

A novel information retrieval method based on R-tree index for smart hospital information system

Xinlu Wang^{1*}, Weiming Meng¹ and Mingchuan Zhang²

M.Tech Student, Department of Computer Science and Technology, Henan University of Science and Technology, Luoyang, China¹

Associate Professor, Department of Computer Science and Technology, Henan University of Science and Technology, Luoyang, China²

Received: 30-January-2019; Revised: 25-April-2019; Accepted: 29-April-2019

©2019 Xinlu Wang et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

High speed data retrieval is a main problem which affects the efficiency of hospital management system with internet of things (IoT). For this reason, we propose a new retrieval method, named dynamical clustering center (DCC) method, which dynamically determines the optimized clustering center when constructing the R-tree. By choosing an optimized clustering center, the method allows the spatial data in the same subspace to be organized into the same sub-tree, and builds an efficient R-tree index layer by layer from the root to the leaves. The experiments show that the proposed method can improve system stability and retrieval efficiency for the smart hospital information system.

Keywords

Hospital system, IoT, Retrieval, R-tree.

1.Introduction

With continuous development of the internet of things (IoT) technology in various fields, it plays an increasingly important role in medical technology [1]. Smart medical care is a new medical model characterized by medical information and "medical maintenance integration"[2]. "Internet +" medical has made the diagnosis and treatment services more convenient, more efficient and smarter [3]. The core is a smart hospital information system for wise information technology, whose kernel is internet of things. IoT include sensor, body area network [4], radio frequency identification (RFID) [5], wireless communication, data processing, network, video detection and recognition and global positioning system (GPS). It's the fact that these comprehensive technologies are applied for exchanging information and communication in the whole medical administrative system, in order to implement a kind of network technology of intelligent identification, location, trace, surveillance, management. Therefore, the hospital information system established through the IoT is more real-time, accurate and efficient than the system without IoT.

The smart hospital information system established by IoT makes medical data has the characteristics of a large number, diverse structure and timeliness. However, traditional data processing (such as B-tree and black-white tree) cannot be able to meet expectations and requires constant optimization [6]. There are many ways to index the database, such as the fixed grid [7], quad-tree [8], R-tree (and its variants) indexes [9]. Because of dynamics, aggregation and high efficiency, R-tree becomes one of the most widely used and mature indexing technologies in today. Based on the IoT intelligent hospital information system, a retrieval method based on R-tree index is proposed. This method uses the dynamical clustering center (DCC) algorithm to construct R-Tree model of the smart hospital information system, and improves the data query efficiency through R-tree index.

The main contributions of this paper can be summarized as follows:

1. SHIS-based index model. In order to improve the data retrieval rate of an intelligent hospital information system based on IoT, a new index model based on the R-tree is proposed.
2. Construction method of R-tree. We propose an R-tree clustering model based on DCC algorithm,

*Author for correspondence

which greatly reduces construct time by dynamically determining the optimized clustering center.

3. We implement the method in reality. The result shown that the method increases the retrieval efficiency of the smart hospital information system. In addition, we performed a detailed analysis of the performance of the system which this method was added to, and learned about the changes in the values of the different parameters.

The remainder of this article is organized as follows. Section 2 introduces the development of smart healthcare and R-tree spatial index in recent years. In Section 3, the smart hospital information system architecture is introduced. In addition, the detailed schemes, for clustering and retrieval of distributed data are proposed. The experiments and performance assessments are presented in section 4 to demonstrate the effectiveness of our proposed approach. Finally, we summarize the paper and point out future research topics in section 5.

2.Related work

2.1Smart healthcare

The concept of IoT comes from the RFID system proposed by the Auto-ID center [10], which aims to connect information sensing devices to the internet through RFID technology for intelligent monitoring and management [11]. The development of IoT technology points out a development direction for future medicine, named: "Smart health". Smart healthcare is a new medical model that uses internet technology to realize the interaction between patients and medical personnel, medical institutions and medical devices, promote innovation in medical information mode. At the same time, it is a dynamic service that enables real-time, intelligent and automatic connectivity from these aspects [12, 13]. Smart medical makes city as the main component. From the perspective of urban information construction, it fully considers the intensive use of resources and future expansion needs, builds a fully covered health and intelligent service system by using the existing information service platform in the region [14].

2.2Indexing technology

With the advent of "Internet +" medical era, traditional database technologies such as relational databases have been difficult to meet today's real needs [15]. In order to adapt to the complexity of data and processing, dynamics of the data structure and temporality of data, etc. [16], a new generation of

database technology has emerged. For example, spatial database technology is used to store spatial coordinates of two-dimensional, three-dimensional or higher-dimensional space and spatial range-related spatial data [17]. A non-relational database is used for massive transaction data, massive interactive data and massive data processing [18]. Object-oriented database technology is combined with database technology and object-oriented technology [19]. These new types of databases are fundamentally different from traditional databases in terms of data models and data storage. Therefore, indexing techniques used in traditional databases are no longer suitable for new databases.

2.3R-Tree

R-tree is a highly balanced tree with the object-defining technology, which is a natural extension of B-tree in k-dimensional space. As one of the most popular dynamic spatial index structures [20–22], R-tree is widely used in prototype research and commercial spatial database systems. The R-tree was originally proposed by Guttman in 1984. Then people have proposed various improvements on different spatial operation requirements. After 20 years of development, the ever-increasing R-tree variants have gradually formed a leafy spatial index R-tree family. From the breadth and depth of its coverage, the multidimensional spatial index R-tree is like a one-dimensional linear index B-tree, which is ubiquitous.

In recent years, many scholars have conducted in-depth research and improvement in the R-tree spatial index.

R+-tree [23] and R*-tree [24] use one by one (OBO) to build R tree, improve query efficiency by improving the construction algorithm. However, the overall effect is not ideal. The OBO method has defects in massive spatial data. For example, the constructed tree structure is unreasonable and there are many overlaps between the minimum bounding rectangles (MBRs). Packed R-tree [25] needs to assume that the spatial data is relatively static in order to pre-process the spatial data.

Subsequently, Brakatsoulas et al. [26] proposed the CR-tree. They believe that the essence of the R-tree index construction is a clustering problem, which improves the splitting algorithm to the multipath splitting of the clustering technique and the overall efficiency of the R-tree. Run-tao and Zhong-xiao [27] combined the improved k-means algorithm in

constructing the R-tree spatial index. The constructed R-tree spatial index has a more compact structure and there is a problem of interference from outlier spatial data. Jing-bin [28] improved the k -medoids clustering algorithm by recursively constructing the R-tree spatial index. To reduce the overlap between MBRs, they pre-set the initial cluster number and specified the cluster center. The R-tree index node is more reasonable, the search efficiency is higher. However, the clustering result is affected by the initial cluster center. Huang et al. [29] proposed a recursive clustering method of constructing R-tree spatial indexes from top to bottom. This method adopts k -means algorithm when dividing spatial data and adopts small-tree-large-tree (STLT) method when constructing R-tree, which avoids the defects caused by split nodes in OBO construction mode.

