RABIN-KARP ALGORITHM IN ALGORITHMS

Abdullayeva Dilobar
Department of Computer Engineering, Tashkent University of Applied Sciences ,190111,Uzekistan
E:mail leylamleyla765@gmail.com

Saydazimov Javlonbek Karimovich
Tashkent university of information technologies named after Muhammad al Khorazmi
javlonbek2020@gmail.com

Abstract. The algorithm is rarely used to match a single pattern, but has significant theoretical significance and is very effective in finding the matches of multiple patterns of the same length. for a text of length n and a pattern of length m, its average and best execution time is O(n) with a correctly selected hash function (see below), but in the worst case it has an O(nm) efficiency, one of the reasons why it is not so widely used. For applications where incorrect positives can be tolerated in search, i.e. applications where some instances of a pattern being found may not actually match the pattern, the Rabin-Karp algorithm works in guaranteed O(n) time and with an appropriate choice of random selection.

Introduction. To perform, the P0 process is selected first. The duration of its processor crack is greater than the value of the time quantum, and therefore the process is performed until the quantum expiration, that is, within 4 time units. After that, it is placed at the end of the queue of ready-to-do processes, which takes the appearance of P1, p2, p0. The next P1 process begins to be executed. Its execution time corresponds to the size of the allocated Quantum, so it continues until the process is complete. Now the turn of the processes in the finished state consists of two P2, p0 processes. The processor is reserved for the p2 process.

Rabin-Karp algorithm

Material from Wikipedia - free encyclopedia

The Rabin-Karp algorithm is a row search algorithm that uses hashing to search for a pattern, i.e. a bottom row, in text. It was developed in 1987 by Michael Rabin and Richard Karp .

The algorithm is rarely used to match a single pattern, but has significant theoretical significance and is very effective in finding the matches of multiple patterns of the same length. for a text of length n and a pattern of length m, its average and best execution time is O(n) with a correctly selected hash function (see below), but in the worst case it has an O(nm) efficiency, one of the reasons why it is not so widely used. For applications where incorrect positives can be tolerated in search, i.e. applications where some of the pattern's found States may not actually match the pattern, the Rabin-Karp algorithm is guaranteed in O (n) time and the appropriate choice of random selection

Material from Wikipedia - free encyclopedia

The Rabin-Karp algorithm is a row search algorithm that uses hashing to search for a pattern, i.e. a bottom row, in text. It was developed in 1987 by Michael Rabin and Richard Karp

The algorithm is rarely used to match a single pattern, but has significant theoretical significance and is very effective in finding the matches of multiple patterns of the same length. for a text of length n and a pattern of length m, its average and best execution time is O(n) with a correctly selected hash function (see below), but in the worst case it has an O(nm) efficiency, one of the reasons why it is not so widely used. For applications where incorrect positives can be tolerated in search, i.e. applications where some instances of a pattern being found may not actually match the pattern, the Rabin-Karp algorithm works in guaranteed O(n) time and with an appropriate choice of random selection. hash function (pa

Since the number of lines we are looking for, k, is so large, traditional one-line search algorithms become inefficient.

Content

Substring ceiling search and competing algorithms

The main task of the algorithm is to find a string of length m in a text of length n, which is called a pattern. One of the simplest algorithms for this task is looking for a bottom line in all possible places:

1 function NaiveSearch (string s[1..n], string sub[1..m]).
I for 1 to n-m+1 for 2
3 for j from 1 to m
4 if s[i+j-1] ≠ sub [j
5 proceed to the next iteration of the outer cycle
6 return i
7 return not found

This algorithm is y in many practical cases

Load matching algorithms

Medial

Sooner or later, web applications will exceed the autonomy of one server. Companies must record both their identity or or other acquisition. To do this, they place their applications on multiple servers and put a load balance in front of it to

distribute incoming requests. To cope with the workload, companies working on a web application may require thousands of servers.

In the Ush post, we will consider ways to distribute single load balancer HTTP requests to multiple servers. We start from the bottom and move on to modern algorithms for cleaning loads.

