



String search

Group 4



String Search (Single pattern matching)

Finding exact matches in a document/string that match the needle.

Haystack of length: n

Needle of length: m



Naive solution

2 nested for-loops

$O(mn)$

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

Overview when to choose

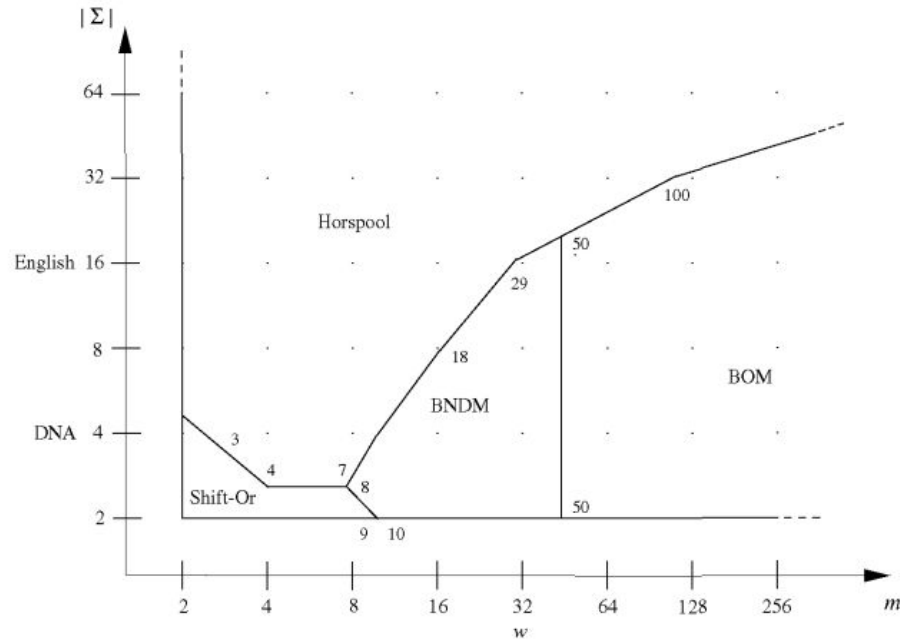


Fig. 2.22. Map of experimental efficiency for different string matching algorithms.



Horspool

- Is a simplification of Boyer Moore algorithm, performance is usually better but not always.
- Works by using the suffix of the needle to find the max jump distance in the haystack.
- Preprocess the needle (length of needle - index in needle - 1) override previous unless the last.
- Start left on haystack but compare string right to left



Pre Processing (COFFEE)

C	O	F	E	*



Pre Processing (COFFEE)

C	O	F	E	*
5				



Pre Processing (COFFEE)

C	O	F	E	*
5	4			



Pre Processing (COFFEE)

C	O	F	E	*
5	4	3		



Pre Processing (COFFEE)

C	O	F	E	*
5	4	2		



Pre Processing (COFFEE)

C	O	F	E	*
5	4	2	1	



Pre Processing (COFFEE)

C	O	F	E	*
5	4	2	1	



Pre Processing (COFFEE)

C	O	F	E	*
5	4	2	1	6



String Matching

C	O	F	E	*
5	4	2	1	6

THECOSTOFFCOFFEEISFAIR
COFFEE



String Matching

C	O	F	E	*
5	4	2	1	6

THECOSTOFFCOFFEEISFAIR
COFFEE



String Matching

C	O	F	E	*
5	4	2	1	6

THECOSTOFFCOFFEEISFAIR
COFFEE



String Matching

C	O	F	E	*
5	4	2	1	6

THECOSTOFFCOFFEEISFAIR
COFFEE



String Matching

C	O	F	E	*
5	4	2	1	6

THECOSTOFFCOFFEEISFAIR
COFFEE