International Journal of Engineering, Science and InformationTechnology
Volume 4, No. 4 (2024) pp. 250-256
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v4i4.662
Research Paper, Short Communication, Review, Technical Paper

Application of Hashing Method in Medical Term Dictionary Application

Moh. Rahmat Irjii Matdoan[1], Rasna[2*], Sahlan M Saleh[3], Andrian Sah[2], Seno Lamsir[4]

[1]*Department of Informatics, Universitas Sains dan Teknologi Jayapura, Indonesia*
[2]*Department of Information System, Universitas Yapis Papua, Indonesia*
[3]*Department of Informatics, Universitas Yapis Papua, Indonesia*
[4]*Department of Theology, STT Samuel Elizabeth, Indonesia*

**Corresponding author Email: razna.anwar@gmail.com*

The manuscript was received on 20 April 2024, revised on 28 August 2024, and accepted on 18 November 2024, date of publication 5 December 2024

Abstract

The number of words or terms that appear and are used in human life makes it difficult for a person to master all of these words. To overcome this, these words or/terms are compiled in a dictionary, one of which is a dictionary of medical terms. This research aims to create software that can perform its function as a medical terminology dictionary program and implement the Hash method into software that can be used for data search. This research applies the hash method to find the hash value using a rolling hash. The hash method uses input text to be transformed into numbers, thus speeding up the computational process. The application of the hash method in the addressing/storage process in the Database is very efficient in terms of time and place. The output appears in the form of word meaning and hash value. The hash value is raised to know the value of the searched term using the rolling hash formula. With this research, it is hoped that searching for the meaning of medical terms in a computerized manner can be consistent with manual searches.

Keywords: *Hashing, Rolling Hash, Computing, Dictionary, Medical Terms.*

1. Introduction

Developing science and technology can create new terms that enrich the language used [1]. This can be seen from the many words or terms that appear and are used in human life. Because the number of these words is quite large and will continue to grow, a person will find it challenging to master all of them [2]. These words or/terms are arranged in a dictionary to overcome this. Many types of dictionaries have been created, such as general words, dictionaries that contain words or terms used explicitly for specific fields, and so on that have not been computerized, one of which is a dictionary of medical terms [3] [4].

A dictionary of medical terms that has existed for a long time is very thick, which makes it difficult for readers, especially students in the health field, to find the desired terms quickly; besides that, the dictionary is also very difficult to carry. For this reason, the author got the idea to create a computerized dictionary application.

The dictionary application the author will create uses the technique of organizing tables with "Hash" organization [5]. With this technique, the data search does not depend on the size or amount of data (N) in the table because the search is done using the key/address obtained from the conversion of the data using the Rolling Hash [6] [7]. Rolling Hash is expected to make searching for the word's meaning from a medical term easier quickly and in a relatively constant/fixed time [8]. Application of the method Hash: The author chooses it to search faster by matching the hash value obtained from the formula Rolling hash.

2. Literature Review

Data search collects data from computers, books, telephones, dictionaries, etc. And then retrieve the necessary data as quickly as possible. Search is a process that is often used in data management. The search process is to find a specific value (data) in a set of data of the same type (either basic type or formation type) [9]. The algorithm can return a record value or a pointer to the Record. Record itself is a type of data that consists of a set of variables called fields. Sequential search is visiting a tree through each node being visited only once, called a tree transverse / tree visit [10].

Hashing is the arithmetic transformation of a String character to a value that represents the original string [11]. According to the language, Hash means to decapitate and then merge. Hashing is a method for storing data in an array (Array) so that data storage, search, addition, and deletion can be done quickly. The basic idea is to calculate the position of records searched for in the array, not compare the records with the array's contents. A function that returns a value or key is called a function Hash, and the variety used is called a table Hash. In theory, the time complexity (T(n)) of the Hash the ideal is O (1). To achieve that, every Record requires a unique key [12].

