Combining Metric Cache and D-cache-enhanced MAMs for Maximizing Efficiency of Similarity Search

K.Karthik, G.Pradeep Reddy Mtech 2nd year , Lecturer Dept of cse, jntuacea,Dept of cse, jntuacea

Abstract

Caching is used in computer applications to speed up the performance data retrieval. It is widely used in database management systems where queries are very frequent for reducing I/O cost. However, the situation is dramatically different in modern complex databases consisting of snapshots of nature (i.e., images, sounds, or other signals), like multimedia databases, bioinformatics databases, time series, etc. Here, people often adopt the similarity search within the content-based retrieval paradigm; (MAMs) are suitable in situations where the similarity measure is a metric distance. In particular, metric access methods (MAMs), used for similarity search in complex unstructured data, have been designed to minimize rather the number of distance computations than I/O cost while making indexing or querying.Recently Skopal, Lokoc, and Hustos proposed D-cache with MAMs such as Pivot tables, *M*-tree and *D*-file to improve performance of query processing. In this paper we propose a combined approach that uses metric cache in front of D-cache enhanced MAMs for boosting the performance of similarity search further. We built a prototype in Java platform to demonstrate the proof of concept. The empirical results reveal that the hybrid approach yields better performance.

Index Terms –MAMs, D-Cache, index-free search, metric cache, similarity search

INTRODUCTION

Efficient data retrieval plays an important role in modern database management systems. In order to reduce I/O operations and thereby eliminate efficiency issues, caching is used in databases. Both caching and indexing strategies [1], [2] are widely used along with storage layouts [3]. These are considered while designing database management systems. RDBMS has become a stable alternative for storing and retrieving valuable business data permanently. In such systems, caching is used to improve query performance. They also realize query performance with various indexing strategies in place. End users need speed in query processing. Delay in processing causes more user wait time that is not expected by end users. Users are too good to see the results as faster as possible. In this context disk caching is highly efficient where queries are very frequent with respect to user sessions. Disc caching has been used for reducing I/O cost incurred while making queries on relational databases. However, in the real world there might be scenarios to use databases containing multimedia content like sounds, images, and other information. For instance time series databases, bioinformatics databases and multimedia databases have already overcome the problem of disk I/O and they focused on distance computations. In such databases similarity search is often used and their needs improvement in processing such queries. Indexing techniques have been used traditionally in order to speed up query processing. The strategies used for indexing are general or domain specific. Whatever be the nature of indexing, it is basically to improve the query processing. Distributed indexing strategies [4] are also developed for speeding up similarity queries made for distributed databases. In is a fact that the data retrieval performance is affected more by CPU cost than I/O cost. Therefore in case of similarity search queries it is important to reduce CPU cost rather than I/O cost [5], [6]. In this context, it is minor issue with I/O cost incurred while performing similarity searches. Here the computation cost is considered to be the number of computations need for answering a specific query. Metric Access Methods (MAMs) are suitable candidates for accessing data based on similarity searches that needs to use distance metrics. The properties of metrics can help to have database organized in memory in order to process queries faster [7], [3], [1], [2]. Later on indexing is also used to answer similarity queries faster. The queries might be k-nearest neighbors (kNN) or other queries like content based image retrieval queries which are modeled after Query by Example (QBE) concept. MAMs are extremely useful in processing such queries as they eliminate unnecessary equivalence classes from the process of search. Reduced computation of query is achieved with MAMs as explored in [8], [9], [5], and [6].

In [10] D-Cache is used along with enhanced MAMs such as D-file, M-tree, and Pivot table in order to improve query processing with respect to similarity searches. In this paper we enhance the system further by using metric cache along with D-cache. Our contributions are as given below.

- 1. We propose metric cache that is placed in front of D-cache for improving the similarity search on metric data assets.
- 2. We propose an architecture which acts as a framework for accommodating D-cache and metric cache for processing similarity searches.
- 3. We built a prototype application in Java platform that demonstrates the proof of concept. The application makes use of huge amount of unstructured data to test the efficiency of the proposed system.

The remainder of the paper is structured as follows. Section II reviews literature pertaining to caching, Dcache, MAMs and other relevant topics. Section III provides information related to the proposed system. Section IV presents experimental results while section V concludes the paper.

