Retrieving Mobile Phone Information Based on Digital Search Tree

Safa A. Ahmed

Department of Chemical Engineering, University of Technology, Baghdad, Iraq

Abstract

   The current study primarily aims to develop a dictionary system for tracing mobile phone numbers for call centers of mobile communication companies. This system tries to save the numbers using a digital search tree in order to make the processes of searching and retrieving customers' information easier and faster. Several shrubs that represent digits of the total phone numbers will be built through following the phone number digits to be added to the dictionary, with the owner name being at the last node in the tree. Thus, by such searching process, every phone number can be tracked digit-by-digit according to a required path inside its tree, until reaching the leaf. Then, the value stored in the node, that represents the name of phone number's owner, is returned. Consequently, the amount of memory required to store data will be reduced and data retrieval will be faster.

Keywords: Mobile phone; Digital tree; Phone storing technique; Manager System

<div dir="rtl">

استرداد معلومات الهواتف المحمولة في مركز الاتصال بناءً على شجرة بحث رقمي

صفا امين احمد

قسم الهندسة الكيمياوية، الجامعة التكنولوجية، بغداد، العراق

الخلاصه

تهدف الدراسة الحالية في المقام الأول إلى تطوير قاموس لتتبع أرقام الهواتف النقالة لمراكز الاتصالات في شركات الاتصالات النقالة. يحاول هذا النظام حفظ الأرقام باستخدام شجرة بحث رقمية لجعل عملية البحث واسترجاع معلومات العملاء أسهل وأسرع. سيتم إنشاء العديد من الشجيرات التي تمثل أرقاماً من إجمالي أرقام الهواتف من خلال إتباع رقم الهاتف الذي سيتم إضافته إلى القاموس مع اسم المالك في العقدة الأخيرة من الشجرة. وبالتالي، من خلال عملية البحث، يمكن تتبع كل رقم هاتف رقماً تلو الآخر وفقاً للمسار المطلوب داخل شجرته حتى الوصول إلى الورقة. ثم تقوم بإرجاع القيمة المخزنة في العقدة التي تمثل اسم مالك رقم الهاتف. وبذلك سيتم تقليل حجم الذاكرة المطلوبة لتخزين البيانات مع سرعة في استرجاع البيانات.

</div>

1 Introduction

   Recently, there were improvements in all technical features for smartphones. Relatively very fast processors and storage units are used for mobile phones. However, they are still considered as slow when compared to the personal computers [1]. Consequently, for smart phones and mobile devices, the algorithms designed to run on desktop computers are not considered as good choice. Some methods are applicable to reduce the computational costs.

_____

*Email: safa.a.ahmed@uotechnology.edu.iq

They are typically employing feature hashing, performing nearest neighbor searches using Hamming distances. In general, they are making use of inverted files, such as hash tables. These tables require large memories to store the hash codes of the features. Moreover, in order to obtain good performance in retrieval, hash codes are relatively large and requiring fairly large memories to store large scale databases [2,3].

Many business managers depend on information technology to select their products; therefore, they use a Customer Relationship Manager System (CRMS), which is a system that can help companies to retrieve all information about their customers [4].

Operational CRMS is one area which is used by the staff in a call center that supports front office processes [5-7]. Also, in CRMS, the technologies and business processes are integrated and adopted to satisfy the requirements of customers during any given interaction [8].

Despite the spread of mobile phone companies and the development of software, especially in call centers, which facilitates the connection to customer information and their retrieval, there is no research available on this subject in the literature.

In 1997, Grover used quantum mechanics to speed up a range of search applications over unsorted data. Quantum mechanical systems can be employed in a superposition of states to instantaneously examine many duplicate names. Successful computations may impose the phases of various operations by an appropriate adjustment, while others randomly interfere. Hence, only the access to the database can be used to obtain the desired phone number [9].

Cortes and Pregibon (2001) built a suite of tools for imagining information that is custom-fitted to display calling motion in order to easily identify the user who has unfamiliar patterns. Then, those tools can be used to detect intercontinental calling fraud [10].

Pregibon and Cortes (2001) also developed signature-based methods to produce a mark from an information stream of call detail records to briefly portray the calling conduct of clients. Furthermore, they utilized peculiarity identification to gauge the bizarreness of another call compared with a specific record [10, 11].

