**（一）知识图谱技术综述**

**Review on Knowledge Graph Techniques**

**--XU Zeng-lin, SHENG Yong-pan,HE Li-rong, and WANG Ya-fang**

**一、科学问题**

**1.1 本文所涉及科学问题**

知识图谱技术是人工智能技术的重要组成部分，其建立的具有语义处理能力与开放互联能力的知识库，可在智能搜索、智能问答、个性化推荐等智能信息服务中产生应用价值。该文在全面阐述知识图谱定义、架构的基础上，综述知识

图谱中的知识抽取、知识表示、知识融合、知识推理四大核心技术的研究进展以及一些典型应用。该文还将评论当前研究存在的挑战。

**1.2 同行专家如何解决**

知识图谱为互联网上海量、异构、动态的大数据表达、组织、管理以及利用提供了一种更为有效的方式，使得网络的智能化水平更高，更加接近于人类的认知思维。目前，知识图谱已在智能搜索、深度问答、社交网络以及一些垂直行业中有所应用，成为支撑这些应用发展的动力源泉。

**1.3 本文所解决的问题**

本文在对知识图谱的定义、架构、大规模知识库等全面阐述的基础上，较为深入地研究了知识图谱中知识抽取、知识表示、知识融合以及知识推理 4大核心技术，并就当前产业界的需求介绍了它在智能搜索、深度问答、社交网络以及一些垂直行业中的实际应用。总结了目前知识图谱面临的主要挑战，并对其未来的研究方向进行了展望。

**1.4 本文解决方案效果**

本文在对知识图谱的定义、架构、大规模知识 库等全面阐述的基础上，较为深入地研究了知识图 谱中知识抽取、知识表示、知识融合以及知识推理4 大核心技术，并就当前产业界的需求介绍了它在智 能搜索、深度问答、社交网络以及一些垂直行业中 的实际应用

**二、研究内容**

**2.1 理论与方法介绍**

知识抽取主要是面向开放的链接数据，通过自 动化的技术抽取出可用的知识单元，知识单元主要 包括实体(概念的外延)、关系以及属性3个知识要素， 并以此为基础，形成一系列高质量的事实表达，为 上层模式层的构建奠定基础。

基于规则与词典的实体抽取方法：早期的实体抽取是在限定文本领域、限定语义 单元类型的条件下进行的，主要采用的是基于规则 与词典的方法，例如使用已定义的规则，抽取出文 本中的人名、地名、组织机构名、特定时间等实体

基于统计机器学习的实体抽取方法

面向开放域的实体抽取方法

**2.2 验证分析与实验效果**

针对如何从少量实体实例中自动发现具有区分 力的模式，进而扩展到海量文本去给实体做分类与 第4期 徐增林，等: 知识图谱技术综述 593 聚类的问题，文献[34]提出了一种通过迭代方式扩展 实体语料库的解决方案，其基本思想是通过少量的 实体实例建立特征模型，再通过该模型应用于新的 数据集得到新的命名实体。提出了一种基于 无监督学习的开放域聚类算法，其基本思想是基于 已知实体的语义特征去搜索日志中识别出命名的实 体，然后进行聚类。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

论文对知识图谱做了简单的介绍，但是并没有很深入地涉及进去知识，知识图谱技术是对语义网标准与技术的一次扬 弃与升华。自Google提出之初至今，其热度依然有 增无减，并随着深度学习、类脑科学等领域的发展， 有逐步演进并发展为智能机器的大脑知识库之趋势。

**3.2 后续研究重点**

在关注到知识图谱在自然语言处理、人工智能 等领域展现巨大潜力的同时，也不难发现知识图谱 中的知识获取、知识表示、知识推理等技术依然面 临着一些困难与挑战，很多重要的开放问题急待学 术界与工业界协力来解决。在未来的几年时间内， 知识图谱仍将是大数据智能的前沿研究问题。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：基于扩展领域模型的有名 属性抽取

**（2）作者介绍**：王宇, 谭松波, 廖祥文

**（3）摘要**: 网页信息抽取是互联网挖掘的重要课题.为了自动化抽取过程,最新的研究利用特定领域的特征,通过机器学习方法对信息抽取过程进行统一建模.但是,对领域特征的依赖使得这类方法难以推广到其他领域中去.因此,对信息抽取问题进行了分析,从中分离出一个可以完全自动化的信息抽取子任务,即有名属性抽取任务.在多个领域的数据集上进行的统计表明,这个子任务覆盖了60%以上的待抽取属性,因此它在整个信息抽取中占有重要地位.并给出了一种基于扩展领域模型的有名属性抽取方法,实验结果表明,这种方法的准确率接近或大于80%,召回率大于90%.