Through the above researches, the key to improving the performance of R-tree spatial indexing is how it is constructed. It is a typical clustering problem that the similar data in space is put under the same sub-tree during the build process. Therefore, it is a common construction method that combined with clustering algorithm when the R-tree spatial index is built. Using clustering during construction can make the R-tree structure more compact, reducing multipath lookups and improving search efficiency. Among them, k -means [30] and k -medoids [31] are the most representative partitioning methods. However, the clustering results obtained by k -medoids and k -means clustering algorithm are affected by the k value, the actual clustering cannot be obtained.

Therefore, in order to solve the problem that the number of selected clusters does not match the actual spatial data distribution and is susceptible to outlier data, the DCC algorithm is proposed. Constructed R-tree in the smart hospital information system through DCC algorithm. Efficient retrieval of data based on R-tree index improves the working speed of the system.

3. Methodology

3.1 System architecture

With the arrival of the information age, the internet of thing which is as a super-large-scale network to promote "big data", has been widely applied in modern intelligent services such as ecological protection, smart home, food safety, energy conservation, emission reduction and logistics transportation, etc. [32]. With the growing infrastructure, we build a smart hospital information system with internet of things technologies such as

RFID. The system is divided into three layers. The first layer is an information perception layer. The second layer is network cognitive layer and the third layer is a smart application layer (as shown in *Figure 1*). Based on the three-layer physical architecture, events are perceived by the nodes and processed locally. The information is sent to the cloud system server through the gateway (base station).

The specific functions of each layer are described below.

- a. Information perception layer is the physical layer. It includes sensors (RFID, WSN, etc.), event parameters or identifiers that exist in the objects. Its task is to complete the acquisition, communication and coordination of data.
- b. Network cognitive layer provides connectivity between objects, network devices, wireless or wired connection and cloud systems, as well as sending and processing data acquired locally. It also includes a gateway component for receiving data sending from the perceptual layer. Starting from information perception, we can obtain massive heterogeneous perception information, and use the interconnection mechanism between networks to distribute and share the perceived information.
- c. Smart application layer is responsible for providing applications and services (for example, machine chassis) to human or non-human users. It can specify and locate various processes, programs, and applications on the internet of things, such as a smart medical record management system.

In *Figure 1*, we can see that the core problem is how to effectively index in the network cognitive layer (for fast data extraction). We use the widely used R-tree indexing method to achieve fast query access to data. The data storage layer of the server database data center is shown in *Figure 2*.

The data storage physically stores spatial data, built indexing structures for various spatial objects based on R-tree and achieve fast retrieval.

The database of the smart hospital information system was based on the relational database. Taking a spatial data set as an example, a set of data tables is used to store spatial geometric data and its corresponding attribute information, and using R-tree to index the important data and attribute information. Firstly, the spatial objects are classified into three types of simple entities: point, line and surface

according to the geometry type. For the same geometric type, the necessary secondary topics are divided according to the nature and needs of the actual spatial entity (as shown in *Figure 3*). Then design a geometric data table and related attribute data table for the hierarchically divided geometric object set. The geometry table consists of a unique

identifier of the geometry object and two fields of binary coordinates (string). Correspondingly, the unique identifier of the object and several attribute fields are designed in the attribute data table, and the connection between the geometric data table and the attribute data table is realized by the unique identifier.