Fashoto *et al*. (2013) used Knee-Point K-means clustering algorithm to discover fraudulent privileges in the National Health Insurance Scheme (NHIS) [12]. Likewise, Adeyemo and Oyeniyi (2015) developed a model in banking sectors using data mining techniques to analyze customer churn. In that model, they utilized simple Knee Means (KMeans) in the clustering stage and a rule-based algorithm for the rule generation stage [13].

This system was built through utilizing Visual Basic.net 2008, which is an object-oriented computer programming language that can be viewed as an evaluation of the Microsoft's Visual Basic (VB) that is generally implemented on the Microsoft.NET Framework. Visual Basic. NET with ADO.net is the best database programming approach since it enables the programmer to work in graphical user interface environment (GUI). The training process was implemented on an Intel core2 dual processor with a speed of 1.80 GHz. The capacity of RAM is 1016 GB. WINDOWS XP is the operating system used for implementation.

1.1 Types of Data in Telecommunication Industry

Fundamentally, data are divided into two types; customer detail records (CDRs) and Profile Data. The CDRs data include "A" number (originator), "B" number (terminator), call type, call duration (total and mean), and count of calls. Profile data include information about mobile number, subscriber tenure, subscriber plan, and region [14].

Each application or system should firstly detect the data in order to enhance and analyze the application. The present study, therefore, attempts to discuss the data used in telecommunication industry.

a.   Call Details Data

This type of data involves very detailed information of each call that is taking place. This includes average call duration, no-answer calls, calls from/ to a different area code, weekday

calls (Monday- Friday), daytime calls (9 am- 5 pm), average calls received per day, average calls originated per day, unique area codes called over given time period (P) [15].

The detailed information is recorded in the call details based on whether it is the source or receiver. Such recodes are used for different purposes, such as billing data [16].

b.   Network Data

Network data represents the data stored in the interconnected components of the telecommunication network in the form of error and status messages, which are to be used for network management. These messages contain, at least, a time-stamp, a string as a unique identifier for the generating hardware or software, and a code containing the reason of the messages [17].

c.   Customer Data

Customer data mainly deals with customer information, e.g. name and address. It also deals with other information like services plan, contact information, family income, and payment history.

Customer data is used for applications like fraud detection and market profiling [18].

1.2 Data Mining Methods that Utilize Data Analytics

Any decision of method, for the most part, relies on the kind of parameters, requirement for examination, and information thickness, to mention just a few. Also, there is a wide scope of strategies for prescient investigation which is utilized in the business. This study will therefore list some of those procedures.

A. Logistic Regression

It is a prescient investigation which is led when the needy variable is twofold. This strategy is utilized to clarify the information itself along with the relationship of the needy paired variable with at least one free factor [18].

B. Linear Regression

It is the most essential kind of relapse which is utilized for condensing and considering the connection between two nonstop factors (one is reliant and the other is free) [18].

C. Decision Tree

It is a graphical portrayal of each conceivable result under any conceivable condition [19, 20].

D. Clustering

It is a strategy which partitions a huge gathering of information into smaller sub-clusters containing comparable sort of information. The principal property of clustering is building the intra-group similitude and diminishing the in-between group comparability [19].

E. Neural Network

It is a methodology that is weakly related to the brain of human being and can be utilized to process data where different components are interconnected so as to understand a specific problem [21].

F. Support Vector Machine

It is a regulated learning technique which uses the idea of choice plane to separate an arrangement of items having distinctive class participation [18].

G. Others

Some of the other techniques that are used for the same purpose include the Discriminate Analysis, Scorecards, Survival Analysis, Market Basket Analysis, Collaborative Filtering, Discrete Choice Modeling, Casual Models, and so on [18].

1.3 Radix Search Tree

Radix Search Tree is a particular data structure dependent on the trie that is utilized to store a bunch of strings. In contrast to a normal trie, the edges of a Patricia trie are marked with successions of characters, as opposed to single characters. These can be series of characters and bit strings, for example, whole numbers or IP addresses, or they are utilized for inverted indexes of text as the information are retrieved. Once in a while, the names radix tree and crit

bit tree are simply applied to trees, putting away numbers, whereas Patricia trie is held for more broad sources of info; yet, the design works in a similar route in all cases [22].

