

Тренажёр «Нормальные алгоритмы Маркова»

<https://kpolyakov.spb.ru/prog/nma.htm>

Нормальные алгоритмы Маркова.

Нормальные алгоритмы Маркова (НАМ) — это строгая математическая форма записи алгоритмов обработки символьных строк, которую можно использовать для доказательства разрешимости или неразрешимости различных задач.

Эти алгоритмы представляют собой некоторые правила по переработке слов в каком-либо алфавите.

При этом исходные данные и результат работы алгоритма являются словами в этом алфавите.

Марков предположил, что любой алгоритм можно записать как НАМ. В отличие от машин Тьюринга НАМ — это "чистый" алгоритм, который не связан ни с каким "аппаратным обеспечением" (лентой, кареткой и т.п.). НАМ преобразует одно слово (цепочку символов некоторого алфавита) в другое и задается алфавитом и системой подстановок.

Для удобства рассуждений допускается пустое слово, которые обозначим Λ . Слова будем обозначать буквами P, Q, R и с индексами.

Формулой подстановки называется запись вида $\alpha \rightarrow \beta$ (читается « α заменить на β »), где α и β – любые слова (возможно, и пустые).

При этом α называется левой частью формулы, а β – правой частью.

Сама **подстановка (как действие) задается формулой подстановки и** применяется к некоторому слову P .

Суть операции сводится к тому, что в слове P отыскивается часть, совпадающая с левой частью этой формулы (т.е. с α), и она заменяется на правую часть формулы (т.е. на β). При этом остальные части слова P (слева и справа от α) не меняются. Получившееся слово R называют **результатом подстановки**.

Условно это можно изобразить так:

P

x	α	y
-----	----------	-----

 \rightarrow R

x	β	y
-----	---------	-----

Правила выполнения НАМ.

Прежде всего, задается некоторое **входное слово** P .

Работа НАМ сводится к выполнению последовательности шагов. На каждом шаге входящие в НАМ формулы **подстановки просматриваются сверху вниз** и выбирается первая из формул, применимых к входному слову P , т.е. самая верхняя из тех, левая часть которых входит в P . Далее выполняется подстановка согласно найденной формуле. Получается новое слово P' .

На следующем шаге это слово P' берется за исходное и к нему применяется та же самая процедура, т.е. **формулы снова просматриваются сверху вниз начиная с самой верхней** и ищется первая формула, применимая к слову P' , после чего выполняется соответствующая подстановка и получается новое слово P'' . И так далее: $P \rightarrow P' \rightarrow P'' \rightarrow \dots$

Следует обратить особое внимание на тот факт, что на каждом шаге формулы в НАМ всегда просматриваются начиная с самой первой.

Необходимые уточнения:

1. Если на очередном шаге была применена обычная формула $(\alpha \rightarrow \beta)$, то работа НАМ продолжается.
2. Если же на очередном шаге была применена заключительная формула $(\alpha \dashv \rightarrow \beta)$, то после её применения работа НАМ прекращается. То слово, которое получилось в этот момент, и есть **выходное слово, т.е. результат применения** НАМ к входному слову.

Пусть дан алфавит $A = \{a_1, a_2, \dots, a_n\}$,
слово $P = a_{i_1} a_{i_2} \dots a_{i_m}$ и слово $Q = a_{j_1} a_{j_2} \dots a_{j_k}$

Под **объединением** слов PQ будем понимать слово:

$$PQ = a_{i_1} a_{i_2} \dots a_{i_m} a_{j_1} a_{j_2} \dots a_{j_k}$$

В частности: $P\Lambda = \Lambda P = P$

Кроме этого: $(P_1 P_2) P_3 = P_1 (P_2 P_3)$

Слово P является **подсловом** слова Q , если слово P является составной частью слова Q , т.е. существуют такие (возможно пустые) слова R_1 и R_2 , что $Q = R_1 P R_2$.

Марковской подстановкой (P,Q) называется следующая операция над словами: в заданном слове R находят первое вхождение слова P и, не изменяя остальных частей слова R , заменяют в нем это вхождение словом Q .