RELATED WORKS

MAMs have been around for some time. However, distance cache for general purpose MAMs is the first time experimented in [10]. In this paper we improve them further with combination of metric cache. There has been caching concept in computer applications which buffers the frequently used data so as to avoid retrieving from hard disk or databases. This will decrease I/O cost dramatically. Caching in the context of databases with various kinds of data is explored in [11] and [12]. Cached distances are used for range queries where M-tree is used as underlying data structure as proposed by Kailing et al. [13]. When the distances are computed from route object to the query, they are cached for future retrievals. This will improve performance of such queries. As the performance gain is more the memory cost of this is tolerable [13]. However, caching of unnecessary distance measures is not done in order to optimize performance further. Batch indexing and querying was also used. Bulk loading f data is part of many applications where good performance is required. To deal with such data two algorithms were proposed in [14]. They made use of index structures for boosting search performance such as Slim-tree and R-tree. The drawback is that the performance is degraded once new objects are inserted into the tree. To overcome this drawback M-tree was proposed which can support parallel insertion of objects. In [15] another approach was proposed to improve efficiency of MAMs. Instead of issuing multiple queries a batch of queries can be used in order to reduce computational andI/O cost. In [10] the D-cache technique proposed reduced I/O cost and also CPU cost through MAMs which have been enhanced to work with D-cache. Knearest neighbor graph concept is used in [16] forperforming kNN queries.

Query result caching plays an important role in speeding up similarity searches. This idea was explored in [17] and [18]. History of similarity queries and answers can be stored in metric cache for performance improvement. In [10] D-cache is used for performanceimprovement of similarity searches. They also hint about the usage of high level metric class in front of D-cache with enhanced MAMs. In this paper we implemented the combined cache where metric cache is in front of D-cache to enhance the query processing speed.

PROPOSED FRAMEWORK

Before presenting our new framework for processing similarity searches we would like to introduce Dcache, enhanced MAMs briefly. However, more information can be found in [10]. D-cache is a data structure which resides in main memory which is meant for storing the distances computed by MAMs. The D-cache stores distance computations in the form of triplets as shown below where the first two are unique identifiers for objects while the third one is the distance between the objects.

$[id(o_i), id(o_j), \delta(o_i, o_j)],$

The content of D-cache is made up of many such triplets. However, it can also be visualized as a sparse matrix as given below where rows and columns represent objects and the cells and their values indicate the distances between objects.

		o_1	o_2	o_3	•••	o_n
	o_1	(δ_{12}	δ_{13}		o_n
	o_2	δ_{21}				δ_{2n}
D =	o_3					
	o_m	δ_{m1}		δ_{m3}		

The operations performed on D-cache include distance retrieval and distance insertion. The enhanced MAMs used by D-cache include D-file which is best used for range query and kNN query, pivot tables and M-tree. More details and algorithms about the usage of D-cache with enhanced MAMs can be found in [10].The proposed framework combines both Metric Cache and D-Cache in order to improve performance of range queries and kNN queries. The architectural overview of the proposed approach is as shown in figure 1.



Fig. 1 –Proposed architecture that combines D-Cache and metric cache

As can be seen in figure 1, it is evident that two kinds of caches are in place for better performance. The metric cache is basically for storing history of similarity queries. The details stored in metric cache include descriptors and ids of database objects that have been returned by queries. When a new query is encountered, the metric cache returns result if exact value is found in metric cache. The metric cache might give a single value or multiple values based on Sometimes the queryrequirements. it may giveapproximate answer by combining a near correct value with other values. In such cases large retrieval error is likely to occur and the metric cache forwards that query to the retrieval system while updating itself. The D-cache on the other hand stores database objects, and their distances. They play a role to work with enhanced MAMs.

PROTOTYPE IMPLEMENTATION

We built a prototype application in Java platform with graphical user interface. The environment used is a PC with 4GB RAM, core 2 dual processor running Window