The radix tree is most straightforward to comprehend as a space-advanced trie, where every hub with only a single youngster is converged with its kid. The outcome is that each inner hub has at any rate two kids. Dissimilar to the case in ordinary attempts, edges can be marked with arrangements of characters, but only as single characters. This makes them considerably more effective for small sets (particularly if the strings are long) and for sets of strings that offer long prefixes. It bolsters the principal tasks (lookup, insert, delete, find archetype, find replacement) which are all O(k), where k represents the lengthiest strings of the set. For an enhancement, it can use two pointers for a string (to the last and first elements) in order to put edge marks away in steady size [23].

1.4 Adaptive Radix Tree

In this section, the Adaptive Radix Tree (ART) is presented. It starts with certain expansive discerns on the potential gains of radix trees over relationship-based trees. This section also shows that the usage of ART could be over the top when using the flexible center point. The portraying, figuring, searching, and expansion of ART are also explained. Finally, the usage of space is analyzed [23].

The most interesting properties of ART can be illustrated as below [23]:

• The complexity and height of the radix tree is depending on keys' length.

• Rebalancing activities are not needed when radix trees are used. In addition, the comparable tree is achieved in all incorporation orders.

• A lexicographic order is adopted to store the keys

• The key of a leaf is represented by the path of that leaf. So, these keys are unconditionally stored and recalled from the related paths.

There are two sorts of center points in a radix tree; the first is the inner center points, where keys are fragmented to various centers. The second is the leaf center points, where the contrasting characteristics are stored with the keys [23].

The most useful representation of an inward hub may be as a variety of two pointers. An "s" bit lump of the key is used as the list into which exhibit during tree crossing. Consequently, the following youngster hub is decided without any further correlations [24].

The call range corresponding to the boundary "s" represents the fundamental to present a radix trees, as it determines the tallness of the tree for a specific key length. A radix tree deducts "k" bit keys and shows dk/se levels of internal hubs. As an example, a radix tree using s=1 has 32 levels with 32 piece keys. On the other hand, only 4 levels are obtained from a range of 8. It is much clearer to distinguish the stature of radix trees as well as the quantity of test in the inquiry trees those are totally adjusted as the pursuit trees based on the correlation are the common structures of ordering in information-based frameworks. More characteristics can be blocked if *s*>1 when radix tree is used, because in the best case each relationship blocks half of all characteristics [25].

From "unadulterated query" execution outlook and the adaptive nodes, it is attractive when a huge range is available. The obstruction of huge range is similarly clear at the point when assortments of pointers are periodically speaking to internal hubs. As most youngster pointers are invalid, the space use is excessive. Such balance explains the stature and space utilization for different estimations of the range boundary, as deducting 1M consistently appropriated 32-piece numbers [25].

2 Research Approach

2.1 Phone Number Database

The dictionary used in the current study is a virtual sample collected from different mobile networks in Jordan-based companies (Zain [079], Orange [077], and Umniah [078]). The following items highlight the properties of the sample:

1. It consists of 5294 phone numbers of students of the Princess Rahmeh College in Al-Balqa' Applied University, along with some related information, such as ID number, address, signature, and nationality.
2. The phone numbers length in this sample is 10 digits.
3. The prefix of each number should begin by one of the following choices: 079, 078, or 077.
4. These phone numbers are used for training and testing the digital search tree method.

