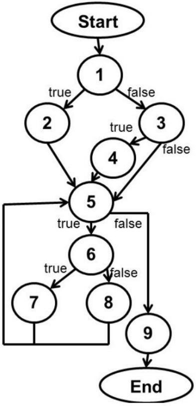
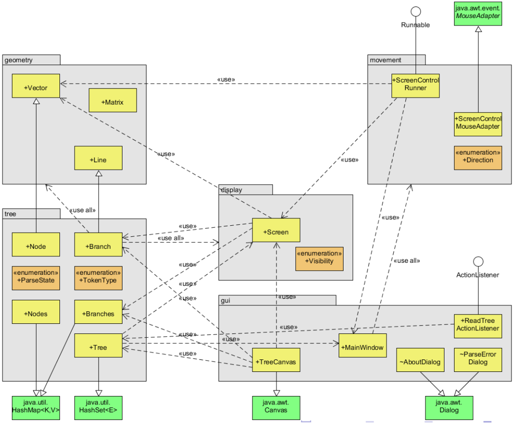
Помеченные орграфы

- Ребра часто представляют взаимодействие между объектами-вершинами

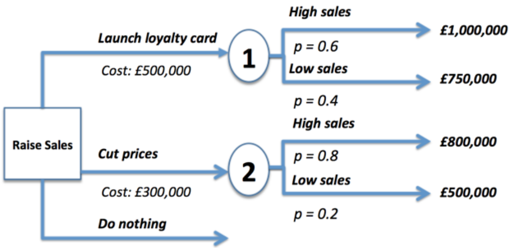
⋆ Когда моделируют несколько видов взаимодействия, ребра **помечают**, задавая функцию $\ell : E \to \Lambda$ из множества ребер в некоторое конечное множество меток
  - ⋆ тройку $G = (V, E, \ell)$ называют **помеченным** (ор)графом
  - ⋆ иногда говорят о **раскрашивании** ребер

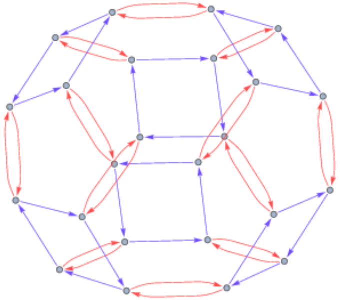
Примеры: слева control flow graph, справа UML class diagram





Примеры: слева business decision tree, справа граф Кэли симметрической группы S_4





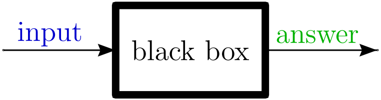
- Самая важная модель помеченного орграфа — конечный автомат

- Детерминированный конечный автомат (ДКА) — это пятерка $\mathcal{A} = (Q, \Sigma, \delta, s, T)$:
  - Q — непустое конечное множество состояний автомата
  - Σ — алфавит автомата (непустое конечное множество)
  - $\delta : Q \times \Sigma \to Q$ — функция переходов
  - $s \in Q$ — начальное (стартовое) состояние
  - $T \subseteq Q$ — множество конечных (терминальных) состояний

⋆ Q, Σ, δ задают помеченный орграф
  - Q — множество вершин
  - δ — множество помеченных ребер (метки — символы из Σ)
  - каждая вершина имеет степень исхода $|\Sigma|$ и ровно одно исходящее ребро с каждой меткой

Функционирование ДКА

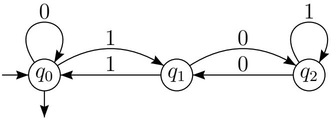
- ДКА — простейший пример математической машины:



  - входом для ДКА является произвольное слово $w \in \Sigma^*$
  - ДКА работает тактами (у него «дискретное время»)
  - перед первым тактом ДКА находится в начальном состоянии s
  - на i-ом такте ДКА обрабатывает символ $w[i]$:
    - переходит из текущего состояния q в состояние $\delta(q, w[i])$
    - ⋆ иначе говоря, ДКА идет из вершины q по исходящему ребру с меткой $w[i]$
  - после обработки всего слова w автомат приходит в некоторое состояние t
  - ⋆ ответом ДКА является булево значение $[t \in T]$
- машины, возвращающие булево значение, называются распознавателями
- Если ДКА на слове w возвращает
  - ⋆ 1, то он читает / допускает w
  - ⋆ 0, то он не читает / отвергает w
- Слова, которые читает ДКА $\mathcal{A}$, образуют множество $L(\mathcal{A}) \subseteq \Sigma^*$, которое называется языком, распознаваемым $\mathcal{A}$
  - (формальный) язык — это произвольное множество слов

Примеры ДКА

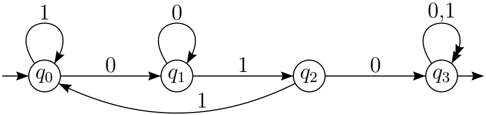
- ДКА, распознающий множество чисел, кратных 3, в двоичной записи:
  - ⋆ ведущие нули игнорируются
  - слева диаграмма (граф) переходов, справа таблица переходов



	0	1	T
q_0	q_0	q_1	1
q_1	q_2	q_0	0
q_2	q_1	q_2	0

! Достройте пример так, чтобы автомат не читал записи с ведущими нулями

- ДКА, распознающий множество строк, в которых есть подстрока 010:



! Перестройте пример так, чтобы распознавалось множество строк, в которых есть подпоследовательность 010

Комментарии и соглашения

- Так как ДКА — орграф, можно говорить о маршрутах
  - при этом вершины и ребра часто называют состояниями и переходами
  - ⋆ каждый маршрут в $\mathcal{A}$ помечен словом $w \in \Sigma^*$, полученным конкатенацией меток составляющих маршрут ребер

★ Для каждого слова $w \in \Sigma^*$ существует единственный маршрут в ДКА $\mathcal{A} = (Q, \Sigma, \delta, s, T)$, помеченный w и начинающийся в s

★ $\mathcal{A}$ читает w $\Leftrightarrow$ этот маршрут заканчивается в вершине из T

- Функцию переходов δ доопределяют на всём множестве $Q \times \Sigma^*$:
  - $\delta(q, a)$ — это конец ребра с меткой a, исходящего из q
  - $\Rightarrow$ конец маршрута с меткой w и началом q обозначим за $\delta(q, w)$
  - часто пишут $q.w$ вместо $\delta(q, w)$, если автомат известен
  - например, $\mathcal{A}$ читает w $\Leftrightarrow$ $s.w \in T$

- Несложно сделать так, чтобы ДКА выдавал k вариантов ответа вместо 2:
  - ⋆ множество T задает разбиение Q на два класса T и $Q \setminus T$, которым соответствуют ответы 1 и 0
  - вместо этого можно задать разбиение Q на k классов, которым соответствуют k возможных ответов
  - в предельном случае $k = |Q|$ ответ — это состояние (или его номер)
- Полученная такой модификацией ДКА машина — это конечный автомат с выходом
  - ⋆ если автомат, проверяющий делимость на 3, будет возвращать номер текущего состояния, он будет вычислять функцию $x \bmod 3$
- Последовательность $\{a_i\}_1^\infty$ называется k-автоматной, если существует автомат с выходом, который по k-ичной записи числа n возвращает a_n (для любого n)
  - ⋆ k-автоматные последовательности интересны тем, что о них можно доказывать теоремы автоматическим построением автоматов
  - → `https://github.com/hamousavi/Walnut`
  - вариант автомата с выходом, возвращающий свое состояние на каждом такте, называют машиной Мура
  - например, каждый элемент дисплея электронных часов управляется машиной Мура, совершающей один такт в секунду

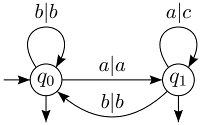
- Детерминированный конечный преобразователь

