**Reference Type:**  Journal Article

**Record Number:** 3

**Author:** 刘峤，李杨

**Year:** 2016

**Title:** <知识图谱构建技术综述.pdf>

**Journal:** 计算机研究与发展

**Short Title:** <知识图谱构建技术综述.pdf>

**'File' Attachments:** internal-pdf://3738049461/知识图谱构建技术综述\_刘峤.pdf

1. 架构

* 各种特点的知识库：百科类网站、通用知识库（全领域）、特定领域知识库，freebase（采用wiki作为来源）、DBpedia（wiki之上的知识网络）、
* 概念：数据层和模式层，事实，三元组，实体（概念）与本体，属性，关系，本体库（管理模式层），自顶而下，自底而上，实体指称项，共现关系，集成实体链接，本体，

1. 构建技术（自底而上）

* 信息抽取:从异构数据源中自动抽取信息得到候选知识单元（自动化从半结构化或无结构数据中抽取实体、关系、实体属性等结构化信息）
* 实体抽取：即命名实体识别（NER）从文本数据集中自动识别命名实体。早期为单一领域，目前是开放域（不限定特定领域，面向互联网，则产生实体分类问题）。一般最终得到离散的命名实体。
* 关系抽取：抽取实体间的关联关系。早期通过人工构造语法和语义规则，采用模式匹配（缺点有二，对人，对工作量对拓展），后期较多采用统计机器学习建模代替人工预定义的语法和语义。近年，逐渐转向半监督和无监督，但是均需先预定实体关系类型。也是从封闭域发展到开放域。
* 属性抽取：从不同的信息源中采集特定实体的属性信息。分为有名属性和无名属性，有名属性可以用关键字来定位。目前两种方法分别是基于百科类网站的半结构化数据和数据挖掘文本信息中属性和属性值的关系模式。
* 知识融合：信息抽取得到的结果是包含冗余和错误信息的，数据间的关系是扁平化，缺乏层次性和逻辑性，所以需要清理和整合
* 实体链接：将抽取得到的实体对象链接到知识库中对应的正确实体对象。
* 实体消歧：用于解决同名实体产生的歧义问题（Apple可指代水果的苹果实体，也可以是智能设备的苹果实体，更可能是公司的苹果实体）。主要方法是聚类法——以实体对象为聚类中心，将所有指向同一目标实体对象的所有指称项都聚集到以该对象为中心的类别下。此方法关键问题是如何定义实体对象和指称项之间的相似度。常用的相似度计算方法有：空间向量模型（词袋模型）（缺点，没有考虑上下文语义信息）、语义模型（包含部分语义特征）、社会网络模型、百科知识模型，新方法：增量证据挖掘算法等。
* 共指消解：用于解决多个指称项对应于同一个实体，表述方式有：对象（实体）对齐、实体匹配、实体同义。可以把共指消解问题看做是分类问题和聚类问题，发展过程是从自然语言处理到统计机器学习。分类问题，早期的NLP主要是以句法分析为基础，代表性方法是Hobbs算法（句法分析树）和向心理论（识别表达模式中的实体），今年来在向心理论基础上加以词性标注和语法分析。聚类问题，基本思想是以实体指称项为中心，通过实体聚类实现指称项和实体对象的匹配，而关键问题在于如何定义实体间的相似性检测。
* 知识合并：即从第三方知识库产品或已有发结构化数据获取知识输入。
* 合并外部知识库：数据层融合（实体的指称、属性、关系以及所属类别等，主要问题是如何避免实例以及关系的冲突），模式层的融合（将新得到的本体融入已有的本体库）。
* 合并关系数据库：一般采用RDF（资源描述框架）作为数据模型，数据和RDF的映射方式有：Direct Mapping、R2RML。
* 合并半结构化数据：如XML，CSV，JSON等格式的数据。
* 知识加工：得到结构化、网络化的知识体系。
* 本体构建：特定领域可用领域专家和众包的方式人工构建本体；跨领域的全局本体从特定领域的现有本体库出发，采用自动构建技术逐步得到。数据驱动的自动化本体构建过程包括3个阶段：实体并列关系相似度计算（方法：模式匹配、分布相似度）、实体上下位关系（IsA）抽取、本体生成。除此之外，还可以采用跨语言知识链接的方法构建本体。
* 知识推理：从知识库中已有的实体关系数据出发，经过计算机推理，建立实体间的新关联。主要分为基于逻辑（一阶谓词逻辑（基于命题）、描述逻辑（基于对象）、基于规则的推理（基于本体的概念层次））和基于图的推理（基于神经网络模型和Path Ranking），近年，还有跨知识库知识推理的趋势。
* 质量评估
* 知识更新：数据层更新（新增或者更新实体、关系、属性，需要考虑数据源的可靠性和一致性），概念层的更新（新增数据后获得了新的概念，需要自动将新的概念添加到知识库的概念层中）。方式：全面更新和增量更新

1. 跨语言的构建

* 跨语言本体构建
* 跨语言知识抽取
* 跨语言知识链接

1. 应用

* 智能语义搜索
* 移动个人助理
* 深度问答系统

1. 新note

* 启发式算法