信息与工程　　　　　　　　　　最新目录 | 下期目录 | 过刊浏览 | 高级检索　　◀◀ Previous Articles | Next Articles ▶▶

基于聚类技术的XML文件代表性结构获取

(吉首大学软件服务外包学院，湖南 张家界 427000)

Representative Structures from XML Documents Based on Clustering Techniques

(Software & Outsourcing Institute,Jishou University,Zhangjiajie 427000,Hunan China)

- 摘要
- 参考文献
- 相关文章

全文: PDF (209 KB)　HTML (1 KB)　**输出:** BibTeX | EndNote (RIS)　　**背景资料**

摘要 XML文件可以利用树状结构来表示，于是把如何将XML文件做聚类看成如何对树状结构的数据作聚类.使用SOM聚类工具搭配上Jaccard 的距离测量公式来对XML 文件做聚类，然后在每个cluster 中利用GST(Graph Search Technique)算法从这些XML文件当中找出他们的最大序列，最后将这些最大序列融合起来成为共同的结构.

关键词:　XML文件　树状结构　聚类　序列挖掘　相同结构

Abstract:　Since an XML document can be represented as a tree structure,the problem how to cluster a collection of XML documents can be considered as how to cluster a collection of tree-structured documents.The author used SOM (Self-Organizing Map) with the Jaccard coefficient to cluster XML documents.Then,an efficient sequential mining method called GST was applied to find maximum frequent sequences.Finally,the author merged the maximum frequent sequences to produce the common structures in a cluster.

Key words:　XML document　tree-structured　clustering　sequential pattern mining　common structure

服务
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- E-mail Alert
- RSS

作者相关文章
- 卓月明

作者简介：卓月明（1970-），男，湖南慈利人，吉首大学软件服务外包学院副教授，硕士，主要从事数据库和智能计算研究.

引用本文:

卓月明. 基于聚类技术的XML文件代表性结构获取[J]. 吉首大学学报自然科学版, 2011, 32(6): 55-58.

ZHUO Yue-Ming. Representative Structures from XML Documents Based on Clustering Techniques[J]. Journal of Jishou University（ Natural Sciences Edit, 2011, 32(6): 55-58.

[1] WEI Jin-mao,WANG Shu-qin,WANG Jing,et al.Fast Kernel for Calculating Structural Information Similarities [C] //Proc. the 3rd IEEE International Conference on Intelligent Systems.Varna,Bulgaria,2006:59-64.

[2] YANG Jian-wu,WILLIAM K CHEUNG,CHEN Xiao-ou.Learning the Kernel Matrix for Xml Document Clustering [C]//Proc. the IEEE International Conference on e-Technology,e-Commerce and e-Service.Hong Kong,China,2005:353-358.

[3] HUANG Yin-fu,LIN Shao-yuan.Mining Sequential Patterns Using Graph Search Techniques [C]//Proc. the 27th Annual International Conference on Computer Software and Applications.Washington,DC,USA,2003:4-9.

[4] ALEXANDRE TERMIER,MARIE-CHRISTINE ROUSSET,MICHèL SEBAG.TreeFinder:A First Step Towards Xml Data Mining [C]//Proc. the IEEE International Conference on Data Mining.Maebashi City,Japan，2002:450-457.

[5] AGRAWAL R,SRIKANT R.Mining Sequential Patterns [C]//Proc. the 11th IEEE International Conference on Data Engineering.Taipei,Taiwan，1995:3-14.

[6] HUANG Yin-fu,LIN Shao-yuan.Mining Sequential Patterns Using Graph Search Techniques [C]//Proc. the 27th Annual International Conference on Computer Software and Applications.Washington,DC,USA, 2003:4-9.

没有找到本文相关文献