



交通大学

城市建筑数据管理中的语义智能

金耀辉

人工智能研究院



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

提纲

- 语义智能
- 案例分享
- 建筑数据管理



关于人工智能的预言：忽悠与危言



► Herbert A. Simon (1965) 人工智能先驱

“20年后，AI可以做任何人类能做的事情”

► 李开复 (2018)

“未来10年，有50%人的工作将被AI取代”

► Stephen Hawking (2016)

“AI可能是人类文明里最糟糕的事件”

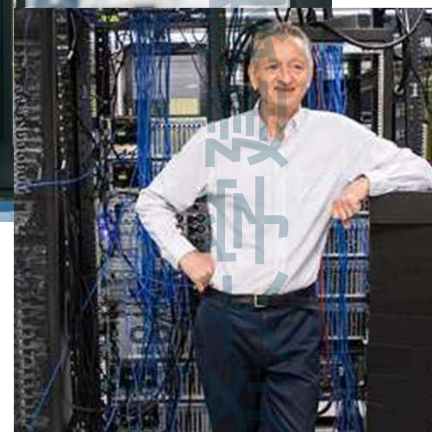
► Elon Musk (2017)

“AI可能毁灭人类”



关于人工智能的冷静思考

- **Doug Lenat (1998) 符号学派的先驱**
“AI甚至连人类最基本的常识都不具备”
- **Yann Lecun (2017) Facebook 首席科学家**
“停止用终结者之类的流言谈论AI”
- **Geoffrey Hinton (2017) Google**
“我们要从头开始”



AI缺乏常识



闵行交大到同济有多远?



百度一下

网页 资讯 视频 图片 知道 文库 贴吧 采购 地图 更多»

百度为您找到相关结果约302,000个

搜索工具

上海交大闵行校区和同济离得远吗?乘车多长时间 爱问知识人

2018年11月6日 - 上海交大闵行校区和同济离得远吗?乘车多长时间越害怕越孤单_ 分享: 全部答案 2018-11-06 01:25:25 十分十分遥远,交大在闵行,同济在杨浦。一个半小时...

<https://iask.sina.com.cn/b/gWN...> - 百度快照

交大闵行校区到同济大学怎么走?- 住哪网

住哪网公交查询到5种乘车线路供您参考,想了解更多交大闵行校区到同济大学的出行信息、附近酒店,查询交大闵行校区到同济大学最佳的交通线路,就请访问住哪网公交!全国...

[www.zhuna.cn/gongjiao/...](http://www.zhuna.cn/gongjiao/) - 百度快照

同济大学到上海交通大学(闵行校区)怎么走_百度知道

2个回答 - 回答时间: 2017年5月13日

最佳答案: 公交线路:地铁10号线 → 地铁1号线 → 地铁5号线,全程约37.2公里 1、从同济大学(四平路校...步行约170米,到达同济大学站 2、乘坐地铁10号线,...

[更多关于闵行交大到同济有多远?的问题>>](#)

[百度知道](#) - [百度快照](#)

交大闵行校区到同济大学怎么走_百度知道

2个回答 - 回答时间: 2018年2月24日

最佳答案: 公交线路:地铁5号线 → 地铁1号线 → 地铁10号线,全程约37.0公里 1、从上海交通大学(闵行...步行约1.0公里,到达剑川路站 2、乘坐地铁5号线,...

[更多关于闵行交大到同济有多远?的问题>>](#)

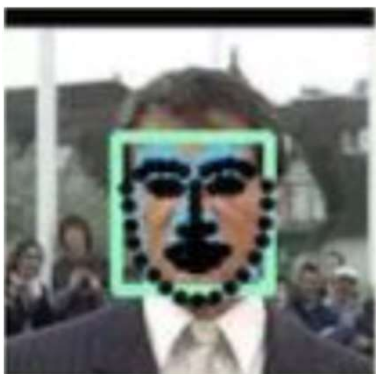
[百度知道](#) - [百度快照](#)



当前的“智能系统”

► 大多是在特定数据集基础上的特定问题

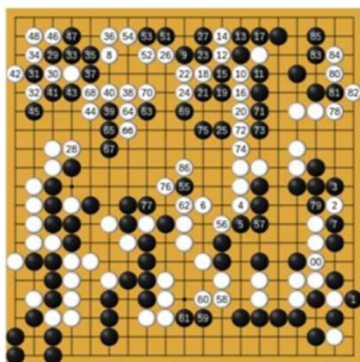
人脸识别



深度学习、
遗传算法

.....