Creating 1000 reference objects. Reference object file already exist! Manually remove it if you want to create a new one. Inserting 10000 objects in the database. It might take a while Performing similarly searches KNN search Result 0 object 13624 Computed distance: 70.0 object 13624 Computed distance: 73.0 Actual distance: 136224 1565424 0 object 13624 Computed distance: 73.0 Actual distance: 136224 1565424 0 object 13624 Computed distance: 73.0 Actual distance: 136224 1565424 0 object 13624 Computed distance: 75.0 Actual distance: 1562094531561424 0 object 13624 Computed distance: 75.0 Actual distance: 156096453165442 0 object 13624 Computed distance: 78.0 Actual distance: 1560964531651478 0 object 13624 Computed distance: 78.0 Actual distance: 1560964531651478 0 object 13624 Computed distance: 78.0 Actual distance: 1742842621617904 KNNetreves earch Result 0 object 13624 Computed distance: 73.0 Actual distance: 17426426272238632 object 13624 Computed distance: 78.0 Actual distance: 1742642621627223632 Actual distance: 17426842621617904 KNNetreves earch Result 0 Actual distance: 1742642621617904 KNNetreves earch Result 0 object 13624 Computed distance: 73.0 Actual distance: 1302914165168424 0 Actual distance: 17426842621617904 Actual distance: 17426842621617904 Actual distance: 17426842621617904 KNNetreves earch Result 0 Actual distance: 17426842621617904 Actual distance: 174208426216179	S Van Result		
1000 reference objects Reference objects file already woldt Manually remove It if you want to create a new one. Inserting 10000 objects in the database. It might take a while Performing similarity searches kNN search Result 0 object 30 Computed distance: 0.0 Actual distance: 0.01 object 3124 Computed distance: 718.0 Actual distance: 19642495272236832 object 3124 Computed distance: 73.0 Actual distance: 10.01 object 3124 Computed distance: 73.0 Actual distance: 10.02 Actual distance: 10.0201414518424 object 127440 Computed distance: 73.0 Actual distance: 10.01 Actual distance: 10.01 object 13124 Computed distance: 72.0 Actual distance: 10.01 Actual distance: 10.01 object 13124 Computed distance: 70.0 Actual distance: 10.01 Actual distance: 10.01 object 13124 Computed distance: 71.00 Actual distance: 10.01 Actual distance: 0.01 object 13124 Computed distance: 72.00 Actual distance: 10.01 Actual distance: 10.04245272236632 <	Creating		
100000 objects in the database. It might take a while Performing similarity searches Nun search Result 0 object 30 Computed distance: 10 Actual distance: 100 Actual distance: 101 object 31524 Computed distance: 7320 Actual distance: 1362/24527223682 object 13624 Computed distance: 730 Actual distance: 1362/24527223682 object 13624 Computed distance: 730 Actual distance: 1362/24527223655 object 12019 Computed distance: 750 Actual distance: 156209453156142 object 12019 Computed distance: 780 Actual distance: 1562094531561478 object 13054 Computed distance: 780 Actual distance: 1562094531561478 object 131524 Computed distance: 780 Actual distance: 156220520110839 object 131524 Computed distance: 780 Actual distance: 156224521673004 Object 131524 Computed distance: 730 Actual distance: 15622452172236532 object 131524 Computed distance: 730 Actual distance: 156245216723065533 object 151674 Computed distance: 730	1000 reference objects Reference object file already exist! Manually re Inserting	move it if you want to create a new one.	
NNN search Result 0 Objert 21524 Computed distance: 0.0 Actual distance: 0.01 Objert 21524 Computed distance: 7180 Actual distance: 119642495272236632 Objert 21524 Computed distance: 7320 Actual distance: 136524115720055833 Objert 12740 Computed distance: 7320 Actual distance: 136024146518424 Objert 12705 Computed distance: 7520 Actual distance: 10562115200302561325465 Objert 250180 Computed distance: 7520 Actual distance: 105625138064825297 Objert 217595 Computed distance: 7500 Actual distance: 105370120110839 Objert 70364 Computed distance: 800 Actual distance: 10527023010839 NNtertives search Result O Actual distance: 10642495272236632 Objert 31524 Computed distance: 730 Actual distance: 10642495272236632 Objert 31524 Computed distance: 730 Actual distance: 0.01 Actual distance: 0.01 Actual distance: 10642495272236632 Objert 131524 Computed distance: 730 Actual distance: 106242652167200655833 Objert 131524 Computed distance: 730 Actual distance: 1300291416518424 Objert 13674 <	100000 objects in the database. It might take a Performing similarity searches	a while	