Замечание:

1) Полученное слово называется **результатом** применения марковской подстановки (P,Q) к слову R

2) Если первого вхождения слова P в слово R нет (и, следовательно, вообще нет ни одного вхождения P в R), то считается что марковская подстановка (P,Q) **не применима** к слову R

Для обозначения марковской подстановки (P,Q) используют запись $P \rightarrow Q$

Эту запись называют **формулой подстановки** (P,Q)

Различают **простые подстановки** $P \rightarrow Q$ и **заключительные подстановки** $P \mapsto Q$

Пример

Данное слово: **521421**

Подстановка: **21 \rightarrow 3**

Результат подстановки **5343**

Пример

Данное слово: **521421**

Подстановка: **25 \rightarrow 7**

Результат подстановки: **не применима**

Пример 1. Пусть алфавит $НАМ$ — это русские буквы и задана система подстановок:

$a \rightarrow н$

$ух \rightarrow ло$

$м \rightarrow с$

Применим эту систему подстановок к начальному слову «муха». Подстановки нужно просматривать по порядку, начиная с первой. Первая подстановка означает: «если в слове есть буквы “а”, заменить первую букву “а” на букву “н”». В слове «муха» есть буква «а», поэтому заменяем её на «н». Получается «мухн».

Начинаем просмотр подстановок сначала. Букв «а» больше нет, поэтому переходим ко второй подстановке. Сочетание «ух» есть в слове «мухн», поэтому вторая подстановка срабатывает, и мы заменяем «ух» на «ло»: получается «млн».

Теперь ни первая, ни вторая подстановки не применимы, а использование третьей даёт в результате слово «слон». Больше ни одну подстановку сделать нельзя, и НАМ заканчивает работу. Таким образом, приведённая система подстановок преобразует слово «муха» в слово «слон».

Пример 2. Построим НАМ для следующей задачи: удалить из строки, состоящей из букв «а» и «b», первый символ. Например, строка «abba» должна быть преобразована в «bba». Казалось бы, здесь нужно использовать систему подстановок: $a \rightarrow .$ $b \rightarrow .$

Однако такой НАМ будет неправильно работать для слов, начинающихся с буквы «b», например для слова «bba», в котором будет удалена последняя буква, потому что первая подстановка выполнится раньше, чем вторая. Перестановка двух строк также не даёт решения — теперь алгоритм неправильно работает для слов, начинающихся с буквы «а». Чтобы решить эту задачу, в алфавит НАМ добавляют еще один специальный символ, например символ «*». Этим символом помечают начало слова, используя подстановку.

$\rightarrow *$

Полный алгоритм выглядит так:

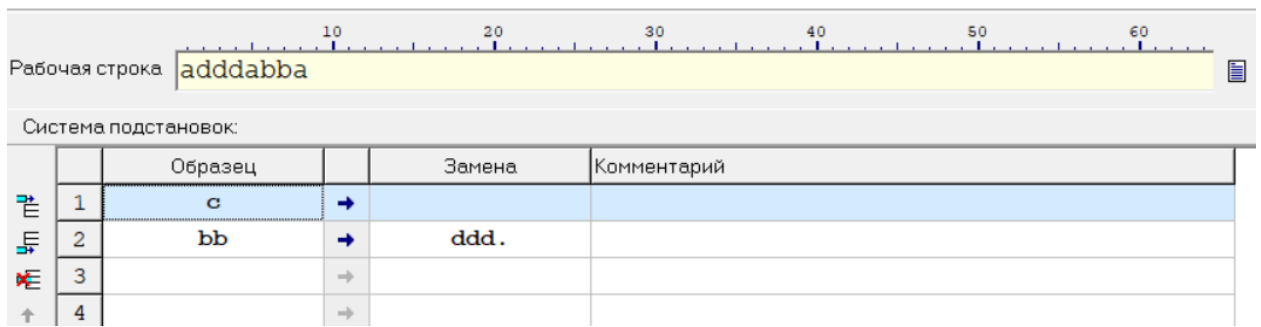
$*a \rightarrow .$ (точка – конец программы)