2.2 Digital Tree Implementation

In the digital tree technique, three trees are built, each representing the prefix of one Telecom Company: (Zain [079], Orange [077], and Umniah [078]), as shown in Figs 1, 2, and 3, respectively. The trees are built by keeping track of the phone number. The search operation is executed by following the phone number digit using suitable track in the appropriate tree of the available trees. When the phone number is completed and the tracking reaches to a node which saves the owner name, the system shows the data in a textbox which represents the name field in a graphical user interface. Based on the radix tree, the Mobile Phone Dictionary includes two phases:
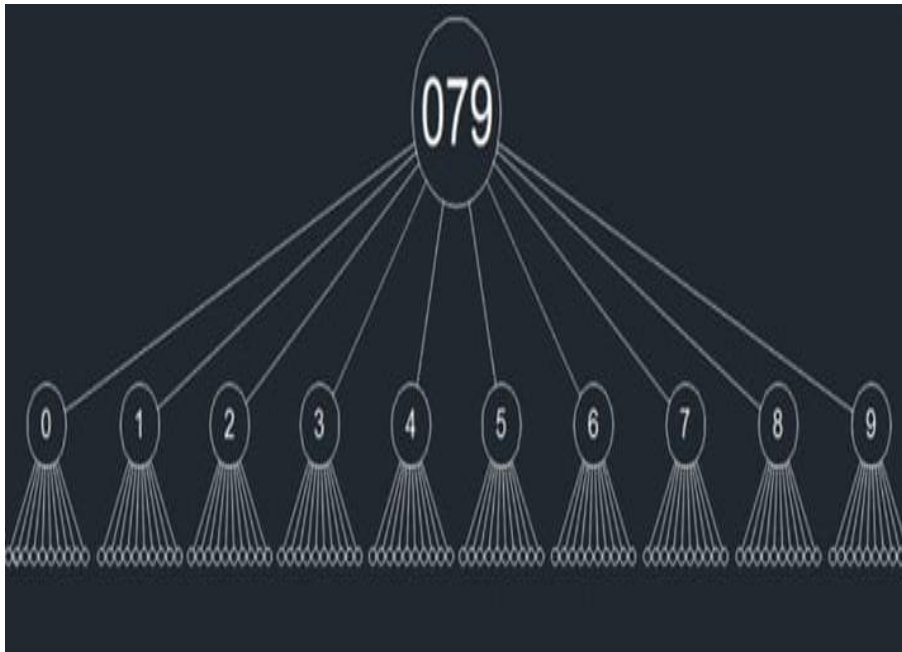
Figure 1-Digital tree used by Zain Mobile Communications Company contain all possible phone number of customers using Radix Tree
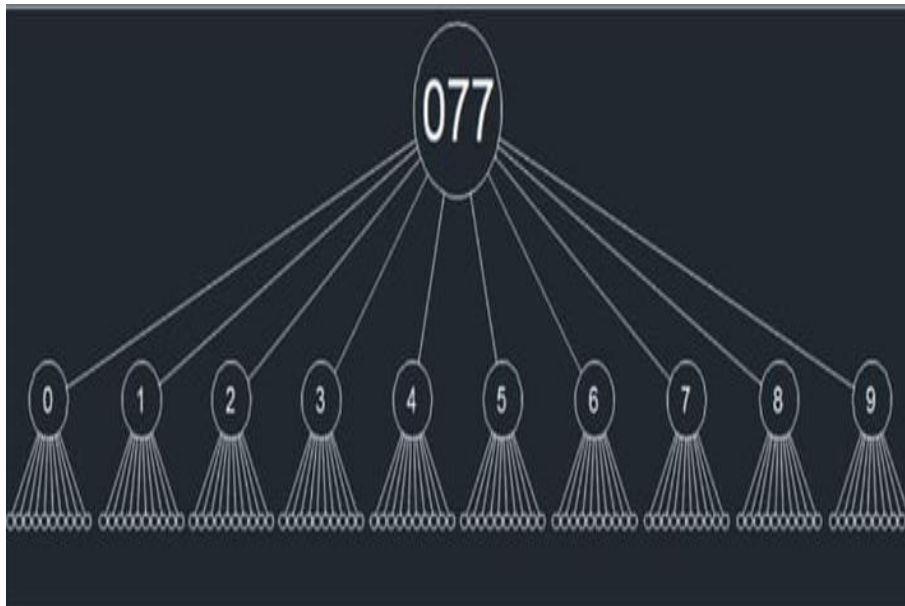
Figure 2-Digital tree used by Orange Mobile Communications Company contain all possible phone number of customers using Radix Tree
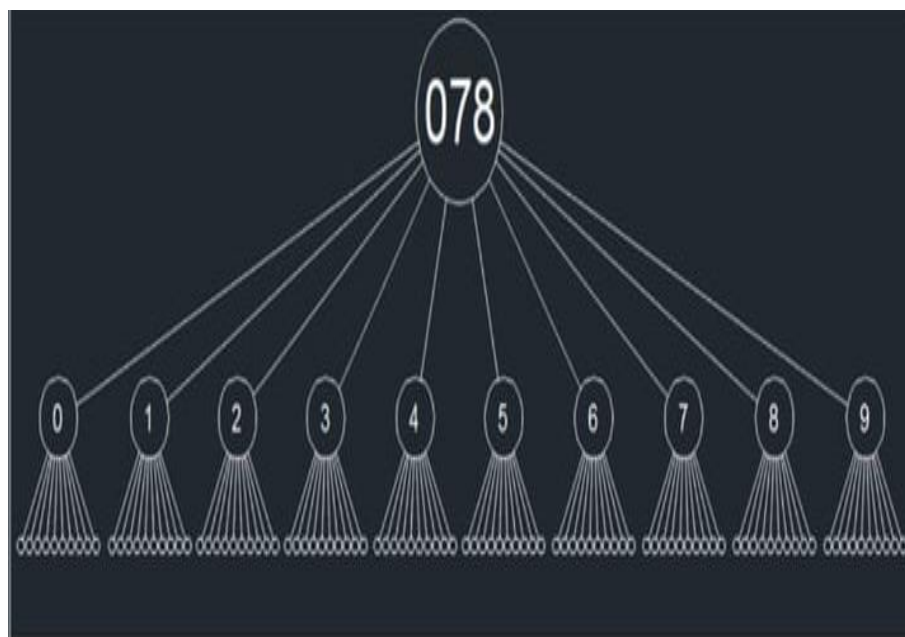

