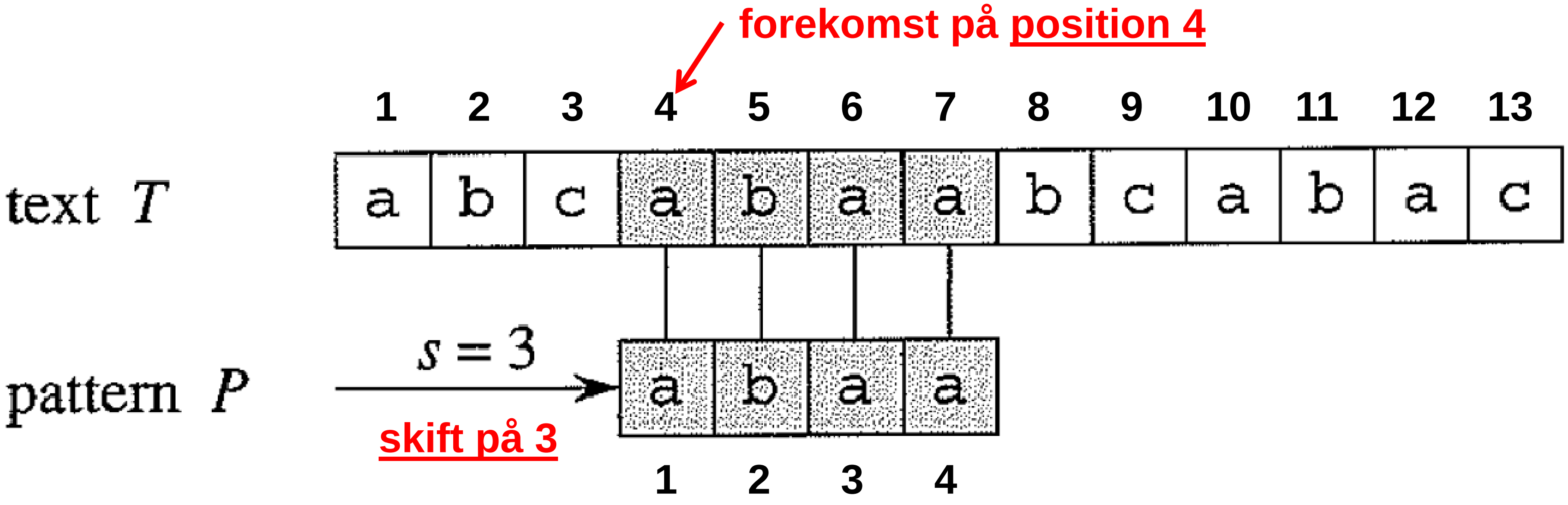
Algoritmer og Datastrukturer 2

Gerth Stølting Brodal

Mønstergenkendelse [CLRS, kapitel 32.1-32.2, 32.4]

