



知识库的构建和管理

报告人：Deepboy

2019年4月22日



知识图谱的架构

整体框架

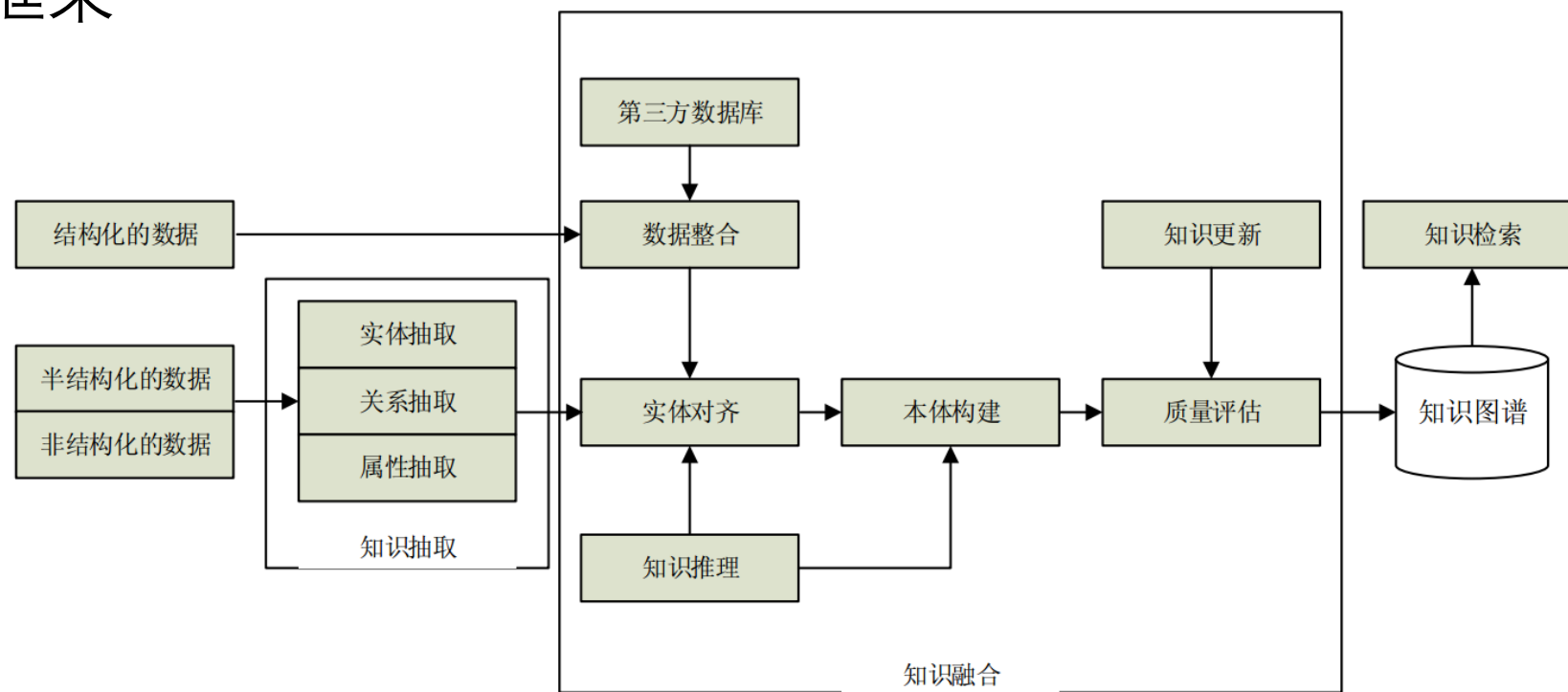


图1 知识图谱的体系架构

知识图谱的架构

知识抽取

从异构数据源中自动抽取信息得到候选知识单元。



实体抽取

从文本数据中自动识别出命名实体（NER）



关系抽取

从文本语料中抽取实体间的关系



属性抽取

从不同信息源中采集特定实体的属性信息

知识图谱的架构

实体抽取

实体抽取研究历史



实体抽取



面向单一领域



面向开放域

关注如何识别文本种的人名、地名等实体信息

- 启发式+人工规则，实现自动抽取实体
- 统计机器学习方法辅助解决命名实体问题
- 有监督学习+先验知识

面向开放的互联网，研究解决全网信息抽取问题

- 人工建立科学完整的实体命名分类体系
- 基于归纳总结的实体命名分类
- 统计机器学习方法对实体进行分类和聚类
- 迭代扩展实体语料库
- 通过搜索引擎的服务器日志，聚类获取新出现的命名实体