American Standard Code for Information Interchange or ASCII (American Standard Code for Information Interchange) [13] is an international standard in the code letters and symbols like Hex and Unicode. Still, ASCII is more universal, e.g., 124 is for the character "|." Computers and other communication tools always use it to show text [14]. The ASCII code has a number composition binary as many as 7 bits. However, ASCII is stored as an 8-bit cipher by adding one digit 0 as the highest significant bit. These additional bits are often used for priority testing. The control characters in ASCII are divided into five groups according to their use: logical communication, Device control, Information separator, Code extension, and physical communication. This ASCII code is widely found on computer keyboards or digital instruments [15].

A database is a collection of information stored in a computer systematically so that it can be checked using a computer program to obtain information from the Database [16]. Database It is used to store information or data that is well integrated inside the computer. To manage Database software called DBMS (Database Management System). DBMS is a software system that allows users to create, maintain, control, and access Databases practically and efficiently. With a DBMS, users will more easily control and manipulate existing data [17]. Meanwhile, RDBMS or relationship Database Management System is a type of DBMS that supports relationships between tables [18]. Besides RDBMS, there are other types of DBMS, such as Hierarchy DBMS, Object Oriented DBMS, etc.

MySQL is the most popular open-source database software globally, and it is used by over 100 million users worldwide. With its reliability, speed, and ease of use, MySQL has become the top choice for many software and application developers on both web and desktop platforms. MySQL users are not limited to individual users and small companies; companies such as Yahoo!, Alcatel-Lucent, Google, Nokia, Youtube, WordPress, and Facebook are also MySQL users. David Axmark, Allan Larsson, and Michael "Monty" Widenius created and developed MySQL in Sweden.

Unlike PHP or Apache, which are software developed by the general community, and the copyright to the source code is owned by their respective authors, MySQL is owned and sponsored by a Swedish commercial company, MySQL AB. MySQL AB holds full copyright on almost all of its source code. The two Swedes and one Finn who founded MySQL AB were David Axmark, Allan Larsson, and Michael "Monty" Widenius [19].

A data flow diagram is a diagram that shows the processes by which data runs in a system. Those pioneered the development of Data Flow Diagrams (DFD) are Tom DeMarco, Edward Yourdon, Chris Gane, and Trish Sarson. This technique only requires 4 (four) symbols [20].

ERD is a conceptual model that describes the relationship between storage (in DFD). ERDs are used to model data structures and relationships between data. With ERD, models can be tested, ignoring the process performed. ERD was first described by Peter Chen as part of the CASE software [21].

3. Methods

Research on the application of the Hash method in the application of medical terminology dictionaries. This research took place from February 2014 to March 2014 to complete the study by collecting some data related to the research.

The data was obtained from a dictionary of medical terms. The author's dictionary came from a medical education institution; the data obtained from this dictionary will be written as input to the application created and stored in the Database.

The author analyzes the problems found in a medical education institution to understand the situation before taking action or deciding to solve the problem.

At this stage, the author designs an application/program. Here, the author created a program to complete the application of the Hash method in medical terminology dictionaries. The first step in designing this program is to design the system work process using DFD, which describes in detail the processes the program will carry out when making a decision. In the coding process, the Pascal Delphi programming language was used with the XAMPP application program, and the Database used was MySQL.

The data needed to process the dictionary of medical terms:

1. Medical terms
2. The word meaning of the term medicine

The system processes data after receiving input data from users. The data is processed to obtain a hash value per input using a rolling hash. The data output carried out by this system displays the word meaning of the medical term searched by the user. Meanwhile, the admin processes the data processing.

