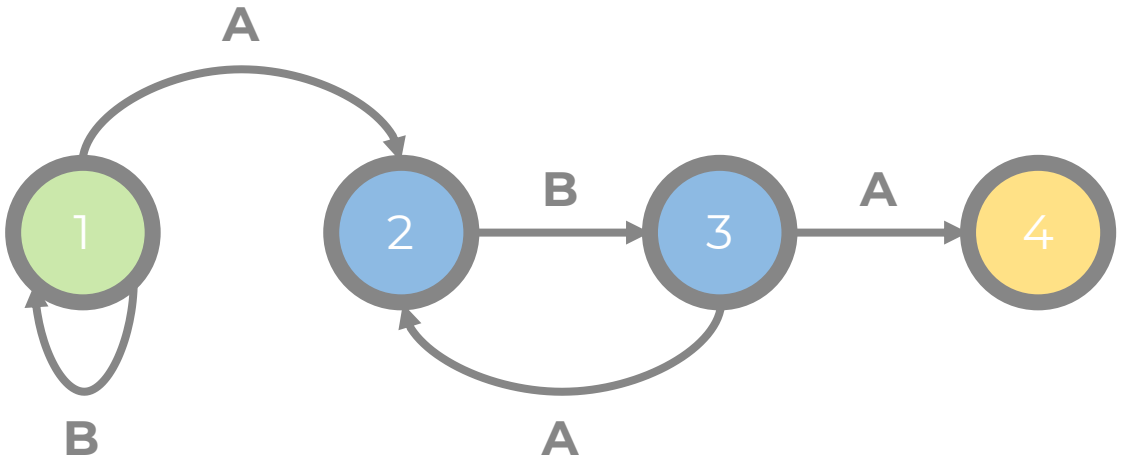


Ищем паттерны в строках по-новому!

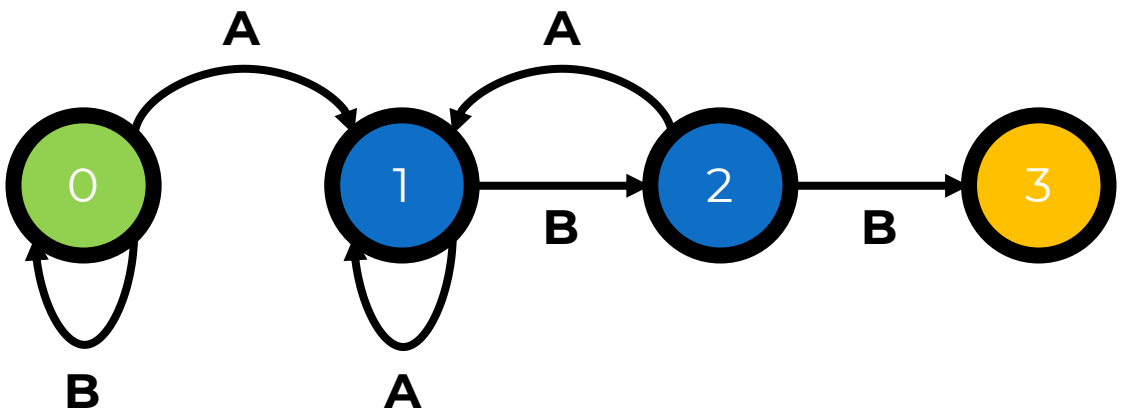
Конечные автоматы

Что у нас сегодня?

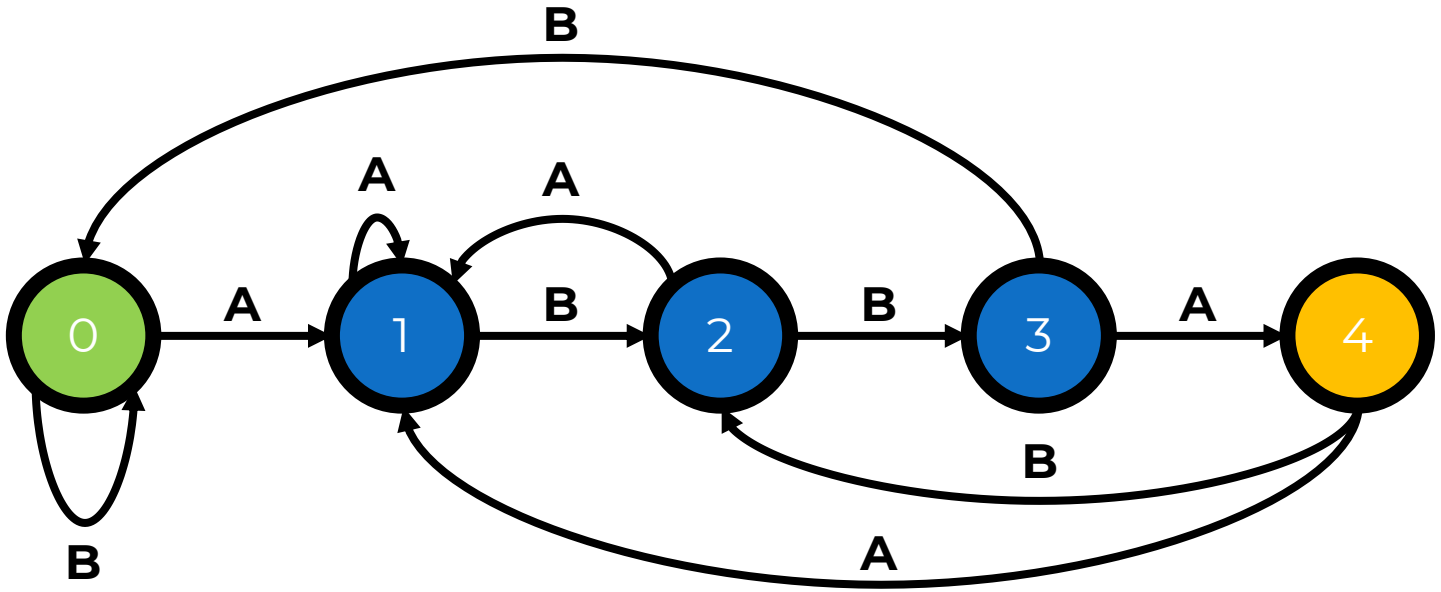
1. Конечный автомат
2. Функция переходов
3. Недетерминированный автомат
4. Построение рег. выражения

Конечный автомат

Конечный автомат – это абстрактное дискретное устройство, имеющее один вход и один выход и в каждый момент находящееся в одном состоянии из конечного набора возможных.