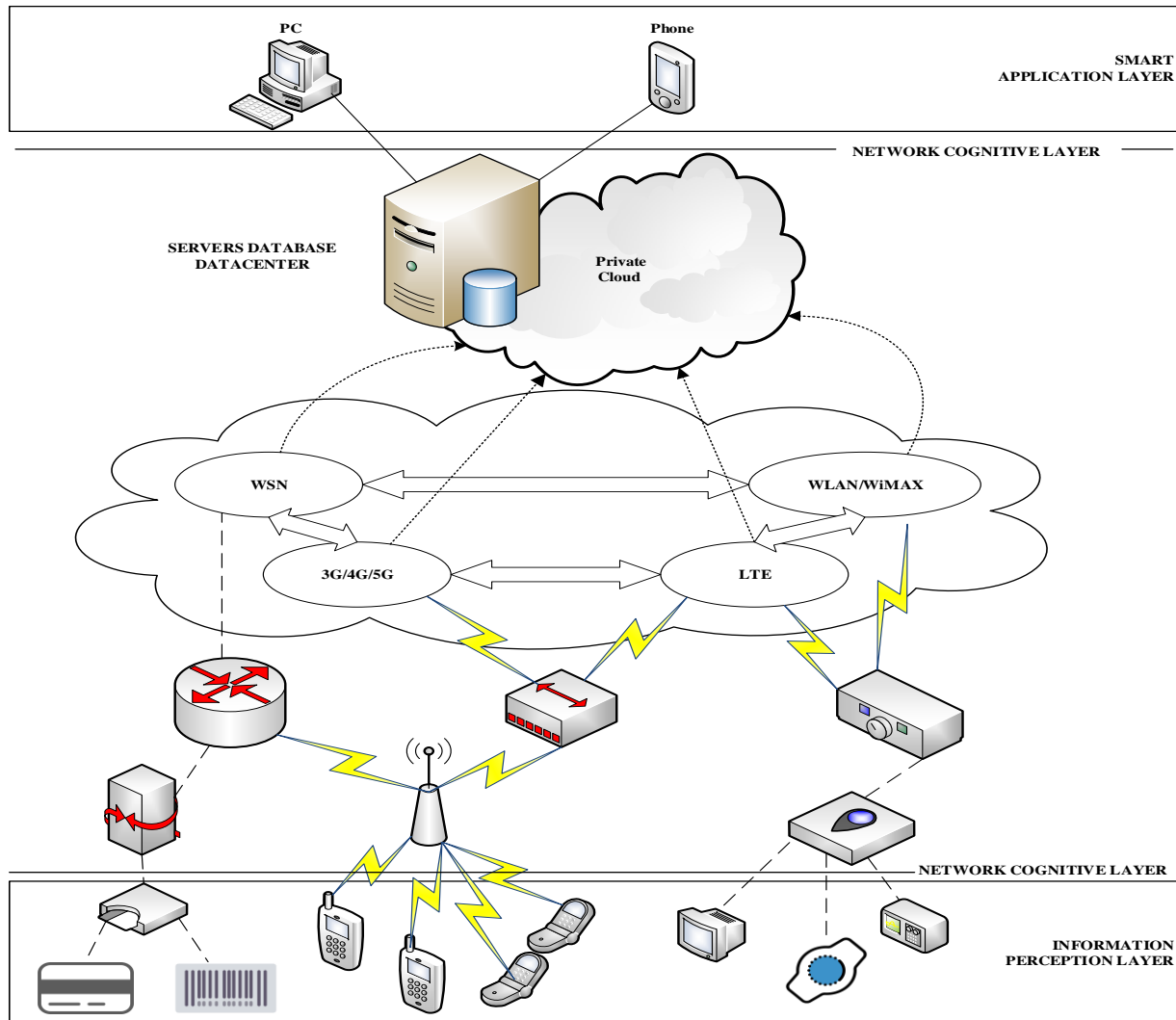


Figure 1 Architecture diagram of the system

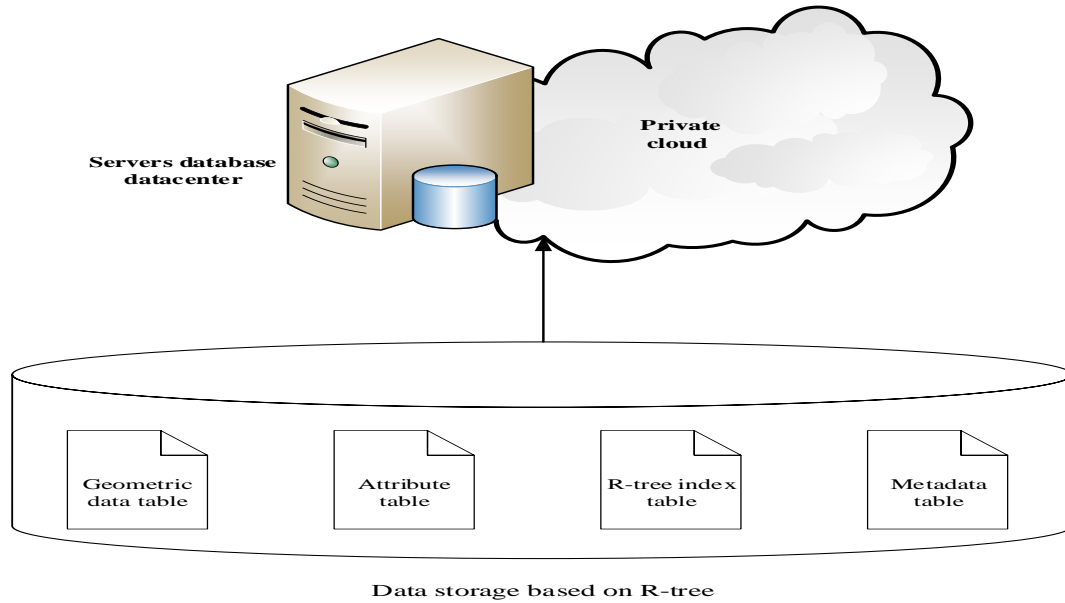
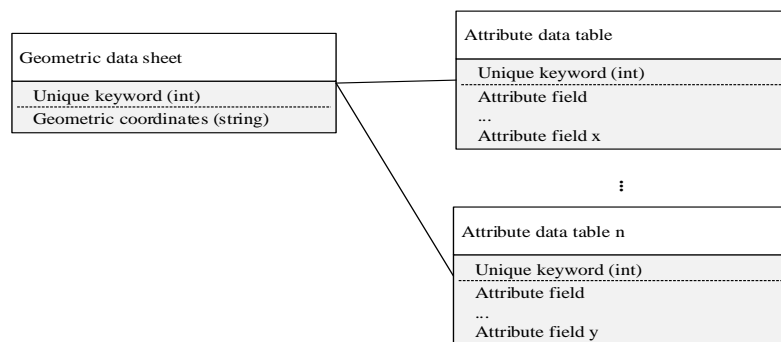
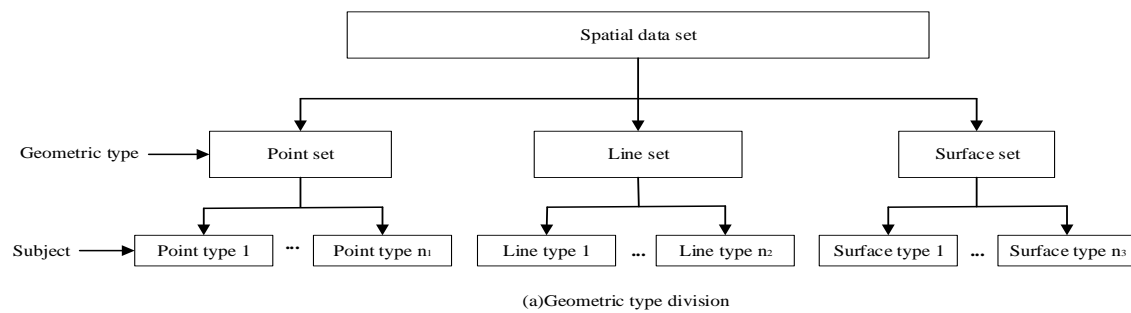


Figure 2 Data storage of the server database data center



(b)Spatial data object logical organization

Figure 3 Spatial dataset hierarchy and its relational table storage

3.2R-tree model

3.2.1DCC

In order to achieve efficient access to large-scale data, the system uses an indexing technique based on R-tree for data. When constructing the R-tree model, if the data distribution law is not known, pre-setting the cluster center will deviate from the actual results

of clustering and affect the efficiency of the constructed R-tree spatial index. In order to determine effectively the clustering center, this paper introduces the *DCC* algorithm to construct the R-tree. Define a distance indicator *R* [33] that measures, proximity object, as follows:

$$R = \frac{1}{\sqrt{m/D}} \quad (1)$$

Where the number of spatial data is m , D is the area of a given spatial extent, and d_i is the distance from data to i . If $d_i \leq R$, marked i is the proximity object of data, and if $d_i > R$, it is recorded as the non-adjacent object of data.

Give r_1, r_2, \dots, r_m is a collection of m spatial data of R^d , supposing C_l is the cluster center of l , then the distance between r_i and C_l (distance function) is:

$$d(r_i, c_l) = \sqrt{(r_i^1 - c_l^1)^2 + (r_i^2 - c_l^2)^2 + \dots + (r_i^d - c_l^d)^2} \quad (2)$$

Suppose the sample in l is $S_l = \{r_{l1}, r_{l2}, \dots, r_{ln}\}$. That is, contains n data, then the mean point of this class is $C_l = (C_l^1, \dots, C_l^k, \dots, C_l^d)$,

Where C_l^k is the k^{th} attribute of C_l , which can be expressed as:

$$C_l^k = (r_{l1}^k + r_{l2}^k + \dots + r_{ln}^k) / n \quad (3)$$

When selecting the cluster center, first obtain the mean point of the data onto the class C_l . Then calculating the distance from C_l to other data, obtain the proximity object of C_l according to the distance indicator R , and calculate the mean point C'_l of the proximity object, taking the spatial data closest to the distance C'_l as the cluster center, and denoted as:

$$r = \arg \min(d(C'_l, r)),$$

$$\text{Where } r = \{r_{l1}, r_{l2}, \dots, r_{ln}\}.$$

Algorithm 1: DCC

Input: m spatial data $S = \{r_1, r_2, \dots, r_m\}$ of R_d ;

Output: Cluster Center r_c .

```

1:  $C_l = (C_l^1, \dots, C_l^k, \dots, C_l^d)$ ;
2: for  $i = 1$  to  $m$  do
3:   Computing  $d(r_i, c_l)$ ;
4:   if  $d(r_i, c_l) \leq R$ ;
5:      $r_i \in M, M = \{r_b \mid d(r_b, C_l)\}$ ;
6:   end if
7: end for
8: Computing  $C'_l$ ;
9:  $r_c = \arg \min(d(r_c, C'_l)), r_i \in S$ ;
10: Output cluster center  $r_c$ .
```

3.2.2 Establish an R-tree index

Using the dynamical R-tree generation algorithm, a reasonable leaf node is inserted into the target object to be accessed by the above clustering center selection algorithm, to implement large-scale dynamic optimization of the R-tree.

