Abstract

With the rapidly growing uses of World Wide Web for various important and sensitive purposes it becomes a sensible necessity to find out the interesting web access patterns from the web access sequences tracked by users frequently. Web access sequential patterns can be used to achieve business intelligence for e-commerce sites and also can be used to analyze system performance. This paper proposes a more efficient web mining algorithm which mines all the sequential patterns from the web access sequences and totally eliminates the concept of linking between nodes. The algorithm uses the aggregate tree structure for mining and then mines from the tree using RST (Root-set of Suffix Trees) for same prefix items. The algorithm finds the frequent sequential patterns by recursively traversing the tree from root-nodes to child-nodes for the length-1 frequent items. The proposed approach doesn’t need to generate any projected tree; it needs only the root-set for each prefix that got in previous step. Experimental results show huge performance gain over the FOF and WAP-tree mining techniques by considerably reducing the mining time.
References


Index Terms

Computer Science
Information Sciences

Keywords

Frequent sequential pattern Web access sequence Web log mining WAP-tree
First-Occurrence Forest (FOF)

and Root-set of Suffix Tree (RST).