- Базовые определения:
  - Σ — конечное множество (алфавит)
  - Σ^* — множество всех конечных последовательностей элементов Σ (слов, строк)
  - $|w|$ — длина слова w, λ — пустое слово (длины 0)
  - на Σ^* задана операция конкатенации (умножения) слов
  - конкатенация ассоциативна, т.е. Σ^* — моноид относительно конкатенации
    - с единицей λ
    - Σ^* называют свободным моноидом или моноидом слов
  - $L \subseteq \Sigma^*$ — язык (конечный или бесконечный)
- Булевы операции над языками:
  - объединение $L_1 \cup L_2$
  - пересечение $L_1 \cap L_2$
  - дополнение $\bar{L} = \{w \in \Sigma^* \mid w \notin L\}$
  - разность $L_1 \setminus L_2 = L_1 \cap \bar{L}_2$
- Умножение языков:
  - $L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\}$
    Пример: $\{ab, abc\}\{ba, cba\} = \{abba, abcba, abccba\}$
  - ⋆ конкатенация ассоциативна $\Rightarrow$ умножение языков ассоциативно
    - и некоммутативно, разумеется
  - ⋆ естественно определяются степени языка: $L^n = L^{n-1}L, L^0 = \{\lambda\}$
- Итерация языка (Kleene star):
  - $L^* = \bigcup_{n=0}^{\infty} L^n$
  - ⋆ Σ^* — это действительно результат применения итерации к языку Σ
    Пример: $\{a, ab\}^* = a\Sigma^* \setminus \Sigma^* bb\Sigma^*$

★ Умножение языков дистрибутивно относительно объединения и пересечения:
- $L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3$, $(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3$
- $L_1(L_2 \cap L_3) = L_1L_2 \cap L_1L_3$, $(L_1 \cap L_2)L_3 = L_1L_3 \cap L_2L_3$

⋆ Для итерации дистрибутивности нет, но верно следующее:
- $(L_1 \cap L_2)^* \subseteq L_1^* \cap L_2^*$
- $L_1^* \cup L_2^* \subseteq (L_1 \cup L_2)^*$
- $L^{**} = L^*$ (идемпотентность)

- Другие полезные операции над языками:
  - $Lw^{-1} = \{u \mid uw \in L\}$ (правое деление на слово w)
  - $w^{-1}L = \{u \mid wu \in L\}$ (левое деление на слово w)
  - $\mathrm{pref}(L) = \{u \mid \exists w : uw \in L\}$ (префиксное замыкание)
  - $\mathrm{fact}(L) = \{u \mid \exists v, w : vuw \in L\}$ (подсловное замыкание)
  - $\mathrm{AD}(L) = \{u \notin L \mid$ все собственные подслова u принадлежат $L\}$ (антисловарь)
- Пусть $\preccurlyeq$ — отношение «быть подсловом» на Σ^*
  - ⋆ $\preccurlyeq$ — отношение порядка
  - ! докажите, что если $L = \mathrm{fact}(L)$, то язык $\mathrm{AD}(L)$ равен множеству минимальных элементов ЧУМа $(\bar{L}, \preccurlyeq)$
- Функция $\phi : \Sigma^* \to \Gamma^*$ — гомоморфизм, если $\phi(uv) = \phi(u)\phi(v)$ для любых $u, v \in \Sigma^*$
  - $\phi(L) = \{\phi(u) \mid u \in L\}$ — взятие гомоморфного образа

- Пусть $\Sigma = \{a_1, \ldots, a_n\}$
- Язык $L \subseteq \Sigma^*$ — регулярный, если он может быть получен применением конечного числа операций объединения, умножения и итерации к языкам $\varnothing, \{\lambda\}, \{a_1\}, \ldots, \{a_n\}$
  - операции $\cup, \cdot, ^*$ также называются регулярными
- Можно взять замыкание любого множества языков $\mathbf{L} \subseteq 2^{\Sigma^*}$ относительно регулярных операций

★ Множество $\mathbf{R} \subset 2^{\Sigma^*}$ всех регулярных языков над Σ совпадает с замыканием множества всех конечных языков над Σ относительно регулярных операций