object 130 Computed distance 100 Actual distance 1.01642495272236832 object 127448 Computed distance 7180 Actual distance 1.16962495272236832 object 127448 Computed distance 7300 Actual distance 1.16962495272236832 object 127448 Computed distance 7300 Actual distance 1.16962495272236832 object 137168 Computed distance 7300 Actual distance 1.169611380868366 object 250882 Computed distance 7760 Actual distance 1.07314027286267 object 150415 Computed distance 7300 Actual distance 1.06031480868366 object 150415 Computed distance 7300 Actual distance 1.06034643151478 object 150415 Computed distance 7300 Actual distance 1.06241627208627 object 131524 Computed distance 710 Actual distance 1.062419720055833 object 131524 Computed distance 7300 Actual distance 1.01 object 11679 Computed distance 7300 Actual distance 1.0024146518422 object 11679 Computed distance 7300 Actual distance 1.0024146518424 object 11679 Computed distance 7300 Actual distance 1.001 object 11679 Computed distance 7300 Actual dista	kNN search Result: 0		
object 131524 Computed distance 7130 Actual distance 119642465272236832 object 131524 Computed distance 7320 Actual distance 156254216723065833 object 131574 Computed distance 7320 Actual distance 130231416158424 object 130174 Computed distance 7520 Actual distance 130231416158424 object 120180 Computed distance 7520 Actual distance 130230416158424 object 20180 Computed distance 7520 Actual distance 13023041253556 object 73054 Computed distance 7800 Actual distance 15500964531851478 object 73054 Computed distance 8080 Actual distance 16500964531851478 object 73054 Computed distance 8080 Actual distance 1652200230110839 object 31524 Computed distance 7320 Actual distance 1054245272230632 object 31524 Computed distance 7320 Actual distance 11642452162180205833 object 131524 Computed distance 7320 Actual distance 11642452172230632 object 131524 Computed distance 7320 Actual distance 130231416516424 object 131524 Computed distance 7320 Actual distance 130231416516424 object 131524 Computed distance 7320	object: 30 Computed distance: 0.0	Actual distance: 0.01	
object 127446 Computed distance 7320 Actual distance 1654216720005683 object 127446 Computed distance 7380 Actual distance 130201416518424 object 12718 Computed distance 7380 Actual distance 130201416518424 object 20180 Computed distance 7300 Actual distance 130201416518424 object 20183 Computed distance 7300 Actual distance 130201416518424 object 21038 Computed distance 7300 Actual distance 130201416518424 object 21038 Computed distance 7300 Actual distance 1303016927386237 object 21038 Computed distance 7320 Actual distance 1405204221817904 Whitedtaree search Result O O object 21524 Computed distance 710 Actual distance 1062452172220652 object 131524 Computed distance 710 Actual distance 101 object 11524 Computed distance 7320 Actual distance 102452157220652 object 11679 Computed distance 7320 Actual distance 130221416518424 object 11679 Computed distance 7320 Actual distance 130221416518424 object 11679 Computed distance 7520 Actual distance 130221416518424 <t< td=""><td>object: 31524 Computed distance: 718.0</td><td>Actual distance: 1.19642495272236632</td><td></td></t<>	object: 31524 Computed distance: 718.0	Actual distance: 1.19642495272236632	
object 181674 Computed distance 7.520 Actual distance 1.520303261325365 object 120180 Computed distance 7.520 Actual distance 1.7620303261325365 object 20180 Computed distance 7.520 Actual distance 1.762661138064836 object 20180 Computed distance 7.520 Actual distance 1.5500964531861478 object 170546 Computed distance 7.520 Actual distance 1.6500964531861478 object 70364 Computed distance 7.520 Actual distance 1.6500964531861478 object 70364 Computed distance 8.08.0 Actual distance 1.645270230110839 object 30 Computed distance 7.00 Actual distance 1.17426842621617904 HNINetrive sert Result 0 Actual distance 1.0542495272230632 object 31524 Computed distance 7.30 Actual distance 1.1942495272230632 object 131524 Computed distance 7.30 Actual distance 1.30291416518424 object 131524 Computed distance 7.30 Actual distance 1.30291416518424 object 131524 Computed distance 7.30 Actual distance 1.30291416518424 object 131524 Computed distance 7.50 Actual distance 1.30291416518424 object 13019 Computed distance 7.50	object: 127448 Computed distance: 732.0	Actual distance: 1.65542167280655833	