游戏对战



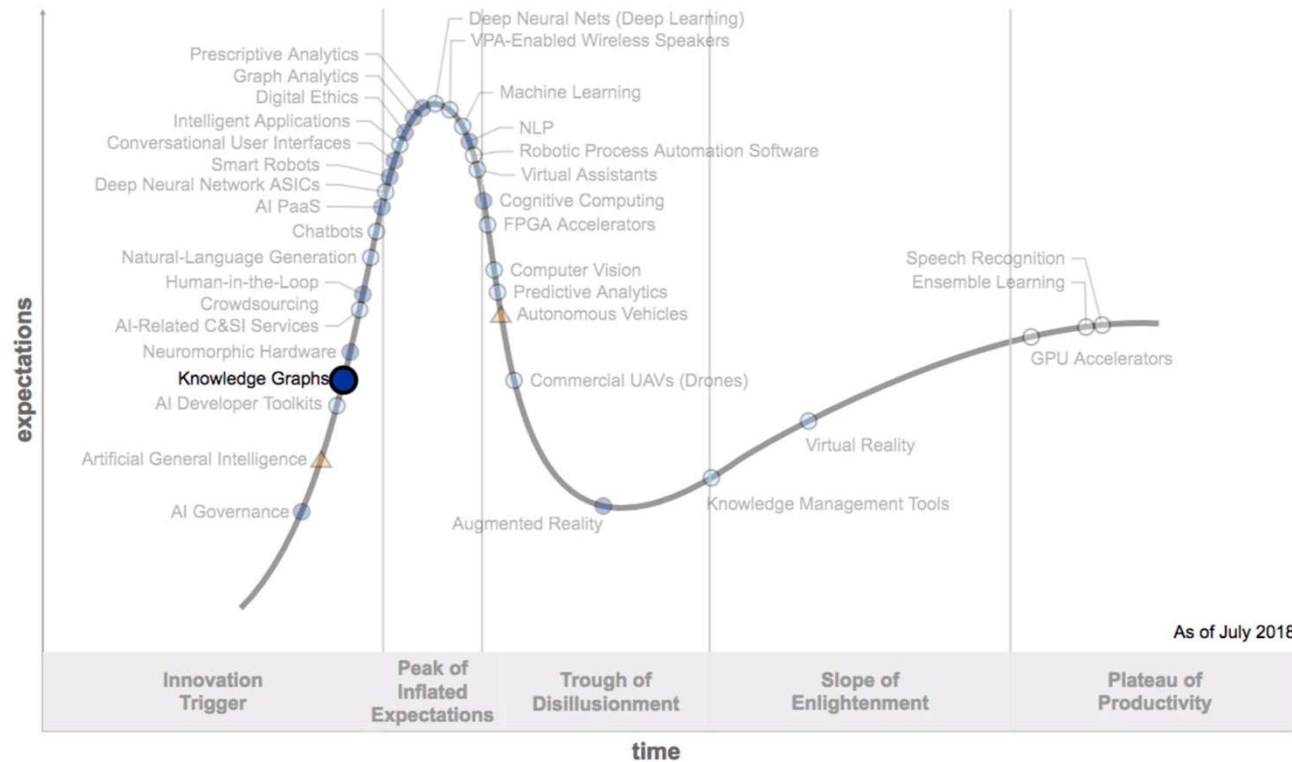
深度学习、
Monte Carlo树搜索

欺诈检测



神经网络、
概率推断

Gartner的AI曲线



“The rising role of content and context for delivering insights with AI technologies, as well as recent knowledge graph offerings for AI applications have pulled knowledge graphs to the surface.”

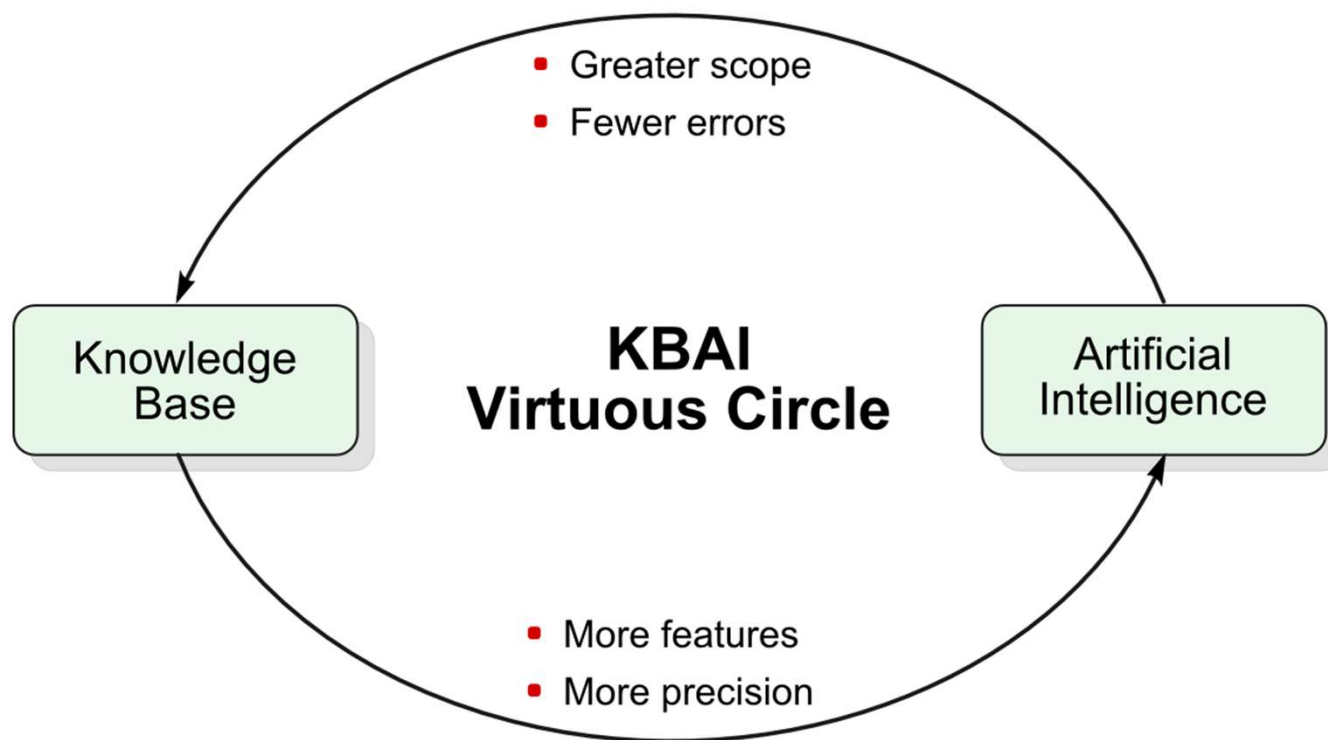
Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ✕ obsolete before plateau

© 2018 Gartner, Inc.