  он же детерминированный конечный трансдьюсер, машина Мили

  получается из ДКА добавлением выходного алфавита Γ и переопределением функции переходов (δ становится функцией из $Q \times \Sigma$ в $Q \times \Gamma$)

  ⋆ каждое ребро графа помечено парой букв (a, b), где $a \in \Sigma$, $b \in \Gamma$
  ⋆ по слову $w \in \Sigma^*$ преобразователь возвращает слово u длины $|w|$ в реальном времени
    - $u[i]$ — это буква, написанная на ребре, по которому автомат идет, читая $w[i]$
  - часто (но не обязательно) $\Gamma = \Sigma$

  Пример: преобразователь изменяет входное слово в алфавите $\{a, b\}$, заменяя в каждой последовательности букв a все буквы, кроме первой, на c

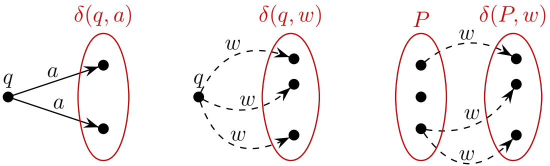


  - *aababaaa* заменяется на *acbabacc*

... Об автоматах есть отдельный курс теории автоматов, а здесь мы обсудим только пару базовых для этой теории теорем

Недетерминированный конечный автомат

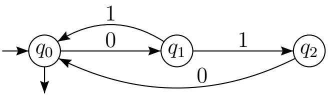
- Что если отказаться от ограничения
  - из каждой вершины исходит ровно одно ребро с данной меткой?
- Если мы возьмем произвольный орграф, в котором выделены множество начальных вершин S и множество терминальных вершин T, а каждое ребро помечено буквой из алфавита Σ, то получится машина, называемая недетерминированным конечным автоматом (НКА)
  - НКА — это пятерка $\mathcal{A} = (Q,\Sigma,\delta,S,T)$, где $\delta \subseteq Q \times \Sigma \times Q$ — множество переходов
  - иногда δ удобно записывать как функцию $\delta : Q \times \Sigma \to 2^Q$
  - $\delta(q, a)$ — множество вершин, в которые из q ведет ребро с меткой a
    - $\delta(q, a)$ может быть пустым
  - доопределим функцию δ:
  - $\delta(q, w)$ — множество вершин, в которые из q ведет маршрут, помеченный w
  - $\delta(P, w)$, где $P \subseteq Q$, — множество вершин, в которые ведет маршрут, помеченный w и начинающийся в вершине из P



- НКА $\mathcal{A} = (Q, \Sigma, \delta, S, T)$ читает/допускает слово $w \in \Sigma^*$, если существует (s, t)-маршрут с меткой w для некоторых $s \in S$, $t \in T$, то есть $\delta(S, w) \cap T \neq \varnothing$
- Язык $L(\mathcal{A})$ состоит из всех слов, читаемых $\mathcal{A}$

Пример НКА и комментарии

Данный НКА читает в точности те слова, которые можно разбить на блоки 01 и 010:



- Недетерминированный выбор связан с состоянием q_1
  - Слово начинается с 1 или содержит 11 — автомат его не читает ($\delta(S, w) = \varnothing$)
  - $\delta(S, 010100) = \{q_1\}$ — автомат не читает 010100
  - $\delta(S, 010101) = \{q_0, q_2\}$ — автомат читает 010101
- Термин «недетерминированный» применительно к алгоритму/машине означает, что вычисление может пойти различными путями
  - если очередной переход можно выбрать несколькими способами, НКА делает недетерминированный выбор
  - определение прочтения слова w означает, что НКА при выборе всегда «угадывает» так, чтобы в конце оказаться в терминальном состоянии
- Определение чтения слова/распознавания языка при помощи НКА включает фундаментальную асимметрию между кванторами $\exists$ и $\forall$

- Похоже, что НКА обладают бо́льшими вычислительными возможностями, чем ДКА; тем не менее, это не так (см. следующий фрагмент)

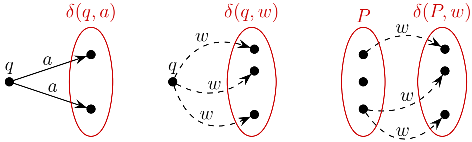
Теорема Рабина–Скотта

Теорема Рабина–Скотта

Для любого НКА существует ДКА, распознающий тот же самый язык.

