

STATISTICAL ANALYSIS AND DATA MINING

Original Article

MiTexCube: MicroTextCluster Cube for online analysis of text cells and its applicationsDuo Zhang , ChengXiang Zhai, Jiawei Han

First published: 05 October 2012

<https://doi.org/10.1002/sam.11159>

Cited by: 1

 About  Access

PDF

Tools



Share

Abstract

A fundamental problem of multidimensional text database analysis is efficient and effective support of various kinds of online applications, such as summarizing the content of a text cell or comparing the contents across multiple text cells. In this paper, we propose a new infrastructure called MicroTextCluster Cube (or *MiTexCube*) to support efficient online text analysis on multidimensional text databases by introducing *micro-clusters* of text documents as a compact representation of text content. Experimental results on real multidimensional text databases show that (i) *MiTexCube* can be materialized efficiently with reasonable overhead in space, and (ii) applications based on the proposed materialized *MiTexCube* are more efficient than the baseline method of direct analysis based on document units in each cell, without sacrificing much quality of analysis, and *MiTexCube* naturally accommodates flexible trade-off between efficiency and quality of analysis. © 2012 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 6: 243–259, 2013

Citing Literature

Number of times cited: 1

Fan Wang and Heping Zhang, Matchings extend to Hamiltonian cycles in k-ary n-cubes, *Information Sciences*, **305**, (1), (2015).

[Crossref](#)

About Wiley Online Library

Help & Support

Opportunities

Connect with Wiley