$*b \rightarrow .$

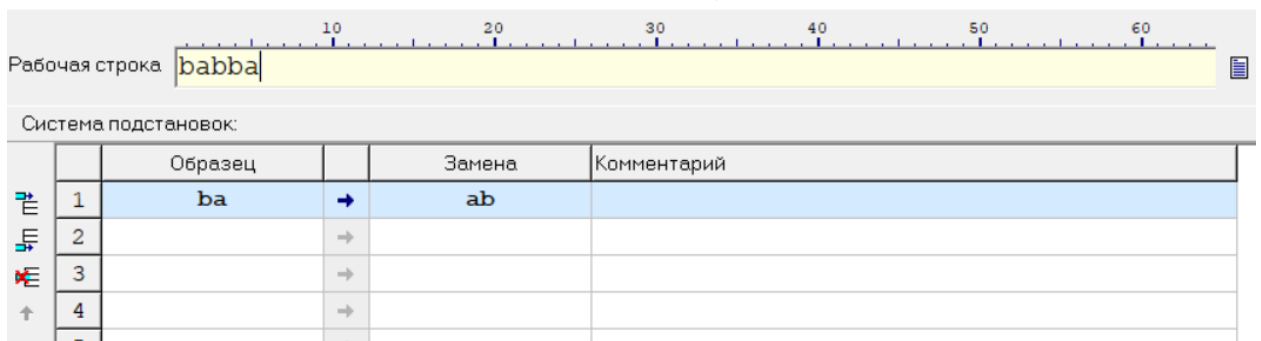
$\rightarrow *$

Сначала срабатывает третья подстановка (ставим «*» в начало строки), затем, в зависимости от первой буквы исходного слова, работает первая или вторая подстановка, и алгоритм заканчивает работу. Дополнительный символ похож на маркер в текстовом редакторе — он отмечает место в тексте, с которым потом будут выполняться какие-то действия.

Пример 3. (удаление и вставка символов). $A = \{a, b, c, d\}$. В слове P требуется удалить все вхождения символа c , а затем заменить первое вхождение подслова bb на ddd .



Пример 4. (перестановка символов) $A=\{a,b\}$. Преобразовать слово P так, чтобы в его начале оказались все символы a , а в конце – все символы b .



Пока в слове P справа хотя бы от одного символа b есть символ a , эта формула будет переносить a налево от этого b . Формула перестает работать, когда справа от b нет ни одного a , это и означает, что все a оказались слева от b .

Пусть в обрабатываемое слово P входит несколько раз подслово a :

$$P \quad \boxed{\dots \quad a \quad \dots \quad a \quad \dots \quad a \quad \dots}$$

и нам надо заменить одно из вхождений a на подслово β . Такая замена делается с помощью формулы $a \rightarrow \beta$. Однако, если мы применим эту формулу к слову P , то будет заменено первое вхождение a . А что делать, если надо заменить какое-то другое вхождение a , скажем второе или последнее? Так вот, чтобы на β заменялось не первое вхождение a , а какое-то другое, это другое вхождение надо как-то выделить, пометить, для чего следует рядом с ним (слева или справа) поставить некоторый символ, скажем $*$, отличный от всех других символов, входящих в P :

$$P \quad \boxed{\dots \quad a \quad \dots \quad *a \quad \dots \quad a \quad \dots}$$

Такой символ будем в дальнейшем называть *спецзнаком* (аналог вспомогательного символа в МТ). Его роль – выделить нужное вхождение α среди других, сделать его уникальным.

Пример 5. (фиксация спецзнаком обрабатываемого символа) $A=\{0,1,2,3\}$. Пусть P – непустое слово. Тракуя его как запись неотрицательного целого числа в четверичной системе счисления, требуется получить запись этого же числа, но в двоичной системе.