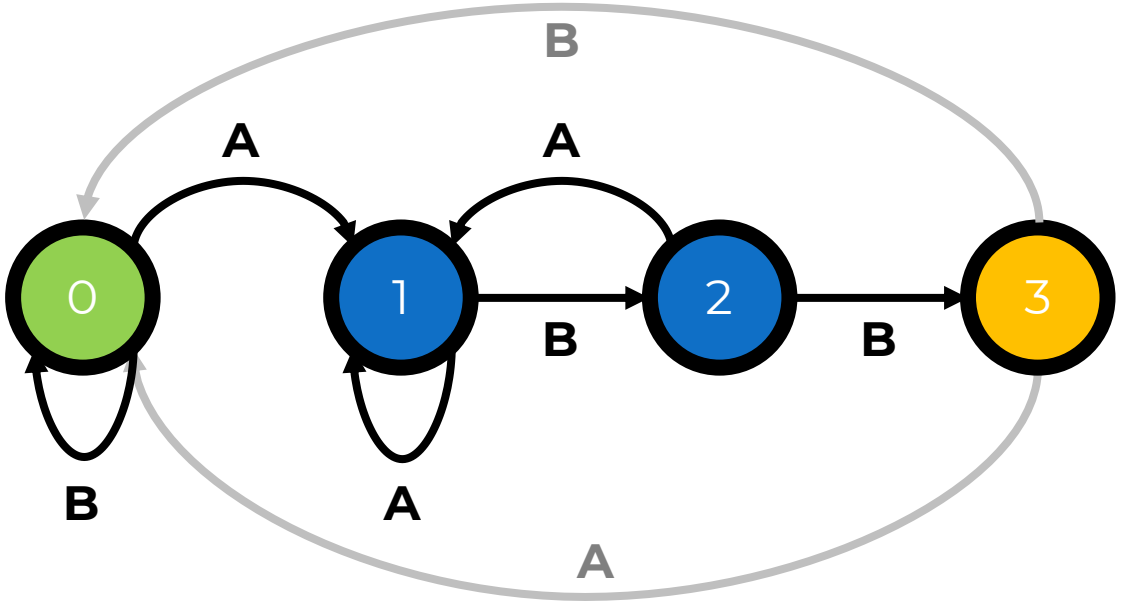
Давайте попробуем построить автомат для паттерна ABBA



Чем задается конечный автомат?

Таблицей переходов

state	A	B
0	1	0
1	1	2
2	1	3
3	1	0



За какое время мы можем наивно вычислить таблицу переходов?

Э...

$O(M^3|\Sigma|)$

???

Сколько мы ищем автоматом паттерн в тексте длины N?

O(N)

Какова асимптотика поиска паттернов в тексте конечным автоматом?

В общем случае?

$O(M^3|\Sigma|) + O(N)$

В лучшем случае?

$O(M^3|\Sigma|) + O(N)$

В худшем случае?

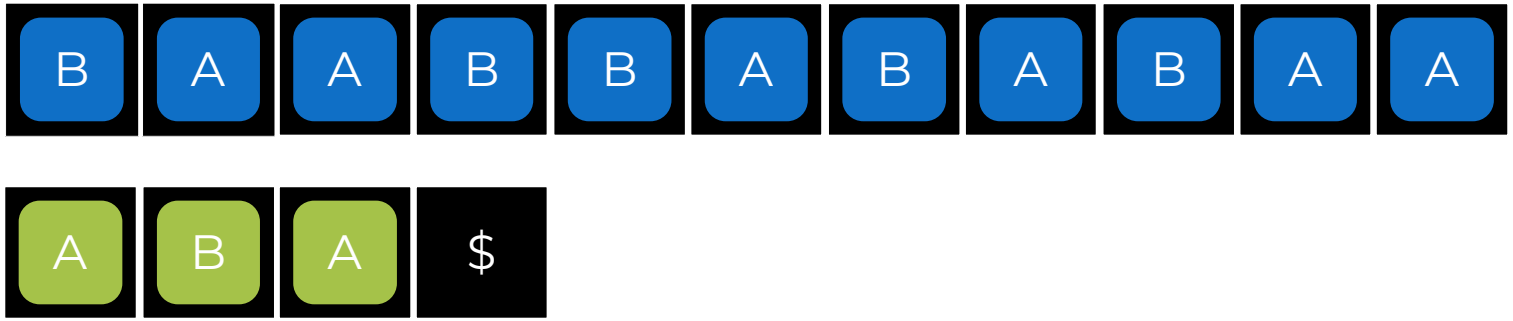
$O(M^3|\Sigma|) + O(N)$

Кто фаворит?