The R-tree index structure corresponding to any set of spatial object sets is shown in Figure 4.

The main process of its establishment is: First, establish the minimum bounding rectangle for all spatial objects. Then according to the dynamic determination of cluster center (DCC) algorithm to group the external rectangular base layer. For example, selected R_{12} , which is closest to the mean point, as initial cluster center and $k=1$. Select the R_{19} with the farthest distance from the cluster center R_{12} and the farthest R_8 from R_{19} as the clustering center and start clustering. R_{13}, R_{14}, R_{17} and R_{18} are divided into R_{19} . $R_9, R_{10}, R_{11}, R_{12}, R_{15}, R_{16}$ are divided into R_8 , so forming two clusters. At this time, $k = 2$, the cluster center and the clustering measure function are calculated. The cluster with the largest radius and its cluster center R_{12} are selected from the two clusters. R_{15} (which is the farthest from R_{12}), R_{11} (which is the farthest from R_{15}) and R_{18} are used as the cluster centers to re-cluster, then evenly distribute and calculate the clustering measure function. In turn, the k value increases continuously until the clustering measure function converges. Finally, the top-level rectangle containing all the spatial objects in the entire region is formed, and the R-tree index structure shown in Figure 3 is obtained [34].

3.3A retrieval method based on R-tree

The R-tree index has a complete index establishment (insertion), query and node deletion algorithm. The following describes only the query algorithm, and performs spatial query based on R-tree index on the geometric table of the database. By establishing an R-tree index in the database, the efficiency of multi-user data retrieval can be greatly improved.

Description: Find all data rectangles intersecting W in the R-tree with root node N .

Algorithm 2: R_Search

Input: N as the R-tree type, W as the request rectangle.

Output: belongs to which rectangle.

```

1: if  $N.\text{level} = 0$  then
2:   return.
3: else
4:   for  $i=1, 2, \dots, n$  do
5:     if  $N.MBR_i$  intersect with  $W$ 
6:       R_search( $N.P_i, W$ ).
7:     end if
8:   end for
9: end if
10: Output the data rectangles of  $W$ .
```

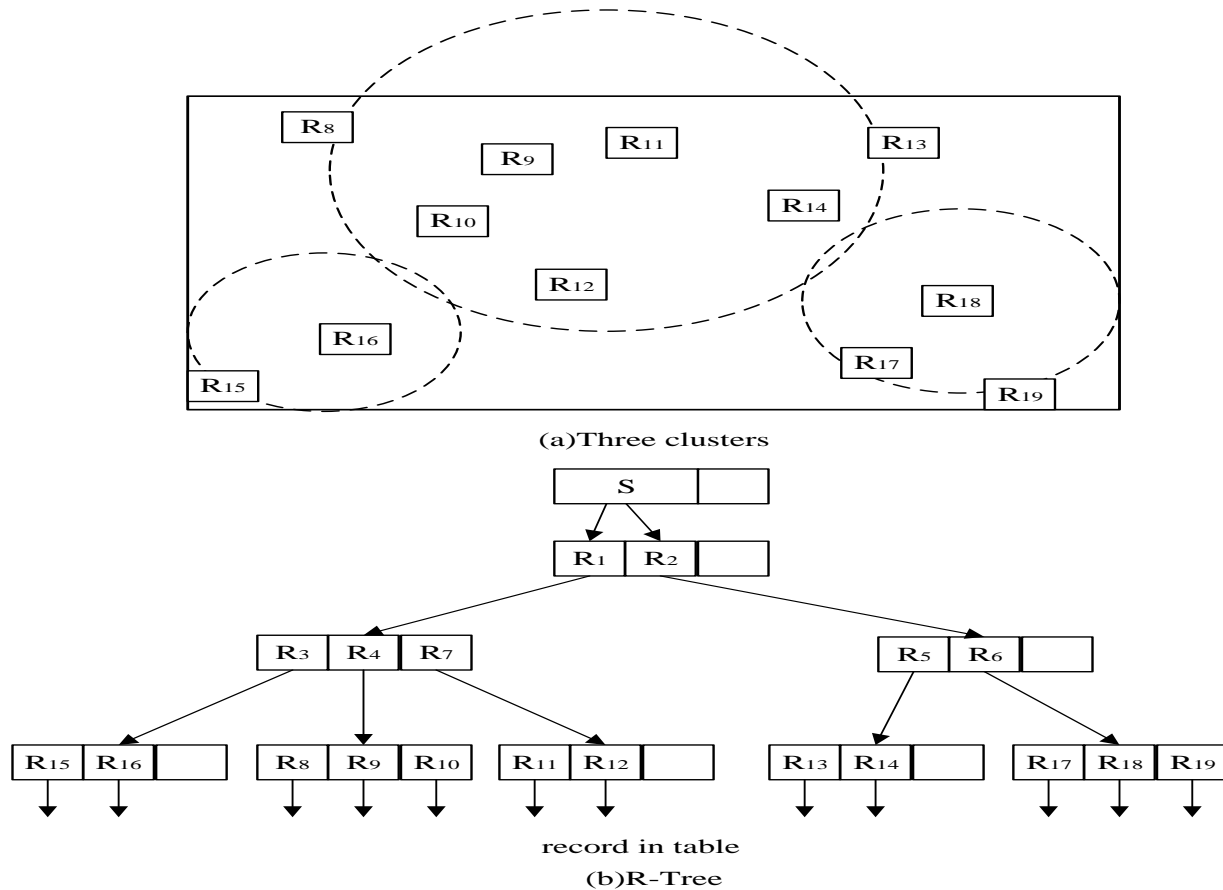


Figure 4 Hierarchy and its r-tree index

4. Experiment

4.1 System implementation

This research developed a smart hospital information system that provides an easy way for users to query information. The system is based on the Internet of things, so it allows users to access it in real time via a personal computer (PC) or smart phone. The era of medical informationization has gradually replaced traditional paper medical records with electronic medical records. However, for special medical records, paper medical records still have an irreplaceable status. When the paper room staff manages these paper medical records, they can quickly query their relevant information through the smart medical system to complete the management of the paper medical records. *Figure 5* shows a screenshot of the query about a medical record. The operating environment of this system is: Inter core i5-4460 3.20GHz processor, 8GB of memory, 31.41GB hard disk and the operating system is Windows 10 x64, programming language is C#.

