

# Применение трансдюсеров в лингвистике

Георгий Алексеевич Мороз

Международная лаборатория языковой конвергенции (НИУ ВШЭ)

«Основные приложения математики», НИУ ВШЭ, 18 февраля 2022

презентация доступна по ссылке: [tinyurl.com/yapvnfm8](https://tinyurl.com/yapvnfm8)



# План доклада

Конечный автомат

Задачи

Трансдюсеры

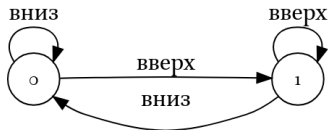
# Автомат(он)

Люди давно мечтали о роботах, но старые версии роботов, напоминавших по своему виду людей или животных называли автомат(он):



Рис. 1: Кукла дровосек (слева). Кукла Каракури, Т. Хисасигэ (справа).

Мы будем считать автоматами абстрактные машины, которые принимают разные состояния, а изменение состояний вызывается некоторым действием.



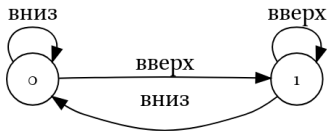
Мы будем считать автоматами абстрактные машины, которые принимают разные состояния, а изменение состояний вызывается некоторым действием.



# Конечный автомат

Конечный автомат — разновидность автоматов у которого есть:

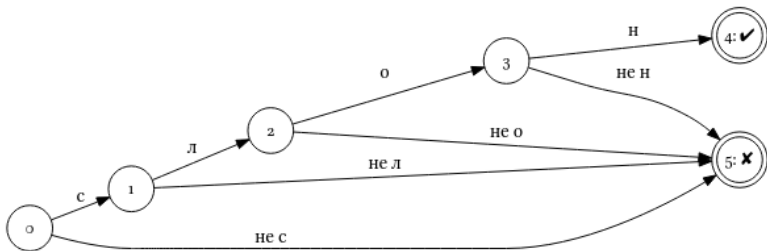
- алфавит, который автомат понимает;
- конечное количество состояний;
- переходы между состояниями;
- одно начальное состояние (часто обозначают нулем);
- набор конечных состояний (часто обозначают двойным кружочком).



# Конечный автомат

Конечные автоматы можно использовать для верификации поданного на вход слов:

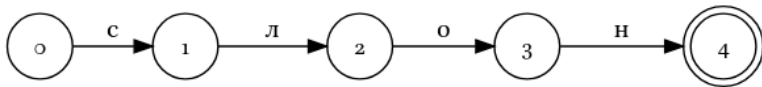
---



# Конечный автомат

Мы будем использовать такую нотацию:

---

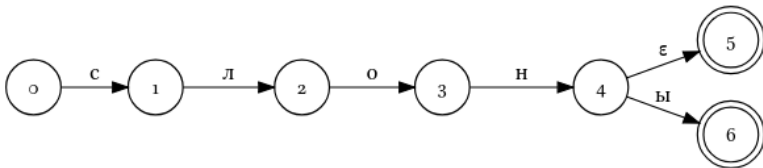




# Конечный автомат

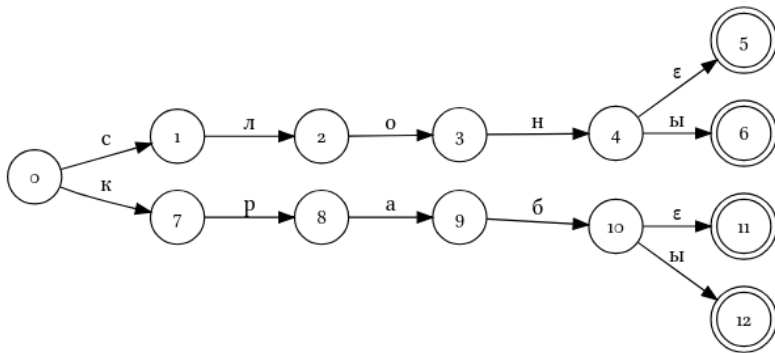
Пустой переход обозначают греческой буквой  $\epsilon$ :