AARHUS UNIVERSITET

Mønster genkendelse



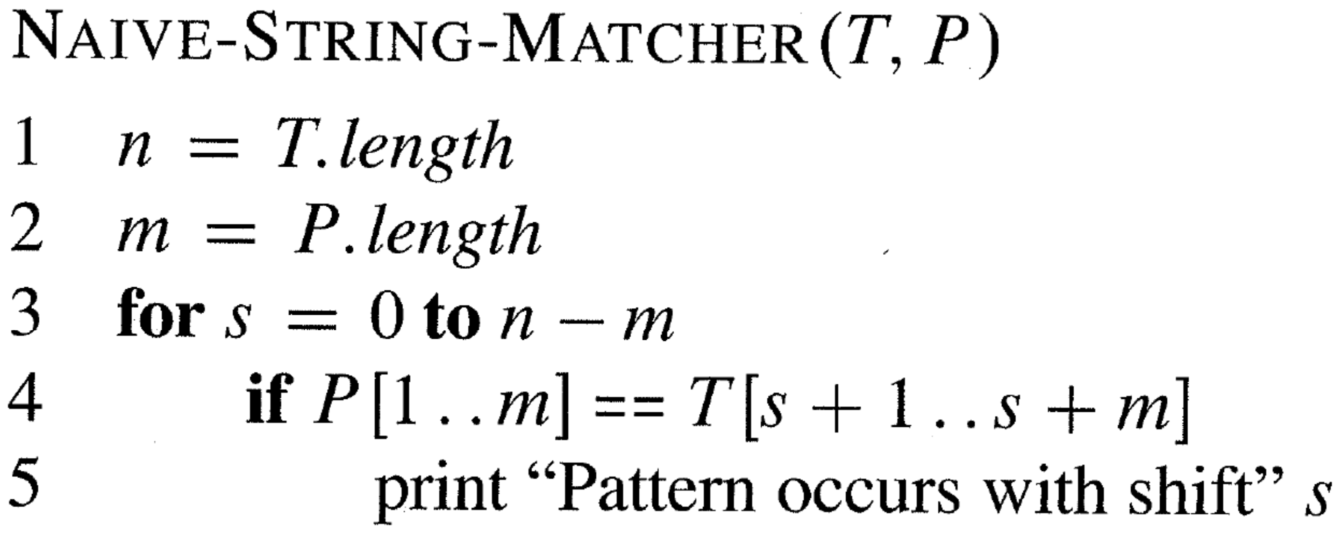
forekomst på position 4

skift på 3

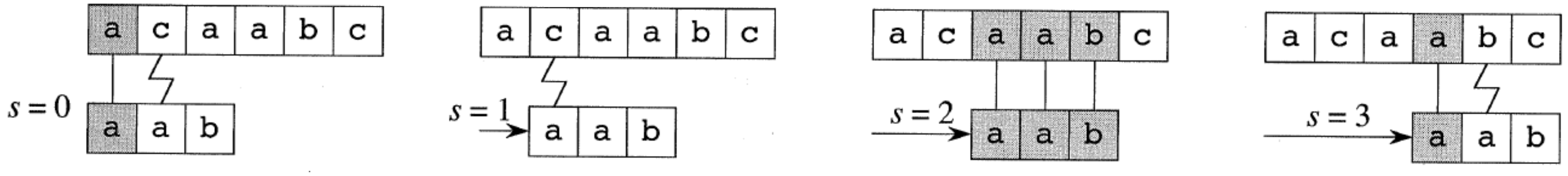
Input: Tekst T af længde n og mønster P af længde m

Output: Alle positioner i T hvor P forekommer