Source: Gartner 2018

语义智能 = Knowledge Graph + Artificial Intelligence



Mike Bergman (2014): Knowledge-based Artificial Intelligence

当前AI应用的痛点



缺乏AI治理

可解释性
算法伦理

数据孤岛

集成、关联
共享、安全

供应商锁定

大厂垄断

缺乏AI战略

规划、路径
验证、落地

数据质量

错、丢
乱、杂

缺少复合人才

领域专家
算法、工程

Source: Gartner 2018

文本/事件抽取

数据轮廓和
数据质量

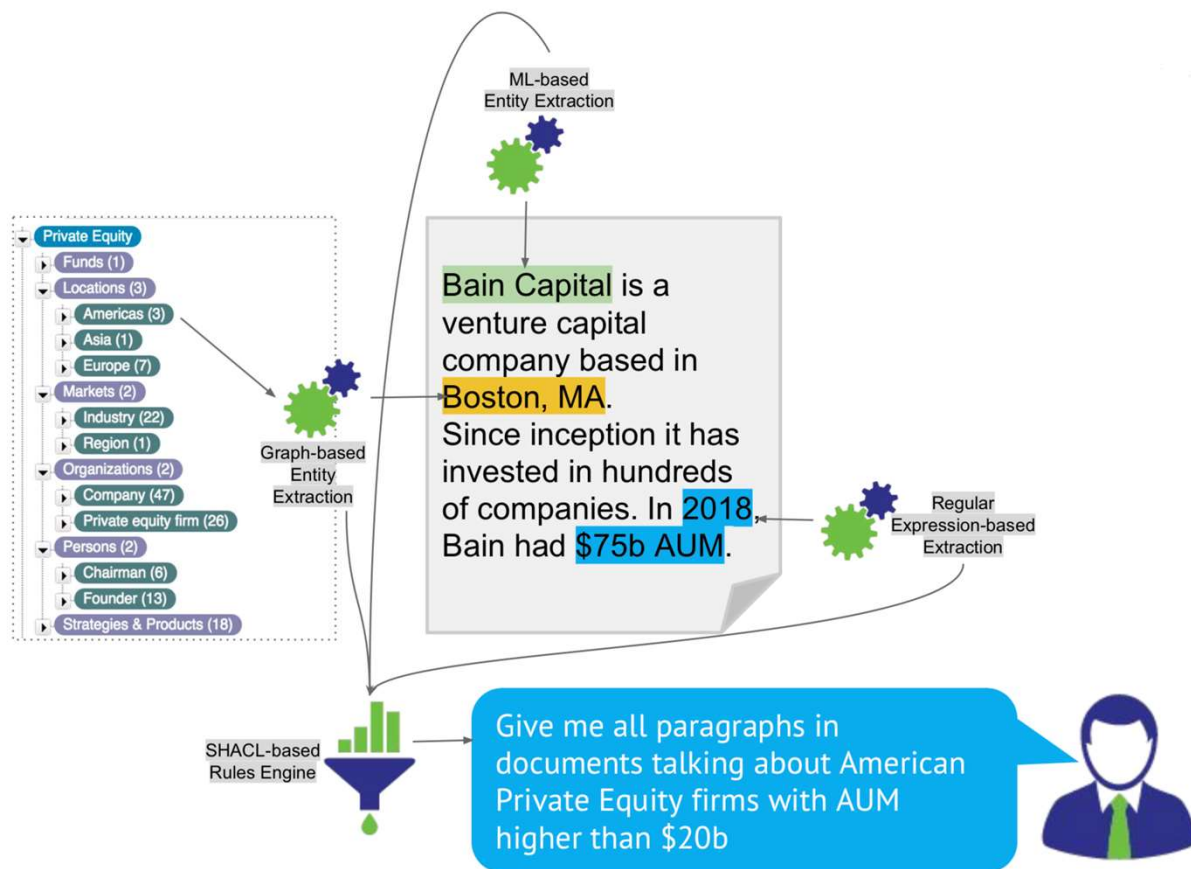
关联数据

可解释
人工智能

相似度量/
推荐系统

人机协同

文本/事件抽取



非结构化数据->结构化数据

三种方式结合：

- 1 正则表达式
- 2 知识图谱
- 3 机器学习

数据轮廓和数据质量



企业数据 原始数据 VIEW_IAC_REG_INFO.csv 数据轮廓

请输入关键字 100 条/页 1 前往 1 页

列名	总行数	数据类型	值分布(频率/值)	标签	主键	独立值	填充率	最小值/最短字段	最大值/最长字段	验证规则	异常值	备注
CORP_UID	12602			0		12602	100.00%	2015082607...	【依法须经...	+	0	+
REG_NO	12602			0		12602	100.00%	07853	4600000001...	+	0	工商注册号
CORP_NAME	12602			1		12523	100.00%	一明(上海...	龚玉国际贸...	+	0	法人名称
ZIP_CODE	12602					94	100.00%	100313	301308		0	邮编
PERSON_NA...	12602					10766	96.00%	ABDULAI BAH	龟岛则充(K...		524	法定代表人...
CORP_TYPE	12602					7	100.00%	个人独资企业	(外国) 承...	+	0	法人类别
ADDRESS	12602			0		11446	100.00%	上海市外高...	中国(上海...	+	0	经营场所
AREA_CODE	12602			0		3	100.00%	外高桥保税区	浦东	+	0	区划
EXP_AREA	12602			0		4	100.00%	10	40	+	0	+
TELEPHONE	12602			0		7351	100.00%	+61421036487	(8621) 53...		42	联系电话

Powered by Seassoon Copyright © 2016 - 2018

数据质量

非领域相关:

空值
重复
类型混乱
.....

领域相关:

字段理解
异常检测
文本聚类
模式识别
规则复用
.....

数据轮廓

收集技术元数据支持数据管理

列或者列集的信息

独立属性

数据挖掘

发现非显著的结果支持业务管理

行或行集的信息

多维属性

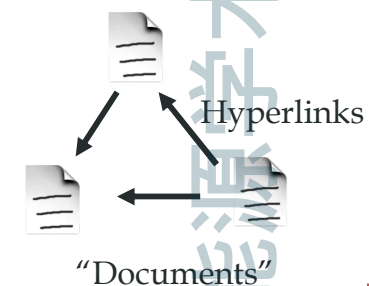
关联数据



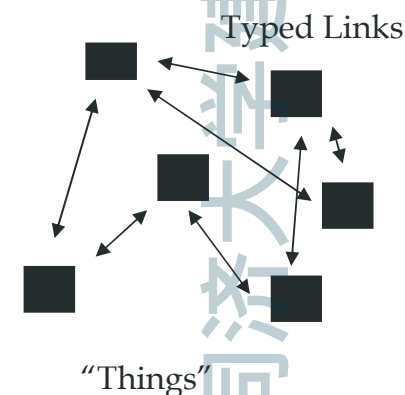
Linked Open Data: 5 Stars (Tim Berners-Lee)

- ★ Make your stuff available on the Web (whatever format) under an open license.
- ★★ Make it available as structured data (e.g., Excel instead of image scan of a table).
- ★★★ Use non-proprietary formats (e.g., CSV instead of Excel).
- ★★★★ Use URIs to denote things, so that people can point at your stuff.
- ★★★★★ Link your data to other data to provide context.

Web of Docs



Web of Data



可解释人工智能



美国国防先进科技局2016科研项目CFP

Explainable Artificial Intelligence (XAI)

DARPA-BAA-16-53

August 10, 2016

arXiv.org

深度学习三大牛之一最近公开的论文

The Consciousness Prior

Yoshua Bengio
Université de Montréal, MILA

September 26, 2017

输入

[海量]非结构化数据

[少量]标注文本语料

[模糊]人工先验逻辑

特征生成

监督泛化

学习推理

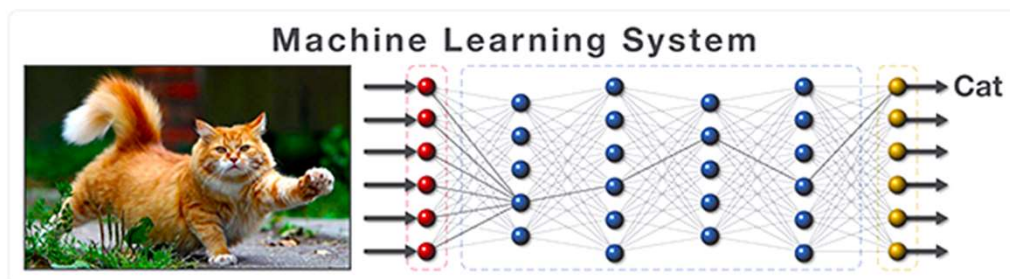
输出

要素提取结构化数据

业务逻辑知识图谱

性能指标

置信度
覆盖度
准确度



This is a cat.