发展历程

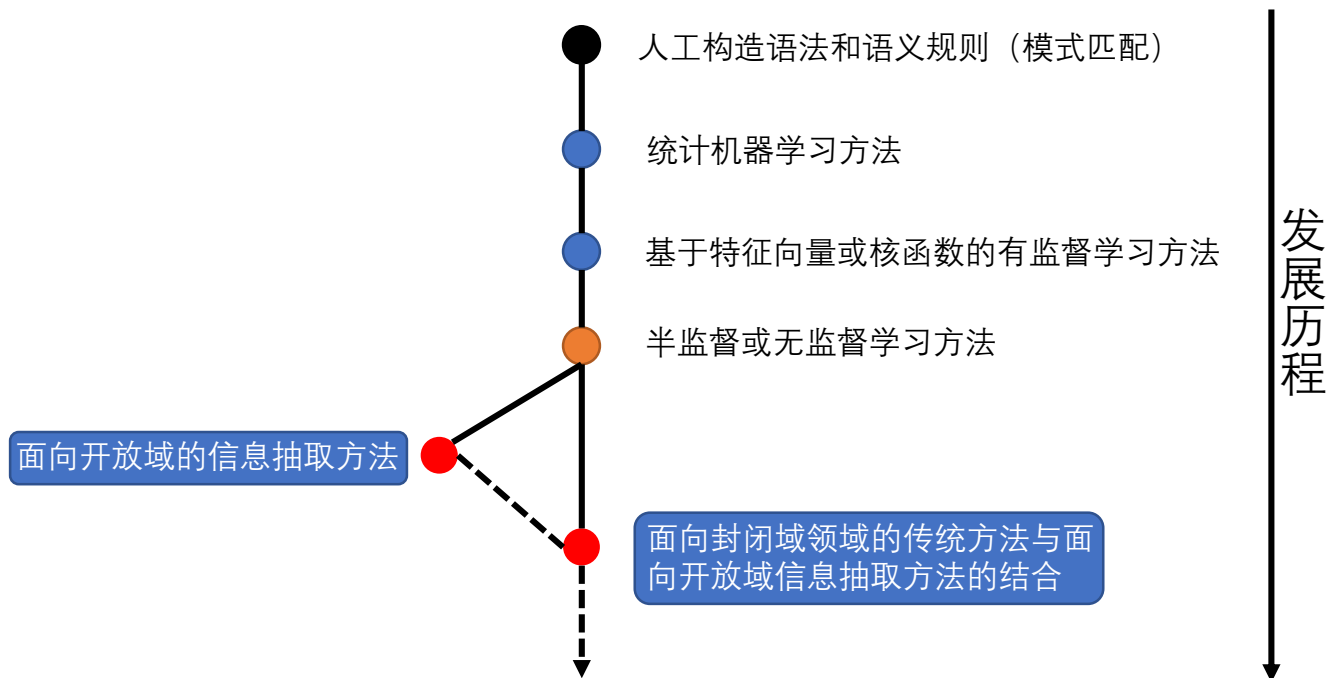
知识图谱的架构

关系抽取

关系抽取研究历史



从相关语料库中提取出实体间的关系，构成网状的知识结构



知识图谱的架构

属性抽取

属性抽取研究历史



从不同信息源中采集特定实体的属性信息

- 将属性抽取任务转化为关系抽取任务
- 基于规则和启发式算法，抽取结构化数据
- 基于百科类网站的半结构化数据，训练实体属性标注模型，从非结构化数据中抽取实体属性
- 基于模式的方法：采用数据挖掘的方法直接从文本中挖掘实体属性和属性值之间的关系模式