**4.2 相关论文二**

**（1）题目**：知识表示学习研究进展

**（2）作者介绍**：刘知远, 孙茂松, 林衍凯

**（3）摘要**：人们构建的知识库通常被表示为网络形式,节点代表实体,连边代表实体间的关系.在网络表示形式下,人们需要设计专门的图算法存储和利用知识库,存在费时费力的缺点,并受到数据稀疏问题的困扰.最近,以深度学习为代表的表示学习技术受到广泛关注.表示学习旨在将研究对象的语义信息表示为稠密低维实值向量,知识表示学习则面向知识库中的实体和关系进行表示学习.该技术可以在低维空间中高效计算实体和关系的语义联系,有效解决数据稀疏问题,使知识获取、融合和推理的性能得到显著提升.介绍知识表示学习的最新进展,总结该技术面临的主要挑战和可能解决方案,并展望该技术的未来发展方向与前景.

**4.3 相关论文三**

**（1）题目**：开放式文本信息抽取

**（2）作者介绍**：赵军, 刘康, 周光有

**（3）摘要**：信息抽取研究已经从传统的限定类别、限定领域信息抽取任务发展到开放类别、开放领域信息抽取。技术手段也从基于人工标注语料库的统计方法发展为有效地挖掘和集成多源异构网络知识并与统计方法结合进行开放式信息抽取。该文在回顾文本信息抽取研究历史的基础上,重点介绍开放式实体抽取、实体消歧和关系抽取的任务、难点、方法、评测、技术水平和存在问题,并结合课题组的研究积累,对文本信息抽取的发展方向以及在网络知识工程、问答系统中的应用进行分析讨论。

**（二）知识图谱的技术实现流程**

**Technology realization process of knowledge map**

**--** **Cao qian, zhao yiming（ITA）**

**一、科学问题**

**1.1 本文所涉及科学问题**

文章剖析知识图谱的实现流程，梳理总结知识图谱在实践中的应用; 为知识图谱的实现流程建立模型，分析它组织、存储、管理和更新大规模知识的过程及相关技术。

**1.2 同行专家如何解决**

随着大数据时代的到来，网络信息资源呈指数增长，万维网成为一个巨大的数据储藏库，但传统的 Web 内容组织结构松散，难以进行智能处理和应用，给大数据环境下的知识组织带来了很大的挑战。同时，用户对知识服务 的需求和期望值越来越高，准确有效地定位相关知识成为 网络用户的迫切需求。因此，根据新环境下知识组织的原则，需要从新的视角探索既符合网络信息资源发展变 化新特点，又适应用户认知需求的知识组织方法，从更深层次上揭示人类知识的整体性和关联性。

**1.3 本文所解决的问题**

本文详细介绍了知识图谱的实现流程，剖析了它是如何有效地组织、存储、管理和更新大规模的知识，并利用 存储的知识进行高效的推理计算和问题求解。同时，总结了知识图谱在商业搜索引擎、问答系统、电商平台、社交 网站中的应用情况。

**1.4 本文解决方案效果**

剖析了知识图谱是如何有效地组织、存储、管理和更新大规模的知识，并利用存储的知识进行高效的推理计算和问题求解。同时在实际应用当中也做了介绍，为初学者提供了很好的了解与学习。

**二、研究内容**

**2.1 理论与方法介绍**

知识图谱的实现流程可总结为 6 个模块，即知识获 取、知识融合、知识存储、查询式的语义理解、知识检索 和可视化展现。常识性知识的获取主要来自百科类站点和各种垂直站 点的结构化数据，如从 DBpedia 中抽取某一主题的知识， 根据一定的抽取策略提取出领域相关的事实，包括主题下的细分知识以及扩展的相关类别知识等。同时还要从一 些半结构化和非结构化数据中抽取实例和属性来丰富相关 实体的描述。

