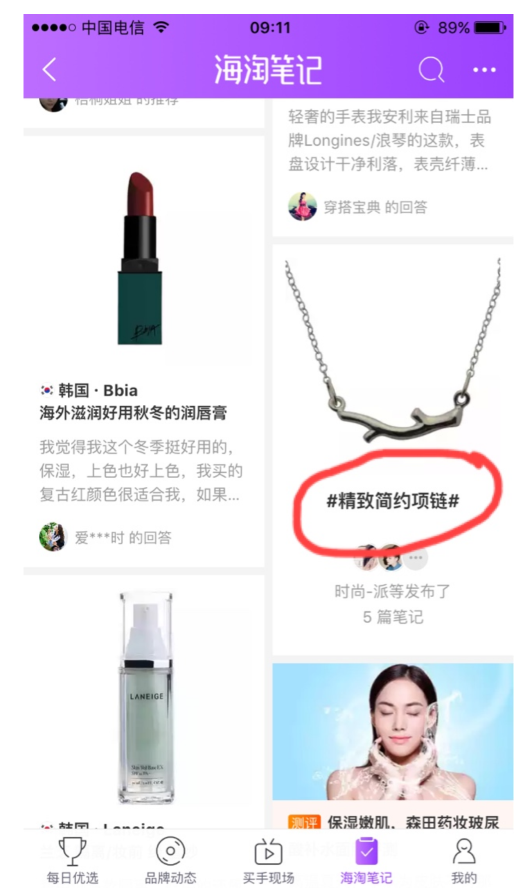
from 阿里巴巴览图

KG辅助智能导购：基于场景的导购

传统基于类目和关键词搜索的导购 → **以场景为中心的智能导购**

横跨类目、组合推荐、高度定制、精准匹配

- 需要更加丰富的商品描述：更多属性，更丰富的商品关联，更全面的知识分类体系
- 场景的描述和定义要灵活可定制，并易于维护和复用
- 场景到商品的匹配要尽肯能精准



三高适宜零食　冬日下午茶　夜间毅行装备　斩男色防干裂的口红

场景知识图谱

- 场景知识图谱用于辅助基于场景的商品导购，即用户在搜索了一个商品时，会自动推荐同一场景下的其他商品。
- 场景知识图谱主要包含场景本体的构建和场景-商品规则知识库的构建。
- 场景本体包含场景词、场景上下位层次关系、同义近义、时序等关系，以及描述场景的五要素补全：when, where, who, how ,why。
    - 场景本体的构建是先通过从商品用户评论中挖掘潜在场景词，并通过算法每周自动添加新挖掘的场景
    - 挖掘出的场景质量低，且无层次结构，通过辅以人工进一步提升质量和构建初始的层次关系。
    - 进一步通过关系推理方法挖掘新的场景上下位、同义及时序关系。
    - 场景的五要素补全更加困难，目前主要考虑引入外部知识库来通过数据融合手段进行自动补全
- 场景规则知识库主要定义场景与商品的挂载关系，构建过程也是先通过数据挖掘候选规则，再通过人工审核校验并扩展，再辅以链接预测与关系推理手段进一步补全。

场景知识库的构建及应用：场景智能导购与自动归集推理

算法自动挖掘

去海边，下午茶，…

知识表示体系

场景规则

下午茶 =
类目 包含 坚果 碳酸饮料 巧克力
产品词 包含 松饼
属性 包含 口味 = 甜

规则运营维护

下午茶 =
类目 包含 坚果 碳酸饮料 巧克力 牛肉干
产品词 包含 松饼 慕斯
属性 包含 口味 = 甜

场景商品关联

智能识别，推理引擎，规则引擎

场景知识审核

外包团队人工审核

一个场景就是一个或一组包含时间、地点、人物、事件等在内的概念描述：Semantic Query or Undefined Class。

场景就是领域知识，运营人员需要更加灵活可定制的方式定义、重构、扩展、组合、复用这些场景描述：场景要能批量生产、能形成场景规则库沉淀领域知识、自动根据热点推荐新的场景。

一个场景本质上由一组类属性描述组成，从场景到相关商品的查询搜索可以刻画为自动分类推理的问题（Subsumption + Instance Checking）。

知识图谱的扩展、任意商品属性的添加、新的数据关系的增加，都不应该影响场景的描述定义方式。场景的定义和知识图谱的松耦合将极大的降低数据提供方的工作负担。

KG辅助决策分析：利用知识洞察消费趋势

去年YSL口红，因为韩剧《太阳的后裔》大火，行业运营观察到口红GMV提升，却未洞察背后的原因，导致错过了打爆该商品的机会。

太阳的后裔

领域知识

最新韩剧相关服装会热卖

热点场景推荐
智能选品
竞品分析

反过来，从新出现的热点回溯找原因（*Abductive Reasoning*）

KG辅助智能管控：不一致性推理

怎样更加灵活方便的检测商品描述之间的不一致性：Inconsistence Checking.

增加商品描述和维护不一致规则库。

这种不一致性可能存在于国家规定与商家提供的产品描述之间，可能存在于不同商家之间的相同商品之间（错误、虚假的描述），可能存在于同一商品不同信息来源中。

KG辅助智能问答：查询重写与问答推理

利用推理对自然语言查询或逻辑查询进行逻辑扩展，放大查询和问答的结果

秋季韩版修身外套

风衣 *is* 外套

秋季 *subsumedBy* 秋冬季

修身 *equalTo* 瘦身

叠加知识推理
生成更多查询

秋冬季韩版修身风衣

秋季韩版修身风衣

秋冬季韩版瘦身风衣

（1）一级翻译器

将自然语言转换成逻辑语言

“产地为某核污染区域的食品”

规则/神经网络模型

（2）推理单元

基于商品知识做逻辑展开

∀ x: 食物(x) ⊓ (产地（x，核污染区））

利用同义词和地理位置图谱展开

∀ x: 食物(x)
⊓ (∀ y: 同义词(y，产地))
(x, (∀ z: 包括下位实体(核污染区, z)))

同款：∀ t: (∃ c:属于产品 (x, c) ⊓属于产品(t,c))