Figure 3-Digital tree used by Umniah Mobile Communications Company contain all possible phone number of customers using Radix Tree

2.2.1 The phase of building the digital trees

As was mentioned earlier, the final nodes (leaves) in the tree must be set up to be put away in the trees. The strategy is applied through the following associated phases:

1. When actuality is associated through the database, each phone number and its owner name are entered; [Owner name] and [Number]; as shown in Figure 4.

2. Each number is split into isolated digits.

3. Each digit is exclusively entered to be kept in the tree depending on its order/ position in the phone number. Thus, the initial 3 digits are put away as a root, while the fourth digit is put away as an offspring of a similar root at the first level in the tree. Digit five is put away as an offspring of the previous digit node, etc.
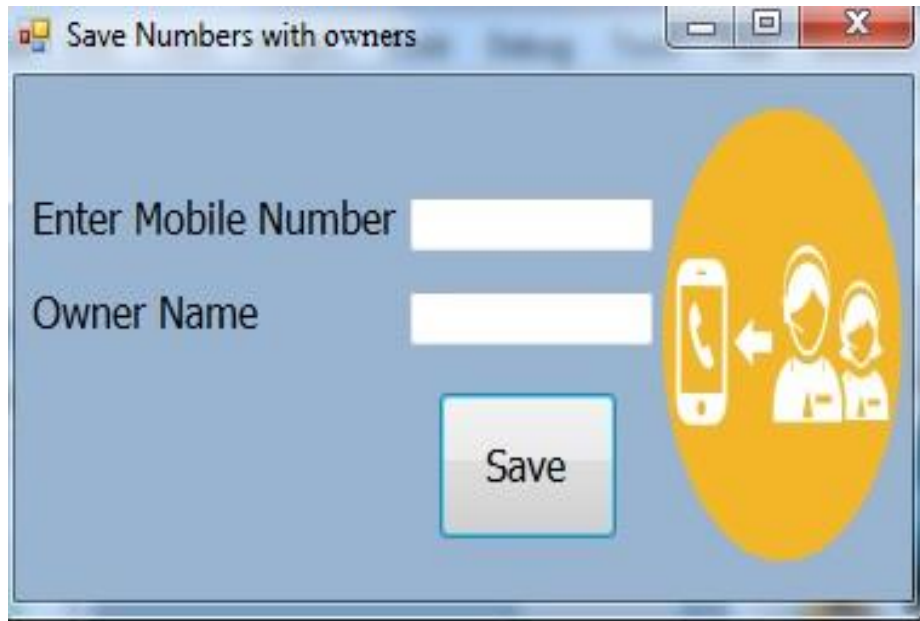
Figure 4- Saving phone number with its owner name

When reaching the leaf, the name of the owner is saved at level seven in the tree. Accordingly, the root may be created through the storing process by some numbers. These numbers are correlated with the same first 3 digits of the current number. Here, it will not be stored again as a new number. Instead, it is going to the fourth digit of the phone number to search for it in an offspring of that root. On the off chance that it is not found, it will be put away. Otherwise it will go to digit five and proceeds till arriving at the last digit in the phone number.