Object 114/19 Computed distance /750 Actual distance 14/2018/02/611360868366 Object 20188 Computed distance /750 Actual distance 1760 Object 20188 Computed distance /750 Actual distance 1760 Object 20188 Computed distance /750 Actual distance 1760 Object 20189 Computed distance /750 Actual distance 16/1820/202011053 Object 20189 Computed distance /750 Actual distance 16/1820/202011053 Object 20180 Computed distance /750 Actual distance 10/1820/202011053 Object 21524 Computed distance /750 Actual distance / 10/2020 Actual distance / 10/2020 Object 21524 Computed distance / 710 Actual distance / 10/2020 Actual distance / 10/2020 Object 21524 Computed distance / 720 Actual distance / 10/2020 Actual distance / 10/2020 Object 11679 Computed distance / 720 Actual distance / 10/2020 Actual distance / 10/2020 Object 11679 Computed distance / 720 Actual distance / 720 Actual distance / 10/202030261352056 Object 12017 Computed distance / 720 Actual distance / 720030261352565 Object 120190 Object 700	object: 81679 Computed distance: 738.0	Actual distance: 1.3802914165188424	
object 200180 Computed distance 780 Actual distance 10/9801 1300960360 object 217592 Computed distance 780 Actual distance 1650009451861478 object 127592 Computed distance 7960 Actual distance 1658009451861478 object 120164 Computed distance 8080 Actual distance 1658009451861478 object 120164 Computed distance 8080 Actual distance 1645201230110639 object 31524 Computed distance 8080 Actual distance 117426842621617904 HNIvefteve Sector Result O object 31524 Computed distance 700 Actual distance 1045246270230632 object 131524 Computed distance 7320 Actual distance 119642495272230632 object 131524 Computed distance 7320 Actual distance 130291416518424 object 131524 Computed distance 7320 Actual distance 130291416518424 object 131524 Computed distance 7320 Actual distance 130291416518424 object 131524 Computed distance 7520 Actual distance 13029141518424527555 object 120190 Computed distance 7520 Actual distance 176955113809808036 object 20100 Computed distance 7780 Actual distance 17695513504069263707	object 143/19 Computed distance: 752.0	Actual distance: 1.45208302561352585	
object 12/058 / Computed distance / 780 Actual distance : 10/03/4/42/786/29 object 15/041 / Computed distance / 780 Actual distance : 16/0410/27/2011/030 object 15/041 / Computed distance : 780 Actual distance : 16/0410/27/2011/030 object 21/042 Computed distance : 10 object 21/042 Computed distance : 710 object 11/042 Computed distance : 710 object 11/042 Computed distance : 710 object 11/042 Computed distance : 720 object 11/042 Computed distance : 720 object 11/042 Computed distance : 750 object 12/042 Computed distance : 750 object 12/0412 Computed distance : 750 object 12/0142 Computed distance : 750 object 12	object: 260180 Computed distance: 776.0	Actual distance: 1.76965113869686366	
object 17/95L Computed distance 756D Actual distance 1.652/0250110639 object 17/95L Computed distance 756D Actual distance 1.658270250110639 object 17/95L Computed distance 756D Actual distance 1.658270250110639 object 17/95L Computed distance 750D Actual distance 1.658270250110639 object 30 Computed distance 700 object 15/24 Computed distance 7.00 Actual distance 1.19642495272230632 object 15/24 Computed distance 7.20 Actual distance 1.3962416516424 object 15/24 Computed distance 7.20 Actual distance 1.3962416516424 object 15/24 (Computed distance 7.50 Actual distance 7.500302561352565 object 200100 Computed distance 7.70 Actual distance 7.570 Actual distance 7.570 Actual distance 7.570400726507	object 250883 Computed distance: 788.0	Actual distance: 1.60/304082/89529/	
