

[论文名称] 基于聚类分析的数据挖掘方法在中国边疆研究中的实践与应用

[作者]: 常永宽, 网络室, 中国社会科学院中国边疆研究所

[摘要] 中国边疆研究所积累了大量的一手调研资料、学术文献、科研成果等异构数据, 这些数据相对分散、种类繁多, 给科研人员的检索利用带来了极大的不便。此外, 随着数据的不断增加, 内容交叉重复的冗余信息也随之增多, 降低了科研人员获取信息的效率和质量。因此, 整合这些异构数据, 并通过数据挖掘技术来充分利用这些数据势在必行。

本文将中国边疆研究所的实际应用为例来阐述和分析, 通过整合海疆研究、新疆研究、西藏研究、东北与北部边疆研究等各方向的异构数据资料, 构建统一的《中国边疆研究数据库》。并通过数据挖掘的聚类分析技术, 来抽取并分析各数据间的关联, 实现数据的聚合, 从而为科研人员提供更具有针对、专业性的数据资源。