

KMP Algorithm

Knuth - Morris - Pratt (KMP) Algorithm

$\bar{n} \propto 1ps$

P_1 : $\overset{1}{a}\overset{2}{b}c\overset{3}{d}a\overset{4}{b}e\overset{5}{a}\overset{6}{b}\overset{7}{f}$
0 0 0 0 1 2 0 1 2 0

P_2 : $a\overset{1}{b}c\overset{2}{d}e\overset{3}{a}\overset{4}{b}\overset{5}{f}\overset{6}{a}\overset{7}{b}c$
0 0 0 0 0 1 2 0 1 2 3

P_3 : $a\overset{1}{a}b\overset{2}{c}a\overset{3}{d}a\overset{4}{a}b\overset{5}{e}$
0 1 0 0 1 0 1 2 0 0

P_4 : $a\overset{1}{a}a\overset{2}{a}b\overset{3}{a}a\overset{4}{c}d$
0 1 2 3 0 1 2 0 0



Knuth-Morris-Pratt (KMP) Algorithm

String: a b a b c a b c a b a b a b d
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

pattern: 0

	1	2	3	4	5
a	a	b	a	b	d
0	0	0	1	2	0



Knuth-Morris-Pratt (KMP) Algorithm

String: a b a b c a b c a b a b a b dⁱ
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Pattern: ⁰

	¹	²	³	⁴	⁵ ^j
a	b	a	b	d	
0	0	1	2	0	



Rabin Karp Algorithm

Rabin-Karp Algorithm for string matching

— Hash Table

→ Hash function

$T = a a a a a b$

$p = \underline{a a b}$

$a = 97 \quad 1$

$b = 98 \quad 2$

$c \quad 3$

$d \quad 4$

$e \quad 5$

\vdots

\vdots

\vdots

$x \quad 24$

$y \quad 25$

$z \quad 26$

Rabin-Karp Algorithm for string matching

- Hash Table

→ Hash function

T = a a a a a b

P = a a b

(Addition is my hash function)

$$1 + 1 + 2 = 4$$

$$a a a = 3$$

$$1 + 1 + 1 = 3$$

$$1 + 1 + 1 = 3$$

$$1 + 1 + 1 = 3$$

$$1 + 1 + 2 = 4$$

$$a = 97 \quad 1$$

$$b = 98 \quad 2$$

$$c \quad 3$$

$$d \quad 4$$

$$e \quad 5$$

.

.

.

$$x \quad 24$$

$$y \quad 25$$

$$z \quad 26$$

T = c c a c c a d b a

P = d b a

$$4+2+1 = \underline{7}$$

$$3+3+1 = \underline{7}$$

spurious hit

$$d=4$$

$$b=2$$

$$a=1$$

c a c

$$3+1+3 = \underline{7} \text{ SH,}$$

a c c

$$1+3+3 = \underline{7} \text{ SH,}$$

c c a

$$3+3+1 = \underline{7} \text{ SH,}$$

c a d

$$3+1+4 = \underline{8} \text{ not a match}$$

a d b

$$1+4+2 = \underline{7} \text{ SH,}$$

T = c c a c c a d b a

P = d b a

$$4+2+1 = \underline{7}$$

$$3+3+1 = \underline{7}$$

Spurious hit.

c a c

$$3+1+3 = \underline{7} \text{ SH}$$

$$a c c = \underline{7} \text{ SH}$$
$$1+3+3$$

$$c c a = \underline{7} \text{ SH}$$
$$3+3+1$$

$$c a d = \underline{8} \text{ not a match}$$
$$3+1+4$$

$$a d b$$
$$1+4+2 = \underline{7} \text{ SH}$$

$$d=4$$

$$b=2$$

$$a=1$$

$$\begin{array}{ccc} d & b & a \\ 4 * 26^2 & 2 * 26^1 & 1 * 26^0 = x \end{array}$$

$$\begin{array}{ccc} c & c & a \\ 3 * 26^2 & 3 * 26^1 & 1 * 26^0 = y \end{array}$$

RK- Algo (T, P, d, q)

$n \leftarrow \text{len}(T)$

$m \leftarrow \text{len}(P)$

$h \leftarrow \frac{d^{m-1} \bmod q}{}$

$p \leftarrow 0$

$t_0 \leftarrow 0$

for $i \leftarrow 1$ to m

do $p \leftarrow (dp + P[i]) \bmod q$

$t_0 \leftarrow (d \cdot t_0 + T[i]) \bmod q$

for $s \leftarrow 0$ to $n-m$

do if $\underline{p} = \underline{t_s}$

then if $\underline{P[1 \dots m]} = \underline{T[s+1 \dots s+m]}$

then "Pattern because with shift " s

if $s < n-m$

then $t_{s+1} \leftarrow (d (t_s - T[s+1] h) T[s+m+1]) \bmod q$

Ex 7

$T = 31415926535 \dots$

$P = 26$

$T \text{ Length} = 11$ so $Q = 11$

$P \bmod Q = 26 \bmod 11 = \underline{4}$

$O(\underline{n-m+1})$

$T = \begin{array}{ccccccccccc} 3 & 1 & 4 & 1 & 5 & 9 & 2 & 6 & 5 & 3 & 5 \end{array}$

Shift 1: $\begin{array}{ccccc} 3 & 1 & 4 & 1 & 5 \\ 2 & 6 & & & \end{array}$ $31 \bmod 11 = 9$ (not a match)

Shift 2: $\begin{array}{ccccc} & 3 & 1 & 4 & 1 \\ & 2 & 6 & & \end{array}$ $14 \bmod 11 = 3$ (not a match)

Shift 3: $\begin{array}{ccccc} & & 3 & 1 & 4 \\ & & 2 & 6 & \end{array}$ $41 \bmod 11 = 8$ (not a match)

Shift 4: $\begin{array}{ccccc} & & & 3 & 1 \\ & & & 2 & 6 \end{array}$ $15 \bmod 11 = 4$ (match but spurious hit)

Shift 5: $\begin{array}{ccccc} & & & & 3 \\ & & & & 2 \end{array}$ $59 \bmod 11 = 4$ (spurious hit)

Shift 6: $\begin{array}{ccccc} & & & & & 3 \\ & & & & & 2 \end{array}$ $92 \bmod 11 = 4$ (spurious hit)

Shift 7: $\begin{array}{ccccc} & & & & & & 3 \\ & & & & & & 2 \end{array}$ $26 \bmod 11 = 4$ (Exact match)

Pattern occurs with shift 7

Shift 8: $\begin{array}{ccccc} & & & & & & & 3 \\ & & & & & & & 2 \end{array}$ $65 \bmod 11 = 10$ (not a match)

Shift 9: $\begin{array}{ccccc} & & & & & & & & 3 \\ & & & & & & & & 2 \end{array}$ $53 \bmod 11 = 9$ (not a match)