Naive Algoritme

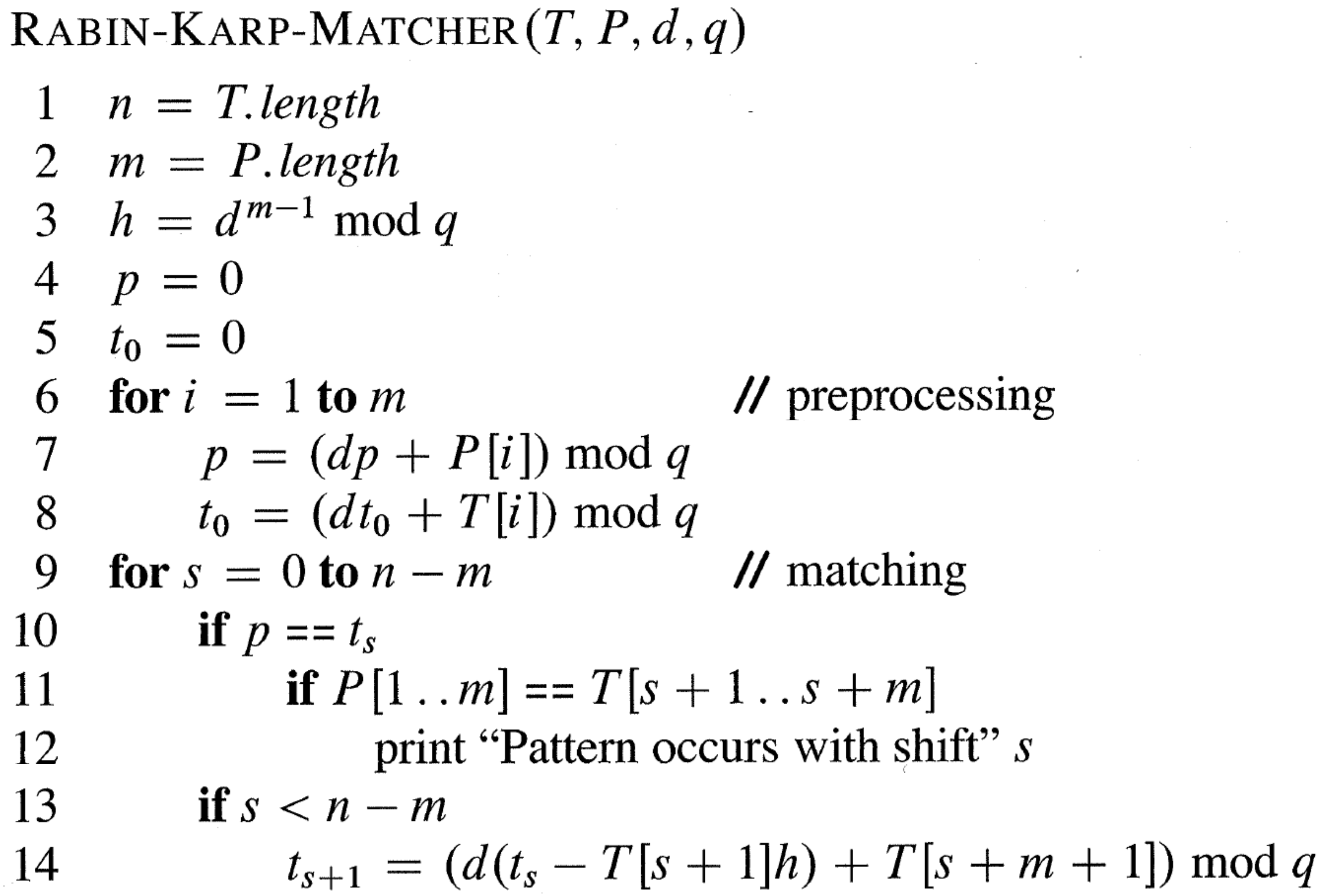


```
NAIVE-STRING-MATCHER($T$, $P$)
$n = T.length$
$m = P.length$
for $s = 0$ to $n - m$
        if $P[1\,..\,m] == T[s+1\,..\,s+m]$
                print "Pattern occurs with shift" $s$
```