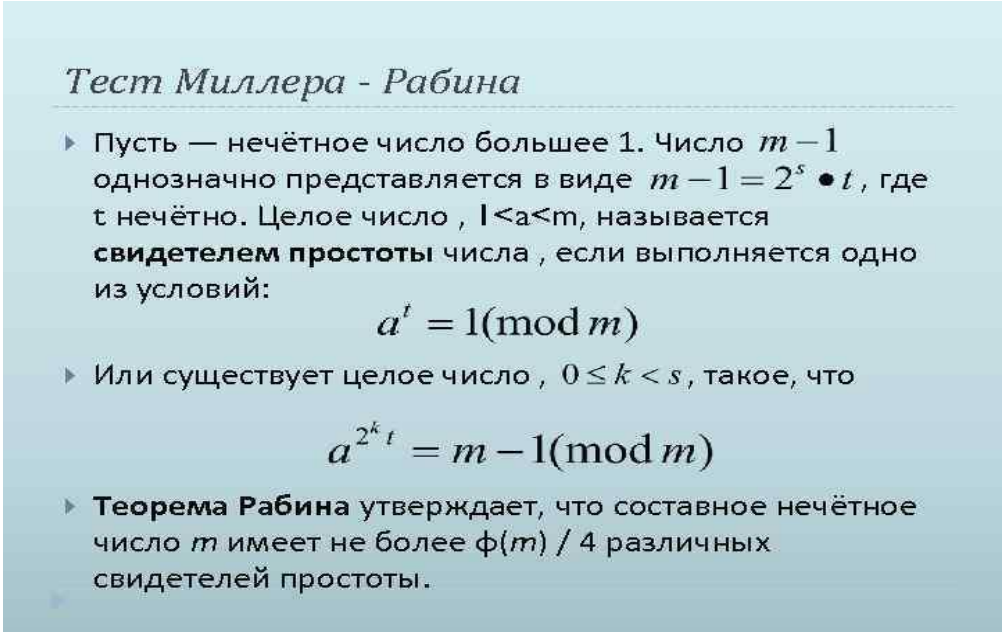
Visualization of the problem

Let's start from the very beginning: requests are sent to one server with one balance. Queries ask (RPS) per second

To perform, the P0 process is selected first. The duration of its processor crack is greater than the value of the time quantum, and therefore the process is performed until the quantum expiration, that is, within 4 time units. After that, it is placed at the end of the queue of ready-to-do processes, which takes the appearance of P1, p2, p0. The next P1 process begins to be executed. Its execution time corresponds to the size of the allocated Quantum, so it continues until the process is complete. Now the turn of the processes in the finished state consists of two P2, p0 processes. The processor is reserved for the p2 process. The allocated processor is completed by the time it runs out, and the next Quanta are allocated for processing p0 - the only one that has not finished its work at the moment. the waiting time for the P0 process (the number of characters "G" in the corresponding line) is 5 time units, for the P1 process - 4 time units, for the P2 process - 8 time units.

Rabin-Karp algorithm methods

The original algorithm proposed by Miller was deterministic and consisted of checking all {\displaystyle 70\ln(m)^{2}}. Miller's algorithm is guaranteed to recognize prime and composite numbers, provided that the generalized Riemann hypothesis is satisfied. The Miller-Rabin algorithm does not depend on the validity of the generalized Riemann hypothesis, but is probabilistic



Another noteworthy algorithm for text search is the Rabin-Karp algorithm, which is based on the hashing method. I will try to explain in detail how it works.

Let us have the following text: FINDINAHAYSTACKNEEDLEINA

Pattern: NEEDLE

a hash number is generated based on characters from 0 to M-1 in the pattern. Here, M is the number of characters in the pattern.

in each [i..N-1] cycle, a hash number is generated for i to M + i - 1 characters in the text. Here, N is the number of characters in the text

If the pattern hash and the text substring hash are equal to each other, the pattern is found in the text.

One of the fastest and easiest ways to generate a hash number is to take the code of each character of the pattern, add them together, and divide it by a large prime number, leaving a remainder. In Javascript, this is done as follows

```
const primeNumber = 997 // prime number
const pattern = "NEEDLE" // search term
const hash = pattern.split('').map(char => char.charCodeAt()).join('') % primeNumber
// hash number for pattern is 769
INDINAHAYSTACKNEEDLEINA
FINDIN  // 5 != 769
 INDINA  // 877 != 769
  NDINAH  // 105 != 769
   DINAHA  // 259 != 769
    INAHAY  // 541 != 769
     NAHAYS  // 414 != 769
      AHAYST  // 271 != 769
       HAYSTA  // 963 != 769
        AYSTAC  // 805 != 769
         YSTACK  // 535 != 769
          STACKN  // 507 != 769
           TACKNE  // 178 != 769
            ACKNEE  // 98 != 769
             CKNEED  // 615 != 769
              KNEEDL  // 317 != 769
               NEEDLE  // 769 == 769
```

A very simple and ingenious method!

Discussion

The hash generation method above may run slower due to the use of split() and join() methods. Therefore, we will perform hash generation using the proven method - Horner's method. The general formula is as follows::

xi = (ti * RM-1 + ti+1 * RM-2 + … + ti+M-1 * R0) % Q

Bu yerda:

xi – matndagi [i..M + i – 1] substring uchun hash soni.

ti – str.charCodeAt(i) – matn/pattern belgisining ASCII jadvalidagi raqami.
R – radix soni. ASCII uchun 128 yoki 256.
M – patterndagi belgilar soni.
Q – tub son.

