

Brute Force

$T[1..n]$: ez lesz a szöveg

$P[1..m]$: ez lesz a keresendő szöveg

(X) ha eltérést találunk

(T) ha találatunk van

(V) ha kifutottunk a szövegből és túlindexeltünk

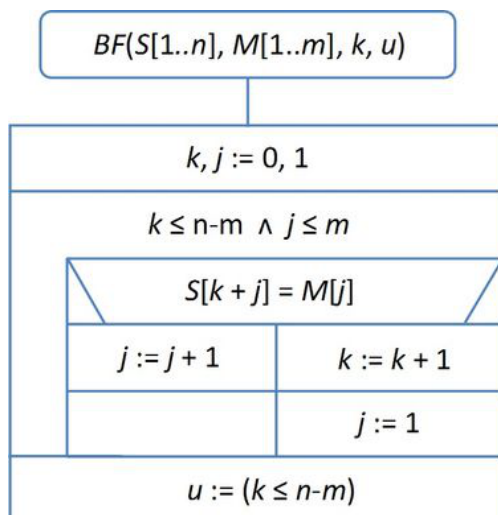
	1	2	3	4	5	6	7	8	9	10	11	
$T[1 \dots 11] =$	A	B	A	B	B	A	B	A	B	A	B	
$P[1 \dots 4] =$	B(X)	A	B	A								
		B	A	B	A(X)							
			B(X)	A	B	A						
				B	A(X)	B	A					
					B	A	B	A(T)				
						B(X)	A	B	A			
							B	A	B	A(T)		
								B(X)	A	B	A	
									B	A	B	A(V)

Ha n határozottan nagyobb mint m , akkor a műveletigények:

$$mT(n) \in \theta(n)$$

$$MT(n) \in \theta(n * m)$$

Stuktogram:



Quick-search

$T[1..n]$: ez lesz a szöveg

$P[1..m]$: ez lesz a keresendő szöveg

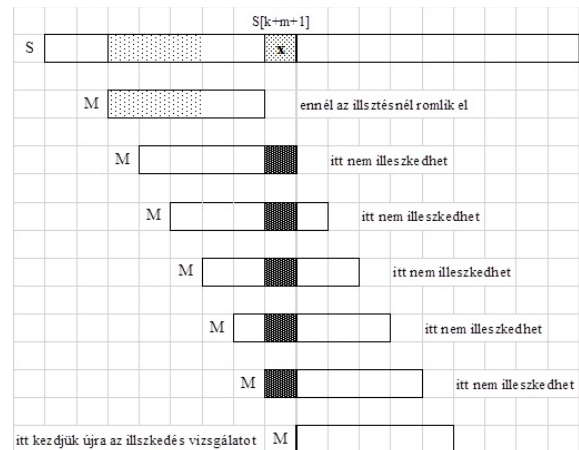
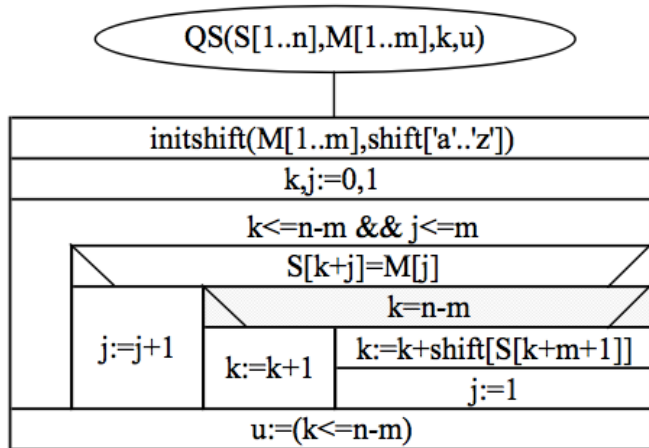
(X) ha eltérést találunk

(T) ha találatunk van

Az algoritmus két fontos lépésből áll.

1. Ha a szöveg, minta utáni első karaktere nem fordul elő a mintában, akkor átugorhatjuk teljesen, hiszen biztosan nem lesz egyezés, ameddig azt a katasztrófát vizsgáljuk.
2. Ha a szöveg, minta utáni első karaktere előfordul a mintában, akkor az előfordulások közül a jobbra lévőre ugrunk.