КМП

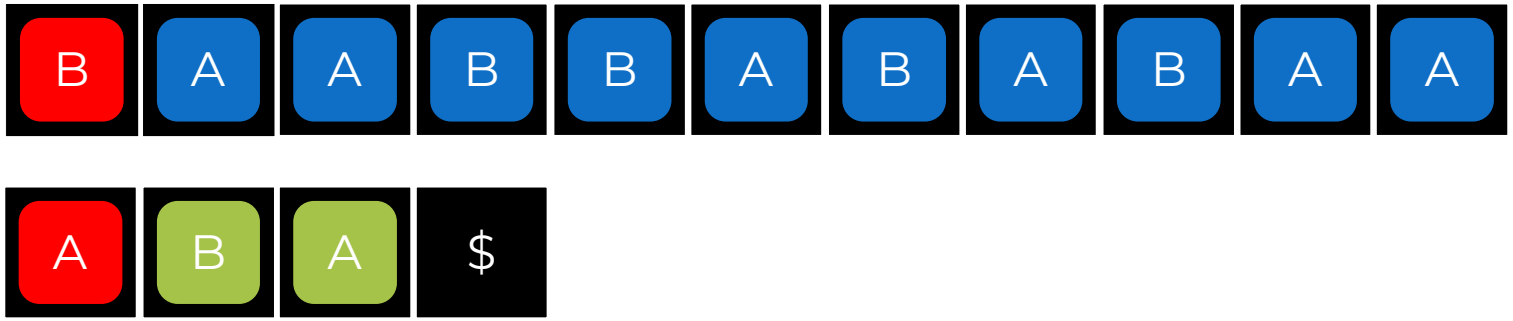
Улучшаем алгоритм КМП

Будем кэшировать информацию о том, куда переходить каждый раз



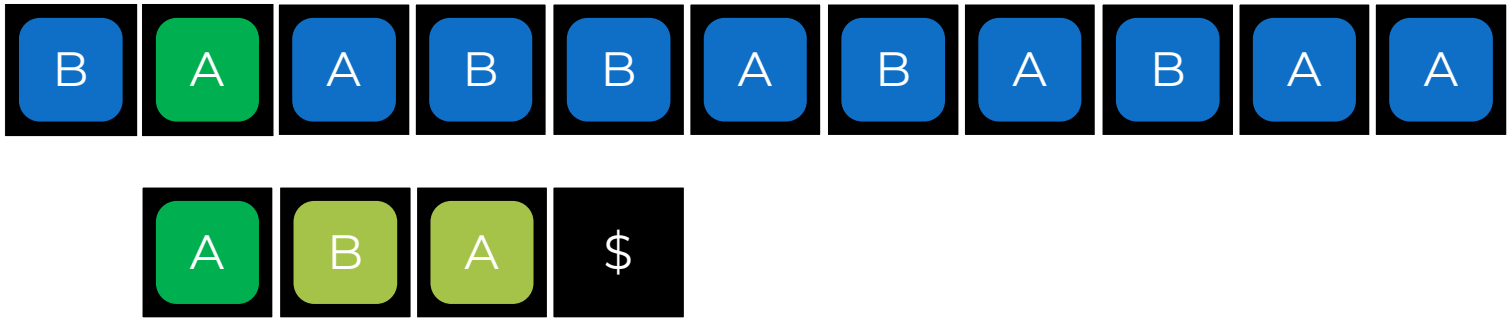
ind	0	1	2	3
π	0	0	1	
A				
B				

Улучшаем алгоритм КМП



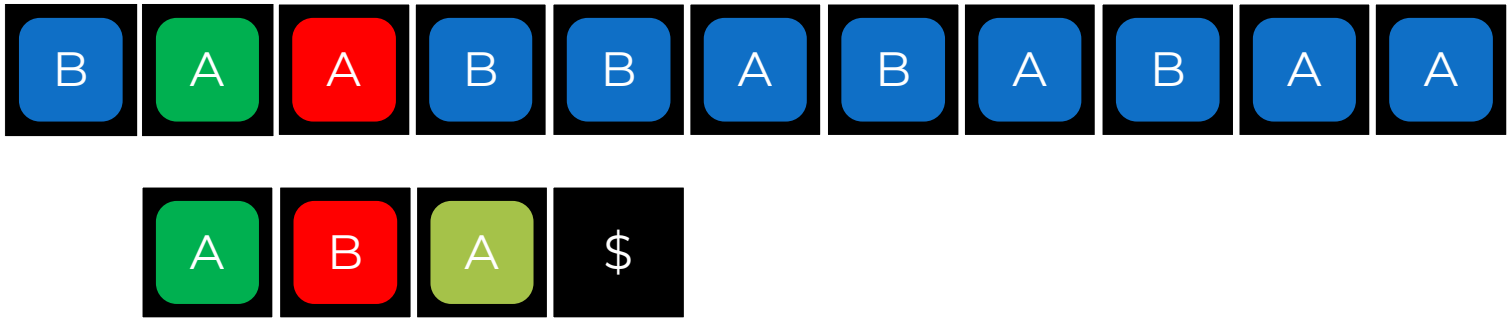
ind	0	1	2	3
π	0	0	1	
A				
B	0			

Улучшаем алгоритм КМП



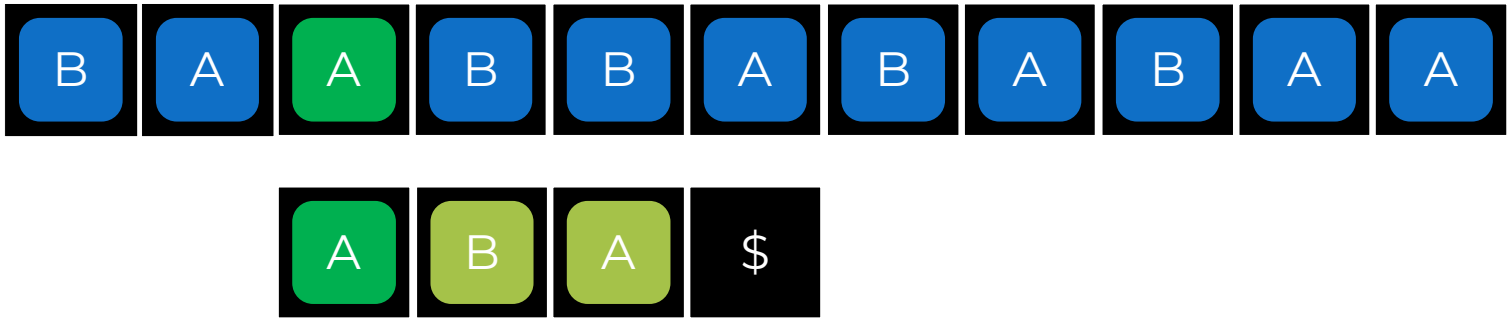
ind	0	1	2	3
π	0	0	1	
A	1			
B	0			

Улучшаем алгоритм КМП



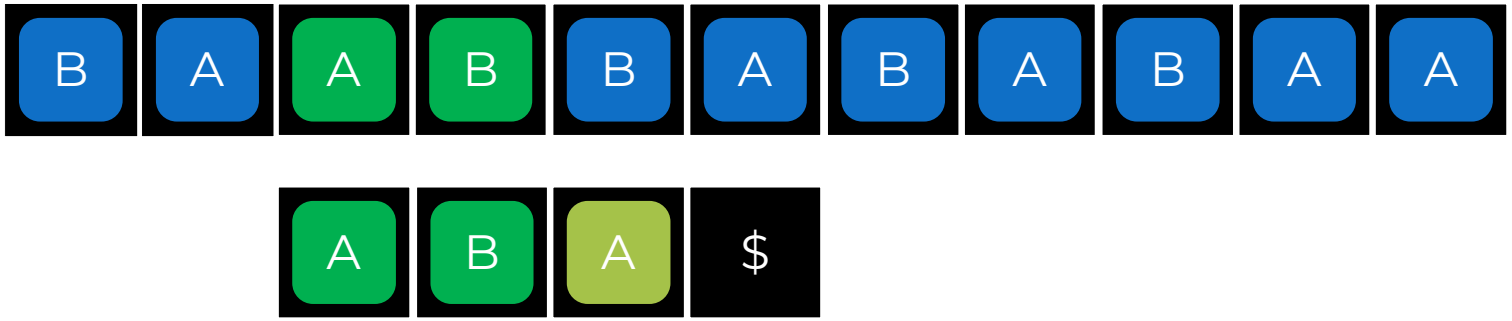
ind	0	1	2	3
π	0	0	1	
A	1			
B	0			

Улучшаем алгоритм КМП



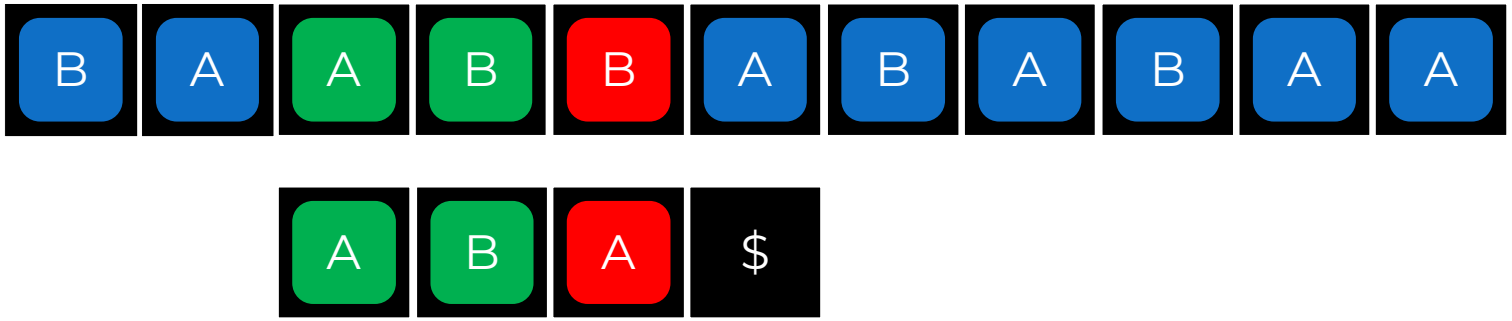
ind	0	1	2	3
π	0	0	1	
A	1	1		
B	0			

Улучшаем алгоритм КМП



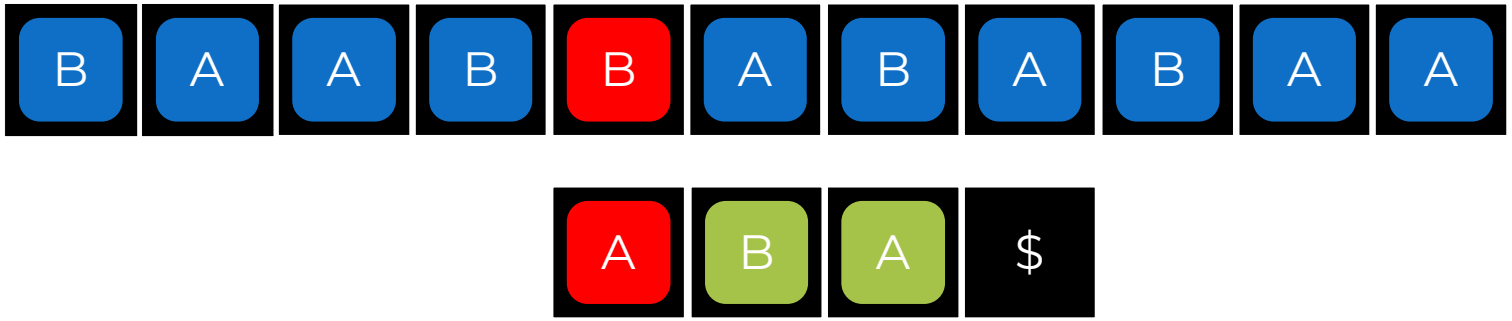