4.2 Parameter setting and result

In order to verify the performance of the system after constructing the R-tree index, this paper conducts experiments through the medical record management system. The multi-dimensionally analysis the performance of the system integrated into R-tree, comparing with hash index [35–37] in various cases.

4.2.1 Parameter setting

The experimental related parameters are setting as follows. The IoT devices is the hospital's infrastructure for collecting data. Each group of devices has its own management, such as data collection through a medical bracelet. The size of R-tree is determined by the volume of data. Sets W as the network bandwidth, and T as the volume of data. Let's N as the number of queries and Q as complexities of queries. Sets α as the available bandwidth, and β as the volume of task. The specific parameter settings are shown in *Table 1*.

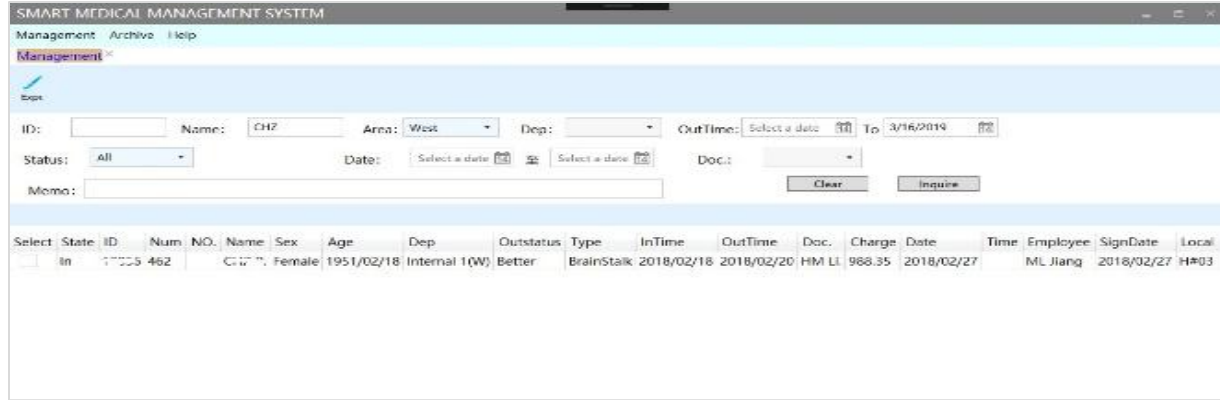


Figure 5 Medical management

Table 1 Parameter chart

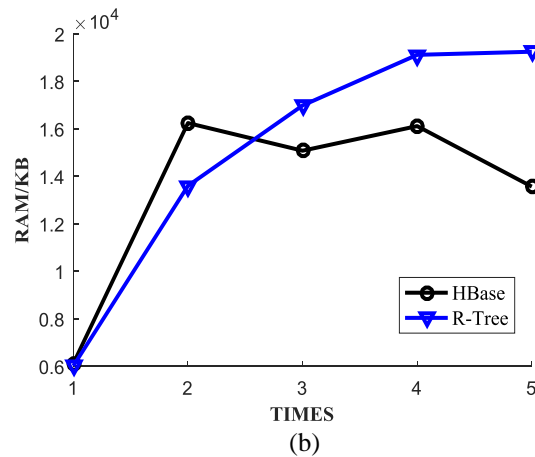
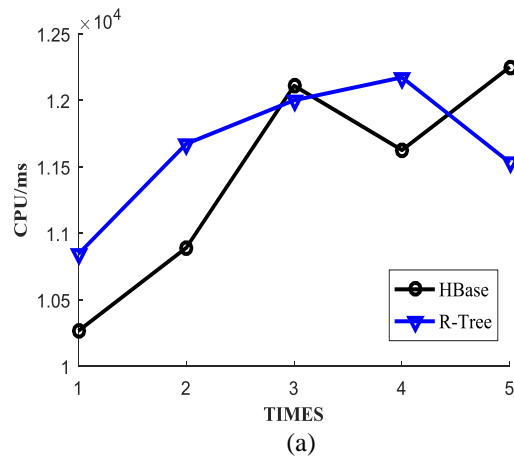
W	100MB
T_1, T_2	0.47MB, 4.48MB
N_1, N_2	5549, 12068
Q_1, Q_2	1, 4
$\alpha = 0$	100MB
$\alpha = 1$	50MB
$\beta = 0$	$T_1 + N_1 + Q_1$
$\beta = 1$	$T_2 + N_2 + Q_2$

4.2.2 Result

The experimental results are shown in Figure 6, Figure 7, Figure 8 and Figure 9. The ordinate is the number of experiment times. The effect of the network state and task volume on R-tree index can be seen from the experimental results. When the number of tasks is small, the system retrieval time for deploying the R-tree is slower than a hash index at the beginning. After a period of time, the system speed increased significantly (Figure 6 and Figure 8). This is because the R-tree is built at the

beginning. The time complexity of hash is $O(1)$ and the R-tree constructed using this method is an $O(n^k \times t)$, that k is the number of clusters, t is the number of iterations, and n is the number of spatial data. The R-tree constructed in this paper reduces the overlap and coverage area between MBRs, which makes the generated R-tree compact and with fewer multi-path lookups. So, it is more efficient than hash index.

When the number of tasks is large, the system performance by deploying the R-tree is better than hashing (Figure 7), especially when the network is busy (Figure 9). This is because for the R-tree, the larger the number of lookups, the more data it will contain and appears more search paths, which affects the lookup response time. The establishment and maintenance of the hash table impose a large burden on the computing performance of the computer.



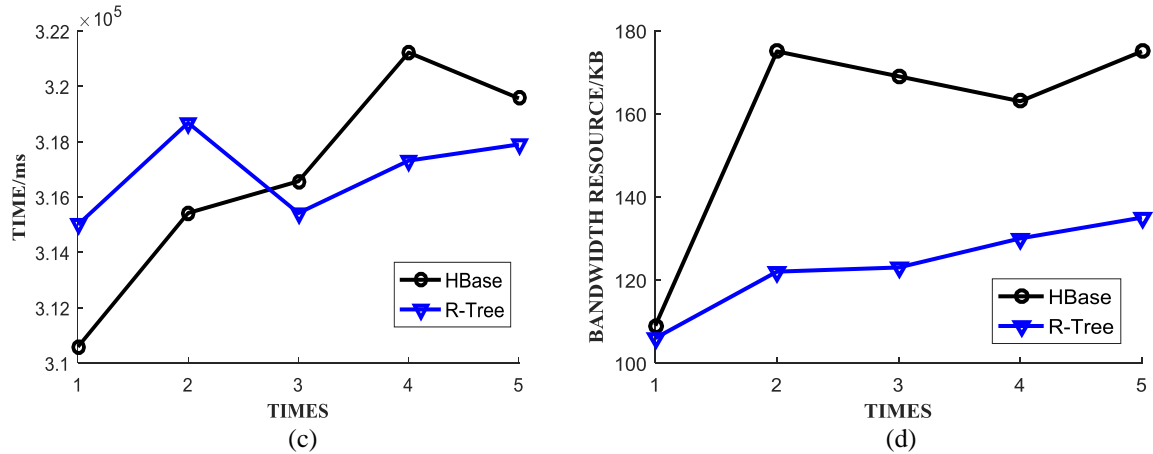


Figure 6 Compares the system's performance in $\alpha = 0, \beta = 0$ (a) CPU time, (b) RAM, (c) runtime, (d) bandwidth resources