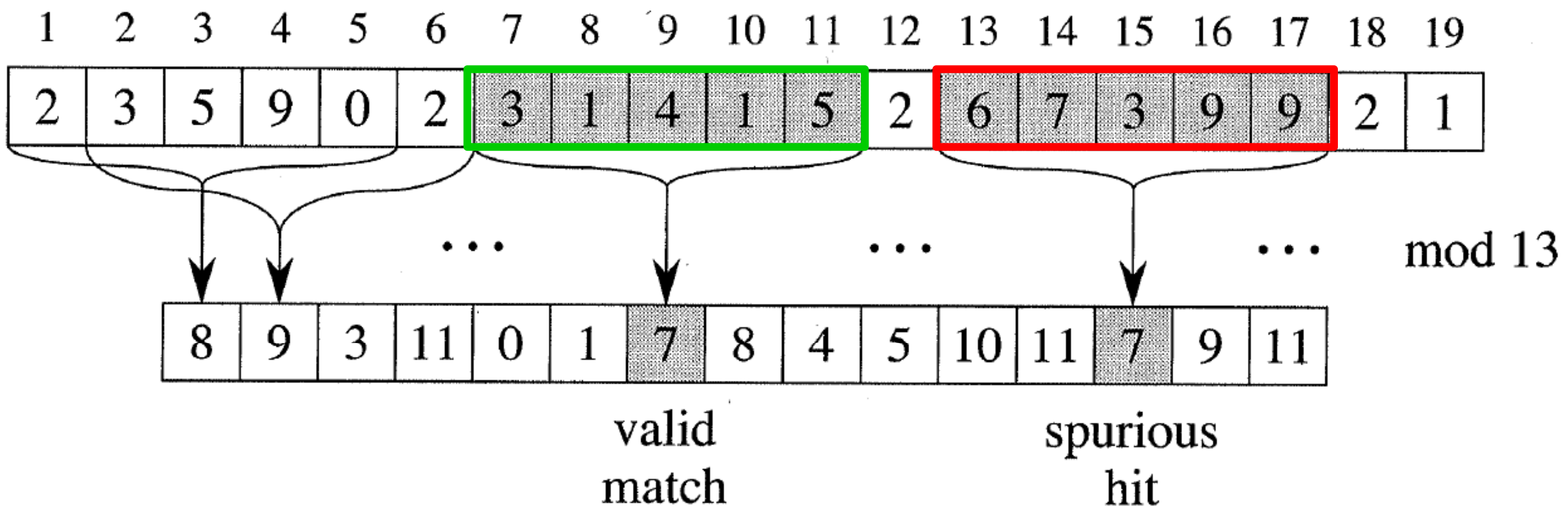
1981

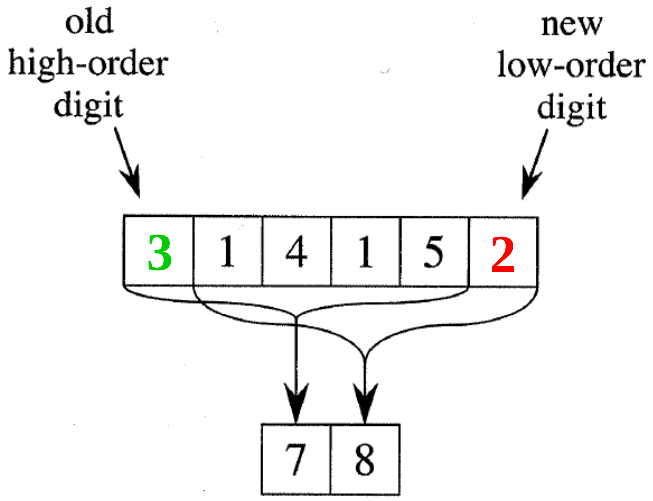
Rabin-Karp



```
RABIN-KARP-MATCHER$(T, P, d, q)$
$n = T.length$
$m = P.length$
$h = d^{m-1} \bmod q$
$p = 0$
$t_0 = 0$
for $i = 1$ to $m$                          // preprocessing
    $p = (dp + P[i]) \bmod q$
    $t_0 = (dt_0 + T[i]) \bmod q$
for $s = 0$ to $n - m$                      // matching
    if $p == t_s$
        if $P[1\,..\,m] == T[s+1\,..\,s+m]$
            print "Pattern occurs with shift" $s$
    if $s < n - m$
        $t_{s+1} = (d(t_s - T[s+1]h) + T[s+m+1]) \bmod q$
```