For example, let us trace the first number, e.g. [Person1][0795822376], which is split into two parts: the prefix part (079) and the remaining digits part (5822376). Part 1 is saved as a root of Zain tree network and part 2 is stored as a child of this root. Digit (5) is stored in level 1 and digit (8) in level 2, while digit (2) in level 3, and so on until reaching to the point that the leaf of the owner name is stored, as shown in Fig. (5).
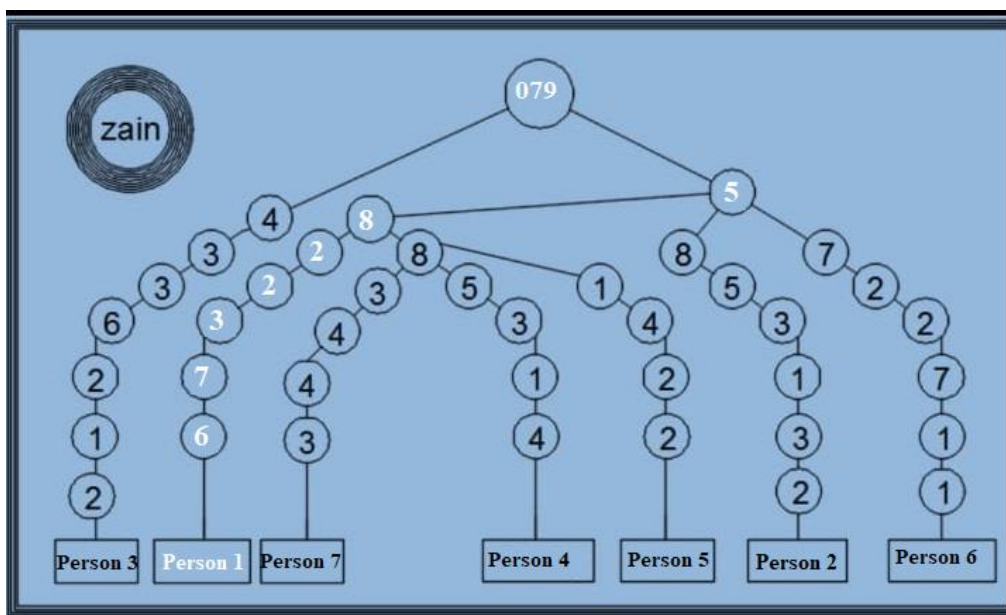
Figure 5-Example of saving phone number in Zain Mobile Communications Company Tree

In this number, i.e. [Person2][795853132], the prefix part (079) has already been saved in the previous case and, thus, there is no need to restore it in each process. The same is true for digit four in the current number, which has been already saved as a child of the previous root. However, the second digit is not saved and, hence, a new node is created to be a child of the parent node, which is (5), as shown in Figure 5.

2.2.2 The phase of testing the digital trees

After the implementation of the digital search tree for all the numbers in the database, the numbers are tested by entering the required numbers in a text box, as shown in Figure 6.

Figure 6-Mobile number search system

Each number is separately processed and the search process is implemented when the [Find] button gets on. Then, the textbox will show the owner name. If it is not found, then an Exclamation box will be shown with "Not Found" massage, as shown in Figure 7.

Figure 7-Owner mobile number is not found in search system

The searching process is performed as follows:

1. The first prefix consisting of three digits – which represent the root of the tree – is extracted to determine the suitable digital search tree.

2. Digit four is examined for determining if it is one of the root's offspring or not. On the off chance that it is not, then the tested phone number is not found and the message box "Not

Found" will appear. Otherwise, the examination process continues for the other digits till it arrives at the last node in the leaf and returns to the owner name.

As shown in Figure 5, assume that a sample is entered in the text-box. The first number is [0795853132], split into 2 parts. The first part, representing the prefix (079), is compared by all values of the root previously generated in the trees. The implementation of the processing with the tree of the root key (079) in the present example is shown in Figure 5.