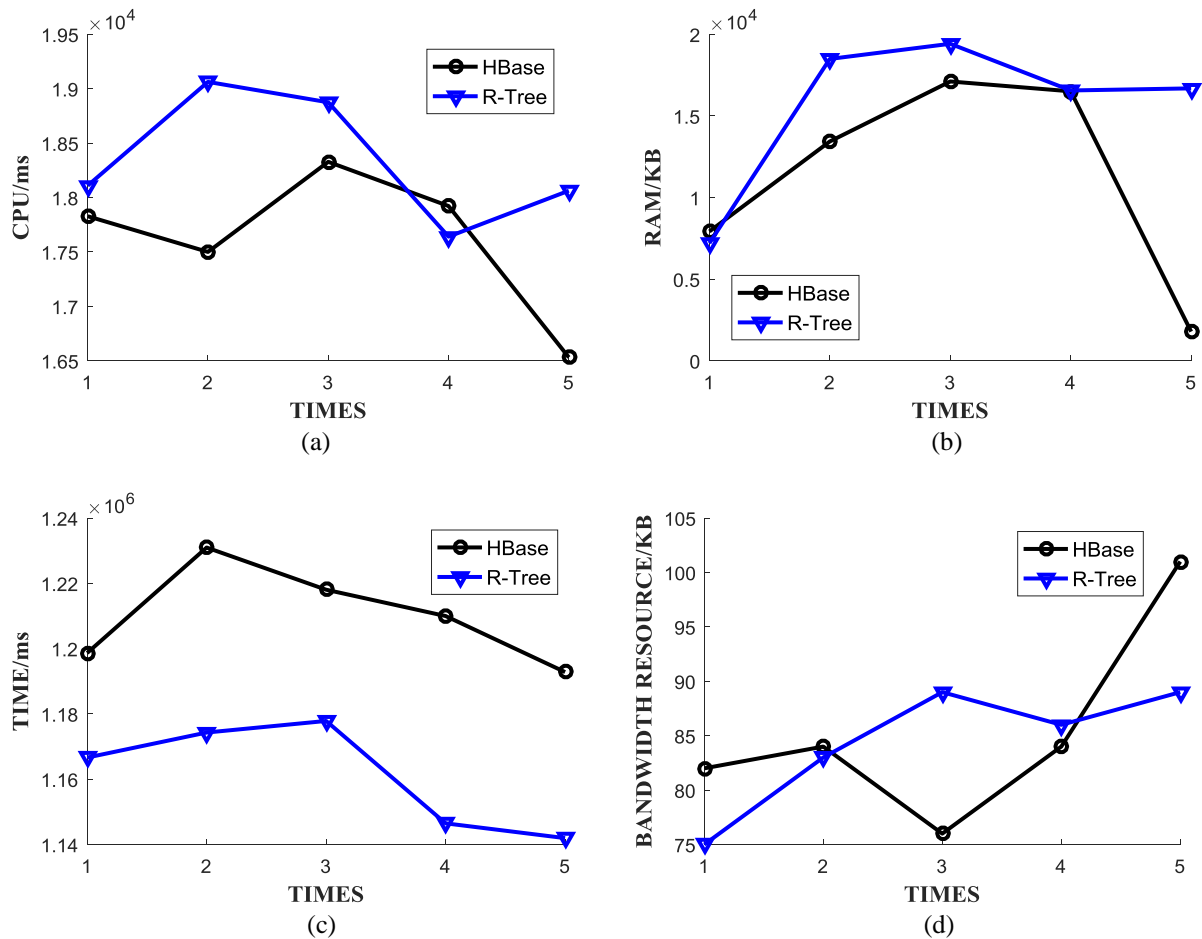


Figure 7 Compares the system's performance in $\alpha = 0, \beta = 1$ (a) CPU time, (b) RAM, (c) runtime, (d) bandwidth resources

