一种JSON和扩展JSON文档集合的模式提法方法

由于NoSQL面向文档数据库的日益普及，抽象-JSON文档正在成为数据表示的通用格式。这种流行的原因之一是它们在没有明确的数据模式的情况下处理大量数据的能力。但是，例如在数据检索、集成和分析任务期间，模式信息有时对应用程序至关重要。鉴于此背景，本文提出了一种方法——从存储在NoSQL面向文档的数据库或其他文件存储库的JSON或扩展的JSON文件集合中提取模式。考虑聚合操作以便为每个不同的结构获取模式，并且提出层级数据结构以对这些结构进行分组从而以JSON模式的格式生成全局模式。在真实数据集上进行的实验，如DBPedia和Foursquare，证明了生成的模式的准确性与相关工作相当甚至更优。

关键词：NoSQL；JSON；扩展的JSON；模式提取；JSON模式；面向文档的数据库

1. 简介

大数据市场