Rabin-Karp: Eksempel

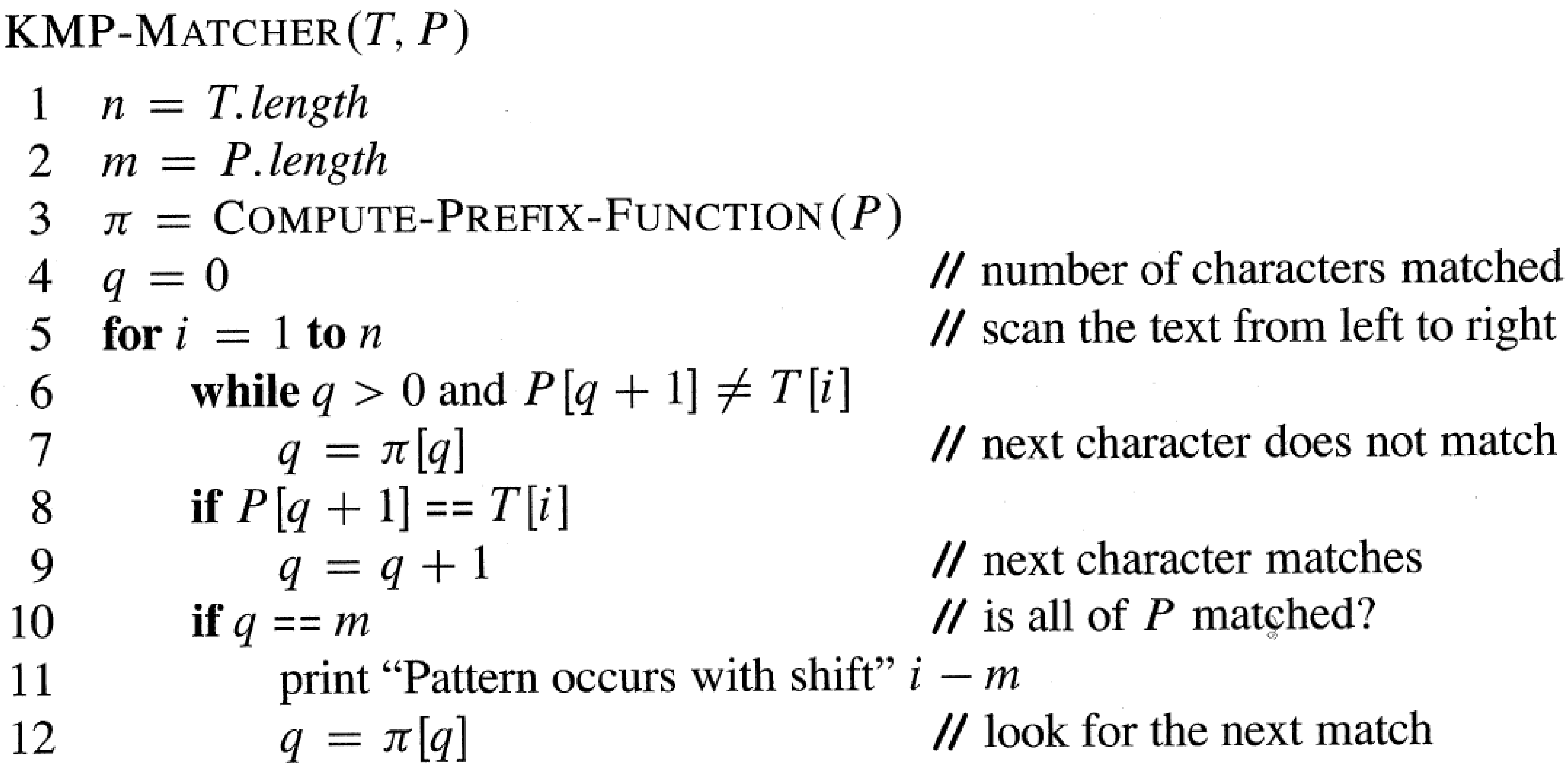




$$\begin{aligned}14152 &\equiv (31415 - \overset{\substack{\text{old}\\\text{high-order}\\\text{digit}\\\downarrow}}{3}\cdot 10000)\cdot \overset{\substack{\text{shift}\\\downarrow}}{10} + \overset{\substack{\text{new}\\\text{low-order}\\\text{digit}\\\swarrow}}{2} \pmod{13}\\ &\equiv (7 - 3\cdot 3)\cdot 10 + 2 \pmod{13}\\ &\equiv 8 \pmod{13}\end{aligned}$$