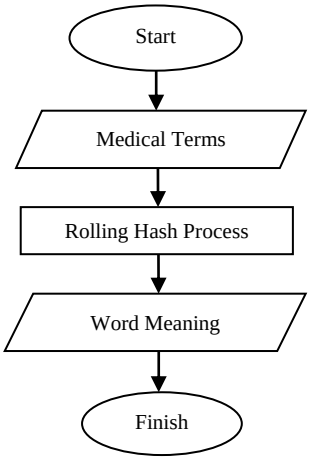

Fig 1. Flow Diagram of Calculation Function

(a) (b)

Fig 2. (a) Flow Diagram of Vocabulary Management (b) Flow Chart of Word Search

4. Result and Discussions

In designing a computer-based system, problem analysis plays a vital role in determining the details of the application to be developed. Problem analysis is a step in understanding the problem before taking action or final resolution decisions.

Searching for the meaning of medical terms at this time is generally still done manually, using a dictionary or searching through a search engine on the internet. The current system is still felt to be ineffective and efficient. Students are still bothered to bring a thick dictionary, and even finding the meaning of the desired term is quite complex and lengthy.

With the support of this system, users, especially medical students, can search for the meaning of medical terms effectively and efficiently because the system has been computerized using the hash method in its application.

4.1. System Planning

1. Context Diagram

   Context diagrams are part of the Data Flow Diagram (DFD) that maps the environment model, which is presented with a single circle representing the entire system at the highest level in the data flow diagram. It contains only one process and shows the system as a whole. The explanation of the mechanism is as follows:

a. The admin logs in to be able to enter the system and process data in the form of word input and the meaning of medical terms. The system then processes the data that has been input by using a rolling hash to address the hash value and store it in the Database.
b. Users log in to be able to enter the system and search for the meaning of medical terms. Users will get information from the results they input into the system so that the output given to the user is the word meaning of the term being sought.

2. DFD Level 0
DFD level 0 is an extension of the context diagram; at this level, every activity that can be carried out by each entity in the system is explained in more detail. The following is an explanation of each activity that occurs at level 0:
a. Process 1.0 is the process of logging in admin and user data
At this stage, the admin or user logs in to be able to enter the system to be used.
b. Process 2.0 is the process of processing user data.
At this stage, the admin inputs or edits user data, including user ID, password, and user level, which will be stored in the user table in the Database.
c. Process 3.0 is a vocabulary processing process.
In this process, the admin inputs the words and their meanings, which will then be processed using a rolling hash to get the hash value and stored in the database data table.
d. The 4.0 process is searching for the meaning of medical terms.
At this stage, the user inputs in the form of words that they want to find their meaning. Furthermore, it is processed by rolling Hash to match the address of the hash value and will produce an output in the form of the meaning of the term desired by the user.

3. DFD Level 1 Login Validation Process
a. Process 1.1 is the process of entering the username and password
The admin and user input the user ID and password at this stage.
b. Process 1.2 Verification process.
This is the stage of checking the username and password input by the admin and user using the user table.
c. Process 1.3 The process of displaying the main menu according to the access rights
If the user ID and password entered by the admin and user are correct, the main menu will appear according to their access rights. If the username and password are incorrect, go back to process 1.1.

4. DFD Level 1 Process of Processing User Data
a. Process 2.1 is the process of adding user data
At this stage, the admin adds user data and saves the data into the table.
b. Process 2.2 is the process of editing participant data
The admin edits the user data at this stage and saves it to the user table.
c. Process 2.3 is the process of deleting participant data.
At this stage, the admin deletes the user data and saves it in the user table.

5. DFD Level 1 Vocabulary Management Process
a. Process 3.1 is the process of adding vocabulary
The admin adds vocabulary and saves it in the data table at this stage.
b. Process 3.2 is the process of editing vocabulary.
The admin edits the vocabulary at this stage and saves it in a data table.
c. Process 3.3 is the process of deleting vocabulary.
At this stage, the admin processes deleting vocabulary and storing it in a data table.

6. DFD Level 1 Word Meaning Search Process
1. Process 4.1 is the process of changing in the form of a hash
At this stage, the user inputs the term they want to search for its meaning, and the system converts the word into a hash value, which will then be processed again for matching.
2. Process 4.2 is the process of matching hash values
At this stage, the system matches the newly entered hash value with the hash value contained in the Database, resulting in an output of the meaning of the term sought.