- Обычный способ записи регулярных языков — регулярные выражения:
  - символы $\varnothing, \lambda, a \in \Sigma$ являются регулярными выражениями
  - r, s — регулярные выражения $\Rightarrow$ $(r)|(s)$, $(r) \cdot (s)$, $(r)^*$ — регулярные выражения
  - других регулярных выражений нет
    - ⋆ $|$ — стандартный символ для перечисления альтернатив (соответствует операции $\cup$)
    - ⋆ иногда вместо $|$ пишут $+$
- Для упрощения записи выражений договорились о приоритете операций:
  - * приоритетнее $\cdot$, $\cdot$ приоритетнее $|$
  - скобки, не меняющие порядок выполнения операций, опускаются
  - знак умножения также опускается
  - Пример: вместо $(((a) \mid (b)) \mid ((b) \cdot (c)))^*$ пишут $(a \mid b \mid bc)^*$

Теорема Клини

Язык регулярен тогда и только тогда, когда он распознается некоторым конечным автоматом.

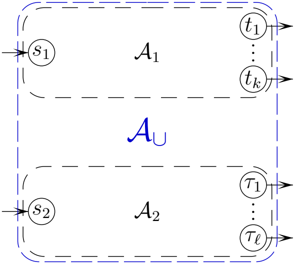
Докажем, что любой регулярный язык распознается конечным автоматом

⋆ теорема Рабина–Скотта дает использовать ДКА и НКА вперемешку

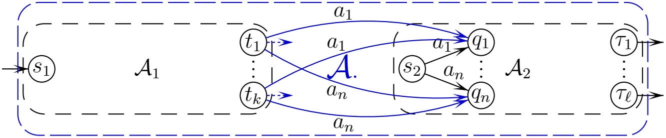
План:

1. построить автоматы, распознающие языки $\varnothing, \{\lambda\}, \{a\}$
   ! постройте самостоятельно
2. по ДКА $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, s_1, T_1)$ и $\mathcal{A}_2 = (Q_2, \Sigma, \delta_2, s_2, T_2)$ построить автоматы, распознающие языки
   - $L(\mathcal{A}_1) \cup L(\mathcal{A}_2)$
   - $L(\mathcal{A}_1) \cdot L(\mathcal{A}_2)$
   - $(L(\mathcal{A}_1))^*$

$$\mathcal{A}_\cup = (Q_1 \cup Q_2, \Sigma, \delta_1 \cup \delta_2, \{s_1, s_2\}, T_1 \cup T_2)$$
$$L(\mathcal{A}_\cup) = L(\mathcal{A}_1) \cup L(\mathcal{A}_2)$$



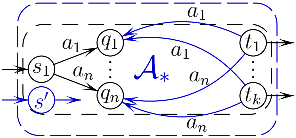
$\mathcal{A}_{\cdot} = (Q_1 \cup Q_2, \Sigma, \delta, \{s_1\}, T_2)$ при $\lambda \notin L_2$,
$\mathcal{A}_{\cdot} = (Q_1 \cup Q_2, \Sigma, \delta, \{s_1\}, T_1 \cup T_2)$ при $\lambda \in L_2$,
где $\delta = \delta_1 \cup \delta_2 \cup \{(t, a, q) \mid t \in T_1, q \in Q_2, (s_2, a, q) \in \delta_2\}$
$L(\mathcal{A}_{\cdot}) = L(\mathcal{A}_1) \cdot L(\mathcal{A}_2)$



$\mathcal{A}_* = (Q_1 \cup \{s'\}, \Sigma, \delta', \{s_1, s'\}, T \cup \{s'\})$, где
$\delta' = \delta_1 \cup \{(t, a, q) \mid t \in T, q \in Q_1, (s_1, a, q) \in \delta\}$

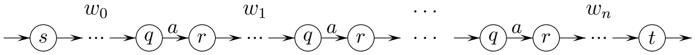
- ⋆ s' нужно только для распознавания λ

$L(\mathcal{A}_*) = (L(\mathcal{A}_1))^*$



- Пусть $\mathcal{A} = (Q, \Sigma, \delta, s, T)$ — автомат; докажем, что $L(\mathcal{A}) \in \mathbf{R}$ индукцией по $|\delta|$
  База индукции: $|\delta| = 0$
  - $L(\mathcal{A}) = \{\lambda\} \in \mathbf{R}$ при $s \in T$ и $L(\mathcal{A}) = \varnothing \in \mathbf{R}$ при $s \notin T$