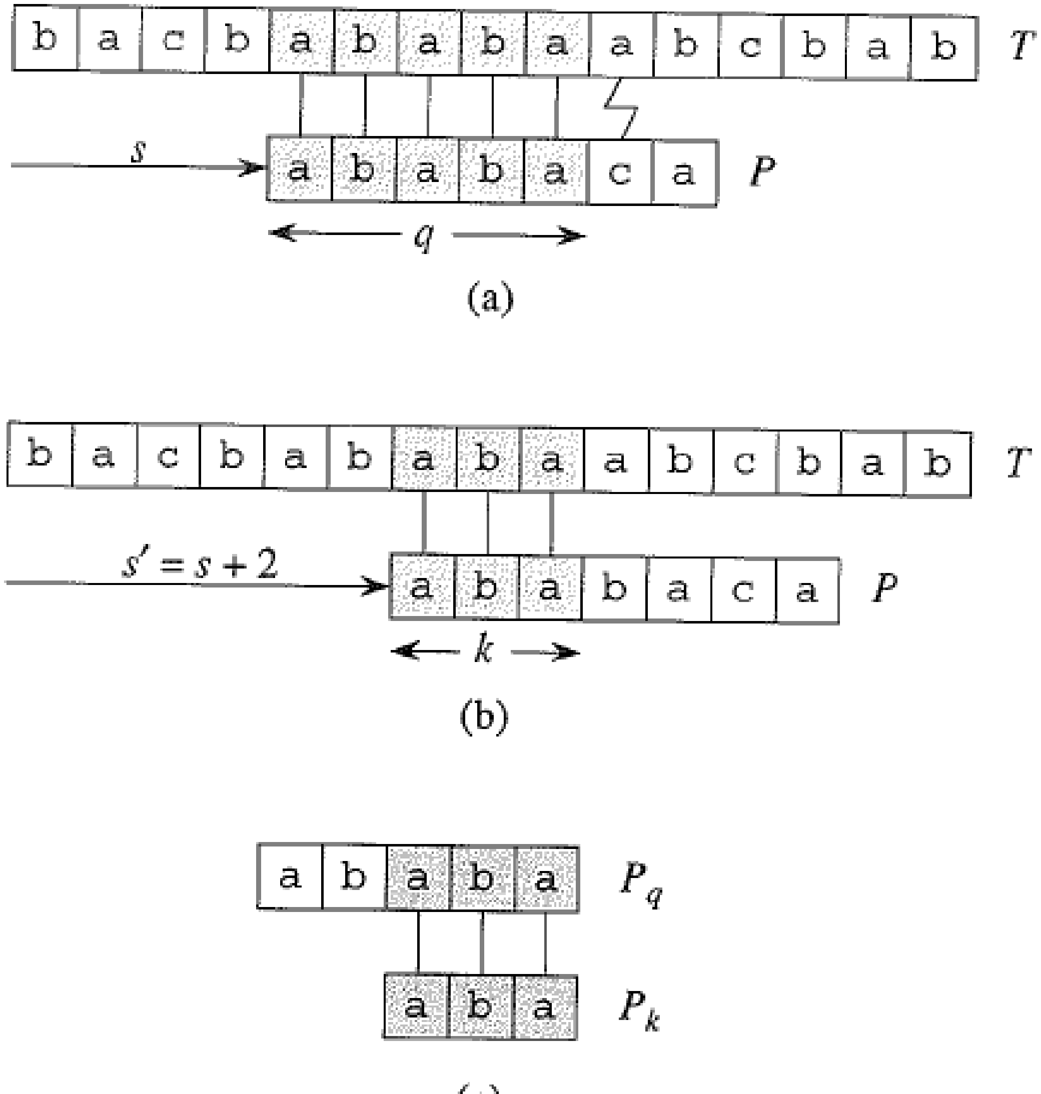
1977

Knuth-Morris-Pratt



```
KMP-MATCHER($T, P$)
$n = T.length$
$m = P.length$
$\pi$ = COMPUTE-PREFIX-FUNCTION($P$)
$q = 0$                                        // number of characters matched
for $i = 1$ to $n$                             // scan the text from left to right
    while $q > 0$ and $P[q+1] \neq T[i]$
        $q = \pi[q]$                           // next character does not match
    if $P[q+1] == T[i]$
        $q = q + 1$                            // next character matches
    if $q == m$                                // is all of $P$ matched?
        print "Pattern occurs with shift" $i - m$
        $q = \pi[q]$                           // look for the next match
```