3 Experimental Results

In order to evaluate the results, as well as to measure the system performance, they were compared to those obtained in previous works [26-30]. Also, a comparison between linear regression phone dictionary and digital tree system is drawn. In this experiment, the same amount of data is stored in both systems and then a group of numbers is entered and searched for the names of their owners. The results show that both systems are identical in terms of the correctness of the results, i.e. both systems give the same results. However, in terms of the time needed to show the results, the digital tree system needs shorter time than that needed in the linear regression system.

5 Causes of the Errors

All training and testing data have the same length (phone number that consists of 10 digits as mentioned above) and there is only one owner for each number. The accuracy of the system in the testing phase is 100% when using training data, because all data are saved and the number can be retrieved with its owner name. While the accuracy of the system is 0% when using testing data because they are not saved and it is impossible to retrieve the number with owner name. This indicates that the number is either not used or used but not recorded for an owner at the telecom company. The accuracy of the system will be 100% when all phone numbers with their owners are saved. Therefore, the accuracy difference between training and testing stages assigns the number of phone numbers in each dataset.

When the size of the dataset is increased, the accuracy will be automatically increased [31]. Hence, it can be supposed that the system has the potential to identify a large amount of phone numbers, while the error rate is automatically decreased. The system may have a "Not Found" result only in the case that the number is not in the system.

6 Conclusions

In conclusion, a mobile phone dictionary was developed through utilizing the technique known as radix search tree. A training dataset was used to train this dictionary and the accuracy was accordingly calculated for the developed number-checker. An overall accuracy of 100% was determined. These results may sufficiently guarantee that the developed dictionary can be efficiently built using the radix search tree. The proposed system tries to save the numbers using a digital search tree in order to make the processes of searching and retrieving customer's information easier and faster. Accordingly, these results may be successfully used by mobile communication companies as the amount of memory required to store data will be reduced and data retrieval will be faster.

Competing Interests

The author declares that she does not have any conflict of interest regarding the publication of this article.

References

[1] K. S. Noori and A.A. Fahad "Monitoring and enhancement of mobile system performance", Iraqi Journal of Science, 61(9) (2020)2418-2425.

[2] S. Ercoli, M. Bertini and A. Del Bimbo, "Compact hash codes and data structures for efficient mobile visual search", Proc. of IEEE Int. Conf. on Multimedia and Expo, 29 June – 3 July 2015 (Torino, Italy).

[3] A. K. Abdul Hassan and M. J. Hadi "Automatic Query Expansion for Arabic Text Retrieval", Iraqi Journal of Science, 58(4C) (2017) 2447-2457.

[4] M. Bahrami, M. Ghorbani and S. Mohammad Arabzad, "Information Technology (IT) as an Improvement Tool for Customer Relationship Management (CRM)", Procedia – Social and Behavioral Sci., 41 (2012) 59-64.

[5] A. Rainer and T. Puschmann, "Successful Practices in Customer Relationship Management", in Proc. 37th Hawaii Int. Colloq. on Sys. Sci. (2004).

[6] A. Mishra and D. Mishra, "Customer Relationship Management: implementation process perspective", Acta Polytechnica Hungarica, 6(4) (2009) 83-99.

[7] M. Srivastava, "Customer Relationship Management (CRM): A Technology Driven Tool", Symbiosis Institute of Business Management Pune (India), Vol. IV (March 2012) 14-25.

[8] R. Bose, "Customer Relationship Management: Key Components for IT Success", Ind. Manag. Data Sys., 102(2) (2002) pp. 89-97.

[9] L.K. Grover, "Quantum mechanics helps in searching for a needle in a haystack", Phys. Rev. Lett., 79 (1997) pp. 325-328.

[10] C. Cortes and D. Pregibon, "Signature-based methods for data streams", Data Min. Knowl. Discov., 5(3) (2001) 167-182.

[11] U.F. Eze, C.J. Onwuegbuchulam, C.H. Ugwuishiwu and S. Diala, "Application of data mining in telecommunication Industry", Int. J. Phys. Sci., 12(6) (2017) pp. 74-88.

[12] S.G. Fashoto, O. Olumide, J. Sadiku and J.A. Gbadeyan, "Application of data mining technique for fraud detection in health insurance scheme using knee-point k-means algorithm", Aust. J. Basic Appl. Sci., 7(8) (2013) pp. 140-144.