The probability of a collision in a hash function depends on how big the prime number Q is. For example, if the number Q is greater than M * N2, the probability of collision is equal to 1/N. So we have generated the hash number of the pattern. Now it generates a hash number from each i to M + i - 1 substrings in the text and compares it to the hash number of the pattern. Hash is generated in the same way for both substring and pattern

References

1. Urinov, K. O., Jumanov, K. A., Khidirov, A. M., Urinov, S. K., Abdiyev, J. M., Jumaboyev, T. A., & Eshmirzayev, M. R. (2020, April). Magnetocaloric effect in polycrystalline cobalt. In *Journal of Physics: Conference Series* (Vol. 1515, No. 2, p. 022079). IOP Publishing.
2. Saidov, A. S., Saparov, D. V., Usmonov, S. N., Kutlimratov, A., Abdiev, J. M., Kalanov, M., ... & Akhmedov, A. M. (2021). Investigation of the Crystallographic Perfection and Photoluminescence Spectrum of the Epitaxial Films of (Si2) 1-x (GaP) x \left(0≤x≤1\right) Solid Solution, Grown on Si and GaP Substrates with the Crystallographic Orientation (111). *Advances in Condensed Matter Physics*, *2021*, 1-8.
3. LEYDERMAN, A., Saidov, A. S., USMONOV, S., Abdiyev, J. M., & Suyarov, Q. T. (2021, February). INFLUENCE OF WEAK GRADED GAP ON INJECTION DIFFUSION REGIMES OF CURRENT TRANSPORT IN SEMICONDUCTOR PN-STRUCTURES. In *Congress Dates* (p. 40).
4. Leiderman, A. Y., Saidov, A. S., & Abdiyev, J. M. (2021). Injection diffusion processes in the weak linear graded-band semiconductor pn-structures. *Euroasian Journal of Semiconductors Science and Engineering*, *3*(1), 3.
5. Abdiev, J., Safarov, O., & Julanov, H. (2022). Study of the properties of polymer composites–reinforcement based on glass and basalt fibers. *Eurasian Scientific Herald*, *7*, 77-88.
6. Abdiev, J., & Safarov, O. (2022). Basalt fiber-basic (primary) concepts. *Web of Scientist: International Scientific Research Journal*, *3*(4), 212-240.
7. Abdiev, J., Abdieva, N. M., & Khasanova, D. Y. (2022). Physical Terms, Problems and Their Solutions. *Problems and Their Solutions (March 25, 2022)*.
8. Qo'chqarovna, M. U., & Safarali, D. (2022). Theory of dependence of ultrasound absorption (amplification) in semiconductors on electron scattering mechanism. *Web of Scientist: International Scientific Research Journal*, *3*(7), 480-489.
9. Saidov, A. S., Usmonov, S. N., Karshiev, A. B., & Abdiev, J. M. (2022, December). Influence of the varizonal Si1-xGex solid solution composition

on the thermovoltaic effect in n–Si–p–Si (1-x) Gex structure. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1112, No. 1, p. 012040). IOP Publishing.