Knuth-Morris-Pratt: Eksempel

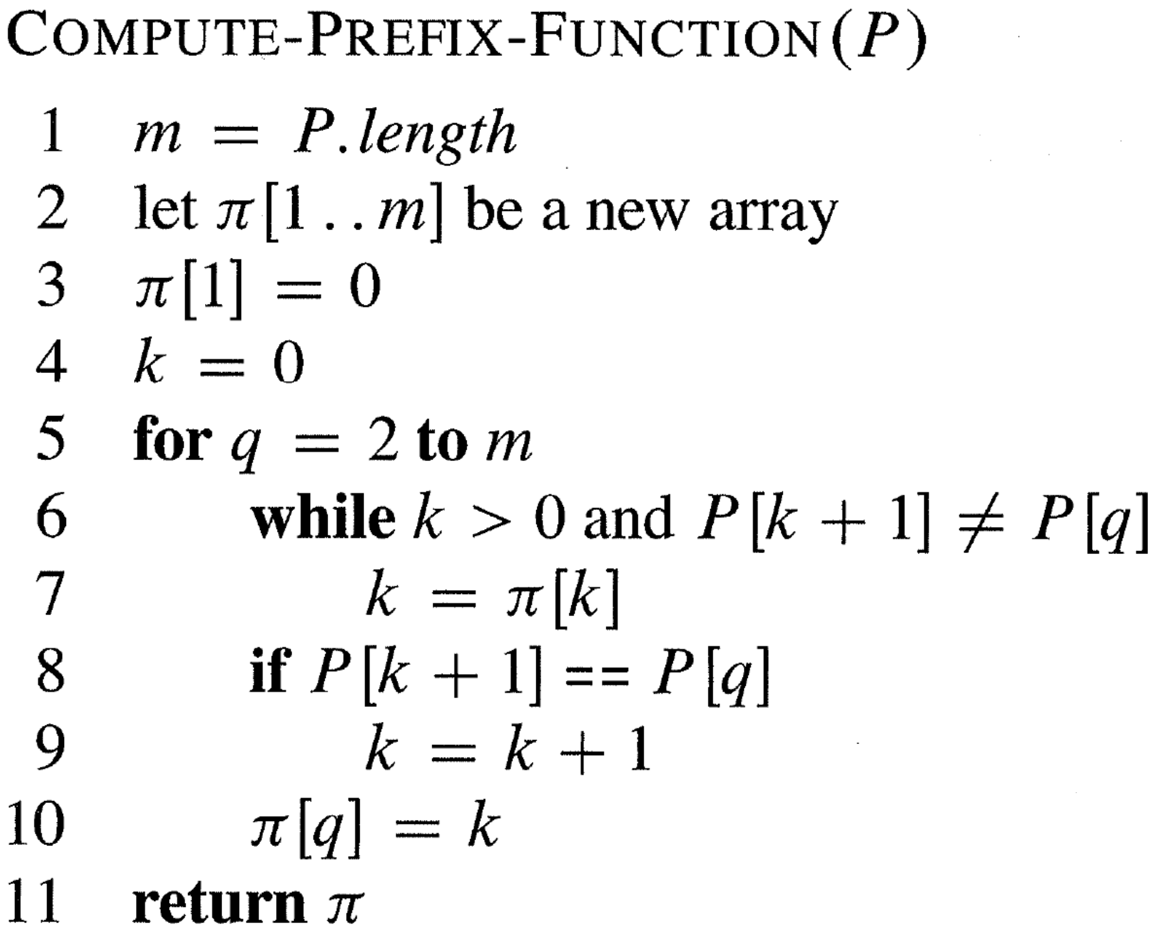


(a)

(b)

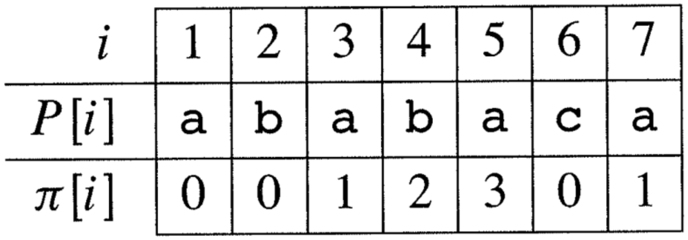
(c)

Knuth-Morris-Pratt:
Beregning af prefix funktionen

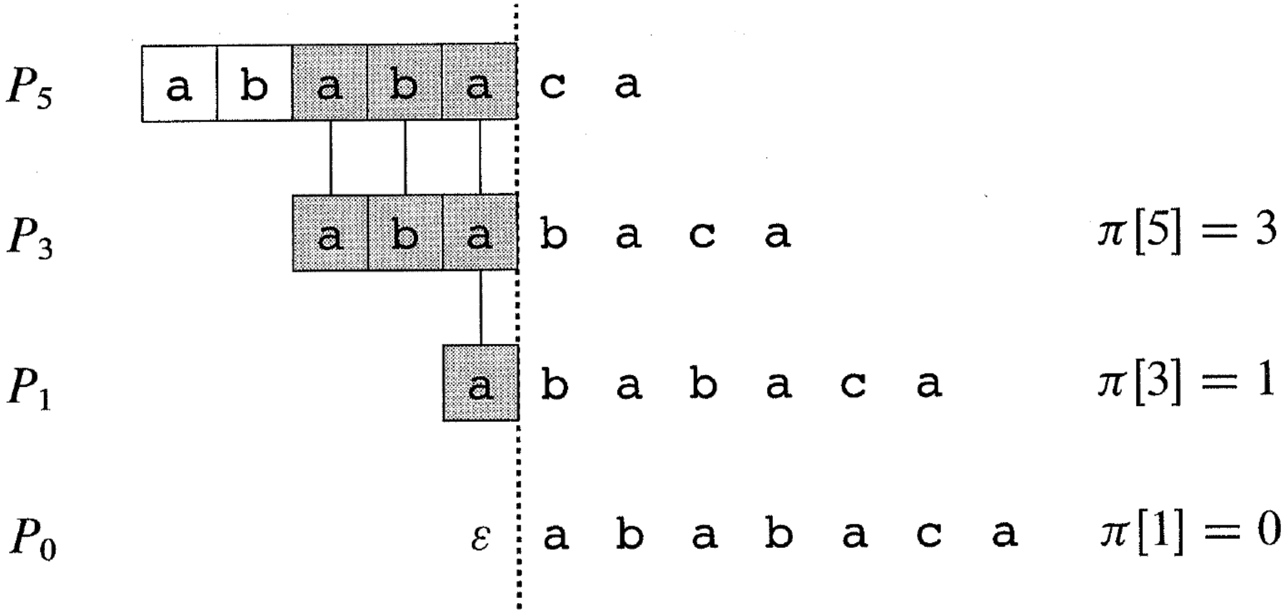


```
COMPUTE-PREFIX-FUNCTION($P$)
$m = P.length$
let $\pi[1 \mathinner{.\,.} m]$ be a new array
$\pi[1] = 0$
$k = 0$
for $q = 2$ to $m$
    while $k > 0$ and $P[k+1] \neq P[q]$
        $k = \pi[k]$
    if $P[k+1] == P[q]$
        $k = k + 1$
    $\pi[q] = k$
return $\pi$
```