Current Explanation

This is a cat:
• It has fur, whiskers, and claws.
• It has this feature:



XAI Explanation

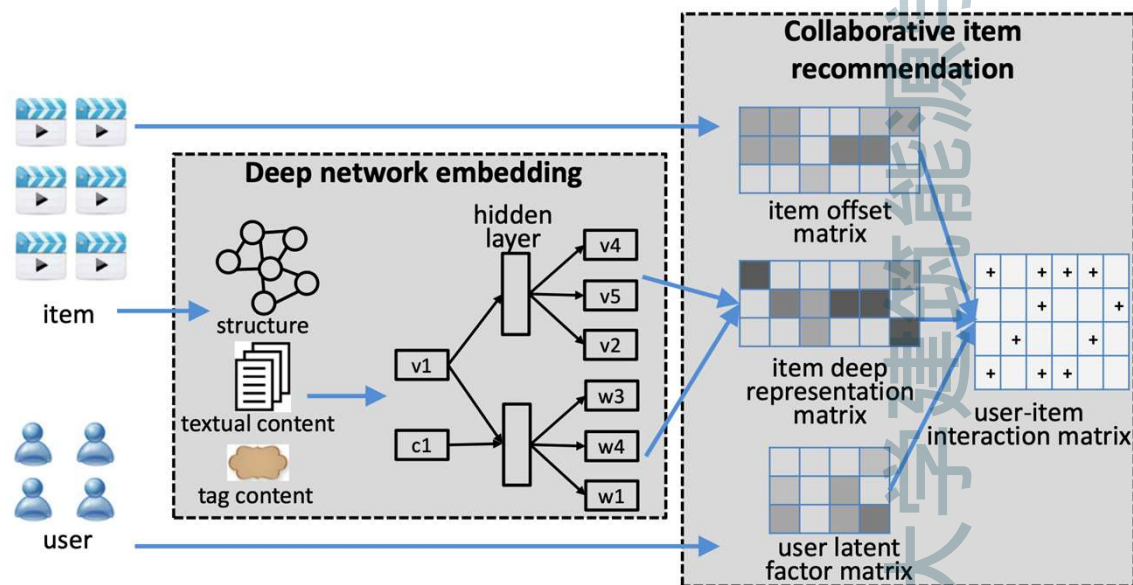
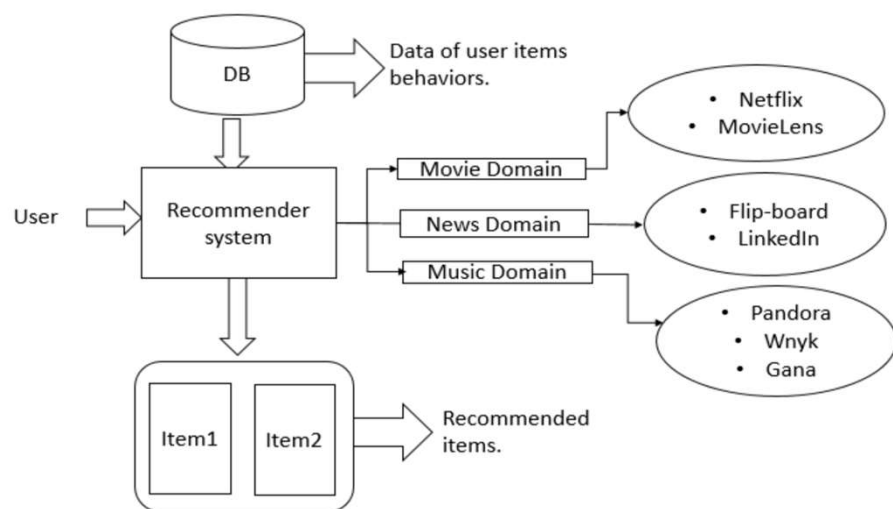
问题：人工参与导致算法性能下降



相似度度量(推荐系统)



深度学习结合协同过滤



人机协同 (主动学习) Human in the loop



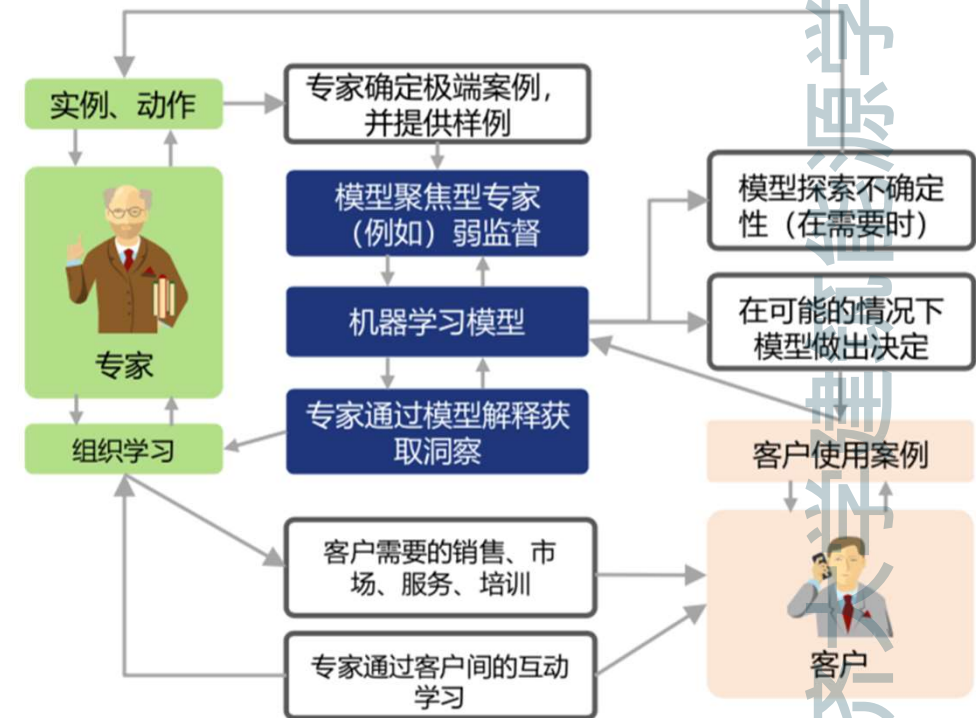
Peter Norvig,
Director of Research,
Google

2017

Why are AI programs different?

- Content: models not programs
- Process: training not debugging
- Release: retraining not patching
- Uncertainty: of objective
- Uncertainty: of action/recommendation
- Uncertainty: propagates through model

非确定性计算、多重反馈循环



案例1：科创产业地图



本项目基于科创中心当前业务需求，开发徐汇科创中心“互联网+科创”数据集成与智能服务系统，实现对徐汇区内生命健康、信息科技两大领域的企业的总体情况管控与高效精准服务。