object 17 0364 Computed distance: 808.0 Actual distance: 1.7426842621617904 NNNetlives saarch Result Actual distance: 0.01 object 31524 Computed distance: 7.02 Actual distance: 1.16424952722236632 object 131524 Computed distance: 7.32 Actual distance: 1.1642495272236632 object 131524 Computed distance: 7.32 Actual distance: 1.302614165168424 object 12017 Computed distance: 7.52 Actual distance: 7.52 Actual distance: 1.3695611369066336 object 120170 Computed distance: 7.52 object 120180 Computed distance: 7.52 object 120180 Computed distance: 7.52 object 120190 Computed distance: 7.50 object 120190	object 150415 Computed distance: 798.0	Actual distance: 1.54162270220110620	-
object 100-0 Computed distance: 0.0 Actual distance: 1.0 Actual distance: 0.0 object 23 Computed distance: 7.10 Actual distance: 1.064/2495272236632 object 13/54 Computed distance: 7.20 Actual distance: 1.3654/216720065833 object 15/54 Computed distance: 7.50 Actual distance: 1.30021416518424 object 12/54 Computed distance: 7.50 Actual distance: 1.30021416518424 object 12/61 Computed distance: 7.50 Actual distance: 1.7696513309696336 object 12/61 Computed distance: 7.50 Actual distance: 1.7696513096968366 object 12/61 Computed distance: 7.50 Actual distance: 1.7696513096968366 object 12/61/2010 Computed distance: 7.60 Actual distance: 1.7696513096968366 object 12/61/2010 Computed distance: 7.60 Actual distance: 1.7696513096968366	object 70264 Computed distance: 730.0	Actual distance: 1.34163276236116635	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	kNNretrieve search Result	100010000000000000000000000000000000000	
object 30 Computed distance: 718.0 Actual distance: 0.1 object 31524 Computed distance: 732.0 Actual distance: 1.18642495272238632 object 127448 Computed distance: 732.0 Actual distance: 1.386241657800655833 object 14749 Computed distance: 732.0 Actual distance: 1.38024146518424 object 14719 Computed distance: 732.0 Actual distance: 1.38024146518424 object 14719 Computed distance: 732.0 Actual distance: 1.380254145518424 object 14719 Computed distance: 732.0 Actual distance: 1.380254145518424 object 14719 Computed distance: 732.0 Actual distance: 1.380325615565 object 260180 Computed distance: 73.0 Actual distance: 1.76085113898680366 object 747014027040027040270 T T	0		
object 11524 Computed distance 7180 Actual distance 11964/249527236832 object 12144 Computed distance 7320 Actual distance 16624(1972005693) object 12147 Computed distance 7320 Actual distance 16624(1972005693) object 12170 Computed distance 7320 Actual distance 130201416518424 object 12170 Computed distance 7520 Actual distance 13020141518425 object 2010 Computed distance 7520 Actual distance 14763003261352695 object 2010 Computed distance 7760 Actual distance 17608611380968386 object 14171 Computed distance 7760 Actual distance 17608611380968386	object 30 Computed distance: 0.0	Actual distance: 0.01	
object 127448 Computed distance: 732.0 Actual distance: 1.6542167208656833 object 1879 Computed distance: 738.0 Actual distance: 1.300214165189424 object 14719 Computed distance: 752.0 Actual distance: 1.4520802561325665 object 26010 Computed distance: 750.0 Actual distance: 1.7695611380968386 object 26010 Computed distance: 770.0 Actual distance: 1.7695611380968386 object 74017 Computed distance: 770.0 Actual distance: 1.7695611380968386	object: 31524 Computed distance: 718.0	Actual distance: 1.19642495272236632	
object 19179 Computed distance 73.00 Actual distance 1.3802914165168424 object 143719 Computed distance 75.20 Actual distance 1.42603032681352695 object 280180 Computed distance 77.80 Actual distance 1.42693032681352695 object 12017 Computed distance 77.80 Actual distance 1.76865113806968386 object 12017 Actual distance 1.76302101027362707 Telephone	object: 127448 Computed distance: 732.0	Actual distance: 1.65542167280655833	
object:143719 Computed distance:752.0 Actual distance: 1.45208302561352585 object:260180 Computed distance:776.0 Actual distance: 1.76965113869666366 object:250882 Computed distance: 1.602300.0273862027	object: 81679 Computed distance: 738.0	Actual distance: 1.3802914165188424	
object: 260180 Computed distance: 776.0 Actual distance: 1.76965113869686366 object: 250883 Computed distance: 788.0 Actual distance: 1.8073040827896287	object: 143719 Computed distance: 752.0	Actual distance: 1.45208302561352585	
object: 250883. Computed distance: 788.0 Actual distance: 1.6073040837895307	object: 260180 Computed distance: 776.0	Actual distance: 1.76965113869686366	
	_ohiert 250883_Computed distance: 788.0	Actual distance: 1 6073040937996397	4