4.2. Table Design

In the system that will be created, two tables will be used in the Database to apply to the dictionary of medical terms. First, the user uses the table to enter the system, which contains the user ID, password, and level. The function of this table itself is to store user login data.

Table 1. User Table

Name	Type	Size	Information
Uid	VarChar	20	Userid to login
Upass	VarChar	15	User password to login
Ulevel	VarChar	20	User level in applying the system

Second, the data table accommodates vocabulary data on medical terms. This table contains the fields d_id, d_hash, d_kata, and d_ meaning.

Table 2. Data Table

Name	Type	Size	Information
d_id	VarChar	20	Data id
d_hash	VarChar	10	The hash value obtained from the process
d_word	VarChar	30	Medical terms
d_ meaning	VarChar	30	Meaning of the word medical term

4.3. System Stages

a. Lowercase

At this stage, the system will perform lowercase, that is, to change all letters to lowercase. See the following example:

Catching cold ⟶ catching cold

b. Separating Words into Characters

At this stage, separate the words into per characters to find out the ASCII value of each character, which will determine the hash value in the rolling hash calculation. See the following example in Table 3.

Table 3. ASCII Values Per Character

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}
c = 99	a = 97	t = 116	c = 99	h = 104	i = 105	n = 110	g = 103	space = 32	c = 99	o = 111	l = 108	d = 100

After getting the ASCII value per character, the hash value will be calculated using a rolling hash. The ASCII value is obtained from the ASCII table in the appendix, a stipulation.

c. Hash Value Calculation

After the ASCII value per character (c) is obtained from the previous process, the next stage calculates the hash value. Previously, we had to first specify the base value (b) and the character length value (k).

b =3, k =13

$$H_{(c,k)} = C_1 * b^{(k-1)} + C_2 * b^{(k-2)} + \cdots + C_{k_i} * b^{(k-k_i)}$$

$$= (99*)+(97*)+(116*)+(99*)+(104*)+3^{(13-1)}3^{(13-2)}3^{(13-3)}3^{(13-4)}3^{(13-5)}$$
$$(105*)+(110*)+(103*)+(32*)+(99*)+3^{(13-6)}3^{(13-7)}3^{(13-8)}3^{(13-9)}3^{(13-10)}$$
$$(111*)+(108*)+(100*)\,3^{(13-11)}3^{(13-12)}3^{(13-13)}$$

$$= 52612659 + 17183259 + 6849684 + 1948617 + 682344 + 229635 + 80190 + 25029 + 2592 + 2673 + 999 + 324 + 100$$

$$= 79618105$$

Information:

c = character

b = base, a prime number that is taken arbitrarily

k = length of character

From the calculation above, the value of Hash from String "catching cold" is 79618105. Value Hash obtained is the benchmark for matching the similarity with value Hash, which is in the Database using SQL syntax to make the search process faster and more precise.

4.4. User Interface

The final stage is the implementation of the system. The implementation of the system's interface is a test benchmark from the program's results made for the system's development. The implementation of the interface can be seen in Figure 3.

Fig 3. User Interface Display

5. Conclusion

From the results of the application of the hash method in the application of the dictionary of medical terms where the hash value is obtained from the hash rolling process, the author can conclude:

1. The application of the hash method in the medical term dictionary application carried out in this study can run well according to manual search.
2. Measuring the degree of similarity between the hash value of the input term and the hash value of the term contained in the Database produces a valid finding. The suitability of the level of similarity is assessed if one document to be compared is also in the set of documents that are compared. Otherwise, the term is not found.
3. The data sample used amounted to 1500 medical terms, which could be searched with 100% accuracy, and the number of input characters hardly affected the search time.
4. With this system, the search for medical terms in health students, in particular, becomes more efficient.

