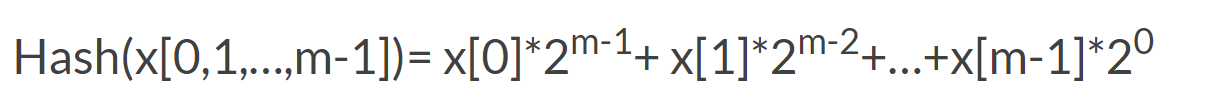
**Thuật toán Rabin-Karp**

1. **Ý tưởng thuật toán**

Thuật toán Rabin-Karp là một thuật toán được sử dụng để tìm kiếm hoặc so khớp chuỗi trong đoạn văn bản bằng cách sử dụng một hàm băm. Không giống như thuật toán so khớp chuỗi thông thường, nó sẽ không đi qua mọi ký tự trong giai đoạn đầu mà nó sẽ lọc các ký tự không khớp và sau đó thực hiện so sánh.

Hàm băm là một công cụ để ánh xạ giá trị đầu vào lớn hơn với giá trị đầu ra nhỏ hơn. Giá trị đầu ra này được gọi là giá trị băm.

Hàm băm:



Hàm băm tốt:

+ Các thao tác cơ bản được thực hiện hiệu quả

+ Khi băm hai xâu con khác nhau có cùng độ dài mm, xác suất hai giá trị băm giống nhau là nhỏ.

1. **Trình bày thuật toán**

Input:

* T[0…n-1]: là văn bản có n ký tự
* P[0…k-1]: là pattern có k ký tự với k ≤ n
* Hash\_T: là giá trị băm của chuỗi con tuần tự T[s…s+k-1] trong T với độ dịch chuyển là s, trong đó 0 ≤ s ≤ n-k.
* Hash\_P: là giá trị băm của P.

Output:

Khi này thuật toán so sánh lần lượt giá trị hash\_T với hash\_P với s chạy từ 0 đến n-k , bước tiếp theo của thuật toán sẽ xảy ra với hai trường hợp như sau:

* TH1: hash\_T = hash\_P, thực hiện phép so khớp chuỗi giữa T[s…s+k-1] và P[0…k-1]
* TH2: hash\_T ≠ hash\_P, nếu s ≤ k , tính gán s = s+1 và tính tiếp giá trị băm hash\_T.

Mã giả:

Text

Description automatically generated

1. Độ phức tạp thuật toán

Độ phức tạp về thời gian trong giao đoạn tiền xử lý là O(M)

Khi tính giá trị băm cho T[i+1…i+m] ta mất thời gian là O(M), do công việc này được thực hiên trong (n-m+1) lần. Do đó độ phức tạp trong giao đoạn tìm mẫu là **O(M\*N)**

Tài liệu tham khảo:

<https://tek4.vn/khoa-hoc/cau-truc-du-lieu-va-giai-thuat/thuat-toan-rabin-karp-ung-dung-trong-doi-sanh-mau>

<https://www.baeldung.com/cs/rabin-karp-algorithm>