shown in figure 2.

Fig. 2 - Result of kNN query

Range query	Type a Word:	combining	
dexing			
1000 words indexed			
IUUU Words indexed			
1000 Words Indexed			
1000 words indexed			
1000 words indexed			
000 words indexed			
1000 words indexed			
1000 words indexed			
10000 words indexed			
0000 words indexed			
0000 words indexed			
UUUU words indexed			
20000 words indexed			
iOR44 words indexed			
MES: %.2freal			
034201070			
D combining			
, contraining			
D coining			
0 compiling			
0			
u curiining			
0 recombining			
0 combing			

Fig. 3 - Result of range query



Fig. 4 – Data and corresponding pivot table

D-Cache KNN	Dcache Range Enter a Word : combined
Creating 1000 reference objects	Indexing
Reference object file already exist! Manually remove it if you want to crea	15000 words indexed
Inserting 100000 objects in the database. It might take a while	30000 words indexed
java.lang.Exception: Cannot sart bulk loadanother bulk load already in	45000 words indexed
Performing similarity searches	60000 words indexed
	75000 words indexed
kNN search Result	90000 words indexed
0 object: 30 Computed distance: 0.0 Actual distance: 0.0	105000 words indexed
1 object: 31524 Computed distance: 359.0 Actual distance: 1.1964	120000 words indexed
2 object 127448 Computed distance: 366.0 Actual	135000 words indexed
3 object: 81679 Computed distance: 369.0 Actual distance: 1.380.	150000 words indexed
4 object 143719 Computed distance: 376.0 Actual	150844 words indexed
5 object: 260180 Computed distance: 388.0 Actual	TIMES: 26.718122398
6 object: 250883 Computed distance: 394.0 Actual	
7 object 217592 Computed distance: 396.0 Actual	
8 object: 150415 Computed distance: 398.0 Actual	combined
9 object: 70364. Computed distance: 404.0 Actual distance: 1.7426	0.0 combined
kNNretrieve search Result	1.0 combiner
0 object: 30 Computed distance: 0.0 Actual distance: 0.0	
1 object 31524 Computed distance: 359.0 Actual distance: 1.1964	1.0 combines
2 object 127448 Computed distance: 366.0 Actual	
3 object: 81679 Computed distance: 369.0 Actual distance: 1.380.	1.0 combine
4 object 143719 Computed distance: 376.0 Actual	
5 object: 260180 Computed distance: 388.0 Actual	2.0 cabined
6 object: 250883 Computed distance: 394.0 Actual	
7 object: 217592 Computed distance: 396.0 Actual	2.0 coffined
8 object: 150415 Computed distance: 398.0 Actual	
9 object: 70364. Computed distance: 404.0 Actual distance: 1.7428	2.0 compiled
kNNretrieveAndSort search Result	2.0 uncombined
0 object: 30 Computed distance: 0.0 Actual distance: 0.0	
1 object: 31524 Computed distance: 1.1964249527223663 Actual	2.0 combings
2 object 140166 Computed distance: 1.267932174910	
3 object: 239389 Computed distance: 1.283057798503	2.0 combated
4 object 79276 Computed distance: 1.3119201211021558 Actual	125msec
5 object 207992 Computed distance: 1.318614755637	
6 object: 65111 Computed distance: 1.3308077550163306 Actual	

Fig. 5 –Result of kNN and range queries that use D-

cache and metric cache

D-Cach	ie MTree			
Enter a word:	combined	Show D-	Cache MTree	
Loading words 150844 words loaded Fest words: 'unplace"				
Creating M-Tree with m \dding words \5000 words added \5000 words added	iinNodeCapacity=2			
CREATE-MTREE 4-Tree created Festing minNodeCapa estMord="upplace"	minNodeC city=2, limit=1	≿apacity=2	realTime=14.	632623
rEST minNo resting minNodeCapa estWord="unplace"	deCapacity=2 city=2, limit=2	testWord="ur	iplace"	limit
rEST minNo resting minNodeCapa estWord="unplace"	deCapacity=2 city=2, limit=4	testWord="ur	iplace"	limit
rEST minNo resting minNodeCapa estWord="unplace"	deCapacity=2 city=2, limit=8	testWord="ur	iplace"	limit
TEST minNo	deCapacitv=2	testWord="ur	nplace"	limit-

Fig. 6 – D –Cache and metric cache with M-tree MAM $% \left({{{\rm{M}}_{\rm{M}}}} \right)$

EXPERIMENTAL RESUTLS

We made experiments without prototype application. The application demonstrates kNN and range queries with enhanced MAMs with D-cache. It also supports the same queries with the proposed combination of metric cache and D-cache. The experimental results are presented in the following graphs.



Fig. 2 – Construction of M-tree using single-way leaf selection

As shown in figure 2, horizontal axis represents D-Cache size while the vertical axis represents construction real time. The results reveal that the proposed shows higher performance than the existing approach.



Fig. 6 - Construction of D-M-tree using single-way multi-way leaf selection.

As shown in figure 3, horizontal axis represents D-Cache size while the vertical axis represents construction real time. The results reveal that the proposed shows higher performance than the existing approach.



Fig. 4 - NN queries on growing database

As shown in the above figure horizontal axis represents database size while vertical axis represents query real time.With respect to NN queries against databases that grow over time, the proposed approach outperforms existing approaches.



Fig. 5 - Negative results on the CoPhIR database

As shown in the above figure horizontal axis represents number of dynamic pivots while vertical axis represents query real time. The results of proposed system are negative in this case due to the nature of database.



Fig. 6 - Impact of the growing number of dynamic pivots

As shown in the above figure horizontal axis represents K while vertical axis represents query real time.The results reveal that the proposed system outperforms existing approaches.

CONCLUSION AND FUTURE WORK

In this paper we studied the similarity search and its performance in the presence of D-Cache with enhanced MAMs such Pivot table. M-tree and D-file proposed by Skopal, Lokoc, and Hustos[10]. We have envisaged that the performance of similarity search can be improved further by using a metric cache in front of D-cache. By using the reusable content in both metric cache and D-cache, the proposed system is capable of improving the processing speed of similarity search. This is achieved by using metric cache that works in tandem with D-cache for higher performance. Our prototype application demonstrates the usability and performance of the hybrid approach we proposed in this paper. Empirical results are encouraging and it can be used in real world applications where similarity searches are made for metric data. One important future direction we have is to optimize the combined approach further by improving underlying MAMs. Cloud based solution is another future direction we have.

