Abstract

This paper identifies the scope for improvement in the execution of baseline kNN join algorithms in a distributed environment. Improvements are suggested and the improved methods are applied in performing kNN joins on R*-Trees. The effectiveness of the proposed improvements have been experimentally verified and presented.

Index Terms

Computer Science
Distributed Systems
Keywords

R*-Trees  kNN join  Hadoop  Mapreduce  z-values.