  Шаг индукции: $|\delta| = k$
  - по предположению индукции, языки, распознаваемые автоматами с менее чем k переходами (ребрами), регулярны
  - возьмем произвольный переход $(q, a, r) \in \delta$, пусть $\delta' = \delta \setminus \{(q, a, r)\}$; положим
    $\mathcal{A}_0 = (Q, \Sigma, \delta', s, T)$
    $\mathcal{A}_1 = (Q, \Sigma, \delta', s, \{q\})$
    $\mathcal{A}_2 = (Q, \Sigma, \delta', r, \{q\})$
    $\mathcal{A}_3 = (Q, \Sigma, \delta', r, T)$
  - ⋆ языки $L(\mathcal{A}_0), L(\mathcal{A}_1), L(\mathcal{A}_2), L(\mathcal{A}_3)$ регулярны по предположению индукции
- Докажем, что $L(\mathcal{A}) = L(\mathcal{A}_0) \cup L(\mathcal{A}_1)a\big(L(\mathcal{A}_2)a\big)^* L(\mathcal{A}_3)$
  - пусть $w \in L(\mathcal{A})$ помечает (s, t)-маршрут W в $\mathcal{A}$, $t \in T$
  - ⋆ если $(q, a, r) \notin W$, то $w \in L(\mathcal{A}_0)$
  - ⋆ если $(q, a, r) \in W$, то $w = w_0 a w_1 \ldots a w_n$, где a отмечают все случаи использования перехода (q, a, r):



  - ⇒ $w_0 \in L(\mathcal{A}_1)$, $w_1, \ldots, w_{n-1} \in L(\mathcal{A}_2)$, $w_n \in L(\mathcal{A}_3)$ ⇒ $w \in L(\mathcal{A}_1)a\big(L(\mathcal{A}_2)a\big)^* L(\mathcal{A}_3)$
  - ⇒ $w \in L(\mathcal{A}_0) \cup L(\mathcal{A}_1)a\big(L(\mathcal{A}_2)a\big)^* L(\mathcal{A}_3)$
  - $L(\mathcal{A}_0) \subseteq L(\mathcal{A})$ — очевидно
  - $w \in L(\mathcal{A}_1)a\big(L(\mathcal{A}_2)a\big)^* L(\mathcal{A}_3)$ ⇒ $w = w_0 a w_1 \ldots a w_n$ как на рисунке ⇒ $w \in L(\mathcal{A})$ □

Замкнутость R относительно операций

- Теорема Клини позволяет доказывать замкнутость **R** относительно операций
  - ⋆ **R** замкнуто относительно дополнения:
  - если $\mathcal{A} = (Q, \Sigma, \delta, s, T)$ — ДКА, $L = L(\mathcal{A})$, то $\bar{L} = L(\bar{\mathcal{A}})$, где $\bar{\mathcal{A}} = (Q, \Sigma, \delta, s, Q \backslash T)$
  - ⋆ **R** замкнуто относительно пересечения, потому что $L_1 \cap L_2 = \overline{(\bar{L}_1 \cup \bar{L}_2)}$
    - формулы де Моргана
  - ⋆ **R** замкнуто относительно разности, потому что $L_1 \setminus L_2 = L_1 \cap \bar{L}_2$
  - ! докажите, у регулярного языка регулярными являются
    левые и правые частные
    гомоморфные образы
    префиксное и подсловное замыкание
    антисловарь

★ Объединение, пересечение и разность регулярных языков можно распознавать при помощи декартова произведения ДКА:

- пусть $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, s_1, T_1)$, $\mathcal{A}_2 = (Q_2, \Sigma, \delta_2, s_2, T_2)$ — ДКА
- ⋆ $\mathcal{A}_\cup = (Q_1 \times Q_2, \Sigma, \delta_1 \times \delta_2, (s_1, s_2), Q_1 \times Q_2 \setminus (Q_1 \backslash T_1) \times (Q_2 \backslash T_2))$
- ⋆ $\mathcal{A}_\cap = (Q_1 \times Q_2, \Sigma, \delta_1 \times \delta_2, (s_1, s_2), T_1 \times T_2)$
- ⋆ $\mathcal{A}_\setminus = (Q_1 \times Q_2, \Sigma, \delta_1 \times \delta_2, (s_1, s_2), T_1 \times (Q_2 \backslash T_2))$