ind	0	1	2	3
π	0	0	1	
A	1	1		
B	0	2		

Улучшаем алгоритм КМП



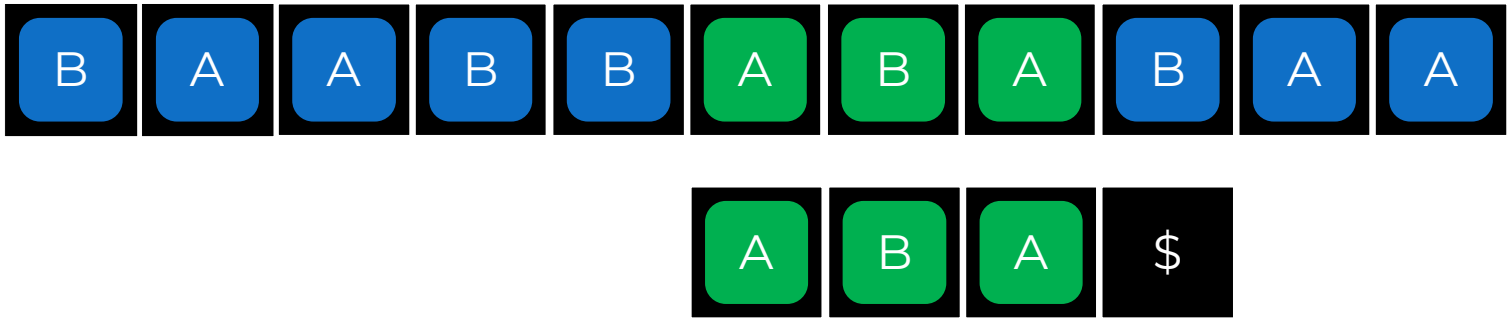
ind	0	1	2	3
π	0	0	1	
A	1	1		
B	0	2		

Улучшаем алгоритм КМП



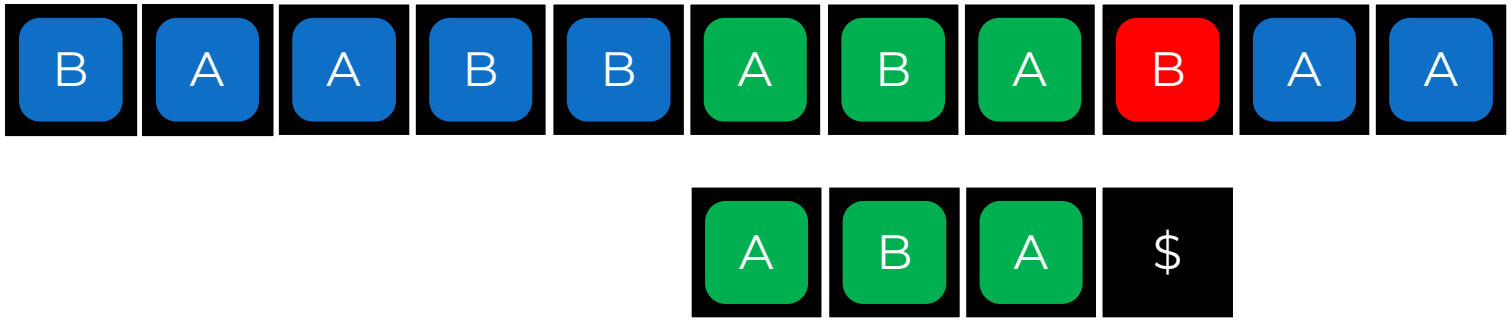
ind	0	1	2	3
π	0	0	1	
A	1	1		
B	0	2	0	

Улучшаем алгоритм КМП



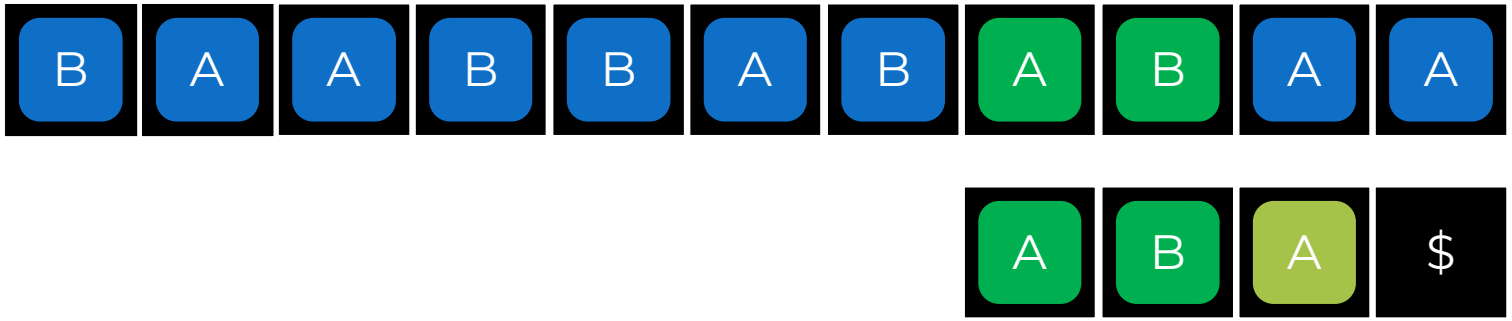
ind	0	1	2	3
π	0	0	1	
A	1	1	3	
B	0	2	0	

Улучшаем алгоритм КМП



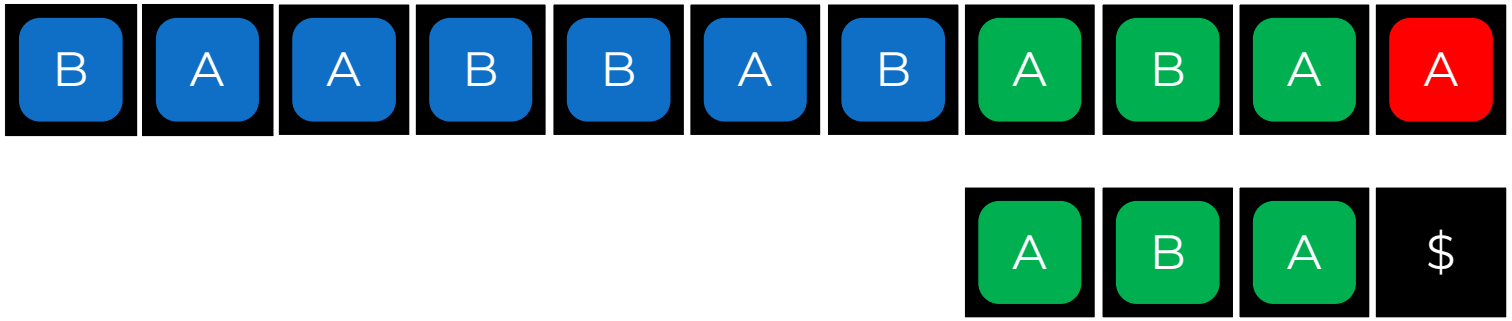
ind	0	1	2	3
π	0	0	1	
A	1	1	3	
B	0	2	0	

Улучшаем алгоритм КМП



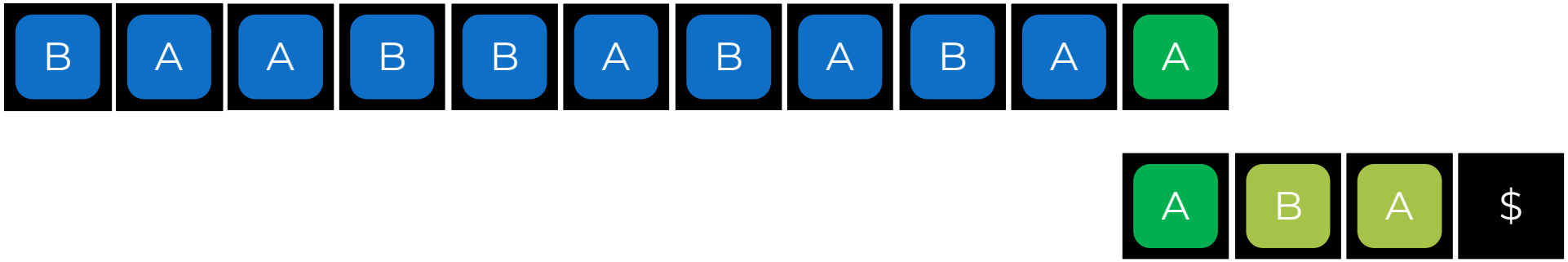
ind	0	1	2	3
π	0	0	1	
A	1	1	3	
B	0	2	0	2

Улучшаем алгоритм КМП



ind	0	1	2	3
π	0	0	1	
A	1	1	3	
B	0	2	0	2

Улучшаем алгоритм КМП



ind	0	1	2	3
π	0	0	1	