企业，人才，载体的垂直搜索功能

通过标签快速查看企业---通过首页标签快速进入重点企业，查看详情。



科创产业分布情况地图展示---地图热力图形式展示区内企业的分布情况、分析区内行业、功能区的分布关系。

融资专题页面---分析区内投融资企业情况，可依据时间、融资额排名。



人工智能专题页面---区内人工智能企业布局分析、产值分析（后期新打的人工智能标签也会一起分析）；全国人工智能企业排名，增量人工智能企业。



自定义标签与人工智能标签---可给企业增加标签，可以是人工智能体系标签也可以是非体系标签，人工智能标签体系可以修改完善。



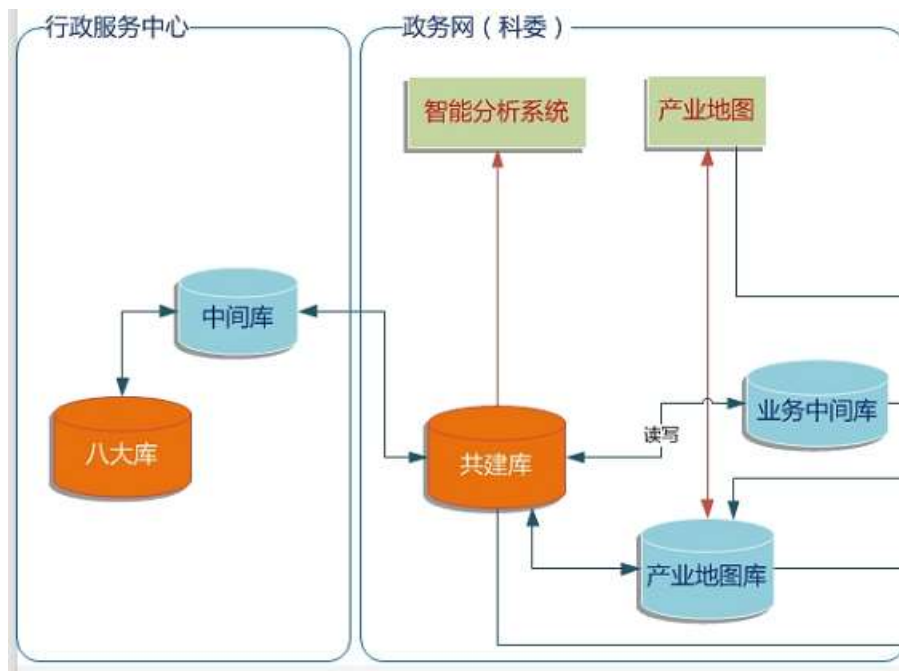
数据集成



• 内部数据



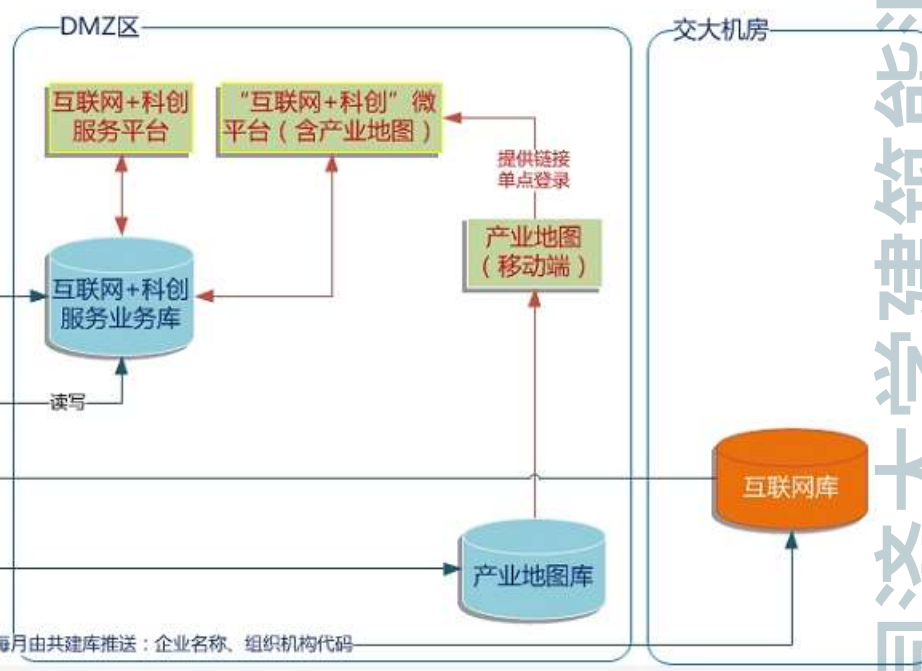
政府内部跨部门数据共享



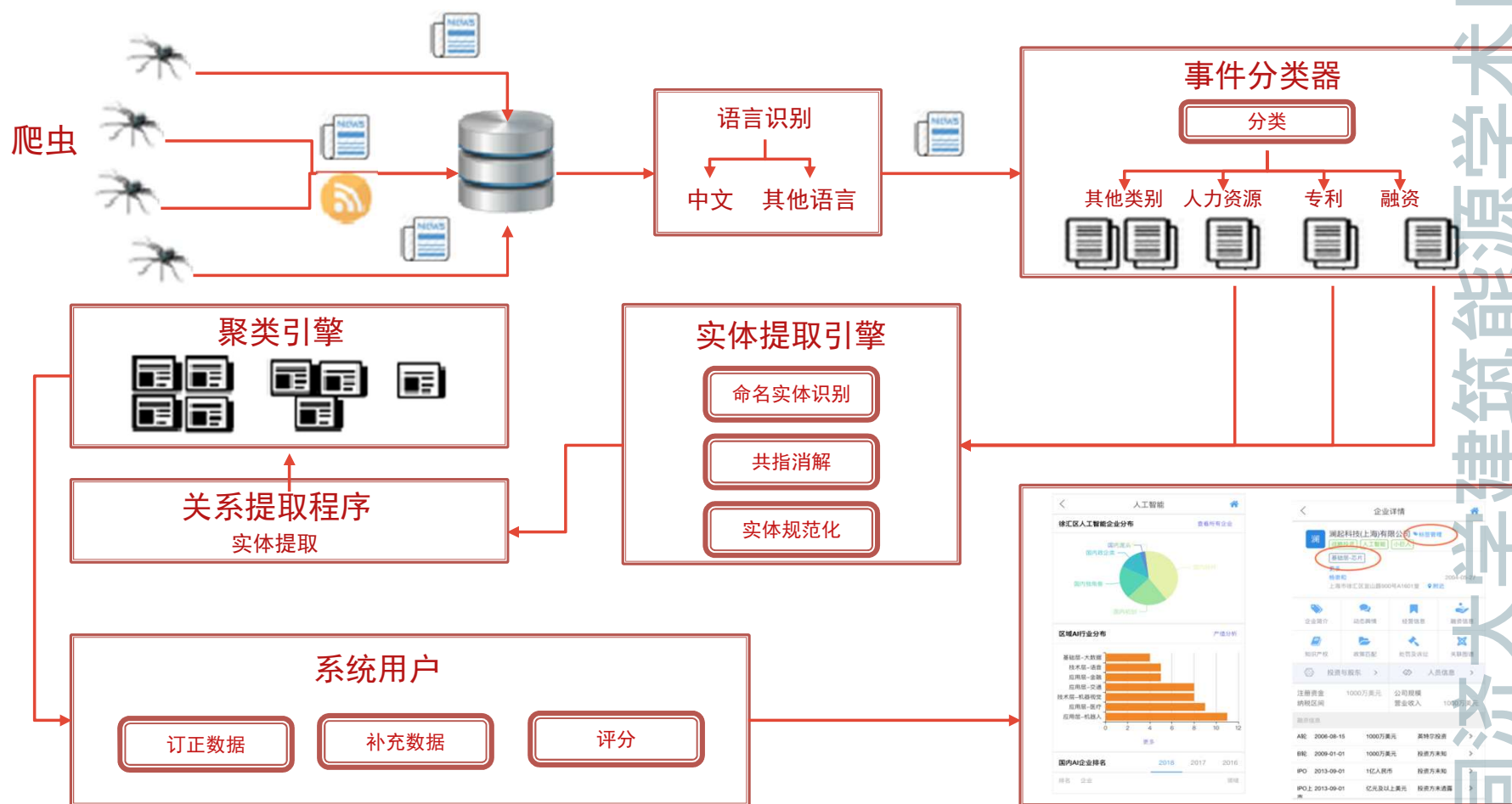
• 外部数据



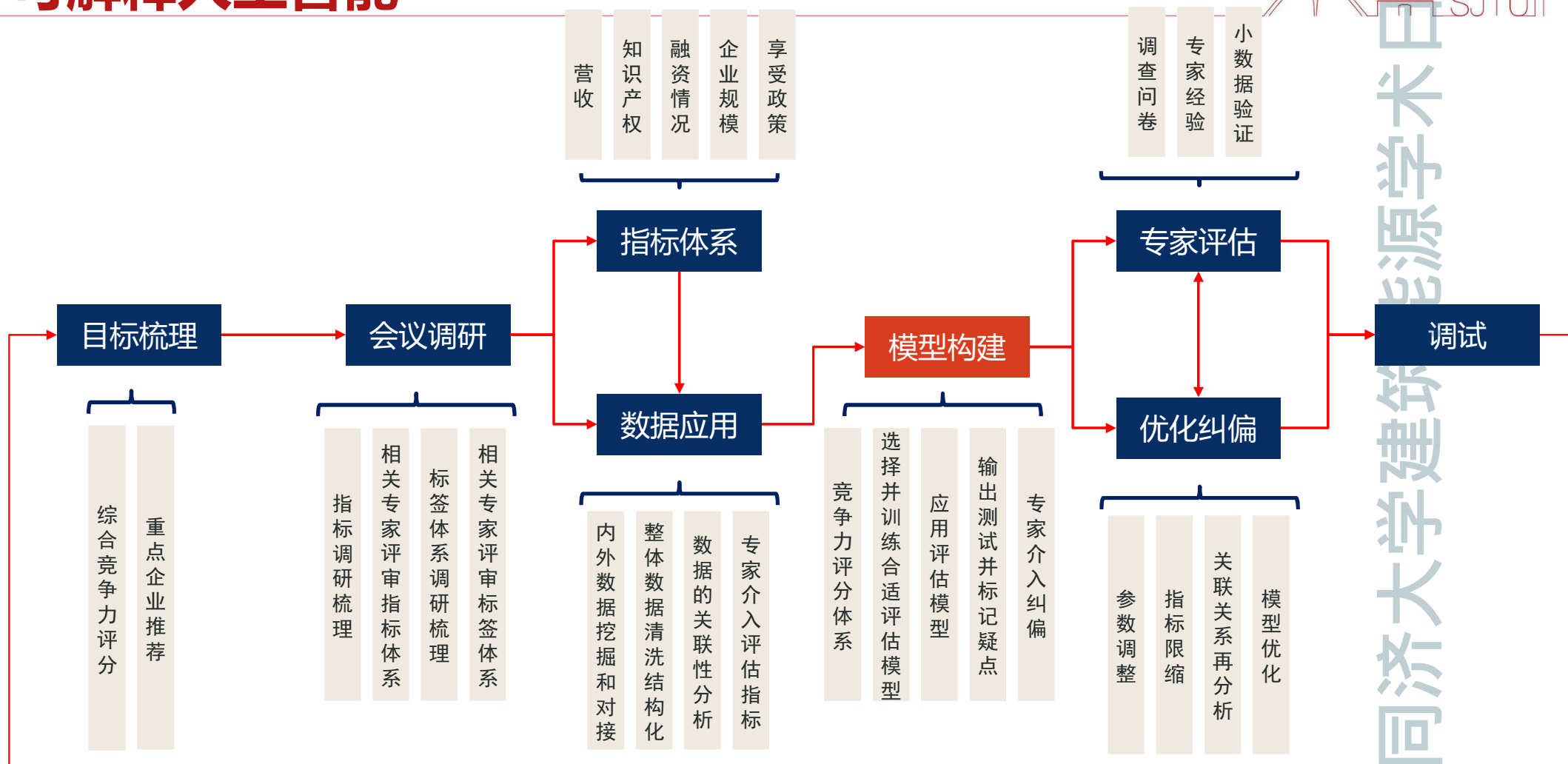
来自互联网的外部数据补充