Stuktoqram:



Példa:

$T =$	A	D	A	B	A	D	C	A	D	A	B	C	A	B	A	D	A	C	A	D	A	D	A
$P =$	C_x	A	D	A																			
		C_x	A	D	A																		
				C_x	A	D	A																
					C_x	A	D	A															
$s = 6$							C	A	D	A_T													
												C	A	D_x	A								
														C_x	A	D	A						
$s = 17$																	C	A	D	A_T			
																		C_x	A	D	A		

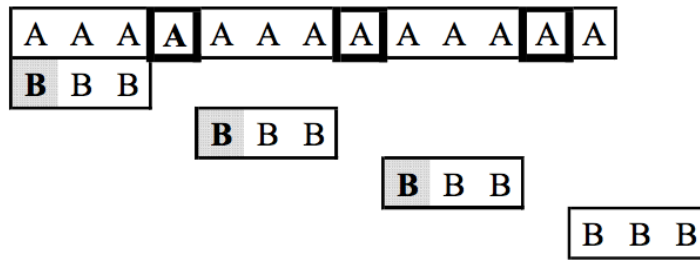
Műveletigény:

$$m\ddot{O}(n, m) = \Theta\left(\frac{n}{m+1}\right)$$

$$M\ddot{O}(n) = \Theta(n * m)$$

Legjobb eset: A minta olyan karakterekből áll, amelyek nem fordulnak elő a szövegben, így a minta első karakterénél már elromlik az illeszkedés, továbbá a minta utáni karakter sem fordul elő a mintában, így azt "átugorhatjuk".

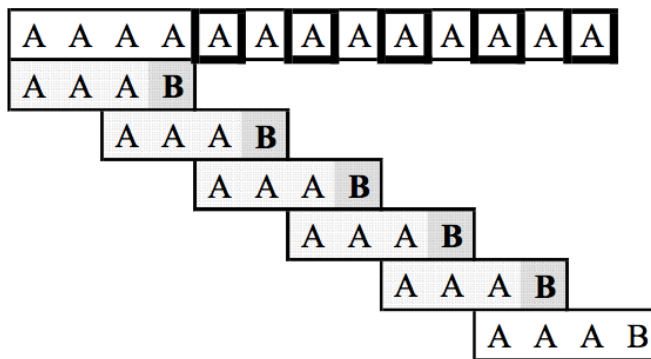
Pl:



$$m\ddot{O}(n, m) = \Theta\left(\frac{n}{m+1}\right)$$

Legrosszabb eset: A minta végén romlik el az illeszkedés és csak kicsiket tudunk "ugrani".

Pl:



$$M\ddot{O}(n) = \Theta(n * m)$$

Knuth-Morris-Pratt

Amennyiben az illeszkedés elromlik, akkor egy hibás kezdetünk van, de ez a kezdet ismert, mivel az elromlás előtti karakterig egyezett a mintával. Ezt az információt használjuk fel, hogy elkerüljük az állandó visszalépést a szövegben a minta kezdetére

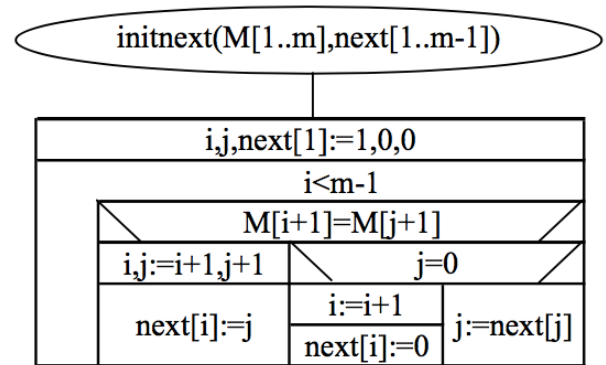
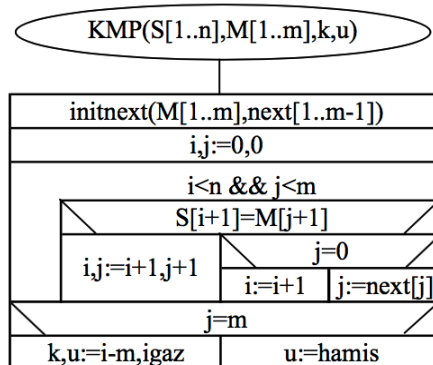
Látható, hogy a minta illeszkedő részének ($M[1..5]$) van egy olyan valódi kezdőszelete (valódi **prefixe**), amely egyezik ezen illeszkedő rész egy valódi végszeletével (valódi **szuffixével**), azaz $M[1..3] = M[3..5]$ ('ABA'='ABA').

A	B	A	B	A	B	A	C
A	B	A	B	A	C		
	A	B	A	B	A	C	

$P[1 \dots 8] =$	B	A	B	A	B	B	A	B									
$T[1 \dots 18] =$	A	B	A	B	A	B	A	B	B	A	B	A	B	A	B	B	A
	B_X																
		B	A	B	A	B	B_X										
$s = 3$				B	A	B	A	B	B	A	B_T						
									B	A	B	A	B	B_X			

$P =$	B	A	B	A	B	B	A	B
j	1	2	3	4	5	6	7	8
$next(j)$	0	0	1	2	3	1	2	3
			B	A	B	A		
				B	A	B		
					B	A		
						B	A	B

Stuktogram:



Műveletigény:

$initNext() : \Theta(m)$

$keresés: \Theta(n)$