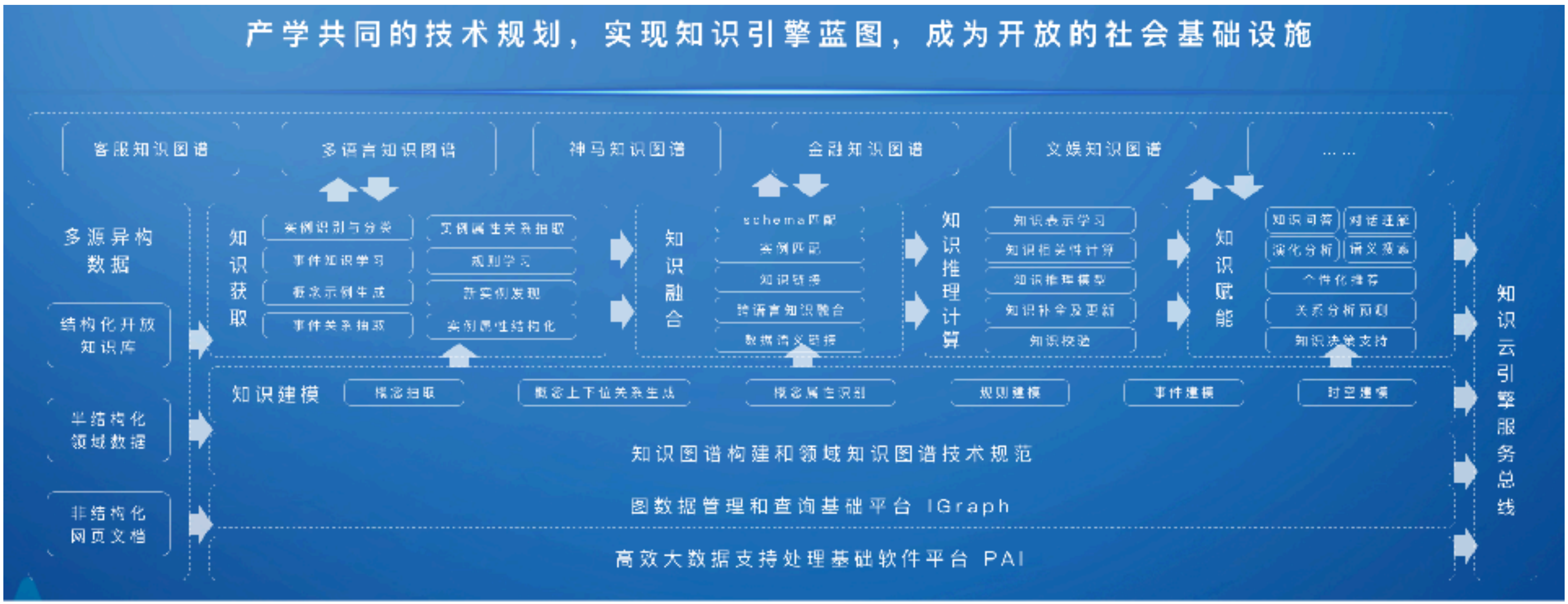
（3）二级翻译器

逻辑语言转换为数据库语言

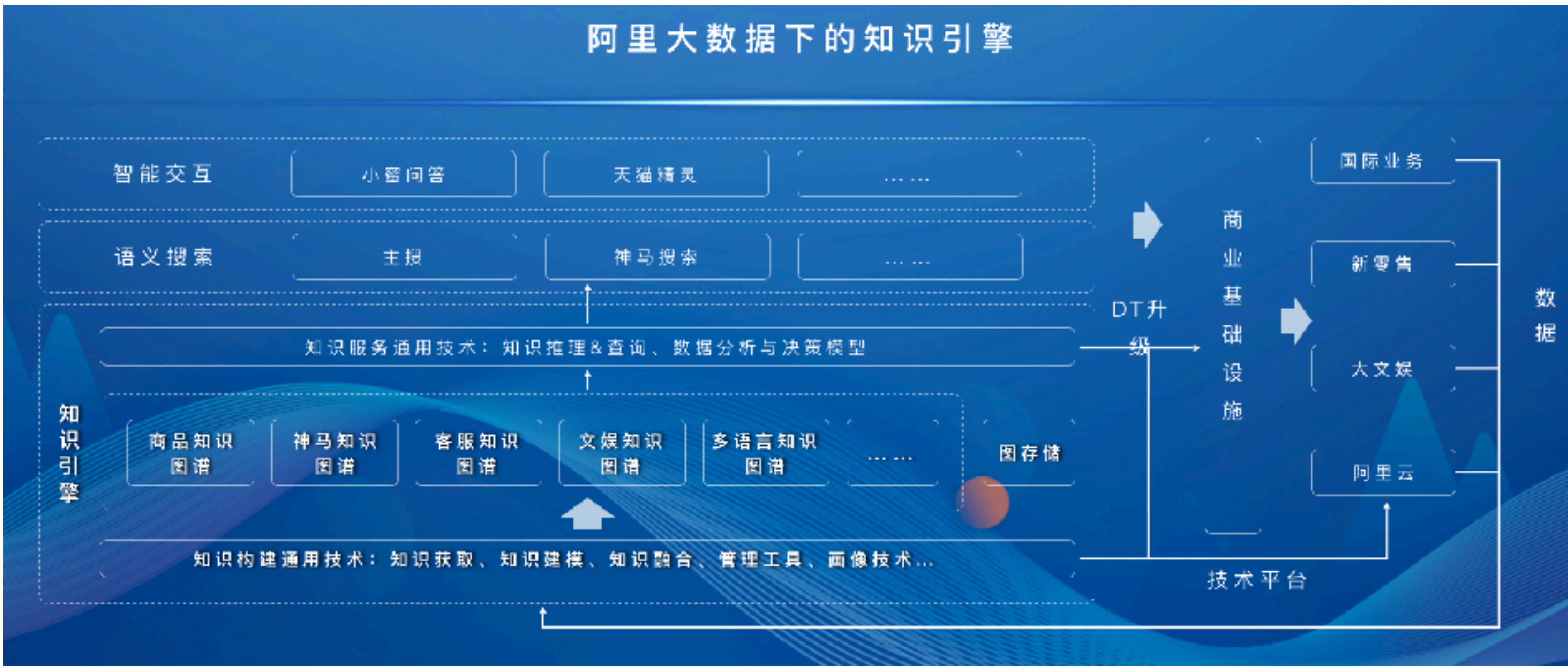
Select * from item where
xxx or xxx

输出满足条件的商品列表

阿里巴巴藏经阁（知识引擎）专项研究计划



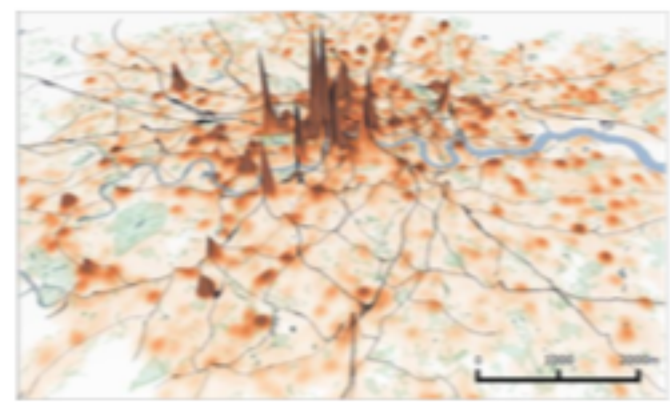
阿里巴巴藏经阁（知识引擎）专项研究计划



城市知识图谱： Dynamic KG and Knowledge Graph for Predictive Analytics

- Stream learning:
  - Extracting and predicting knowledge from temporal evolution of data
    - Learning from the past time-series, and predict the missing or future
  - Wide applications

Bus delay and traffic congestion forecasting



AIR QUALITY INDEX

Air Quality Index (AQI) Values	Levels of Health Concern
0 to 50	Good
51-100	Moderate
101-150	Unhealthy for Sensitive Groups
151-200	Unhealthy
201-300	Very Unhealthy
301 to 500	Hazardous

Spatial inference and time-series forecasting of Air Quality Index (AQI)

Jiaoyan Chen, Freddy Lécué, Jeff Z. Pan, Huajun Chen: Learning from Ontology Streams with Semantic Concept Drift. IJCAI 2017: 957-963

Structured Knowledge Base as Prior Knowledge to Improve Urban Data Analysis. International Journal of Geo-Information. 2018

Concept Drift

- Concept drift in stream learning
  - Unexpected changes in data distribution [Coble and Cook, 2000]
    - The model is built on old data and becomes inconsistent with the new data as time passes.
  - Current solutions [Gama et al. 2014]:
    - Recent sample priority
    - Weighted model ensemble
    - Dynamic sliding window
    - . . .