10. Taylanov, N. A., Urinov, S. X. O. G. L., & Abdiev, J. M. O. G. L. (2022). A Fourth-Order Runge-Kutta Method for Numerical Solution of the Kuramoto-Sivashinsky Equation.
11. Abdullayeva, D. O. T., Ismoilov, M. M., qizi Abdieva, N. M., & Khasanova, D. Y. (2022). Peach signal control software is an important part of our lives. *Eurasian Scientific Herald*, *7*, 145-152.
12. Taylanov, N. A. (2022). A Fourth-Order Runge-Kutta Method for Numerical Solution of the Kuramoto-Sivashinsky Equation. *Eurasian Scientific Herald*, *7*, 58-61.
13. Mayinova, U. (2023). ENHANCING ENERGY EFFICIENCY IN INDUSTRIAL ENTERPRISES THROUGH ADVANCED AUTOMATION AND CONTROL SYSTEMS. *ECONOMIC AND SOCIAL ASPECTS OF THE DEVELOPMENT OF ALTERNATIVE ENERGY, IMPACT ON THE ECOLOGY*.
14. Payzullaev, A. N., Allaev, B. A., Mirzaev, S. Z., Abdiev, J. M., Urinov, J., & Parkash, A. (2023). The Impact of Silicon Dioxide Nanoparticle Size on the Viscosity and Stability of Nanofluids: A Comprehensive Study. *ECS Advances*.
15. Mayinova, U. A., & Abdiev, J. M. (2023). THE TEMPERATURE-DEPENDENT SUPERFLUID DENSITY AND RELATED LONDON PENETRATION DEPTH. *International journal of advanced research in education, technology and management*, *2*(6).
16. Urinov, J., & ugli Abdiev, J. M. (2023, January). TECHNOLOGY FOR OBTAINING VO2 FILMS FROM THE GAS PHASE DURING THERMAL DECOMPOSITION. In *3rd International Conference on Material Science, Smart Structures and Applications*. Scienceweb-National database of scientific research of Uzbekistan.
17. Xusanbek Ulug'bek o'g, M., al-Khwarizmi, M., Abdullayeva, D. O. T., qizi Abdieva, N. M., & Khasanova, D. Y. (2022, June). DEVELOPMENT OF ALGORITHMS AND PROGRAMS FOR PROCESSING SPEECH SIGNALS ON VISUAL DSP++ PLATFORM. In *E Conference Zone* (pp. 83-99).
18. Urinov, K. (2023, January). TECHNOLOGY FOR OBTAINING VO2 FILMS FROM THE GAS PHASE DURING THERMAL DECOMPOSITION. In *3rd International Conference on Material Science, Smart Structures and Applications*. Scienceweb-National database of scientific research of Uzbekistan.
19. Xusanbek Ulug'bek o'g, M., al-Khwarizmi, M., Abdullayeva, D. O. T., qizi Abdieva, N. M., & Khasanova, D. Y. (2022, June). DEVELOPMENT OF ALGORITHMS AND PROGRAMS FOR PROCESSING SPEECH

SIGNALS ON VISUAL DSP++ PLATFORM. In *E Conference Zone* (pp. 83-99).

20. Jo'lanov Hasan Komilovich, Azimov Alisher Mirzo ugli, Omadjon Safarov, Abdiev Jurabek Muzaffar ugli, & Yaxshimurodov Ahmad Asror ugli. (2022). UNDERSTANDING THE MAGNETO CALORIC EFFECT IN SUPER SPIN GLASS COBALT-BASED NANOPARTICLES. *Web of Scientist: International Scientific Research Journal*, *3*(7), 545–563. Retrieved from https://wos.academiascience.org/index.php/wos/article/view/2243
21. Raufovich, M. R., & Muzaffar o'g'li, A. J. THE ROLE OF TEACHER AND STUDENT COOPERATION IN THE MODERN EDUCATIONAL PROCESS OF GOOGLE DOCS.
22. KUŞÇULU, N. (2021). Congress Dates.
23. Qo'chqarovna, J. H., & Safarov, O. UNDERSTANDING THE MAGNETO CALORIC EFFECT IN SUPER SPIN GLASS COBALT-BASED NANOPARTICLES.
24. Ortikovich, U. J., Muzaffar o'g'li, A. J., & Rafikovich, M. R. **Fex V**3− x **O4**MAGNETIC AND ELECTRICAL PROPERTIES OF SPINEL.
25. Mayinova, U. (2022, September). DISTINCTIVE SUPERCONDUCTING STATES AND PROPERTIES OF DOPED HIGH-TC CUPRATES. In *IV International Scientific Forum "NUCLEAR SCIENCE AND TECHNOLOGIES" dedicated to the 65th anniversary of the Institute of Nuclear Physics*. Scienceweb-National database of scientific research of Uzbekistan.
26. Mayinova, U. (2022). CALCULATION OF CORRELATION FUNCTIONS FOR A MODEL OF LIMITED ROTATIONAL MOTION OF MOLECULES IN CONDENSED MATTER. *Scienceweb academic papers collection*.
27. Umidullayev, S. U., & Mayinova, U. K. (2021). MOLEKULALAR ICHKI CHEKLANGAN AYLANMA HARAKATINI HISOBGA OLGAN HOLDA KONDENSIRLANGAN MUHITLARDA YORUG'LIKNING RELEYCHA SOCHILISHI NAZARIYASI. In *Вопросы технических и физико-математических наук в свете современных исследований* (pp. 128-136).
28. Mayinova, U. (2023). ENHANCING ENERGY EFFICIENCY IN INDUSTRIAL ENTERPRISES THROUGH ADVANCED AUTOMATION AND CONTROL SYSTEMS. *ECONOMIC AND SOCIAL ASPECTS OF THE DEVELOPMENT OF ALTERNATIVE ENERGY, IMPACT ON THE ECOLOGY*.
29. Mayinova, U. A., & Abdiev, J. M. (2023). THE TEMPERATURE-DEPENDENT SUPERFLUID DENSITY AND RELATED LONDON PENETRATION DEPTH. *International journal of advanced research in education, technology and management*, *2*(6).