摘要

为改进传统检索方式没有有效利用专利图片的不足，以及结合图片检索大数据量的特点，本文介绍了一种基于Hadoop的分布式外观设计专利图像检索系统。本系统除了采用基本的文本检索方式以外，加入了图像检索方式。采用的基于内容的图像检索，即对图像的颜色、形状、纹理等特征进行特征融合，并结合其他特征提取，相似度计算的图像检索算法，对两张图片进行相似度计算，并将这一过程搭建在Hadoop分布式系统上，提供大数据量，高效率，高并发的检索过程。系统在对图像进行检索作业时，对将要进行检索的图像专利特征库进行分割，分割后的数据块传递给分布式系统中的节点，节点中的Map任务在对接收到的数据块进行计算检索处理并得到结果，最后Reduce任务对 Map任务结果进行最终处理输出到HDFS中。

关键词：分布式,Hadoop,外观设计专利,图像检索

Abstract

In order to improve the shortage of traditional search methods which doesn’t use the image of patents, and taking into account large amount of data during image retrieval, this paper introduces a design patent image retrieval distributed system based on Hadoop.The system use the basic text retrieval methods, also use image retrieval methods. Content-based image retrieval which used in this system, extract features include the color, shape, texture and other characteristics of the image, combined with other feature extraction and similarity calculation algorithm for image retrieval.The System builded on Hadoop support large data volume, high efficiency and high concurrent retrieval process. While retrieving image , the image feature database will be splited into blocks,these data block is transmitted to the distributed nodes in the system.Map tasks on tasknode process these data blocks and send result to reduce tasks.Finally reduce tasks collect result and output to HDFS system.

**Keywords:** distributed,Hadoop,design patent , image retrieval