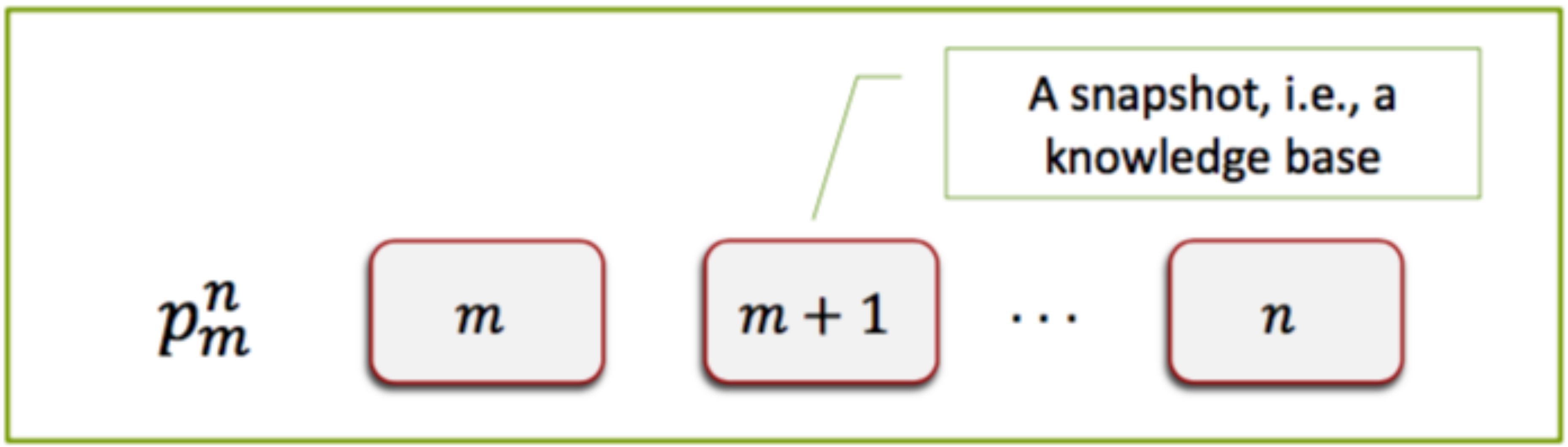
Assumption: the neighboring data are similar.
They can manage gradual changes, but fail for sudden, abrupt changes

Coble, J., & Cook, D. J. (2000, May). Real-Time Learning when Concepts Shift. In *FLAIRS Conference* (pp. 192-196).

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, *46*(4), 44.

Streaming Knowledge Graph

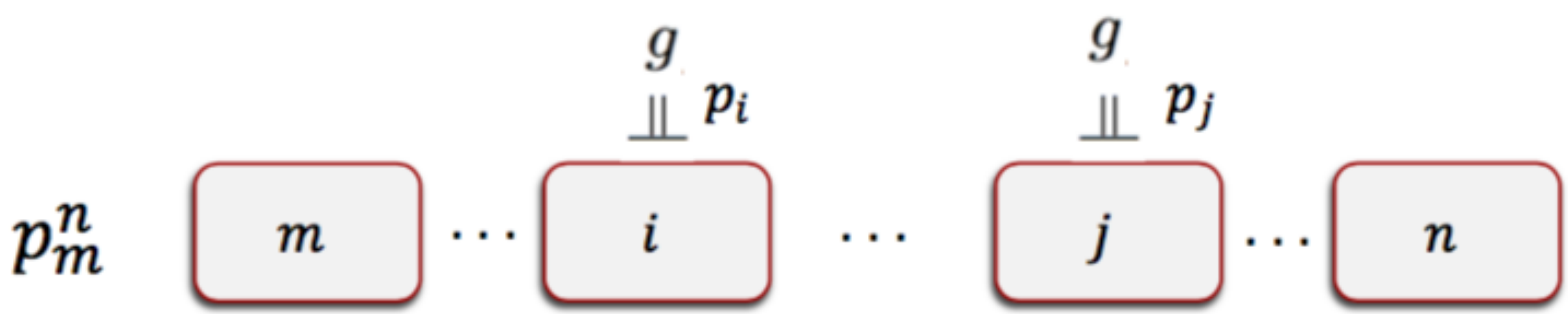
- TBox $\mathcal{T}$ (terminological component) and ABox $\mathcal{A}$ (assertion axioms)
- DL $\mathcal{L}$ Ontology Stream



Definition 1. *(DL $\mathcal{L}$ Ontology Stream)*
A DL $\mathcal{L}$ ontology stream $\mathcal{P}_m^n$ from point of time m to point of time n is a sequence of (sets of) Abox axioms $(\mathcal{P}_m^n(m), \mathcal{P}_m^n(m+1), \cdots, \mathcal{P}_m^n(n))$ with respect to a static TBox $\mathcal{T}$ in a DL $\mathcal{L}$ where $m, n \in \mathbb{N}$ and $m < n$.

Streaming Knowledge Graph

- Prediction Change: two snapshots have large enough probability difference for some entailments



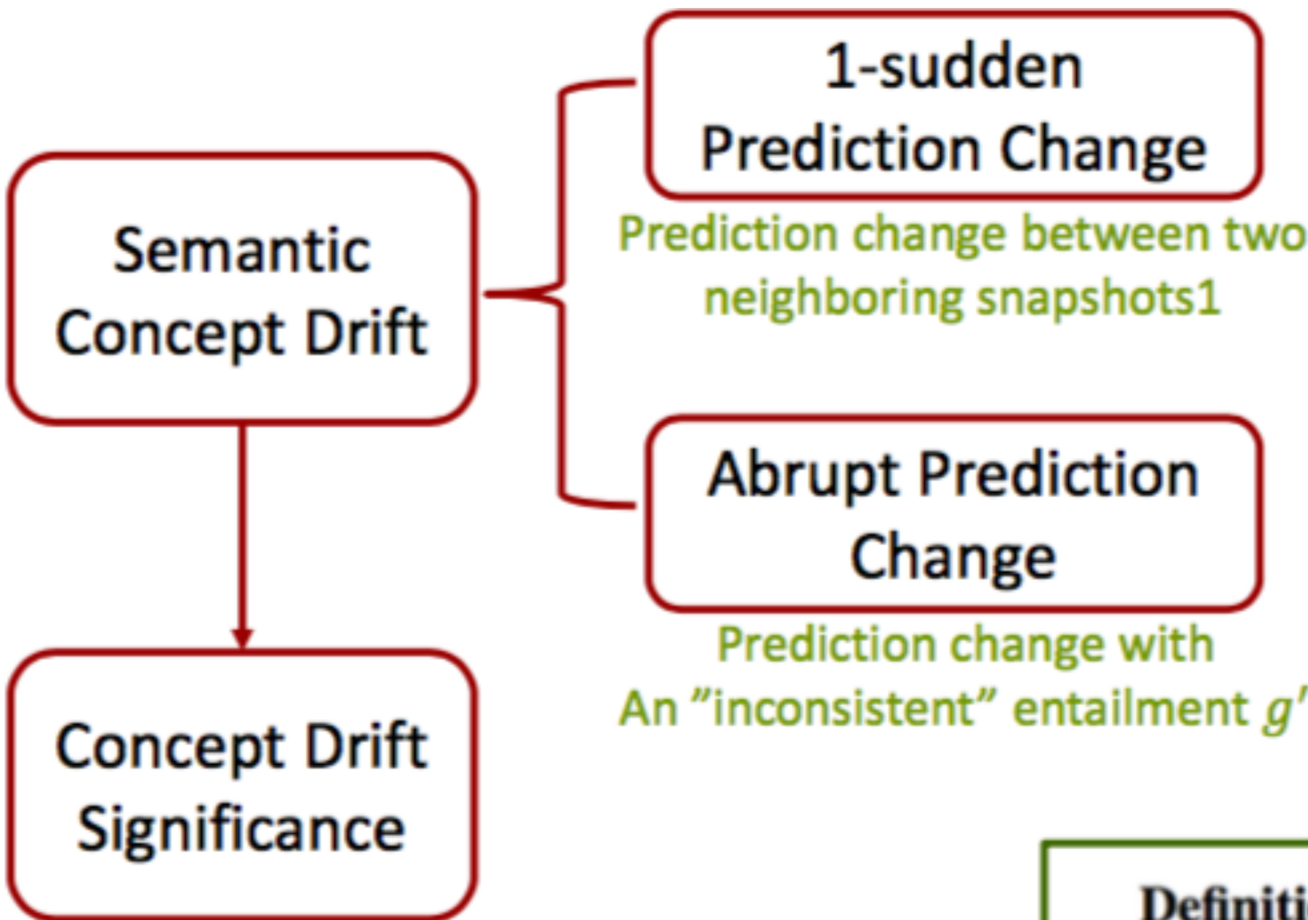
- If there is an entailment g that $||p_i - p_j||$ is large, snapshots i and j contain a prediction change.
- g is called an evidence entailment.
- The set of all evidence entailments is denoted as $\mathbb{C}_{|\mathcal{T}\cup\mathcal{A}}(S_0^n, i, j, \varepsilon)$