文本/事件提取与人机协同



可解释人工智能



案例2：行政审批



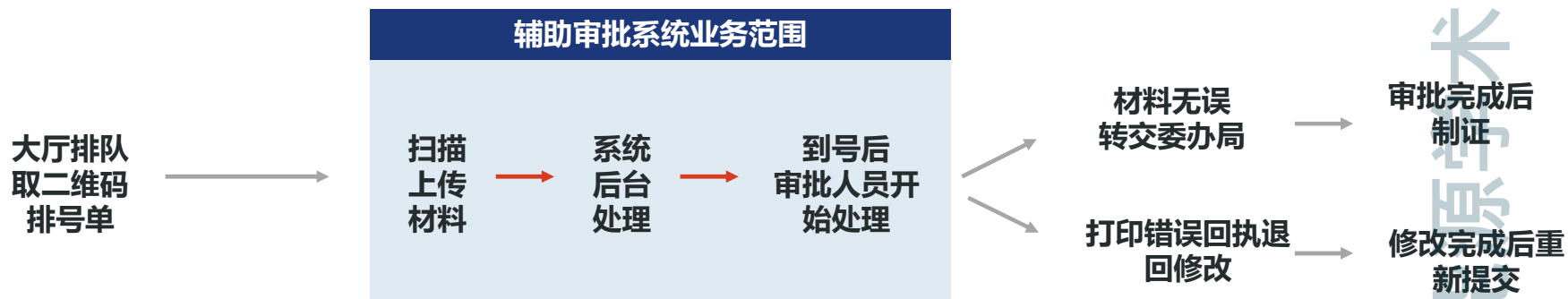
“一窗通办”的三大障碍



人工智能流程嵌入

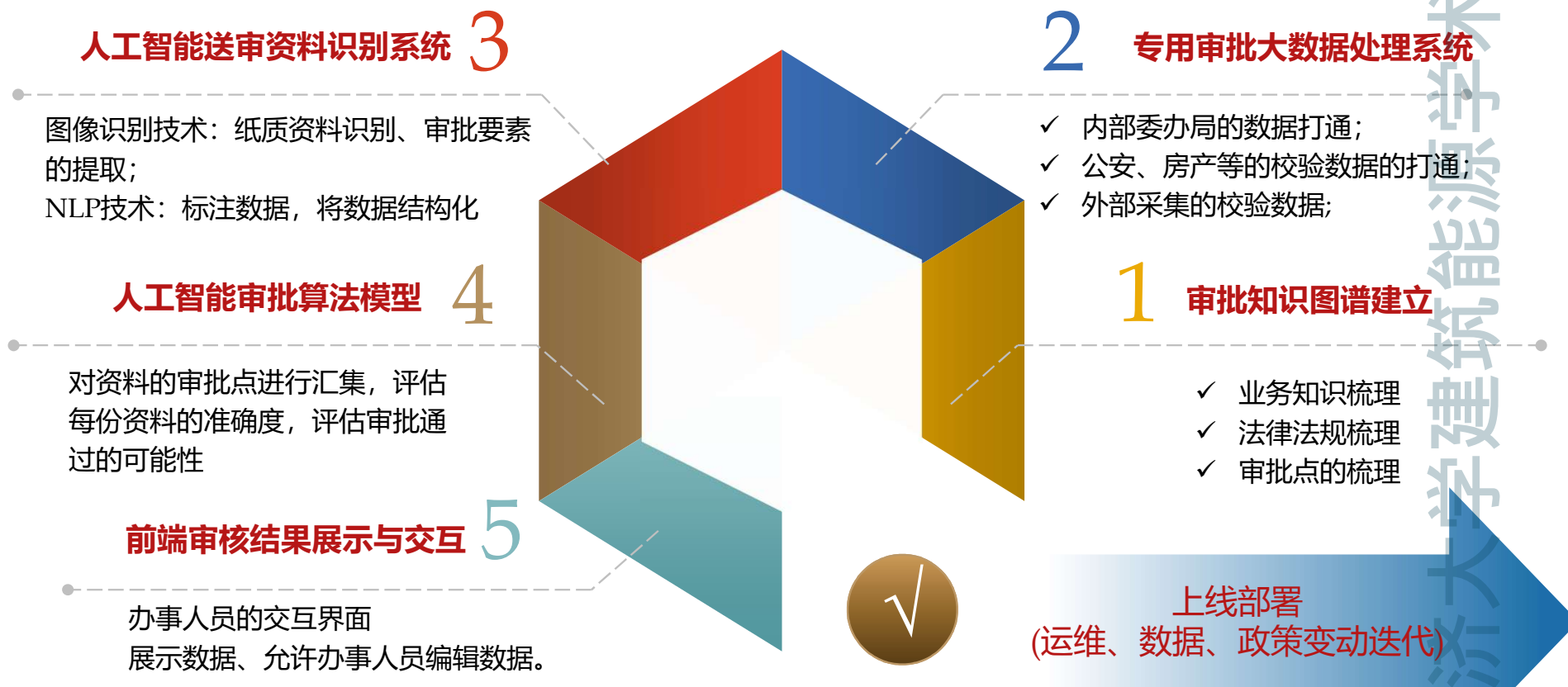


实际业务流程



审批系统使用过程

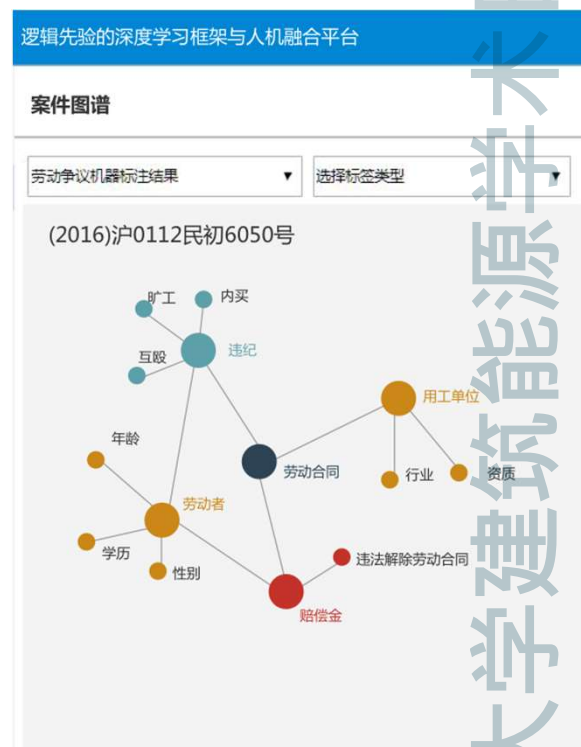
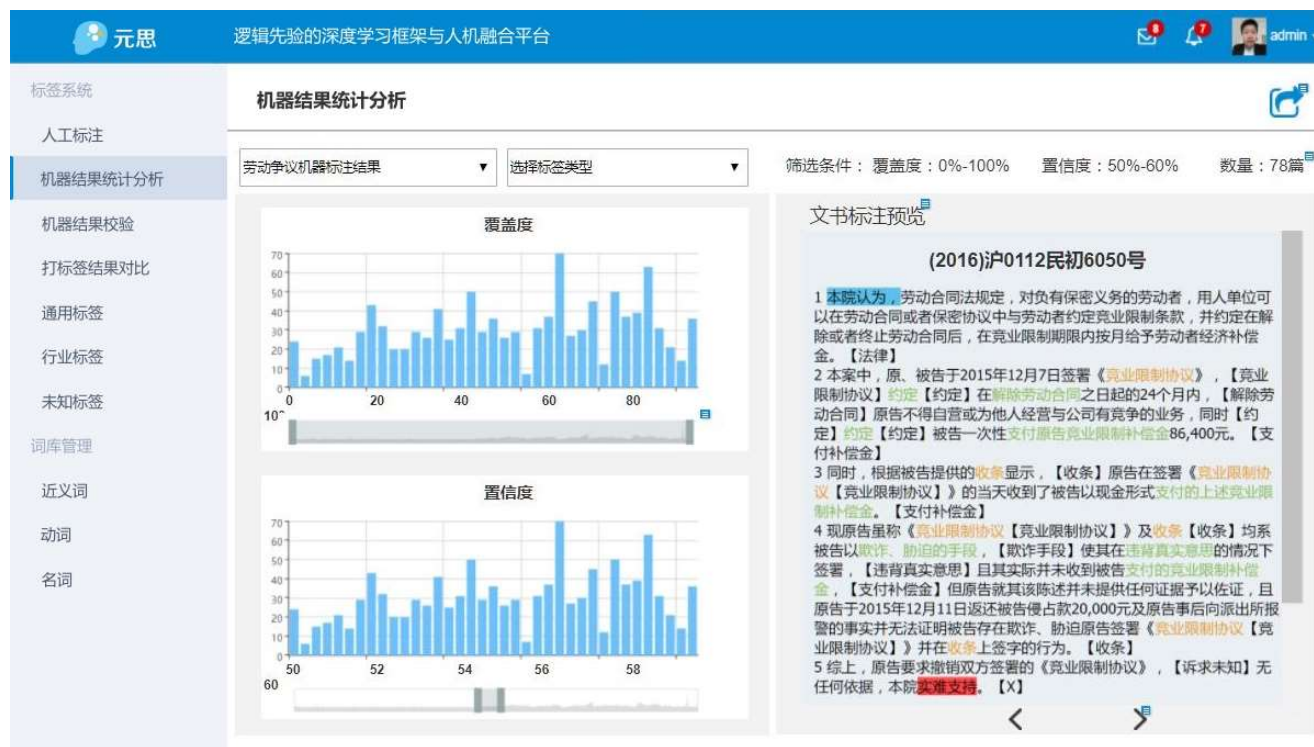