A	1	1	3	1
B	0	2	0	2

Какая стала асимптотика?

Все еще
O(N)

Мы улучшили КМП на константу. Что мы посчитали в процессе?

Таблицу переходов

X	**0**	**1**	**2**	**3**
π	0	0	1	
A	1	1	3	1
B	0	2	0	2

Формальная функция переходов

1. Считаем π-функцию для паттерна;

2. Считаем функцию переходов для состояния 0;

3. $\sigma(\text{state}, \text{symbol}) = \begin{cases} \text{state} + 1 & \text{если } s[\text{state}] = \text{symbol} \\ \sigma(\pi[\text{state} - 1], \text{symbol}) & \text{иначе} \end{cases}$

Какая стала асимптотика подсчета функции переходов?

Почти. **O(M|Σ|)**

O(M)?

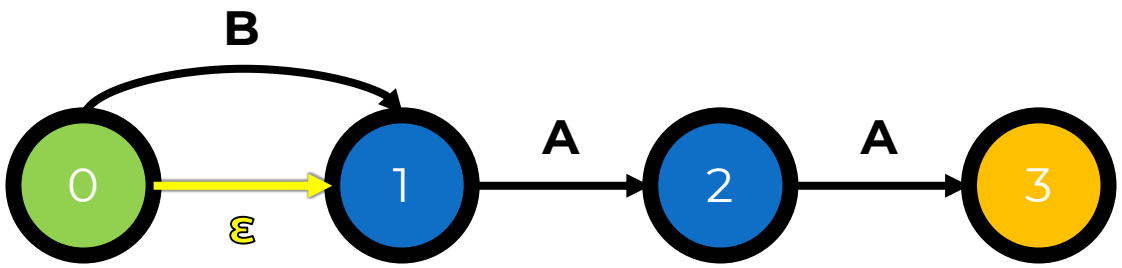
Детерминированность конечного автомата