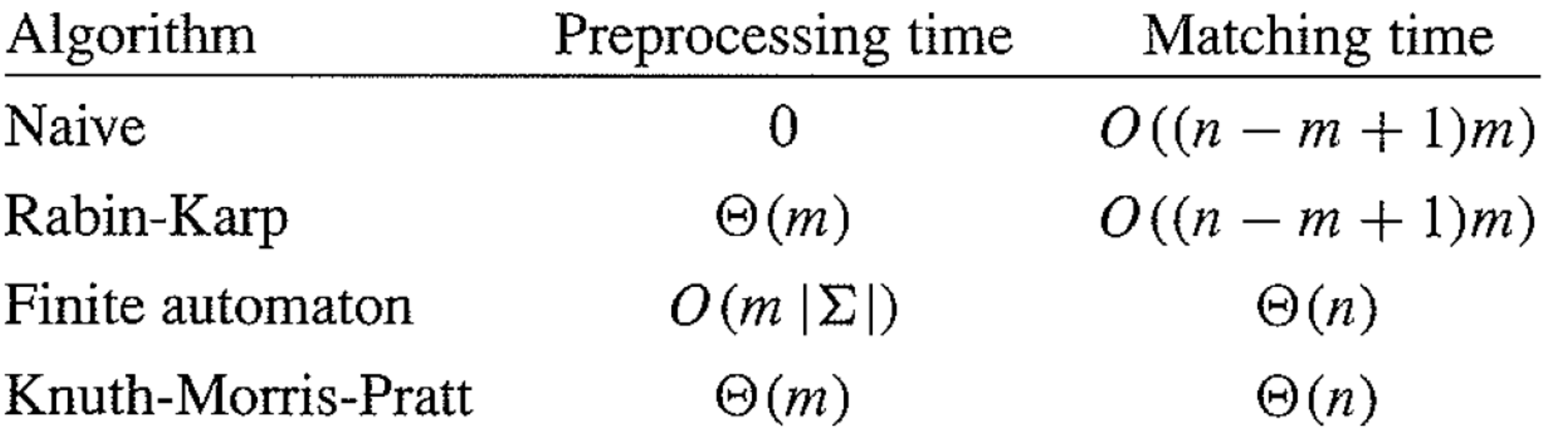
Knuth-Morris-Pratt: Beregning af prefix funktionen



i	1	2	3	4	5	6	7
$P[i]$	a	b	a	b	a	c	a
$\pi[i]$	0	0	1	2	3	0	1



Worst-case tider



Algorithm	Preprocessing time	Matching time	[CLRS]
Naive	0	$O((n-m+1)m)$	32.1
Rabin-Karp	$\Theta(m)$	$O((n-m+1)m)$	32.2
Finite automaton	$O(m\|\Sigma\|)$	$\Theta(n)$	(32.3)
Knuth-Morris-Pratt	$\Theta(m)$	$\Theta(n)$	32.4