References

[1] B. Megawati, "Perkembangan padanan istilah bahasa asing dalam bahasa indonesia akibat perkembangan ipteks," 2019.

[2] W. Fitriantiwi and others, *Esai penerapan ejaan bahasa indonesia*. Guepedia, 2020.

[3] Y. Ardi, D. Andreswari, and Y. Setiawan, "Rancang Bangun Aplikasi Kamus Istilah Kedokteran Dengan Menggunakan Algoritma Boyer-Moore Berbasis Android," *Rekursif J. Inform.*, vol. 5, no. 3, 2017.

[4] R. Durling, *A dictionary of medical terms in Galen*, vol. 5. Brill, 2018.

[5] Z. Munawar *et al.*, *Konsep Dasar Pengenalan Ilmu Komputer*. Cendikia Mulia Mandiri, 2023.

[6] H. Jiang and S.-J. Lin, "A rolling hash algorithm and the implementation to LZ4 data compression," *IEEE Access*, vol. 8, pp. 35529–35534, 2020.

[7] D. DiPietrantonio, "Fast Rolling Hashes for Data Deduplication," New York University Tandon School of Engineering, 2022.

[8] S. Roller, S. Sukhbaatar, J. Weston, and others, "Hash layers for large sparse models," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 17555–17566, 2021.

[9] N. Chafid and S. Alfian, "Penerapan Algoritma Pencarian Sequential Pada Aplikasi Kamus Tiga Bahasa Indonesia-Jawa-Jawa Banten," *J. Satya Inform. Vol. 4 No. 1, Mei 2019 Hal*, vol. 4, no. 1, pp. 53–62, 2019.

[10] A. Anna, "Aplikasi Kamus Bahasa Indonesia--Bugis Berbasis Web Dengan Metode Sequential Search," *J. Inform.*, vol. 3, no. 2, 2016.

[11] L. J. Purba and L. Sitorus, "Perancangan Aplikasi Untuk Menghitung Persentase Kemiripan Proposal Dan Isi Skripsi Dengan Algoritma Rabin-Karp," *J. Tek. Inform. UNIKA St. Thomas*, vol. 3, no. 1, pp. 17–25, 2018.

[12] C. A. Karim and others, "APLIKASI KAMUS BAHASA INDONESIA MOI MENGGUNAKAN METODE HASHING BERBASIS ANDROID," STMIK AKAKOM YOGYAKARTA, 2018.

[13] P. Loshin, "ASCII (American Standard Code for Information Interchange). What Is, TechTarget," 2024.

[14] R. Frost, "American Standard Code for Information Interchange," *Gale Encycl. Sci.*, vol. 1, pp. 174–175, 2020.

[15] A. Tantoni and M. T. A. Zaen, "Implementasi Double Caesar Cipher Menggunakan ASCII," *J. Inform. dan Rekayasa Elektron.*, vol. 1, no. 2, pp. 24–32, 2018.

[16] K. Modhiya, "Introduction to DBMS, RDBMS, and NoSQL Database: NoSQL Database Challenges," *RDBMS, NoSQL Database NoSQL Database Challenges (March 6, 2021)*, 2021.

[17] A. Nordeen, *Learn DBMS in 24 Hours*. Guru99, 2022.

[18] E. Setyawati, H. Wijoyo, N. Soeharmoko, and others, "Relational Database Management System (RDBMS)," 2020.

[19] B. Rawat, S. Purnama, and others, "MySQL Database Management System (DBMS) On FTP Site LAPAN Bandung," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 173–179, 2021.

[20] W. Wulandari and A. D. Y. Widiantoro, "Design data flow diagram for supporting the user experience in applications," *Des. data flow Diagr. Support. user Exp. Appl.*, vol. 25, no. 2, pp. 14–20, 2017.

[21] W. S. Davis and D. C. Yen, "Entity-relationship diagrams," in *The Information System Consultant's Handbook*, CRC Press, 2019, pp. 195–204.