Definition 4. ***(Prediction Change)*** *Let S_0^n be a stream; $\mathcal{T}$, $\mathcal{A}$ and $\mathcal{G}$ be TBox, Abox and its entailments. A prediction change in S_0^n is ocuring between time i and j in $[0, n]$ with respect to $\mathcal{T}$, $\mathcal{A}$ and its entailments iff:*

$$\exists g \in \mathcal{G} : \|p_{|\mathcal{T}\cup\mathcal{A}}(S_0^n(i) \models g) - p_{|\mathcal{T}\cup\mathcal{A}}(S_0^n(j) \models g)\| \geq \varepsilon \quad (30)$$

where $\varepsilon \in (0, 1]$ is a variable bounding the difference of estimation, $\|v\|$ refers to the absolute value of v, and $j > i$

Streaming Knowledge Graph



An indicator to evaluate the severity of a semantic concept drift, i.e., the percentage of evidence entailments

Definition 5. *(α-Sudden Prediction Change)* *A prediction change at point of time i in stream $\mathcal{S}_0^n$, satisfying* (30), *is defined as α-sudden, with $\alpha \in (0, n-i]$ iff $j = i + \alpha$.*

Definition 6. *(Abrupt Prediction Change)* *A prediction change, satisfying* (30), *is abrupt iff $\exists g' \in \mathcal{G}$ s.t.*

$$\mathcal{T} \cup \mathcal{A} \cup g \cup g' \bigcup_{k=0}^{\max\{i,j\}} \mathcal{S}_0^n(k) \models \bot \tag{31}$$

where $\bigcup_{k=0}^{\max\{i,j\}} \mathcal{S}_0^n(k)$ captures all axioms from any snapshot $\mathcal{S}_0^n(k)$ of stream $\mathcal{S}_0^n$ with $k \in [0, \max\{i,j\}]$.

Definition 8. *(Semantic Concept Drift Significance)* *The significance of a semantic concept drift, defined between points of time $i \in (0, n)$ and $i+1$ of $\mathcal{S}_0^n$ with ε, $\mathcal{T}$, $\mathcal{A}$, $\mathcal{G}$ as difference, TBox, ABox, and entailments, is:*

$$\sigma_{|\mathcal{T}\cup\mathcal{A}}(\mathcal{S}_0^n, i, \varepsilon) \doteq \frac{|\mathbb{C}_{|\mathcal{T}\cup\mathcal{A}}(\mathcal{S}_0^n, i, i+1, \varepsilon)|}{|\{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n(i) \models g \vee \mathcal{T} \cup \mathcal{S}_0^n(i+1) \models g\}|} \tag{32}$$

where the expression in between | refers to its cardinality.

Consistent Vector

- Consistent vector (weights):
  - Encodes percentage of invariant entailments;
  - Positive if two snapshots are consistent, negative elsewise

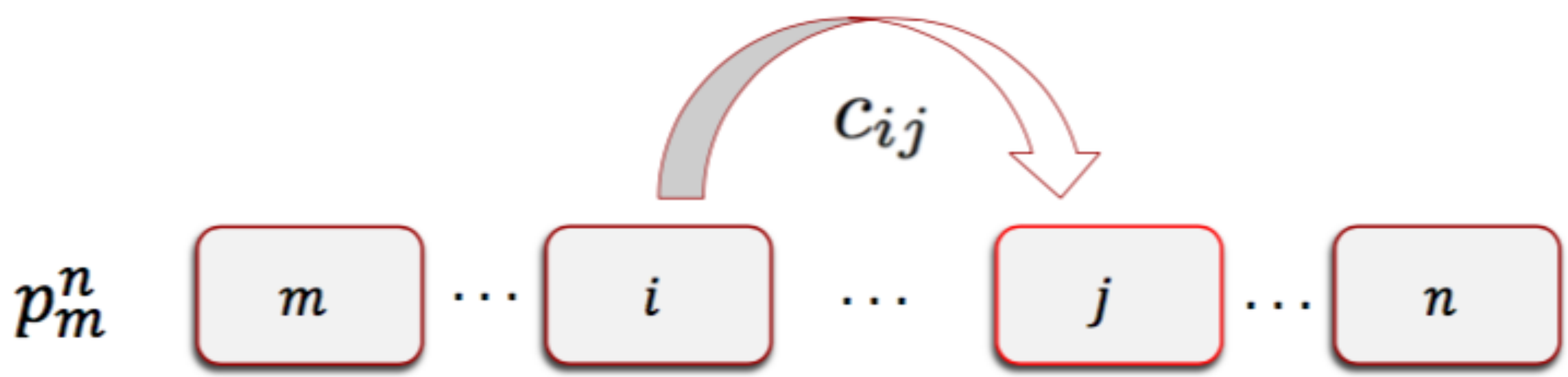
Definition 9. *(Consistency Vector)*
A consistency vector of snapshot $\mathcal{S}_0^n(i)$ *in* $\mathcal{S}_0^n$*, denoted by* $\mathbf{c}_i$*, is defined* $\forall j \in [0, n]$ *by* c_{ij} *if* $i < j$*;* c_{ji} *otherwise such that:*

$$c_{ij} \doteq \begin{cases} \frac{|\mathcal{G}_{inv}^{i,j}|}{|\mathcal{G}_{new}^{i,j}|+|\mathcal{G}_{inv}^{i,j}|+|\mathcal{G}_{obs}^{i,j}|} & \text{if } \mathcal{T} \cup \mathcal{S}_0^n(i) \cup \mathcal{S}_0^n(j) \not\models \perp \\ \frac{|\mathcal{G}_{inv}^{i,j}|}{|\mathcal{G}_{new}^{i,j}|+|\mathcal{G}_{inv}^{i,j}|+|\mathcal{G}_{obs}^{i,j}|} - 1 & \text{otherwise} \end{cases} \quad (33)$$

