Thuật toán Horspool

Phát biểu thuật toán

- Là dạng đơn giản của thuật toán Boyer – Moore

- Chỉ dùng ký tự xấu để tính toán bước nhảy

- Dễ thực hiện

- Giai đoạn tiền xử lý độ phức tạp O(m+ α) thời gian và O(α ) không gian

- Giai đoạn tìm kiếm phức tạp O(m x n ) thời gian

- Chỉ số trung bình của sự so sánh cho 1 ký tự text là giữa 1/ α và 2/( α +1)

b.Code demo thuật toán

public class Horspool ***{*** public static boolean cmp***(***char***[]*** x, char***[]*** y, int y1***) {*** for ***(***char c : x***) {*** if ***(***c != y***[***y1++***]) {*** return false;  
 ***}  
 }*** return true;  
 ***}*** //tao mang danh dau khoang cach tu tung tu ki tu cuoi cung cua xau mau den ki tu cuoi cung cua xau (khong tinh ki tu cuoi cung)  
 //vi du xau mau: GCAGAGAG  
 //mang danh dau: A(1) C(6) G(2) T(8)  
 public static int***[]*** preBc***(***char***[]*** x***) {*** int***[]*** bc = new int***[***255***]***;  
 int m = x.length;  
 for ***(***int i = 0; i < 255; i++***) {*** bc***[***i***]*** = m;  
 ***}*** for ***(***int i = 0; i < m - 1; i++***) {*** bc***[(***int***)*** x***[***i***]]*** = m - i - 1;  
 ***}*** return bc;  
 ***}*** public static void search***(***char***[]*** x, char***[]*** y***) {*** int m = x.length;  
 int n = y.length;  
 int***[]*** preBc = *preBc****(***x***)***;  
 int i = 0;  
 while ***(***i <= n - m***) {*** char c = y***[***i + m - 1***]***;  
 if ***(****cmp****(***x, y, i***)) {*** System.*out*.println***(***"Các vị trí xuất hiện trong văn bản của xâu mẫu là: " + i***)***;  
 ***}*** i = i + preBc***[***c***]***;  
 ***}  
 }*** public static void main***(***String***[]*** args***) {*** char***[]*** x = "GCAGAGAG".toCharArray***()***;  
 char***[]*** y = "GCATCGCAGAGAGTATACAGTACG".toCharArray***()***;  
 *search****(***x, y***)***;  
 ***}***

c.Kiểm nghiệm

Bảng ký tự tồi72

Bước 1:

G C A T C G C A G A G A G T A T A C A G T A C G

1

G C A G A G A G

Dịch : 1 (bmBc[A])

Bước 2:

G C A T C G C A G A G A G T A T A C A G T A C G

2 1

G C A G A G A G

Dịch: 2 (bmBc[G])

Bước 3:

G C A T C G C A G A G A G T A T A C A G T A C G

2 1

G C A G A G A G

Dịch: 2 (bmBc[G])

Bước 4:

G C A T C G C A G A G A G T A T A C A G T A C G

2 3 4 5 6 7 8 173

G C A G A G A G

Dịch: 2 (bmBc[G])

Bước 5:

G C A T C G C A G A G A G T A T A C A G T A C G

1

G C A G A G A G

Dịch: 1 (bmBc[A])

Bước 6:

G C A T C G C A G A G A G T A T A C A G T A C G

1

G C A G A G A G

Dịch: (bmBc[T])

Bước 7:

G C A T C G C A G A G A G T A T A C A G T A C G

2 1

G C A G A G A G

Dịch: 2 (bmBc[G])

Kết thúc th