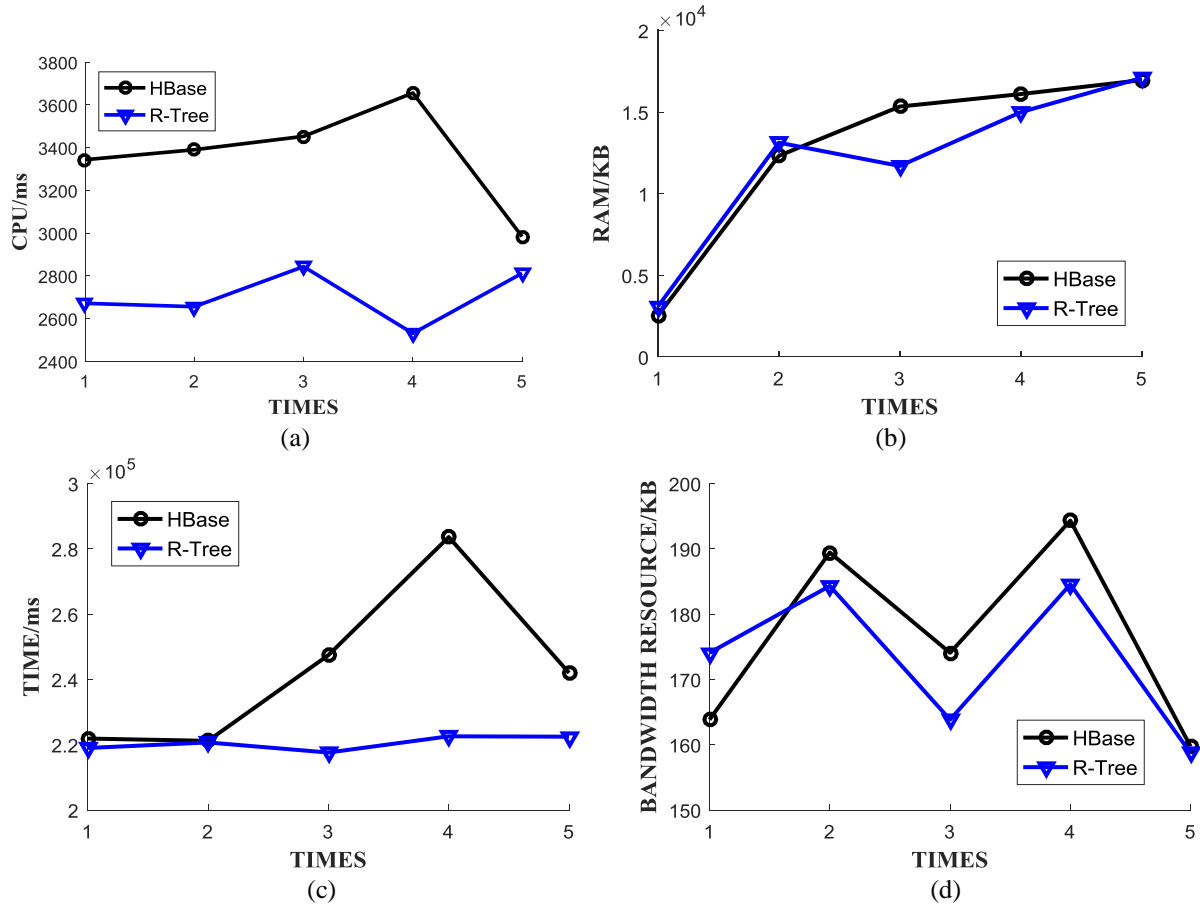
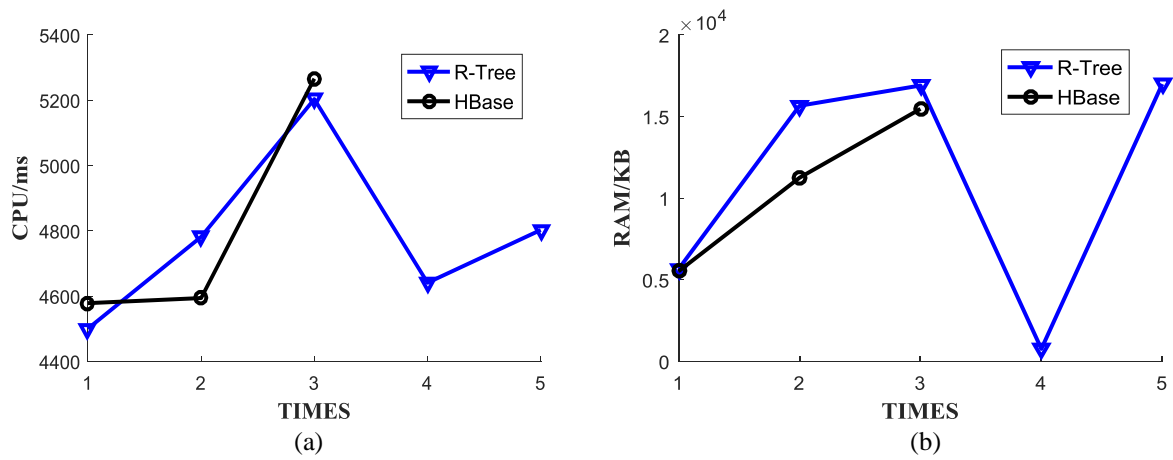


Figure 8 Compares the system's performance in $\alpha=1, \beta=0$ (a) CPU time, (b) RAM, (c) runtime, (d) bandwidth resources



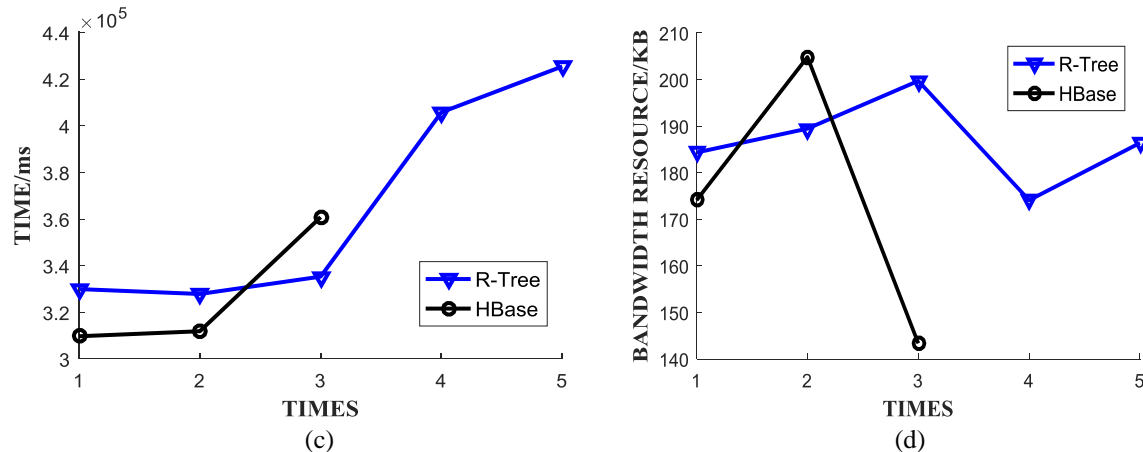


Figure 9 Compares the system's performance in $\alpha = 1, \beta = 1$ (a) CPU time, (b) RAM, (c) runtime, (d) bandwidth resources

5. Conclusion and future work

Internet of things technology has great application potential in hospital personnel, material management and medical technology development. But data acquisition is often a bottleneck. Therefore, the realization of data sharing and efficient query is an important basis for improving the efficiency and effectiveness of the medical information application. This paper proposes an index structure, which dynamically determines the clustering center to cluster medical data, and builds R-tree layer by layer from top to bottom. Through the smart hospital information system experiment, the system performance is evaluated under different network status and tasks, which proves that the smart hospital information system with R-tree index improves the efficiency of information search.

This study used only relatively small data sets because the experiments were based on initial system implementation. With the increasing of data set, you should consider comparing it with other indexing techniques.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants no. 61801171, and in part by the basic research projects at the University of Henan Province under Grants No. 19zx010.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Ge M, Bangui H, Buhnova B. Big data for internet of things: a survey. *Future Generation Computer Systems*. 2018; 87:601-14.
- [2] Abdelgawad A, Yelamarthi K. Internet of things (IoT) platform for structure health monitoring. *Wireless Communications and Mobile Computing*. 2017.
- [3] Biltoft J, Finneman L. Clinical and financial effects of smart pump—electronic medical record interoperability at a hospital in a regional health system. *The Bulletin of the American Society of Hospital Pharmacists*. 2018; 75(14):1064-8.
- [4] Akbulut FP, Akan A. A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics. *Measurement*. 2018; 128:237-46.
- [5] Bibi F, Guillaume C, Gontard N, Sorli B. A review: RFID technology having sensing aptitudes for food industry and their contribution to tracking and monitoring of food products. *Trends in Food Science & Technology*. 2017; 62:91-103.
- [6] Balkir AS, Foster I, Rzhetsky A. A distributed look-up architecture for text mining applications using mapreduce. In *proceedings of international conference for high performance computing, networking, storage and analysis 2011*. ACM.
- [7] Wang C, Zhu Y, Ma Y, Qiu M, Liu B, Hou J, et al. A query-oriented adaptive indexing technique for smart grid big data analytics. *Journal of Signal Processing Systems*. 2018; 90(8-9):1091-103.
- [8] Neubauer K, Haubelt C, Wanko P, Schaub T. Utilizing quad-trees for efficient design space exploration with partial assignment evaluation. In *Asia and South pacific design automation conference 2018* (pp. 434-9). IEEE.
- [9] Balasubramanian L, Sugumaran M. A state-of-art in R-tree variants for spatial indexing. *International Journal of Computer Applications*. 2012; 42(20):35-41.