$\mathcal{G}_{new}^{i,j}$ new entailments

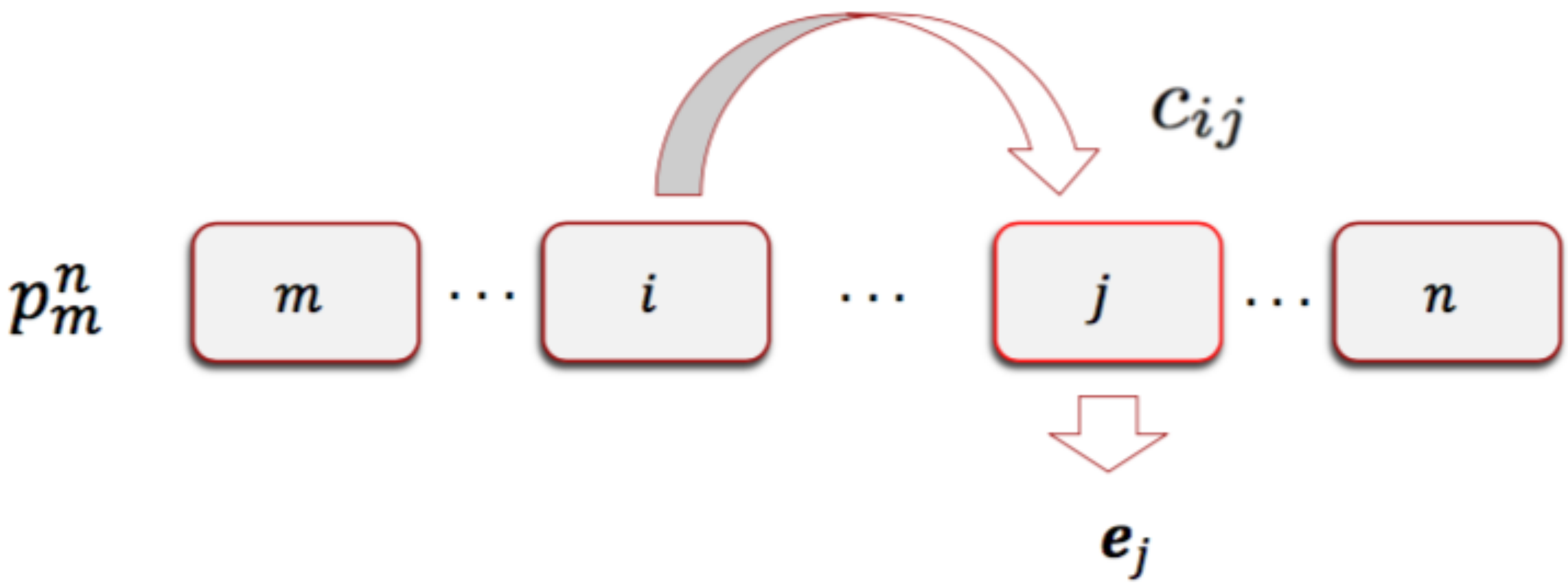
$\mathcal{G}_{inv}^{i,j}$ invariant entailments

$\mathcal{G}_{obs}^{i,j}$ obsolete entailments



Entailment Vector

- Entailment vector : more than features
  - Observed variables + inferred classification entailments



Sampling and Learning

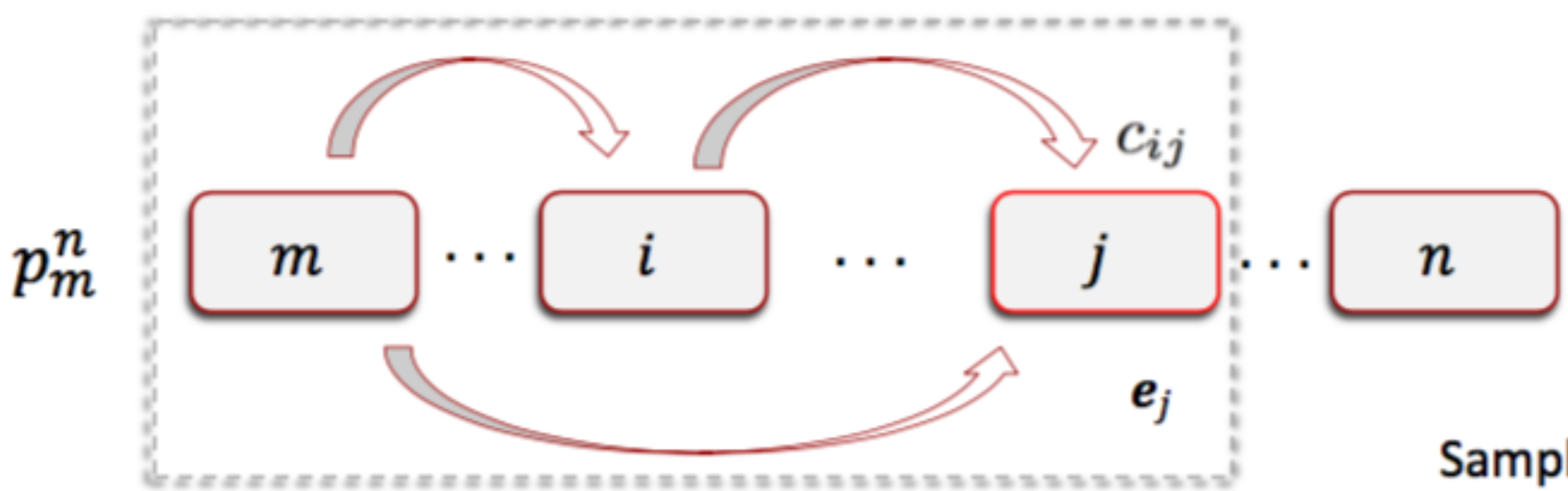
- Linear model: $f(\boldsymbol{e}_i) = a^T \boldsymbol{e}_i + b$, where $a \in \boldsymbol{R}^N$ and b $\in \boldsymbol{R}$ are model parameter to learn

- Samples: $\{(\boldsymbol{e}_i, \boldsymbol{v}(g_i))|i = 1,2, \dots, n\}$, where $\boldsymbol{e}_i$ represents the entailment vector for S_0^n(i), $\boldsymbol{v}(g_i)$ is in [0,1], capturing the estimation of entailment g_i

- Object function with regularization:

$$O_j(a,b) \doteq \sum_{i=1}^{\kappa} \omega_{ij} L(\mathbf{v}(g_i), f(\mathbf{e}_i)) + \alpha R(a), \qquad (35)$$

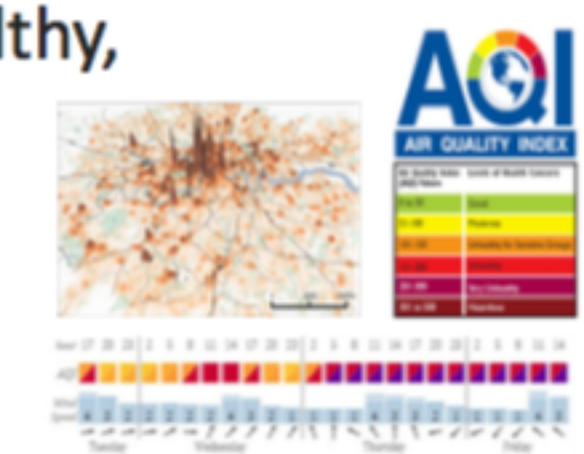
$$\omega_{ij} \doteq \begin{cases} 0, & \text{if } c_{ij} > 0 \\ -c_{ij} & \text{else,} \end{cases} \qquad (36) \qquad \omega_{ij} \doteq \begin{cases} 0, & \text{if } c_{ij} < 0 \\ c_{ij} & \text{else,} \end{cases} \qquad (37)$$