Конечные автоматы подразделяются на **детерминированные** и **недетерминированные**.

Детерминированным называется такой конечный автомат, в котором из любого состояния по любому символу возможен переход не более чем в одно состояние.

Недетерминированный автомат

Недетерминированный автомат может содержать эпсилон-переходы (переход по пустой строке) и может переходить в несколько разных состояний по одним входным данным.



Для чего можно использовать недетерминированные автоматы?

Для построения регулярных выражений

Построение регулярного выражения

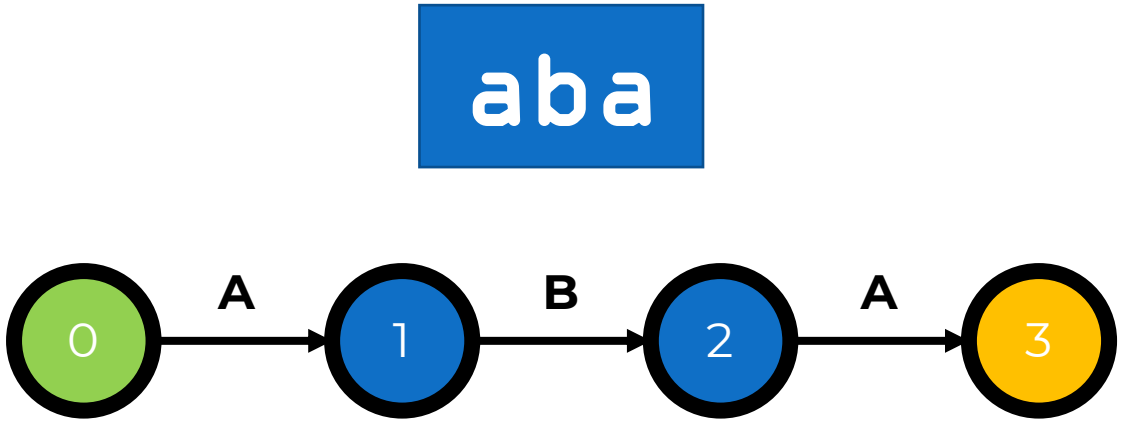
Для того чтобы построить недетерминированный конечный автомат для регулярного выражения:

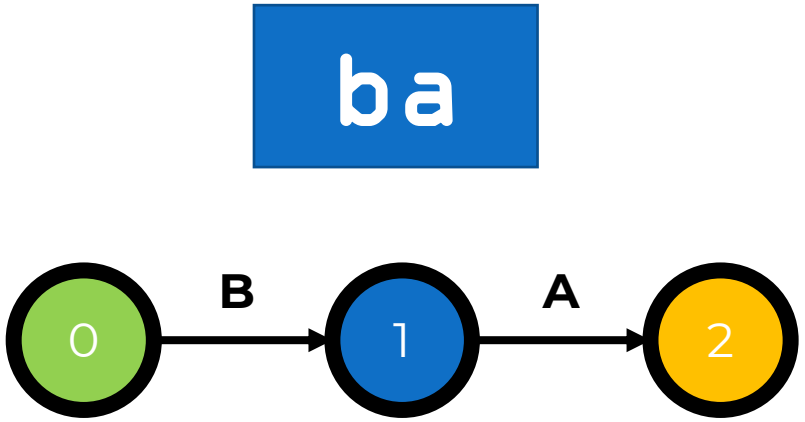
`(aba)+(ba)*(bc)?b`

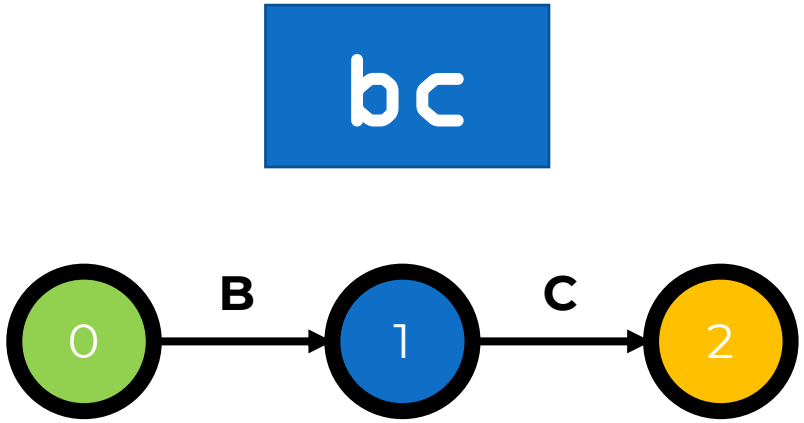
1. Построить отдельные детерминированные автоматы* для частей регулярного выражения:

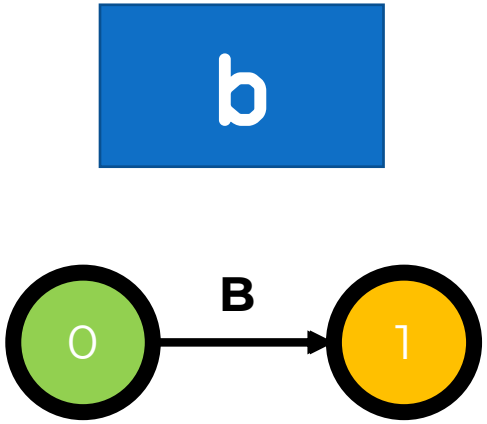
`aba` `ba` `bc` `b`

*считаем, что все переходы кроме основных ведут к смерти состояния


aba
0
A
1
B
2
A
3


ba
0
B
1
A
2


bc
0
B
1
C
2


b
0
B
1

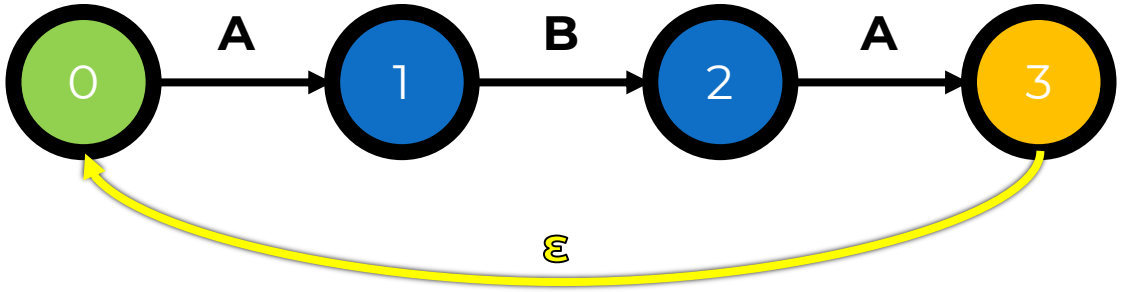
Построение регулярного выражения

2. Добавить к каждой части соответствующие модификаторы

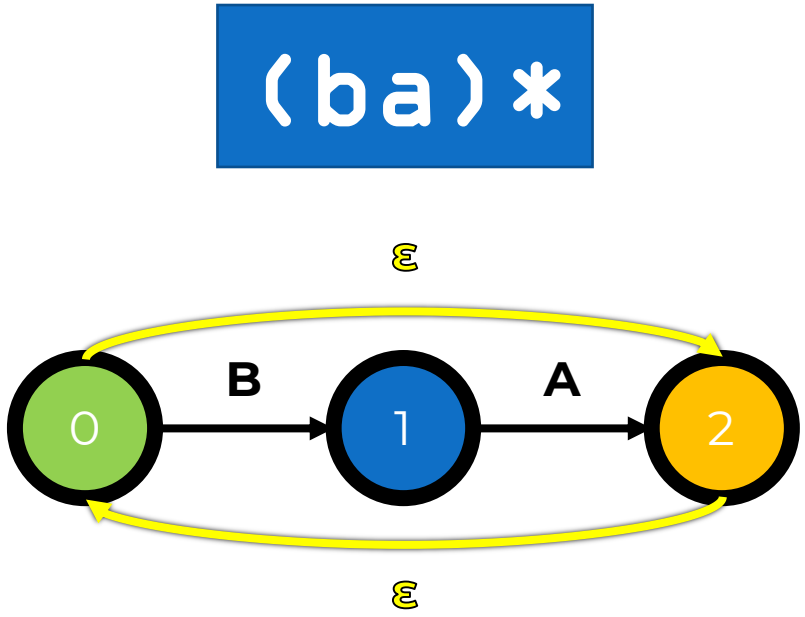
```
(aba)+(ba)*(bc)?b
```

- Для **(...)+** добавить ε-переход из конечного состояния в стартовое:

