

Analyzing the Load Balance of Term-based Partitioning

Ahmad Abusukhon
Faculty of Science & IT
Al-Zaytoonah Private University of Jordan
Amman Jordan
ce4aab@student.sunderland.ac.uk
Mohammad Talib
Department of Computer Science
University of Botswana
Private Bag UB 00704, Gaborone, BOTSWANA
talib@mopipi.ub.bw

Abstract— In parallel (IR) systems, where a large-scale collection is indexed and searched, the query response time is limited by the time of the slowest node in the system. Thus distributing the load equally across the nodes is very important issue. Mainly there are two methods for collection indexing, namely document-based and term-based indexing. In term-based partitioning, the terms of the global index of a large-scale data collection are distributed or partitioned equally among nodes, and then a given query is divided into sub-queries and each sub-query is then directed to the relevant node. This provides high query throughput and concurrency but poor parallelism and load balance. In this paper, we introduce new methods for terms partitioning and then we compare the results from our methods with the results from the previous work with respect to load balance and query response time.