**2.2 验证分析与实验效果**

异构数据整合要进行数据清洗、实体对齐、属性值决 策以及关系的建立。数据清洗包括对拼写错误的数据、相 似重复数据、孤立数据、数据时间粒度不一致等问题进行处理; 实体对齐解决来自不同数据源的相同实体中对 同一特性的描述、格式等方面不一致的问题，对实体描述 方式和格式进行规范统一，如 “籍贯”与 “出生地”的 表述差别，日期书写格式的不同等; 属性值决策主要是针 对同一属性出现不同值的情况下，根据数据来源的数量和可靠度进行抉择，提炼出较为准确的属性值; 关系是知识 图谱中非常重要的知识，任何实体概念都不是孤立的，都 处在和周围概念一定的逻辑关系中，如等同关系、属分关 系和相关关系等。从本质上看，知识图谱建立关系的过程 可以简化为相关实体挖掘，即寻找用户类似查询中共现的 实体或是在同一个查询中被提到的其他实体，通过对链接 的提取统计以及对用户查询日志的分析，发掘查询式的主 题分布，把同一主题中的相关实体进行类型验证并建立关联。 实体的重要度主要通过 PageRank 等算法进行计算， 实体属性和实体间的关系、不同实体和语义关系的流行程 度、抽取的置信度等都会影响实体重要度计算的结果。 用户查询式中的实体被识别后，关于该实体的结构化摘要 就会展现给用户，当查询涉及多个实体时，就需要选择与 查询更相关且更重要的实体展现出来。如查询 “李娜”， 同名实体有超过 20 个，就要根据重要度的计算对这些实 体进行排序。 推理的规则一般涉及两类，针对属性和针对关系的。 通过推理验证可以检测逻辑矛盾，提高知识质量; 也可以 获取属性值和实体间隐含的关系，从而建立更多实体间的 关联。通过推理形成新的数据对知识进行再扩展，提高知识的完整性，并通过知识的聚合、分类等技术把知识 具体化和分类整合。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

知识图谱的发展仍面临很多难题和挑战，比如知识库的自动扩展，异构知识的处理，推理规则的学习等，本论文对知识图谱实现过程中具体技术细节分析的不够充分。

**3.2 后续研究重点**

首先是知识库的自动扩展，其中包括高精确度知识的获取、实时知识获取、长尾知识获取、利用群体智慧的知识获取，异构知识的处理，开放分类知识的建立，推理规则的学习以及知识可信度问题的解决等; 其次，如何更加准确地理解用户的需求、情感和个性化偏好，对搜索结果进行智能整理和展示还有待深入研究。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：基于混合推理的知识库的构建及 其应用研究

**（2）作者介绍**：钟秀琴，刘忠，丁盘苹

**（3）摘要**: 该文提出了基于OWL本体与Prolog规则的平面几何知识库的构建方法,从而可形式化地表示平面几何中丰富的语义信息.一方面,用类型、定义域、值域、分类、属性、实例等本体描述来表达结构化的知识,为领域内概念与概念之间关系的描述提供形式化的语义;另一方面,用Prolog规则来解决本体不能有效表达的诸如属性之间的关系和操作等问题,从而支持复杂关系间的推理.在此基础上,用Protégé和Prolog构建了一个基于本体和规则的平面几何知识库.实验证明:此知识库可实现知识和语义层次上的信息查询,还可进行复杂问题求解,其丰富的语义描述和混合推理能力弥补了传统知识库的不足.

**4.2 相关论文二**

**（1）题目**：本体知识库的自然语言查询重写研究

**（2）作者介绍**：王志，夏士雄，牛强

**（3）摘要**：为了让普通用户也能利用本体知识库来解决问题,提升本体知识库检索的灵活性和改进查询效率.结合本体知识库固有的特征,给出了查询系统的框架和处理过程,并对系统中关键的技术和算法进行描述.提出了基于WordNet的本体词汇扩展和查询三元组映射算法.通过实验,验证了该方法的有效性.

**4.3 相关论文三**

**（1）题目**：关于知识组织体系的若干理论问题

**（2）作者介绍**：马文峰，杜小勇

**（3）摘要**：知识组织体系是一套完整的描述与组织资源的机制,由概念及关系组成的概念体系是知识组织体系的核心。知识组织体系包括概念类聚体系和概念关联体系两个层次。作为新型知识组织体系的本体能够有效实现知识结构的描述与揭示,促使更大范围内资源的知识整合和知识检索成为可能。