---



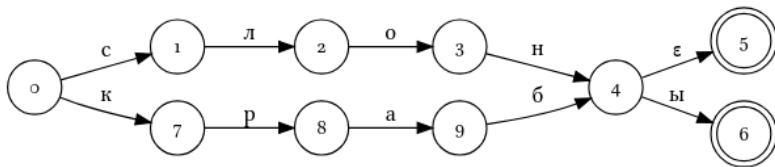
# Конечный автомат

Мы можем соединить два автомата:



# Конечный автомат

Любой конечный автомат можно автоматически минимизировать до оптимального количества состояний:



# План доклада

Конечный автомат

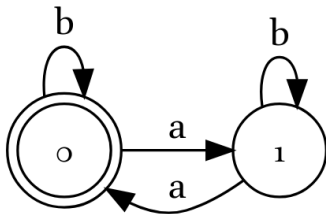
Задачи

Трансдюсеры

Может ли конечный автомат отличить последовательности символов  $a$  и  $b$ , в которых число символов  $a$  четное?

пример	ответ
bbbbbb	да
bbbbba	нет
babbba	да
aa	да
ab	нет
aaaaaa	да

Может ли конечный автомат отличить последовательности символов  $a$  и  $b$ , в которых число символов  $a$  четное?

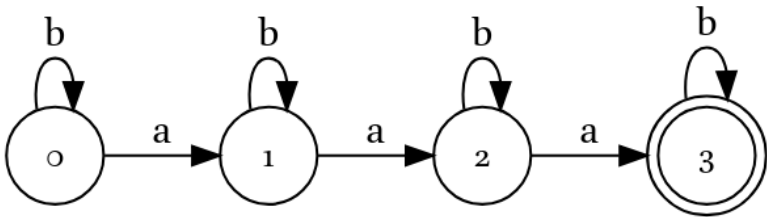


Может ли конечный автомат отличить последовательности символов  $a$  и  $b$ , в которых число  $a$  равно трем?

пример	ответ
bb	нет
ababb	нет
aaabb	да
abbaa	да
ba	нет
babaa	да

## Задачи

Может ли конечный автомат отличить последовательности символов  $a$  и  $b$ , в которых число  $a$  равно трем?





Может ли конечный автомат отличить палиндром от других последовательностей символов?

пример	ответ
довод	да
доход	да
показ	нет
кабак	да
козёл	нет
комок	да
потоп	да
шабаш	да

## Задачи

Может ли конечный автомат отличить палиндром от других последовательностей символов?

- Конечный автомат должен иметь **конечное** количество состояний.
- Палиндромов бесконечное количество.
- Конечные автоматы нигде не хранят информацию о предыдущих состояниях.
- Предположим, что такой автомат существует. Тогда мы сможем становить автомат на середине двух палиндромов  $x$  и  $y$ , перепутать их вторые части и автомат должен будет сказать, что это палиндром.

	первая половина	вторая половина	
$x$	11011	11011	палиндром
$y$	10001	10001	палиндром
$z$	11011	10001	не палиндром

# План доклада

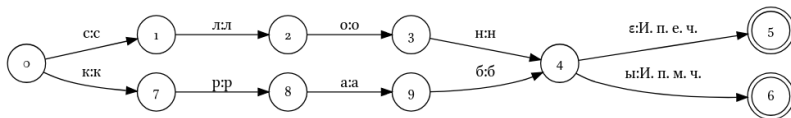
Конечный автомат

Задачи

Трансдюсеры

# Трансдюсеры

Трансдюсеры (в русской [википедии](#) они названы конечными автоматами с выходом) – это такие конечные автоматы, которые имеют дополнительный выходной алфавит и функции выхода.



# Трансдюсеры в лингвистике используют для

- морфологического анализа:

слонами → слон<Тв. п. м. ч>.

- транслитерации/транскрипции:

слонами → [slanamʲɪ]

- предективного ввода

сло... → слово, словно, сложный, сложиться...

- в системах проверки правописания

сланами → слонами?

- в системах автоматического перевода близкородственных языков

слонами → слон<Тв. п. м. ч> → elephant<Тв. п. м. ч> →

→ słoń<Тв. п. м. ч> → słońiami

# Полезные свойства трансдюсеров

- трансдюсеры обратимы
- трансдюсеры можно соединять между собой
- трансдюсеры можно оптимизировать

Спасибо за внимание!

А еще спасибо Нику Хауэллу!

