

基于叙词表的领域知识图谱初始种子集自动生成算法

《中文信息学报 2018年8月》

韩其琛, 赵亚伟, 姚 郑, 付立军

(中国科学院大学 大数据分析技术实验室, 北京 100049)

摘 要: 随着认知计算的飞速发展, 通用知识图谱的自动构建取得了极大的进步, 但在垂直领域由于缺乏本体等语义信息, 导致进展缓慢。叙词表广泛分布于各个专业领域且蕴藏着丰富的语义信息, 如能对这些语义信息进行合理的提取和利用, 必然能在一定程度上帮助领域知识图谱的自动构建。该文提出两个假设, 利用假设可以从叙词表内部结构中提取实体类型和关系类型, 进而设计了一种基于叙词表的领域知识图谱初始种子集自动生成算法。最后, 以地质领域和林业领域的叙词表作为实验对象, 采用Bootstrapping算法, 利用由叙词表自动生成的初始种子集进行抽取工作, 通过对抽取到的结果进行分析, 结果表明利用叙词表得到的初始种子集可以取得同人工设计种子比较接近的效果。此外, 所提模型具有通用性, 为叙词表在构建领域知识图谱中的应用提供了一种新的思路。

关键词: 叙词表; Bootstrapping; 领域; 知识图谱

中图分类号: TP391 **文献标识码:** A