- Доказательство:
  - возьмем произвольный НКА $\mathcal{B} = (Q, \Sigma, \delta, S, T)$
  - построим ДКА $\mathcal{A}$ такой, что $L(\mathcal{A}) = L(\mathcal{B})$
  - пусть $\mathcal{A} = (2^Q, \Sigma, \delta', S, T')$, где $\delta'(P, a) = \delta(P, a)$, $T' = \{P \in 2^Q \mid P \cap T \neq \varnothing\}$

напомним:



  - докажем, что $\delta'(P, w) = \delta(P, w)$ для любого слова w индукцией по $|w|$:
  - база индукции: для $|w| = 0$ имеем $\delta'(P, \lambda) = \delta(P, \lambda) = P$
  - шаг индукции: пусть $w = ua$, $a \in \Sigma$
    $\delta'(P, w) = \delta'(\delta'(P, u), a) = \delta'(\delta(P, u), a) =$
    $\delta(\delta(P, u), a) = \bigcup_{r \in \delta(P,u)} \delta(r, a) = \bigcup_{q \in P} \delta(q, ua) = \delta(P, ua)$
  - осталось заметить, что
    $L(\mathcal{A}) = \{w \in \Sigma^* \mid \delta'(S, w) \in T'\} = \{w \in \Sigma^* \mid \delta'(S, w) \cap T \neq \varnothing\} =$
    $\{w \in \Sigma^* \mid \delta(S, w) \cap T \neq \varnothing\} = L(\mathcal{B})$ $\square$

- Пусть $\mathcal{A} = (Q, \Sigma, \delta, s, T)$ — ДКА
  - состояние $q \in Q$ достижимо, если существует $w \in \Sigma^*$ такое, что $q = s.w$
  - ⋆ т.е. если вершина q достижима из начальной вершины s
  - ⋆ недостижимые состояния можно удалить — это балласт, который занимает лишнее место и не влияет на функционирование автомата
  - ⋆ достижимые состояния находятся поиском из начальной вершины
- При построении ДКА $\mathcal{A}$, распознающего тот же язык, что и данный НКА $\mathcal{B}$, поиск совмещают с построением, получая $\mathcal{A}$ без недостижимых состояний:

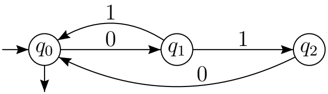
1. **для каждого** $P \subseteq Q$ $\ label(P) \leftarrow 0$
2. $Q' \leftarrow \{S\}$
3. **пока** $(\exists P \in Q' : label(P) = 0)$, **повторять**
4. **для каждого** $a \in \Sigma$
5. $\delta'(P, a) \leftarrow \bigcup_{q \in P} \delta(q, a)$
6. $Q' \leftarrow Q' \cup \{\delta'(P, a)\}$
7. $label(P) \leftarrow 1$
8. $T' \leftarrow \{P \in Q' \mid P \cap T \neq \varnothing\}$

★ Обычно при использовании этого алгоритма Q' получается намного меньше, чем 2^Q; тем не менее, существуют НКА, для которых $Q' = 2^Q$

! Изучив доказательство теоремы Рабина–Скотта, придумайте, как вычислить ответ НКА $\mathcal{B}$ на слове w за время $O(|w| \cdot |Q|^2)$

- это бывает выгоднее, чем построение и хранение большого ДКА

Пример



	0	1	T
q_0	q_1	$\varnothing$	1
q_1	$\varnothing$	q_0, q_2	0
$\varnothing$	$\varnothing$	$\varnothing$	0
q_0, q_2	q_0, q_1	$\varnothing$	1
q_0, q_1	q_1	q_0, q_2	1