Как известно, для перевода числа из четверичной системы в двоичную надо каждую четверичную цифру заменить на пару соответствующих ей двоичных цифр:

$$0 \rightarrow 00, 1 \rightarrow 01, 2 \rightarrow 10, 3 \rightarrow 11$$

Но этот алгоритм неправильный, Ошибка здесь в том, что после замены четверичной цифры на пару двоичных цифр уже никак нельзя отличить двоичные цифры от четверичных, поэтому наш НАМ начинает заменять и двоичные цифры. Значит, надо как-то отделить ту часть числа, в которой уже была произведена замена, от той части, где замены ещё не было. Для этого предлагается пометить слева спецзнаком * ту четверичную цифру, которая сейчас должна быть заменена на пару соответствующих двоичных цифр, а после того как такая замена будет выполнена, спецзнак нужно поместить перед следующей четверичной цифрой. слева от спецзнака всегда находится та часть числа, которая уже переведена в двоичный вид, а справа – часть, которую ещё предстоит заменить. Поэтому никакой путаницы между четверичными и двоичными цифрами уже не будет.

Итак, спецзнак * сначала должен быть размещён слева от первой цифры четверичного числа, а затем он должен «перепрыгивать» через очередную четверичную цифру, оставляя слева от себя соответствующие ей двоичные цифры. В конце же, когда справа от * уже не окажется никакой цифры, спецзнак надо уничтожить и остановиться.

Рабочая строка 0123

Система подстановок:

	Образец		Замена	Комментарий
1	*0	→	00*	
2	*1	→	01*	
3	*2	→	10*	
4	*3	→	11*	
5	*	→	.	
6		→	*	
7		→		

Пример 6. (перемещение спецзнака) $A=\{a,b\}$. Требуется приписать символ a к концу слова P .

Подсказка: пометить конец слова. Для этого следует переместить спецзнак в конец слова и заменить его на a .

Пример 7. $A=\{a,b\}$. В слове P заменить на aa последнее вхождение символа a , если такое есть.

Подсказка: добавить ещё один спецзнак, скажем $\#$, распределив между спецзнаками обязанности: пусть $*$ движется вправо, а $\#$ – влево. Появится же спецзнак $\#$ должен тогда, когда $*$ дойдет до конца слова, т.е. когда справа от $*$ не окажется других символов.

Рабочая строка

Система подстановок:

	Образец		Замена	Комментарий
1	*a	→	a*	
2	*b	→	b*	
3	*	→	#	
4	b#	→	#b	
5	a#	→	aa.	
6	#	→	.	
7		→	*	

Пример 8. (перенос символа через слово)

$A=\{a,b\}$. Перенести в конец непустого слова P его первый символ.

Пустое слово не менять.

Пример 9. (использование нескольких спецзнаков)

$A=\{a,b\}$. Удвоить слово P , т.е. приписать к P (слева или справа) его копию.

Решение

Предлагается следующий план решения задачи:

1. Приписываем к концу слова P символ $=$, справа от которого будем строить копию P .
2. Просматриваем по очереди все символы слова P и, не уничтожая их, переносим копию каждого символа в конец.
3. Удаляем символ $=$, который отделял слово P от его копии, и останавливаем алгоритм.

Теперь уточним этот план.

Как приписать некоторый символ к концу слова, мы уже знаем: надо сначала приписать слева к слову какой-то спецзнак, скажем $*$, затем перегнуть его направо через все символы слова и в конце, когда за $*$ не окажется никакого символа, заменить на символ $=$:

$$abb \rightarrow *abb \rightarrow a*bb \rightarrow ab*b \rightarrow abb* \rightarrow abb=$$

Из предыдущего примера мы также знаем, как переносить символы слова в конец слова. Только теперь сами символы уничтожать уже не надо. Поэтому поступаем так: если надо скопировать символ a , то порождаем за ним новый символ A (заменяем a на aA), после чего этот символ A переставляем с каждым последующим символом (в том числе

и с символом =), перенося тем самым A в конец слова, где и заменяем на a :

$$abb= \rightarrow aAbb= \rightarrow abAb= \rightarrow abbA= \rightarrow abb=A \rightarrow abb=a$$

Аналогично копируются и символы b .