Sampling and Learning

Experiments

- Beijing Air Quality (BAQ) Context:
  - Streams: air pollutants and meteorology elements (B_1), wind speed (B_2), humidity (B_3)
  - 6 Classes: Good, Moderate, Unhealthy, Very Unhealthy, Hazardous, Emergent



- Dublin Bus Delay (DBD) Context:
  - Streams: bus GPS location, delay, congestion status (D_1), weather condition (D_2), road incidents (D_3)
  - 5 Classes: Free, Low, Moderate, Heavy, Stopped

Experiments

City	$\mathcal{ID}$: $Features$	Δ=6 hours		Δ=12 hours		Δ=18 hours	
		✗	✓	✗	✓	✗	✓
Beijing	$\mathcal{B}_1 : B_1$	.351	.398	.344	.441	.261	.342
	$\mathcal{B}_2 : B_1 + B_2$	.398	.449	.350	.453	.279	.371
	$\mathcal{B}_3 : B_1 + B_3$	.421	.508	.373	.459	.282	.379
	$\mathcal{B}_4 : B_1 + B_2 + B_3$	.501	.611	.389	.478	.286	.393
Average Improvement (%)		17.206		25.890		33.954	
Dublin	$\mathcal{D}_1 : D_1$	.455	.514	.387	.441	.321	.387
	$\mathcal{D}_2 : D_1 + D_2$	.534	.688	.499	.553	.361	.497
	$\mathcal{D}_3 : D_1 + D_3$	.601	.701	.513	.645	.371	.547
	$\mathcal{D}_4 : D_1 + D_2 + D_3$	.659	.921	.533	.834	.601	.745
Average Improvement (%)		24.550		26.744		32.408	

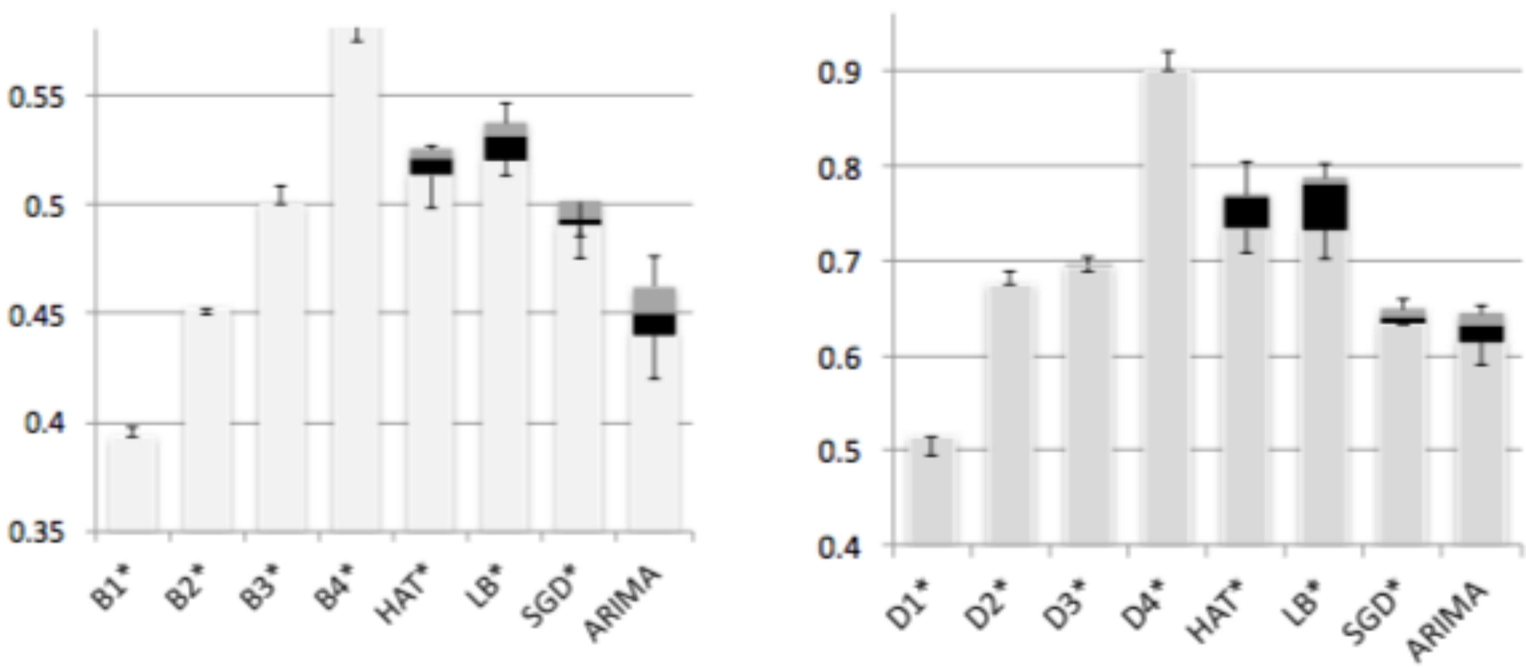
- ✓ columns: using semantic embedding (consistent vector)
- ✗ columns: NOT using semantic embedding

- B_1, B_2, B_3, B_4: data streams for air quality forecasting
- D_1, D_2, D_3, D_4: data streams for bus delay forecasting

- Semantic embedding (✓ columns) has an average improvement of 26.6% over all forecasting tasks. Even for long-term forecasting task ($\Delta = 18$ hours), the improvement is 33.1%

Experiments

- Our approaches: B_i and D_i, $i = 1,2,3,4$, each of which represents a stream subset
- Baselines (without semantic embedding):
  - Stochastic Gradient Descent (SGD)
  - Auto-Regressive Integrated Moving Average (ARIMA);
  - Recentness-based: Hoeffding Adaptive Tree (HAT) and Leveraging Bagging (LB)



Accuracy of BAQ (Left, Beijing) and DBD (Right, Dublin)

- The more features used, the more robust (accurate) the models are (e.g., B_4>B_3>B_2>B_1)
- Our state-of-the-art B_4 and D_4 outperform the baselines
- Semantic consistence matters more than recentness (i.e., HAT and LB) during stream learning

What is Transfer Learning?

I want to train a prediction model to predict departure delay of my flight from LAX to JFK between 4pm and 6pm, but only have historical 112 records …

B6 (JetBlue Airways)

