

摘 要

随着科学技术的进一步发展,各种数据正以前所未有的速度增长着。如何有效存储,索引,搜索图数据成为一个日益显著的问题,因此图搜索已成为一个热门话题,并在大量领域都有应用,如分子化学,药物学,传感器网络,关系网络,XML 文档等等。

本文首先介绍了图搜索的基本概念,并探究了目前几个经典的图搜索算法,分析总结了这些算法优缺点。

然后,针对精确搜索,本文以 **GraphGrep** 为基础提出了一种基于二次哈希开链法的新算法,解决了 **GraphGrep** 在索引定位中哈希容易产生冲突的问题,加快了搜索效率,并通过实验比较了两算法性能。

其次,针对相似性搜索,本文结合 **G-Hash** 中小波匹配核函数和简化包表示提出了一种基于节点相似度的新算法,在效率不逊于 **G-Hash** 的基础上提高了匹配精度,降低了编码难度,并通过实验证明了此算法的正确性与可行性。

关键词: 图搜索 图精确搜索 图相似性搜索 二次哈希开链 节点相似度

ABSTRACT

With the further development of the science and technology, all kinds of data is growing at an unprecedented speed. How to effectively store, index and search figure data become an increasingly important problem, so the graph search has become a hot topic, and are used in many areas, such as molecular chemistry, pharmacology, sensor networks, human network, XML documents, and so on.

This paper first introduces the basic concept of graph search, and explore the current several classic graph search algorithm, these algorithms were analyzed advantages and disadvantages.

Then, in view of the accurate search, this paper put forward a new algorithm based on the secondary hash chain method, on the basis of the GraphGrep, to solve the GraphGrep hash prone to conflicts when index positioning, speed up the search efficiency, and compared the two algorithms performance through experiments.

Secondly, in view of the similarity search, this paper combine G-Hash wavelet matching kernel function and Reduced Bag to node similarity, then puts forward a new algorithm based on node similarity, the efficiency is not inferior to the G-Hash, But improved the precision of matching, reduces the coding difficulty, and through the experiment proves the correctness and feasibility of this algorithm.

Keywords: Graph search Graph accurate Search Graph Similarity Search
twice-hash chain node similarity

