Abstract

This paper describes a method to optimize the plane sweep algorithm. The goal of this paper is to develop a method that reduces the comparison through removing the tandem replicated word comparison and also using partial search technique in a document for escaping from the keywords that are ineffective. The approach introduces Plane Sweep algorithm that is the base algorithm used to search for keywords. Reducing the search area, change the number of keyword's comparisons in a document and speed up our search algorithm. So searching operation is done in a smaller space and we don't need to search all the keywords in a document. In this algorithm, we make a new technique to create the algorithm that detect the number of tandem replicated words in a document and also searching on a target part, thus reducing the number of keywords in a document speed up our search algorithm. In proposed algorithm time complexity with lower order has been created than the basic algorithm. Searching for results occurs in a reduced space and it has led to a better performance without comparing all the keywords in the list. The algorithm is robust, and highly effective especially in a high volume of data.
- Hao Yan, Shuming Shi, Fan Zhang, Torsten Suel, Ji-rong Wen, Efficient Term Proximity Search with Term-Pair Indexes, the 19th ACM conference on Conference on information and knowledge management CIKM 10 (2010).
- Ralf Schenkel, Andreas Broschart, Seungwon Hwang, Martin Theobald, Gerhard Weikum, Efficient Text Proximity Search, String Processing and Informatio Retrieval (2007).
- Hao Yan, Shuming Shi, Fan Zhang, Torsten Suel, Ji-rong Wen, Efficient Term Proximity Search with Term-Pair Indexes, the 19th ACM conference on Conference on information and knowledge management CIKM 10 (2010).

Index Terms

Computer Science

Algorithms

Keywords

Plane sweep algorithm   Replicated data   String matching   Optimized algorithm partial search

Text retrieval

Proximity search