Abstract

Mining of frequent traversal paths in web logs is an application of sequence mining and useful with many applications that include web recommendation, caching, pre-fetching etc. Most of the existing algorithms follow a bottom-up approach to mine sequence patterns in a database. In this paper, a fast top-down algorithm is presented to discover maximal traversal paths which are contiguous sequences in web log session sequences. The algorithm avoids candidate sequence generation and searches only maximal potential patterns in the minimized search space during mining process. Experimental results show that the proposed algorithm
can perform better than an existing approach.

Reference

Index Terms

Computer Science

Information Retrieval

Key words

Sequence Database

Contiguous Sequence

Maximal Potential Pattern

Maximal Traversal Path