REFERENCES

[1] C. Bo"hm, S. Berchtold, and D. Keim, "Searching in High-Dimensional Spaces—Index Structures for Improving the Performanceof Multimedia Databases," ACM Computing Surveys,

vol. 33, no. 3, pp. 322-373, 2001.

[2] J.S. Vitter, "External Memory Algorithms and Data Structures:Dealing with Massive Data,"ACMComputing Surveys, vol. 33, no. 2, pp. 209-271, citeseer.ist.psu.edu/vitter01external.html, 2001.

[3] S.D. Carson, "A System for Adaptive Disk Rearrangement,"Software—Practice and Experience, vol. 20, no. 3, pp. 225-242, 1990.

[4] M. Batko, D. Novak, F. Falchi, and P. Zezula, "ScalabilityComparison of Peer-to-Peer Similarity Search Structures," FutureGeneration Computer Systems, vol. 24, no. 8, pp. 834-848, 2008.

[5] E. Cha'vez, G. Navarro, R. Baeza-Yates, and J.L. Marroqui'n, "Searching in Metric Spaces," ACM Computing Surveys, vol. 33,no. 3, pp. 273-321, 2001.
[6] P. Zezula, G. Amato, V. Dohnal, and M. Batko, Similarity Search:The Metric Space Approach (Advances in Database Systems). Springer, 2005.

[7] W. Effelsberg and T. Haerder, "Principles of Database BufferManagement," ACM Trans. Database Systems, vol. 9, no. 4, pp. 560-595, 1984.

[8] H. Samet, Foundations of Multidimensional and Metric DataStructures. Morgan Kaufmann, 2006.

[9] G.R. Hjaltason and H. Samet, "Index-Driven Similarity Search inMetric Spaces," ACM Trans. Database Systems, vol. 28, no. 4, pp. 517-580, 2003.

[10] TomasSkopal, JakubLokoc and Benjamin Bustos, "D-Cache: Universal Distance Cache

for Metric Access Methods". IEEE Transactions On

Knowledge And Data Engineering, Vol. 24, NO. 5,

MAY 2012.

[11] B. Nam, H. Andrade, and A. Sussman, "Multiple Range QueryOptimization with Distributed Cache Indexing," Proc. ACM/IEEEConf. High Performance Networking and Computing (SC '06), p. 100,2006.

[12] J.M. Shim, S.I. Song, Y.S. Min, and J.S. Yoo, "An Efficient CacheConscious Multi-Dimensional Index Structure," Proc. Int'l Conf.Computational Science and Its Applications (ICCSA '04), vol. 4,pp. 869-876, 2004.

[13] K. Kailing, H.-P.Kriegel, and M. Pfeifle, and S. Schnauer, "Extending Metric Index Structures for Efficient Range QueryProcessing," Knowledge and Information Systems, vol. 10, no. 2,pp. 211-227, 2006.

[14] J.V. den Bercken and B. Seeger, "An Evaluation of Generic BulkLoading Techniques," Proc. 27th Int'l Conf. Very Large Data Bases(VLDB '01), pp. 461-470, 2001.

[15] B. Braunmu["] Iler, M. Ester, H.-P.Kriegel, and J. Sander, "MultipleSimilarity Queries: A Basic DBMS Operation for Mining in MetricDatabases," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 1,pp. 79-95, Jan./Feb. 2001.

[16] R. Paredes, E. Cha'vez, K. Figueroa, and G. Navarro, "PracticalConstruction of k-Nearest Neighbor Graphs in Metric Spaces,"Proc. Fifth Int'l Workshop Experimental Algorithms (WEA '06),pp. 85-97, 2006.

[17] F. Falchi, C. Lucchese, S. Orlando, R. Perego, and F. Rabitti, "Caching Content-Based Queries for Robust and Efficient ImageRetrieval," Proc. 12th Int'l Conf. Extending Database Technology(EDBT '09), pp. 780-790, 2009.

[18] F. Falchi, C. Lucchese, S. Orlando, R. Perego, and F. Rabitti, "AMetric Cache for Similarity Search," Proc. ACM Workshop Large-Scale Distributed Systems for Information Retrieval (LSDS-IR '08),pp. 43-50, 2008.

ARR I