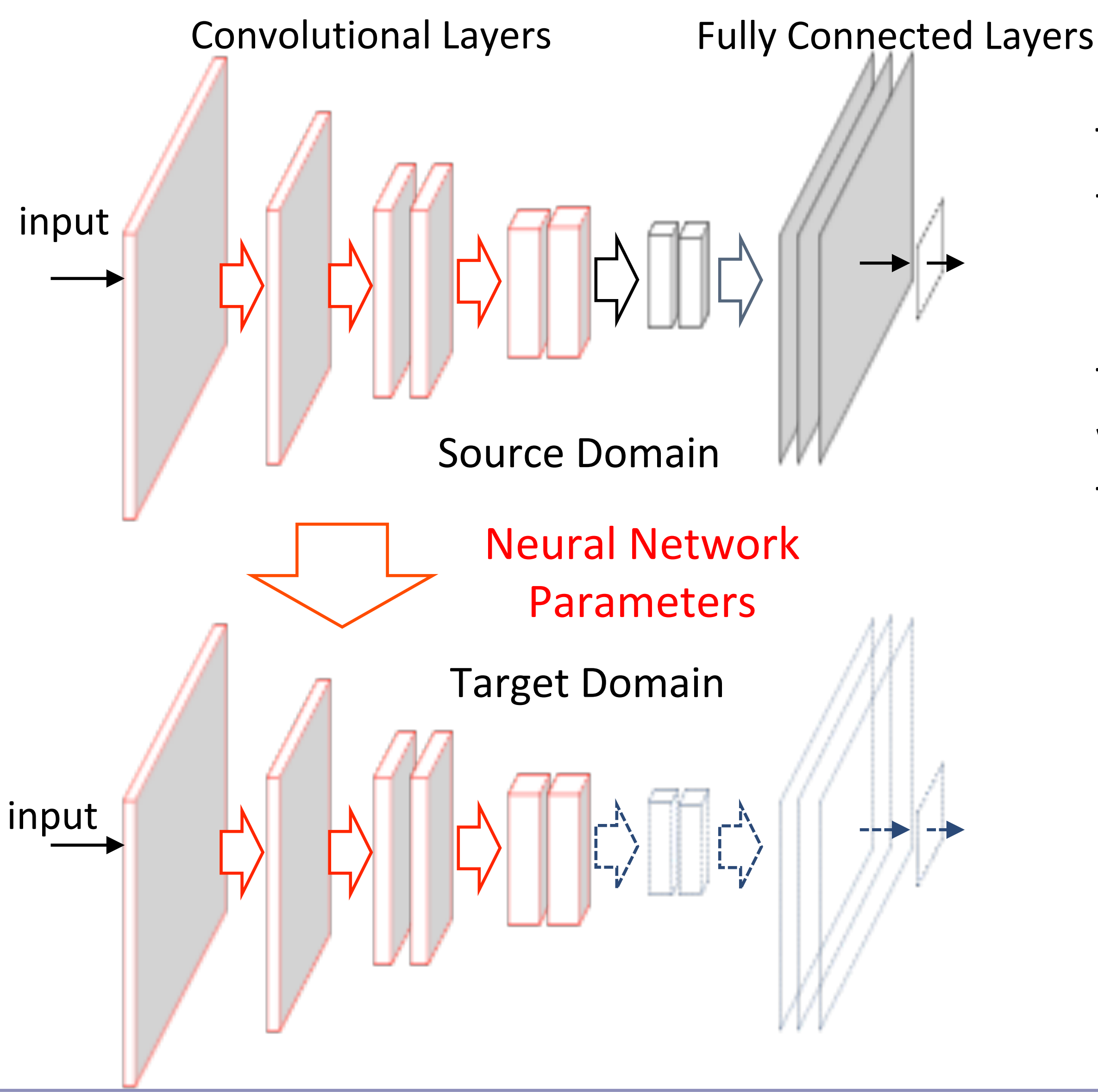
I have thousands, why not use my data, or just take my model ?!

DL (Delta Airline)

- Transfer Learning (TL)
  - Machine learning (ML) algorithms that utilize data/model/parameters of a source domain to help train a prediction model for a target domain
- Domain
  - Identified by data (input distribution) and task (output distribution)

Jiaoyan Chen, Freddy Lécué, Jeff Z. Pan, Ian Horrocks, Huajun Chen. Transfer Learning Explanation with Ontologies. International Conference on the Principles of Knowledge Representation and Reasoning . KR2018.

CNN Feature Transfer



To measure the transferability of CNN features, we first measure Feature Specificity Index (FSI) and Feature Generality Index (FGI) by analyzing the model's performance in target domain, with and without transfer, with and without fine tuning of the transferred layers [1]. Then we define **Feature Transferability Index** (FTI):

$$FTI = \frac{\omega_1 \cdot FGI + \omega_2 \cdot (-FSI)}{\omega_1 + \omega_2}$$

The higher FTI, the higher feature transferability

Reference [1] Yosinski, Jason, et al. "How transferable are features in deep neural networks?." NIPS. 2014.

Knowledge Modeling

- We model each ML sample as a **learning sample ontology (LSO)**, denoted by $\mathcal{O} = \ll \mathcal{T}, \mathcal{A} >, \mathcal{S} >$, where $\mathcal{T}$ represents TBox (terminology axioms), $\mathcal{A}$ represents ABox (assertion axioms) and $\mathcal{S}$ represents the annotation (property-value pairs).

- **Example 1.** ***(An LSO and A Learning Domain on Departure Flights)***
    - *The following axioms are an LSO with the annotation property-value pairs* $\mathcal{S}$:= {date: 01/01/2018, carrier: DL, origin: LAX, dest: JFK }. *It corresponds to one ML input sample that is related to a flight departure from airport LAX to JFK by DL on 01/01/2018.*
    - *LSOs annotated by* $\mathcal{S}'$:= {carrier: DL, origin: LAX, dest: JFK} and *the target entailment* g^t := Delayed(d) *whose existence is to be predicted are a learning domain.* $\mathcal{S}'$ *is the annotation of the learning domain.*

$$Departure \sqcap \exists hasDelMin.\{Pos\} \sqsubseteq Delayed \quad (1)$$
$$Departure \sqcap \exists hasDelMin.\{Neg\} \sqsubseteq OnTime \quad (2)$$
$$hasCar \circ hasCarHub \sqsubseteq hasDepHub \quad (3)$$
$$hasNebApt \circ hasRecDep \sqsubseteq hasRecNebDep \quad (4)$$
$$Departure \sqcap \exists hasOri.\{CA\} \sqcap \exists hasDes.\{CA\} \sqsubseteq inCADep \quad (5)$$

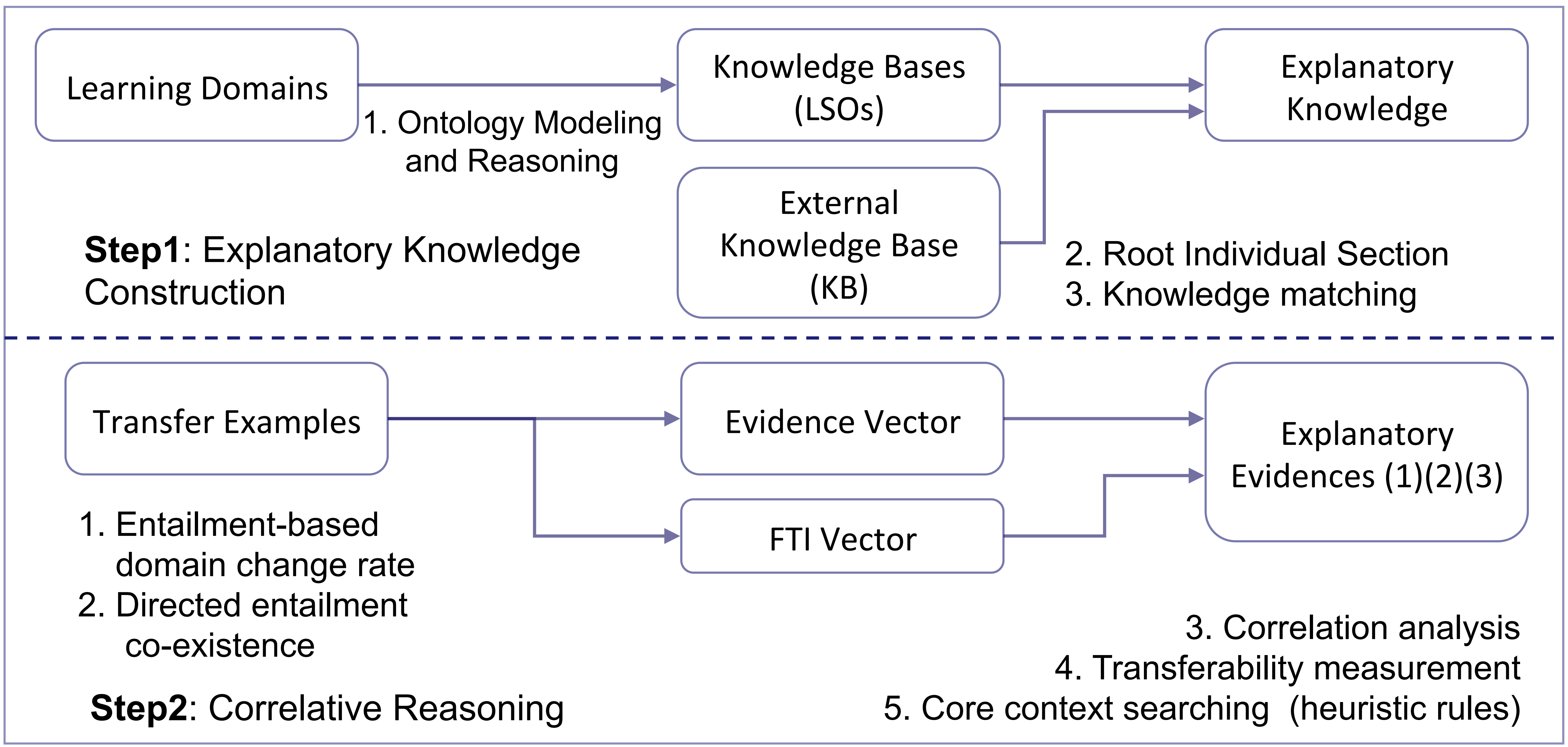
TBox Axioms e.g., concepts and roles (DL $\mathcal{EL}$++)

