T[i]

```
HOCUSPOCUSABRABRACADABRAY
          ABRACADABRAX
           ~~ABRACADABRAX~~
            ~~ABRACADABRAX~~
             ABRACADABRAX
                    ABRACADABRAX
                       ABRACADABRAX
                        ABRACADABRA
```

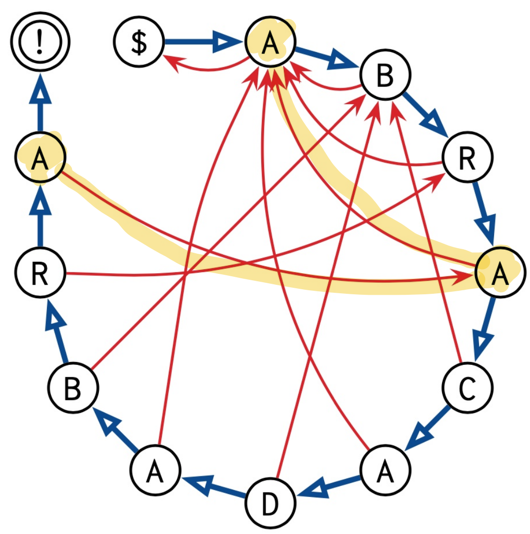
Once we've matched T[i] to P[j]

we don't need to explicitly compare T[i] to anything else

The next reasonable shift is the smallest value of s such that T[s..i-1] is a shorter prefix of P.

A B R A Y

$P[i]$	A	B	R	A	C	A	D	A	B	R	A
$fail[i]$	0	1	1	1	2	1	2	1	2	3	4



```
KnuthMorrisPratt(T[1..n], P[1..m]):
    fail[1..m] ← ComputeFailure(P)
    j ← 1
    for i ← 1 to n
        while j > 0 and T[i] ≠ P[j]
            j ← fail[j]
        if j = m        ⟨⟨Found it!⟩⟩
            return i − m + 1
        j ← j + 1
    return None
```

ABRACADABRA

j

ABRABRACADABRA

i

Running time $= O(n)$ $\qquad$ $\leq 2n$ comparisons

Two types of comparisons:

$T[i] = P[j] \to i{+}{+}\ j{+}{+}$ $\qquad$ $\leq n$ times

$T[i] \neq P[j] \to i$ same, $j \downarrow$ $\qquad$ $\leq n$ times

> $P[1..fail[j]-1]$ is the longest proper prefix of $P[1..j-1]$ that is also a suffix of $T[1..i-1]$.

> $P[1..fail[j]-1]$ is the longest proper prefix of $P[1..j-1]$ that is also a suffix of $P[1..j-1]$.

border of string is proper prefix that is also suffix

										j	
$P[i]$	A	B	R	A	C	A	D	A	B	R	A
$fail[i]$	0	1	1	1	2	1	2	~~1~~ 0	~~2~~ 1	~~3~~ 1	~~4~~ 0

$O(n^3)$ time brute force

$O(n^2)$ time DP

```
KnuthMorrisPratt(T[1..n], P[1..m]):
    fail[1..m] ← ComputeFailure(P)
    j ← 1
    for i ← 1 to n
        while j > 0 and T[i] ≠ P[j]
            j ← fail[j]
        if j = m        ⟨⟨Found it!⟩⟩
            return i − m + 1
        j ← j + 1
    return None
```

```
ComputeFailure(P[1..m]):
    j ← 0
    for i ← 1 to m
        fail[i] ← j          (*)
        while j > 0 and P[i] ≠ P[j]
            j ← fail[j]
        j ← j + 1
    return fail[1..m]
```

if P[j]=P[i]
fail[i]←fail[j]
else
fail[i]←j

ABACABADABACABA C ABACABADABACABA

A border of P is either
— the longest border of P or
— a border of a border of P

ABAABABADAABC

AAAAAAAAAAAAA

$$fail[i] = \begin{cases} 0 & \text{if } i = 0, \\ \max\limits_{c \geq 1} \left\{ \underline{fail^c[i-1]} + 1 \;\middle|\; P[i-1] = P[fail^c[i-1]] \right\} & \text{otherwise.} \end{cases}$$

KMP makes $O(n)$ comparisons (given fail)

<u>not</u> $O(1)$ at each position in T

<u>$O(m)$</u> at each value of i

Optimized fail: $O(\log m)$ failures at each i.

$j \leftarrow 0, i \leftarrow 1$	$\mathbf{\$}^j$	$\mathbf{A}^i$	B	R	A	C	A	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		**0**											…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	$\mathbf{A}^j$	$\mathbf{B}^i$	R	A	C	A	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	**1**										…
$j \leftarrow \mathit{fail}[j]$	$\mathbf{\$}^j$	A	$\mathbf{B}^i$	R	A	C	A	D	A	B	R	X	…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	$\mathbf{A}^j$	B	$\mathbf{R}^i$	A	C	A	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	**1**									…
$j \leftarrow \mathit{fail}[j]$	$\mathbf{\$}^j$	A	B	$\mathbf{R}^i$	A	C	A	D	A	B	R	X	…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	$\mathbf{A}^j$	B	R	$\mathbf{A}^i$	C	A	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	**1**								…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	A	$\mathbf{B}^j$	R	A	$\mathbf{C}^i$	A	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	**2**							…
$j \leftarrow \mathit{fail}[j]$	$\$$	$\mathbf{A}^j$	B	R	A	$\mathbf{C}^i$	A	D	A	B	R	X	…
$j \leftarrow \mathit{fail}[j]$	$\mathbf{\$}^j$	A	B	R	A	$\mathbf{C}^i$	A	D	A	B	R	X	…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	$\mathbf{A}^j$	B	R	A	C	$\mathbf{A}^i$	D	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	**1**						…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	A	$\mathbf{B}^j$	R	A	C	A	$\mathbf{D}^i$	A	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	1	**2**					…
$j \leftarrow \mathit{fail}[j]$	$\$$	$\mathbf{A}^j$	B	R	A	C	A	$\mathbf{D}^i$	A	B	R	X	…
$j \leftarrow \mathit{fail}[j]$	$\mathbf{\$}^j$	A	B	R	A	C	A	$\mathbf{D}^i$	A	B	R	X	…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	$\mathbf{A}^j$	B	R	A	C	A	D	$\mathbf{A}^i$	B	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	1	2	**1**				…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	A	$\mathbf{B}^j$	R	A	C	A	D	A	$\mathbf{B}^i$	R	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	1	2	1	**2**			…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	A	B	$\mathbf{R}^j$	A	C	A	D	A	B	$\mathbf{R}^i$	X	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	1	2	1	2	**3**		…
$j \leftarrow j+1, i \leftarrow i+1$	$\$$	A	B	R	$\mathbf{A}^j$	C	A	D	A	B	R	$\mathbf{X}^i$	…
$\mathit{fail}[i] \leftarrow j$		0	1	1	1	2	1	2	1	2	3	**4**	…
$j \leftarrow \mathit{fail}[j]$	$\$$	$\mathbf{A}^j$	B	R	A	C	A	D	A	B	R	$\mathbf{X}^i$	…
$j \leftarrow \mathit{fail}[j]$	$\mathbf{\$}^j$	A	B	R	A	C	A	D	A	B	R	$\mathbf{X}^i$	…