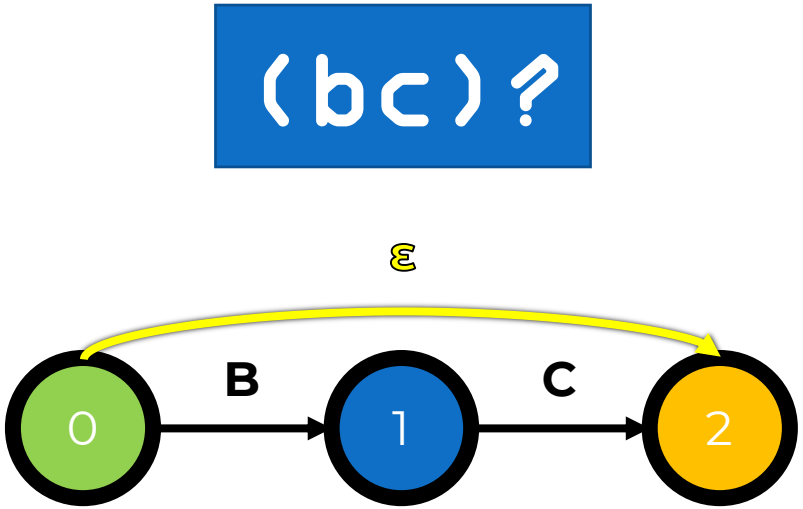
(aba)+



- Для **(...)*** добавить два ε-перехода из стартового состояния в конечное и обратно:



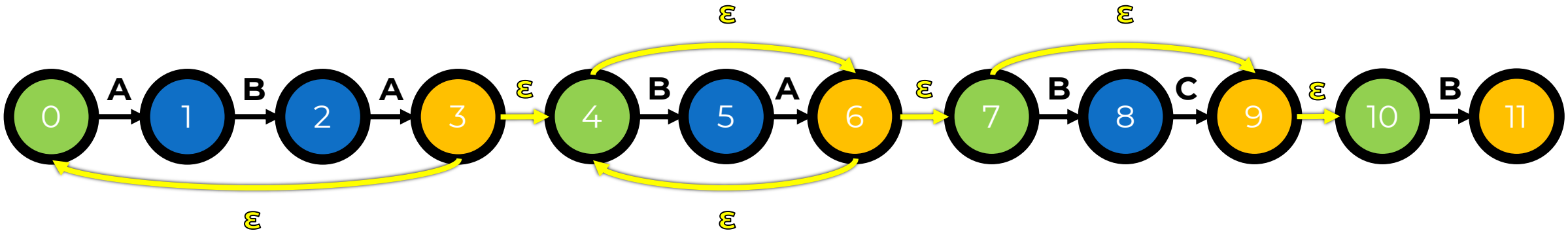
- Для **(...)?** добавить ε-переход из стартового состояния в конечное:



Построение регулярного выражения

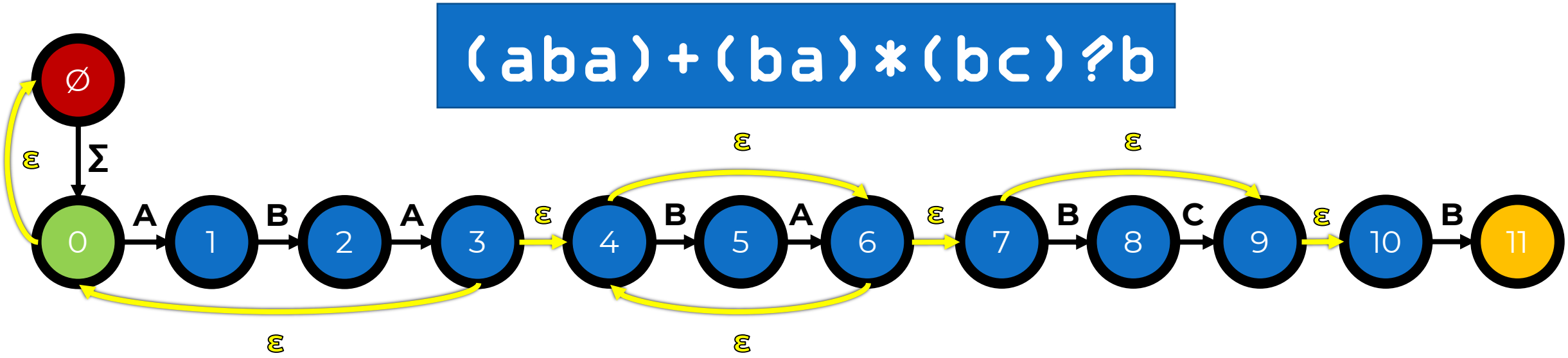
3. Соединяем автоматы. Начальное состояние остается начальным только у первой части, конечное остается конечным только у последней.

```
(aba)+(ba)*(bc)?b
```



Построение регулярного выражения

4. Если хотим искать паттерн во всем тексте – добавить специальное состояние Ø с переходом в стартовое состояние по любому символу.



Асимптотика недетерминированного конечного автомата

За какое время можно проверить, что в последовательности есть паттерн?

$O(M) \cdot N =$
$= O(MN)$

А если нужно найти все вхождения паттерна?

$O(MN\alpha^N)$

Спасибо за внимание!