- [10] Bayani M, Segura A, Alvarado M, Loaiza M. IoT-based library automation and monitoring system: developing an implementation framework of implementation. *E-Ciencias de la Información*. 2018; 8(1):83-100.
- [11] Dehury CK, Sahoo PK. Design and implementation of a novel service management framework for IoT devices in cloud. *Journal of Systems and Software*. 2016; 119:149-61.
- [12] Gelogo YE, Hwang HJ, Kim HK. Internet of things (IoT) framework for u-healthcare system. *International Journal of Smart Home*. 2015; 9(11):323-30.
- [13] Radanović I, Likić R. Opportunities for use of blockchain technology in medicine. *Applied Health Economics and Health Policy*. 2018; 16(5):583-90.
- [14] Kim JC, Chung K. Mining health-risk factors using PHR similarity in a hybrid P2P network. *Peer-to-Peer Networking and Applications*. 2018; 11(6):1278-87.
- [15] Banawan K, Ulukus S. Private information retrieval from coded databases. In *international conference on communications 2017* (pp. 1-6). IEEE.
- [16] Jin T, Wang J, La Rosa M, Ter Hofstede A, Wen L. Efficient querying of large process model repositories. *Computers in Industry*. 2013; 64(1):41-9.
- [17] Guangjun W, Shupeng W, Ming C. Massive structured data oriented storage and retrieve system. *Journal of Computer Research and Development*. 2012; 49(1):1-5.
- [18] Chen J, Chen Y, Du X, Li C, Lu J, Zhao S, et al. Big data challenge: a data management perspective. *Frontiers of Computer Science*. 2013; 7(2):157-64.
- [19] Dittrich J, Quiané-Ruiz JA, Richter S, Schuh S, Jindal A, Schad J. Only aggressive elephants are fast elephants. *Proceedings of the VLDB Endowment*. 2012; 5(11):1591-602.
- [20] Gaede V, Günther O. Multidimensional access methods. *ACM Computing Surveys*. 1998; 30(2):170-231.
- [21] Ahn HK, Mamoulis N, Wong HM. A survey on multidimensional access methods. *University of Science and Technology, Clearwater Bay, Hong Kong*. 2002:1-19.
- [22] Lu H, Ooi BC. Spatial indexing: past and future. *IEEE Data Eng. Bull.* 1993; 16(3):16-21.
- [23] Sellis T, Roussopoulos N, Faloutsos C. The R+-Tree: a dynamic index for multi-dimensional objects. *International conference on very large data bases* 1987.
- [24] Beckmann N, Kriegel HP, Schneider R, Seeger B. The R*-tree: an efficient and robust access method for points and rectangles. In *SIGMOD record* 1990 (pp. 322-31). ACM.
- [25] Roussopoulos N, Leifker D. Direct spatial search on pictorial databases using packed R-trees. *ACM SIGMOD record*. 1985; 14(4):17-31.
- [26] Brakatsoulas S, Pfoser D, Theodoridis Y. Revisiting R-tree construction principles. In *east European conference on advances in databases and information systems* 2002 (pp. 149-62). Springer, Berlin, Heidelberg.
- [27] Run-tao LI, Zhong-xiao HA. Spatial index structure based on R-tree and quadtree: RQOP_tree. *Journal of Harbin Institute of Technology*. 2010; 42(2):323-7.
- [28] Jing-bin W. Optimization algorithm for R-tree combining with spatial-clustering. *Computer Engineering and Application*. 2014; 50(5):112-5.
- [29] Huang Z, Qin Y, Zhang X, Zhao J, Jiang L. A static R-tree organization method based on top-down recursive clustering. In *international conference on geoinformatics 2013* (pp. 1-5). IEEE.
- [30] Saveetha V, Sophia S. Optimal tabu k-means clustering using massively parallel architecture. *Journal of Circuits, Systems and Computers*. 2018; 27(13).
- [31] Narayana GS, Vasumathi D. An attributes similarity-based K-medoids clustering technique in data mining. *Arabian Journal for Science and Engineering*. 2018; 43(8):3979-92.
- [32] Sethi P, Sarangi SR. Internet of things: architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*. 2017.
- [33] Chen X. Fast synchronization clustering algorithms based on spatial index structures. *Expert Systems with Applications*. 2018; 94:276-90.
- [34] Achakeev D, Seeger B. A class of R-tree histograms for spatial databases. In *proceedings of the international conference on advances in geographic information systems 2012* (pp. 450-3). ACM.
- [35] Gao X, Qiu J. Social media data analysis with IndexedHBase and iterative MapReduce. In *proceedings of the workshop on many-task computing on clouds, grids, and supercomputers 2013*.
- [36] Liu B, Zhu Y, Wang C, Chen Y, Huang T, Shi W, et al. A versatile event-driven data model in Hbase database for multi-source data of power grid. In *international conference on smart cloud 2016* (pp. 208-13). IEEE.
- [37] Chen X, Zhang C, Ge B, Xiao W. Spatio-temporal queries in HBase. In *international conference on big data 2015* (pp. 1929-37). IEEE.



Xinlu Wang was born in Luoyang in China on August 1993. She received her BSc in Information Engineering at Zhengzhou Chenggong University of Finance and Economics, Zhengzhou, China in 2016. Currently, she is M.Tech Student in Information Engineering, specialized in Computer Science and Technology at the School of Information Engineering, Henan University of Science and Technology, Luoyang, China. Her area of research and interests lies in the field of Big Data.
Email: 160304300184@stu.haust.edu.cn



Weiming Meng was born in Puyang in China on June 1993. He received his BSc in Information Engineering at Henan University of Science and Technology, China in 2016. Currently he is M.Tech Student in Information Engineering, specialized in Computer Science and Technology at the School of Information Engineering, Henan University of Science and Technology, Luoyang, China. His area of research and interests lies in the field of Cloud Computing.
Email: 170317050189@stu.haust.edu.cn



Mingchuan Zhang is an Associate Professor engaged in the field of Computer Science and Technology Research. He works in the School of Information Engineering, Henan University of Science and Technology, Luoyang, China. His area of research and interests lies in the field of Next-Generation Internet, Routing Mechanism and Protocol, Internet of Things.
Email: zhang_mch@haust.edu.cn