发展历程

知识图谱的架构

知识融合

对知识抽取得到的结果进行清理和整合，消除概念歧义，剔除冗余和错误概念。



实体链接

对于从文本中抽取得到的实体对象，将其链接到知识库中对应的正确实体对象

实体消歧

- 解决同名实体产生歧义问题

共指消解

- 解决多个指称对应同一实体对象的问题



知识合并

根据第三方知识库或已有结构化数据进行知识的合并

合并外部数据库

- 数据层和模式层的融合

合并关系数据库

- RDB2RDF: 将关系数据库的数据转换成RDF的三元组数据

知识图谱的架构

知识加工

将事实转变成结构化、网络化的知识体系。



本体构建

以形式化方式对概念及其之间的联系给出明确定义



知识推理

从知识库中已有的实体关系数据出发，推理并建立实体间的新关联



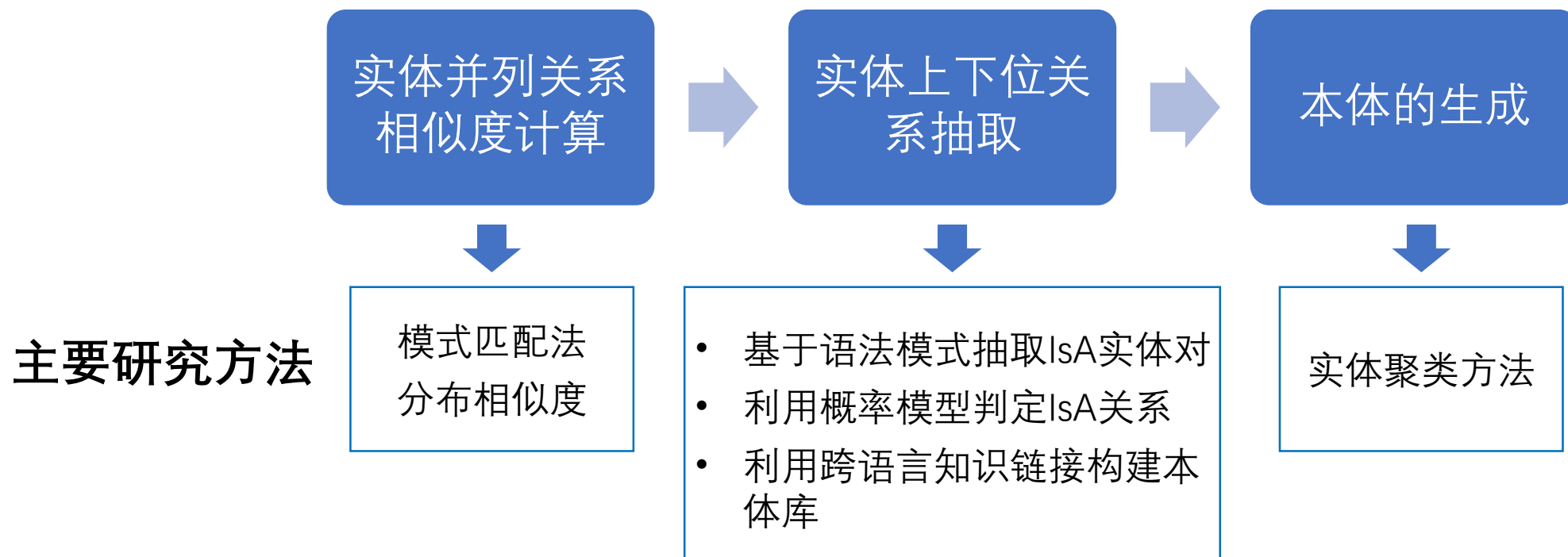
质量评估

对知识的可信度进行量化，通过舍弃置信度较低的知识来保障知识库的质量

知识图谱的架构

本体构建

本体构建的阶段



知识图谱的架构

知识推理

方法分类



基于逻辑的推理方法

采用抽象或具象的Horn子句建模，基于一阶逻辑谓词、描述逻辑以及规则进行推理

基于图的推理方法

将实体和关系映射到一个低维的向量空间中(embedding)，基于知识的语义表达进行推理建模

基于深度学习的推理方法

使用神经网络建模，结合以上两种推理方法进行建模

知识图谱的架构

知识更新

按逻辑分类

概念层的更新

新增数据后获得了新的概念，自动将新的概念添加到知识库的概念层中

数据层的更新

新增或更新实体、关系、属性值，选择在各数据源中出现频率高的事实和属性加入知识库

更新方式

全面更新

从零开始重新构建知识图谱

增量更新

以新增数据为输入，向图谱中添加知识





知识图谱的架构

内容及后续待完善