$Airport(LAX)$	(6)	$locatedIn(LAX, CA)$	(7)
$Carrier(DL)$	(8)	$Departure(d)$	(9)
$hasDelMin(d, Pos)$	(10)	$hasWea(d, wea)$	(11)
$hasOri(d, LAX)$	(12)	$hasCar(d, DL)$	(13)
$Airport(JFK)$	(14)	$hasDes(d, JFK)$	(15)
$LAX = ori$	(16)	$DL = car$	(17)
$hasRecDep(d, d_1)$	(18)	$hasCar(d_1, MU)$	(19)
$hasRecDep(d, d_2)$	(20)	$hasCar(d_2, AA)$	(21)
$Delayed(d)$	(22)	$HeavySnow(wea)$	(23)

ABox Axioms e.g., observed facts

Inferred Entailments

Method


Learning Domains
1. Ontology Modeling and Reasoning
Knowledge Bases (LSOs)
Explanatory Knowledge
External Knowledge Base (KB)
Step1: Explanatory Knowledge Construction
2. Root Individual Section
3. Knowledge matching
Transfer Examples
Evidence Vector
FTI Vector
Explanatory Evidences (1)(2)(3)
1. Entailment-based domain change rate
2. Directed entailment co-existence
Step2: Correlative Reasoning
3. Correlation analysis
4. Transferability measurement
5. Core context searching (heuristic rules)

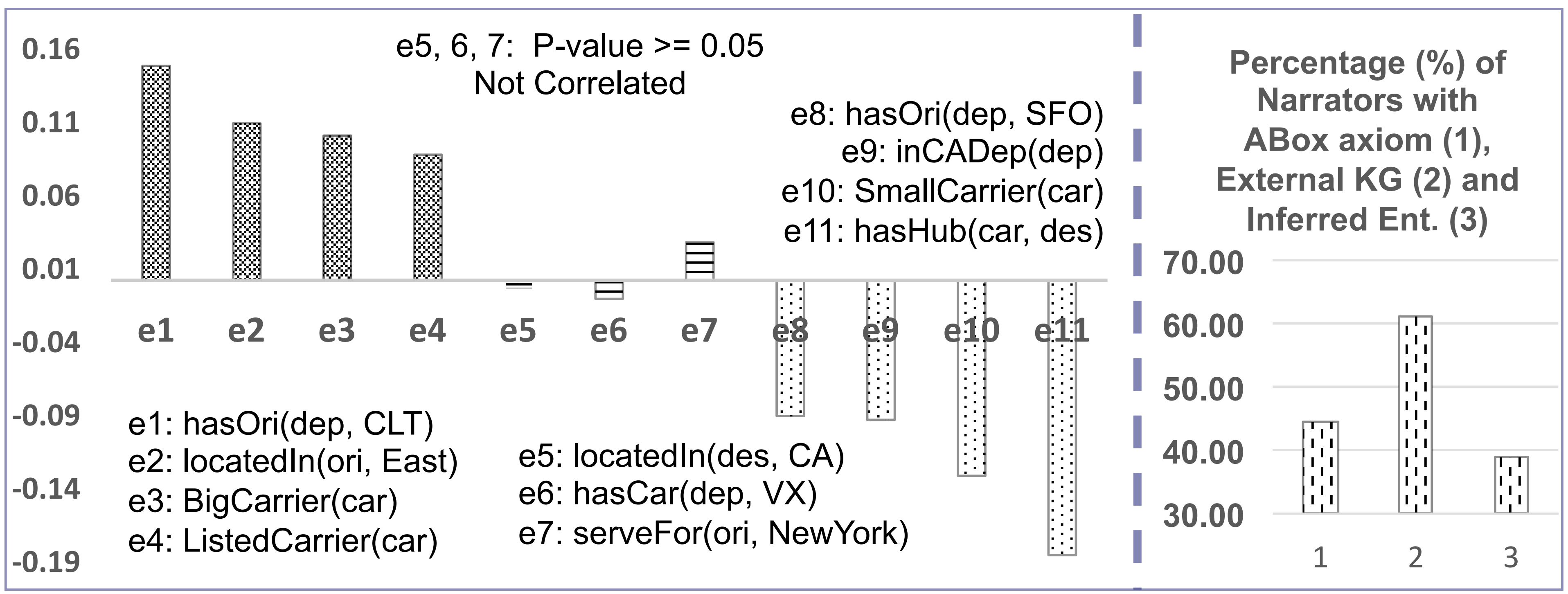
Knowledge-based Explanations

Explanations

- Three kinds of explanations:
  - General factor:
    - statistics index (e.g., new entailment
    - percentage) of source domain and target domain
  - Particular narrator:
    - Particular entailments that affect the transferability
  - Core context:
    - Combinations of entailments that affect the transferability

Experiments and Results

- **Particular Narrator**
  - Examples: 1) explain the positive transfer from (DL, ORD, LAX) to (AA, ORD, SFO) with "*the origin airport of both source domain and target domain are in the east part of US (e2)*"; 2) explain the negative transfer from (DL, ORD, LAX) to (B6, LAX, JFK) with "*the carriers of source and target domains are small companies (e10)".*
  - A large part of entailment narrators depend on external knowledge and semantic reasoning (that is why we need expressive ontology and reasoning)



[Left] Examples of Most (Least) Positive (Negative) Correlated Entailments with Transferability; [Right] Percentage of Entailment Narrators

小结：知识的价值不只是搜索与问答

- Knowledge Graph for Predictive Analytics
  - Decision Intelligence
- Knowledge Graph for Transfer Learning
  - Domain Adaptation
  - Few shot learning
  - Meta-learning
- Knowledge Graph for XAI
  - KR is designed for both human and machine.

HTTP://OpenKG.CN

谢谢！请批评指正