Digital Healthcare Empowering Europeans
R. Cornet et al. (Eds.)

doi:10.3233/978-1-61499-512-8-993

A Novel Cache and SSD-based Index Structure for Health Record Indexing

Yang Du[a,1] and Sule Yildirim-Yayilgan[a]
[a]*Department of Information and Media Technology, Gjøvik University College, Norway*

Keywords. Data Retrieval, Indexing

Introduction

Textual-based indexing has been widely used in various medical databases. It plays increasingly important role in modern medical information systems since it directly affects the diagnosis time. Building an efficient indexing system will greatly improve the quality of treatment and save lives in emergency departments of hospitals. Researchers have set up various algorithms to optimize the query response time. B+ tree is ubiquitous in all kinds of databases as a well-known index structure with superior performance [1]. However, it shows poor performance especially when the database operation involves many retrievals. It has been proven that query performance can be significantly improved if the node size is equal to cache size. Fractal prefetching B+ Tree was developed by a group of researchers to improve the retrieval performance. Cache-first approach produces good results but it cannot simultaneously utilize CPU cache and SSD page. We propose a multi-dimensional tree structure named Cache-SSD-Based tree (Fig.1).

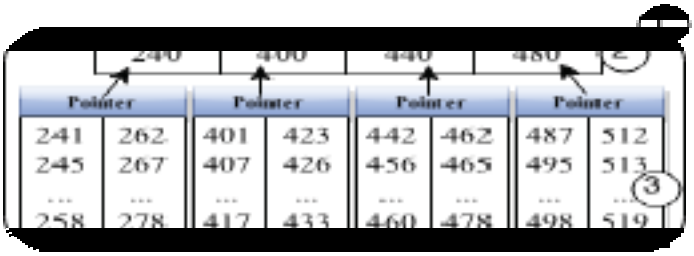

Figure 1. The structure contains: 1. Multi-element node, 2. Array node, 3. Child node. Each child contains 2 elements (left and right).

1. Methods

We designed a multi-element node structure to enlarge the node size and further influence the depth of the tree. This way, the average search length is expected to be

[1] YangDu.

reduced. We will investigate the relationship between the node size, CPU cache size and SSD page size to find the best node size for a reasonable performance.

In our approach, index structure uses CPU cache and SSD page. Each multi-element node ① consists of an array node ② and corresponding child nodes ③, each with left and right elements. We place the constraint that the multi-element node is equal to the cache size and each of left and right elements equals SSD page size.

2. Results

Until now, we have built an indexing algorithm for the proposed indexing structure. Our results will particularly be useful in registration, indexing and retrieval of ski injuries in Norway [2]. We plan to integrate ski injury registration system with hospital system.

The performance of database index is decided by time and space complexity of the algorithm used for indexing. It is affected by hardware restrictions and hence some algorithms with the same time complexity perform differently. Comparing time complexity with the competing algorithms' will help find out the bottleneck of cache-based index structure. We will build a cost model to evaluate the performance.

References

[1] Sippu, Seppo, and Eljas Soisalon-Soininen. "B-Tree Traversals." Transaction Processing. Springer International Publishing, (2014). 159-183.

[2] Jeppesen, J. C., and S. Y. Yayilgan. "Systematic ski injury registration using a digital platform." Studies in health technology and informatics 205 (2014): 1251.