

doi:10.1631/FITEE.1700105

题目: 论文演化图: 学术文献多视角结构化检索

概要: 学术文献检索关注于选取最可能符合用户信息需求的论文。目前大部分检索系统局限于输出相关文献列表, 而这些检出文献相互独立。本文旨在揭示检索结果的相互关系。提出一种为学术文献建立结构化检索结果的方法, 称为论文演化图 (PEG)。PEG 采用多个演化链描述查询输入信息在不同主题方向的演化情况。通过论文作者、参考文献引用、论文内容信息这 3 个视角, PEG 能够发现文献之间各种潜在关系, 并多视角展示文献演化过程。该文献检索系统支持关键词、单篇论文、双论文 3 种查询方式。PEG 构造主要有 3 个步骤: 首先, 采用元图分解法把文献软聚合为多个群落, 获取每篇论文的主题分布; 其次, 从与查询相关的文献群落中提取主题连贯性演化链。每条演化链反映查询信息的某一视角; 最后, 提取的演化链组合形成论文演化图, 可以覆盖查询涉及的所有主题。基于真实文献数据库的实验结果表明, 该方法能够建立对用户有意义的论文演化图。

关键词: 论文演化图; 学术文献检索; 元图分解; 主题连贯性