[13] A.O. Oyeniyi and A.B. Adeyemo, "Customer Churn Analysis in Banking Sector Using Data Mining Techniques", African J. Comput. ICT, 8(3) (2015) pp. 41-48.

[14] R. Assaf, "The Impact of Social Network Analysis on Telecommunication Customers", Int. J. Sci. Technol. Res., 9(1) (2020) pp. 888-896.

[15] G.M. Weiss, "Data Mining in Telecommunications", in Data Mining and Knowledge Discovery Handbook, edited by O. Maimon and L. Rokach, Springer (2005), pp. 1189-1201.

[16] C. S. Hilas, P. A. Mastorocostas, and I. T. Rekanos, "Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A case Study", Appl. Math. Inf. Sci., 9(4) (2015) pp. 1709-1718.

[17] G. Nagarajan, R.I. Minu, V. Vedanarayanan, S.D. Sundersingh Jebaseelan and K. Vasanth, "CIMTEL Mining Algorithm for Big Data in Telecommunication", Int. J. Eng. Technol., 7(5) (2015) pp. 1709-1715.

[18] L. Saha, L. Sahoo and S.K. Routray, "Survey on CRM Analytics in Telecommunication Industry", Int. J. Computer Trends Technol., 45(1) (2017) pp. 21-27.

[19] J. Han, M. Kamber and J. Pei, "Data Mining Concepts and Techniques", 3rd ed., Morgan Kaufmann (2014) p. 121.

[20] R. Al-Tarawneh, H.S.A. Hamatta, H. Muiadi, P. Abdullah and B. Ghazi, "Novel Approach for Arabic Spell-Checker: Based on Radix Search Tree", Int. J. Computer Appl., 95(7) (2014) pp. 975-8887.

[21] O. I. Abiodun, M. U. Kiru, A. Jantan, A. E. Omolara, K. V. Dada, A. M. Umar, O. U. Linus, H. Arshad, A. A. Kazaure, and U. Gana, ''Comprehensive review of artificial neural network applications to pattern recognition,'' IEEE Access, 7(2019), pp. 158820–158846, doi:10.1109/ACCESS.2019.2945545.

[22] K. Kaczmarski and A. Wolant, "Radix Tree for Binary Sequences on GPU", Int. Conf. on Parallel Process. Appl. Math., March 2018, pp. 219-231.

[23] V. Leis, A. Kemper and T. Neumann, "The adaptive radix tree: ARTful indexing for main-memory databases", IEEE 29th Int. Conf. on Data Eng., 8-12 April 2013.

[24] M. Bohm, B. Schlegel, P.B. Volk, U. Fischer, D. Habich and W. Lehner, "Efficient in-memory indexing with generalized prefix trees," in Datenbanksysteme für Business, Technologie und Web (BTW), Kaiserslautern, Germany, 2011, pp. 227-246.

[25] J. Corbet, "Trees I: Radix trees", available online at: http://lwn.net/Articles/175432/.

[26] E. Berenci, C. Carpineto, V. Giannini and S. Mizzaro, "Effectiveness of keyword-based display and selection of retrieval results for interactive searches", Int. J. Digital Lib., 3(3) (2000) 249-260.

[27] E. Schofield and G. Kubin, "On Interfaces for Mobile Information Retrieval. In Human Computer Interaction with Mobile Devices", in Proc. 4th Int. Symp., Mobile HCI 2002 Pisa, Italy, September 18-20 (2002) pp. 383-387 (Springer, Berlin Heidelberg).

[28] P. Borlund. "The IIR evaluation model: a framework for evaluation of interactive information retrieval systems", Inform. Res., 8(3) (2003) p. 152.

[29] C. Carpineto, S. Mizzaro, G. Romano and M. Snidero, "Mobile Information Retrieval with Search Results Clustering: Prototypes and Evaluations", J. Am. Soc. Inform. Sci. Technol., 60(5) (2009) 877-895.

[30] A.A. Alidin and F. Crestani, "Context acquisition in just-in-time mobile information retrieval", in 2012 Int. Conf. on Inform. Retrieval Knowledge Manag., pp. 203-207, March 2012.

[31] X. Zhu, C. Vondrick, C.C. Fowlkes and D. Ramanan, "Do We Need More Training Data?", Int. J. Comput. Vis., 119(1) (2016) pp. 76–92 (2016).