Главный вопрос здесь: как узнать, какой именно символ исходного слова мы только что скопировали и какой символ надо копировать следующим? Для этого используем стандартный приём со спецзнаком – будем помечать новым спецзнаком $\#$ тот символ, который должен копироваться следующим (вначале это первый символ входного слова):

$$\#abb= \rightarrow a\#Abb= \rightarrow a\#bAb= \rightarrow a\#bbA= \rightarrow a\#bb=A \rightarrow a\#bb=a$$

Как только копия очередного символа окажется в конце, спецзнак $\#$ должен «запустить» процесс копирования следующего символа:

$$\begin{aligned} a\#bb=a &\rightarrow ab\#Bb=a \rightarrow ab\#bB=a \rightarrow ab\#b=Ba \rightarrow ab\#b=aB \rightarrow \\ &\rightarrow ab\#b=ab \rightarrow abb\#B=ab \rightarrow abb\#=Bab \rightarrow abb\#=aBb \rightarrow \\ &\rightarrow abb\#=abB \rightarrow abb\#=abb \end{aligned}$$

Когда справа от спецзнака $\#$ окажется символ =, это будет означать, что входное слово полностью скопировано. Осталось только уничтожить символы $\#$ и =, после чего остановиться.

Теперь отметим, что в НАМ, реализующем такое копирование, важен взаимный порядок расположения формул ($A\xi \rightarrow \xi A$, $B\xi \rightarrow \xi B$, $A \rightarrow a$ и $B \rightarrow b$), которые переносят символы A и B в конец и там восстанавливают символы a и b , и формулами ($\#a \rightarrow a\#A$ и $\#b \rightarrow b\#B$), которые «вводят в игру» символы A и B . Поскольку последняя пара формул должна срабатывать только после того, как символ A или B будет полностью перенесён в конец и заменён на a или b , то эта пара формул должна располагаться в НАМ ниже всех первых формул.

И ещё один момент. В этом НАМ используются два спецзнака $*$ и $\#$, первый из которых нужен для приписывания символа = справа к входному слову, а второй – для указания, какой символ слова должен копироваться следующим. Как ввести эти спецзнаки? Отметим, что использовать для этого две формулы $\rightarrow*$ и $\rightarrow\#$ нельзя, т.к. первая из них будет блокировать доступ ко второй. Оба этих спецзнака надо вводить сразу одной формулой $\rightarrow\#*$. При этом надо учитывать, что формулы с $*$ должны применяться самыми первыми, поэтому они должны располагаться в начале НАМ. Формулы же с $\#$, A и B должны распо-

лагаться ниже, чтобы они работали только после того, как исчезнет * и появится символ =.

Рабочая строка		Система подстановок:			
	Образец		Замена	Комментарий	
1	*a	→	a*		
2	*b	→	b*		
3	*	→	=		
4	Aa	→	aA		
5	Ab	→	bA		
6	A=	→	=A		
7	A	→	a		
8	Ba	→	aB		
9	Bb	→	bB		
10	B=	→	=B		
11	B	→	b		
12	#a	→	a#A		
13	#b	→	b#B		
14	#=	→	.		
15		→	#*		

Пример 10. (согласованная работа с различными частями слова)

$A=\{a,b\}$. Пусть слово P имеет чётную длину $(0, 2, 4, \dots)$. Удалить правую половину этого слова.

Подсказка: согласуем просмотр очередной буквы из левой половины с удалением буквы из правой половины.

1. Приписываем спецзнак * (звездочку) слева к входному слову P .
2. Звездочка сигнализирует, что необходимо обработать следующую за ней букву (если она есть) или прекратить обработку слова, если справа от звездочки ничего нет. Обработка буквы состоит в следующем. Звездочка перепрыгивает через букву и одновременно выпускает «стрелу»: спецзнак | появляется справа от звездочки. Далее стрела «летит» через буквы к концу слова и «поражает» (т.е. удаляет) последнюю букву. Сама «стрела» при этом тоже исчезает.
3. Действия пункта 2 продолжаются до тех пор пока звездочка не окажется справа от текущего слова, что будет сигнализировать об окончании процесса удаления правой половины слова. После этого сигнала остается удалить спецзнак * и закончить выполнение алгоритма.