试点效果

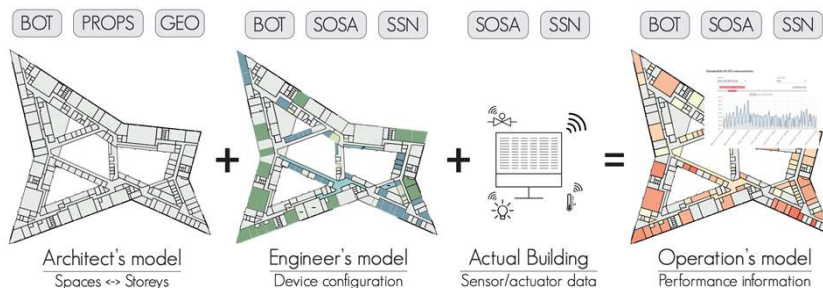


案例3：司法领域——类案推荐

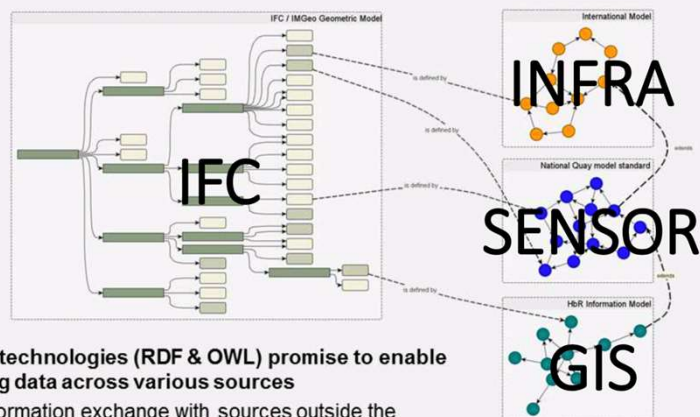


国家重点研发计划项目支持，完成了对5万裁判文书的理解，覆盖劳动争议、道路交通、婚姻家庭、民间借贷等主要民商事，形成对诉请、证据、事实、法条、判决等业务逻辑标注，初步构建了知识图谱和类案推荐系统，在上海、四川、贵州等地法院得到检验。

建筑数据的采集、关联与管理



Goal of Linked Data in Construction



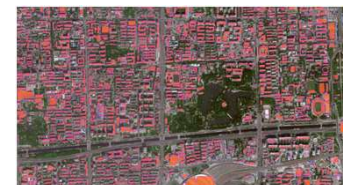
Semantic web technologies (RDF & OWL) promise to enable linking building data across various sources

⇒ improved information exchange with sources outside the traditional BIM environments, additional to the already existing techniques

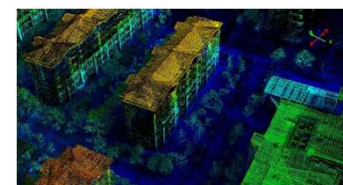
Model Support Group

buildingSMART
International home of openBIM

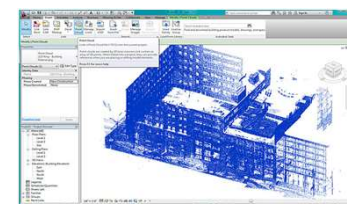
遥感图像



街景+激光点云



BIM



BIM+IoT



Linked Building Data

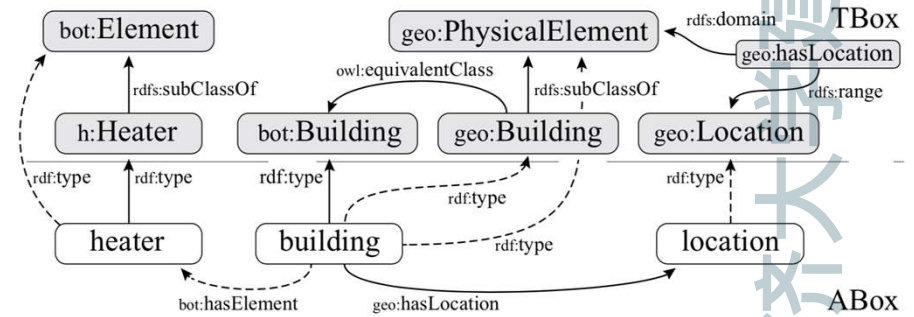
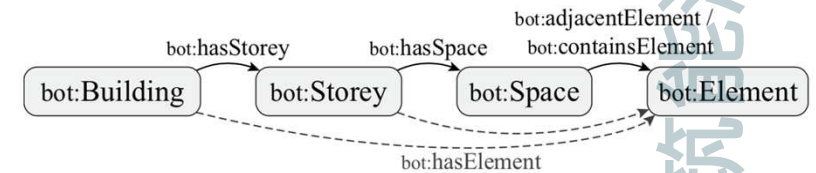
Ontology	Domain	Online	Building	Storey	Room	Elements
DBpedia	World	X	Building	-	-	-
Schema.org	World	X	Civic-Structure	-	-	-
ifcOWL	AEC	X	IfcBuilding	IfcBuildingStorey	IfcSpace	IfcElement
ifcWOD	AEC	-	X	X	X	X
BIMSO	AEC	-	?	?	?	?
CBIM	AEC	X	-	-	-	Object
SEMANCO	Smart cities	X	SUMO: Building	SUMO: Floor	Space	subclasses
cityGML	Smart cities	X	Building	-	-	-
DAREED	Smart cities	-	?	?	?	?
SEAS	Systems	X	Building	Building-Storey	Room	Building-Space-Connection
SAREF	Smart homes	X	-	-	Building-Space	Building-Object
DogOnt	Smart homes	X	Building	Storey	Room	Building-Thing
ThinkHome	Smart homes	X	Building	Building-Storey	Space	Opening, Equipment



W3C Linked Building Data Community Group

Git repository for the LBD Community Group at W3C

<https://www.w3.org/community/lbd/>



Building topology information in domain specific ontologies



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

同济大学建筑能源学术日