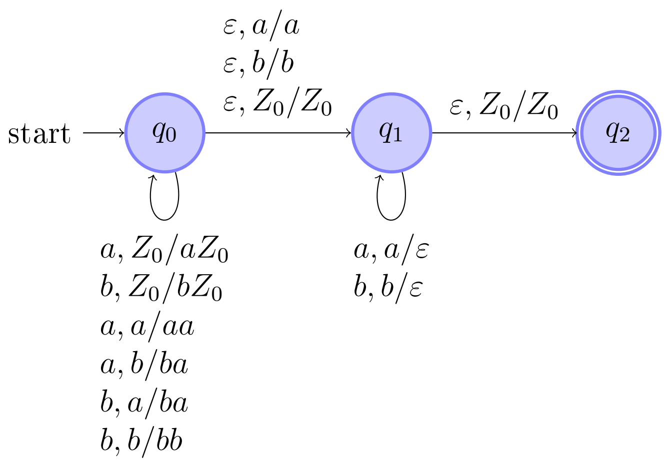
Пример: автомат для языка $L_{ww^R} = \{ww^R \mid w \in \{a, b\}^*\}$

Замечание: w^R означает: «слово w, записанное задом—наперёд». Т. о., L_{ww^R} это язык палиндромов над алфавитом a, b.

Известно, что автомат построить можно, т. к. ранее была построена КС-грамматика для L_{ww^R} и приведено утверждение о совпадении МП- и КС-языков. В общем случае, понять, является ли язык автоматным сложно. Аналогичный вопрос для КС-языков может быть решён с помощью т. н. *леммы о накачке (разрастании)*. Она устанавливает некоторое свойство для скорости роста длины слов в любом КС-языке, опровергнув которое, можно утверждать, что имеем дело с не КС-языком. Таков, например, язык $\{a^n b^n c^n | n \geqslant 0\}$.

Метод. Автомат имеет два «рабочих» состояния, q_0, q_1: в первом он читает часть слова на ленте, относящуюся к w, а во втором — к w^R. В первом состоянии он каждую прочитанную букву кладёт на стек, чтобы потом можно было сравнить её с соответствующей буквой w^R. В завершающее состояние q_2 перейдём после q_1 при опустошении стека (при появлении маркера дна стека). Будем использовать графическое представление автомата.



Опишем автомат с помощью функции перехода:

$\delta(q_0, a, Z_0) = (q_0, aZ_0)$ — первая сверху метка петли на q_0
$\delta(q_0, \varepsilon, a) = (q_1, a)$ — первая сверху метка дуги $q_0 q_1$
$\delta(q_1, a, a) = (q_1, \varepsilon)$ — первая сверху метка петли на q_1
...

Автомат имеет существенно недетерминированную природу. Например, находясь в q_0, он на каждом такте пытается угадать, достигнута ли середина слова: можно сказать, что автомат создаёт своего полного клона (с такой же конфигурацией) и каждый из двух идёт по своей дуге. Если кто-то один из множества клонов дойдёт до q_2, прочтя всё слово, то это слово будет допущено. Рассмотрим это на примере слова $abba$.

1. Неудачный клон:
$$(q_0, abba, Z_0) \vdash (q_1, abba, Z_0) \vdash (q_2, abba, Z_0)$$
2. Удачный клон:
$$(q_0, abba, Z_0) \vdash (q_0, bba, aZ_0) \vdash (q_0, ba, baZ_0) \vdash (q_1, a, aZ_0) \vdash \\ \vdash (q_1, \varepsilon, Z_0) \vdash (q_2, \varepsilon, Z_0).$$

В определение МП-автомата была заложена недетерминированность:
$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to \mathbf{P_{fin}}(Q \times \Gamma^*).$$
Можно определить класс ДМП детерминированных МП-автоматов, которые совпадают с (Н)МП-автоматами во всём, кроме вида функции переходов:
$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to Q \times \Gamma^*,$$
причём, для всех $q \in Q$ выполнено условие:
$$\mathbf{if}\ \exists a \in \Sigma\ \exists Z \in \Gamma \mid \delta(q, a, Z) \neq \emptyset\ \mathbf{then}\ \delta(q, \varepsilon, Z) = \emptyset.$$

Для конечных автоматов выполнялось равенство $L(\text{НКА}) = L(\text{ДКА})$. Оно доказывалось с помощью детерминизации НКА-автомата. Функция переходов НКА выглядела так: $Q \times \Sigma \to P(Q)$, в строящемся ДКА в качестве множества состояний Q' бралось $P(Q)$. Ясно, что эта конструкция не подойдёт для МП-автоматов, так как $\mathbf{P_{fin}}(Q \times \Gamma^*)$ не является конечным множеством. Оказывается, другую конструкцию придумать невозможно.

Теорема. *$L(\text{МПА}) \neq L(\text{ДМПА})$.*

Утверждение. *$L_{ww^R} \notin L(\text{ДМПА})$*

Интуитивно понятно, что на практике желательно иметь дело с детерминированными объектами и алгоритмами. Например, в теории синтаксического анализа и компиляции используется класс языков, порождаемых т.н. LR(k)-грамматиками, причём
$$L(\text{КС}) \supsetneq L(\text{LR}(k)) = L(\text{ДМПА}) \subsetneq L(\text{МПА}).$$

Историческая справка

- МП-автоматы впервые были введены в работах Эттингера (Oettinger), 1961 и Шютценберже (Schutzenberger), 1963.
- Эквивалентность МПА- и КС-языков установлена Хомским (Chomsky) во внутреннем докладе MIT, 1961, и впервые опубликована в печати Эви (Evey), 1963.
- ДМПА введены Фишером, 1963 и Шютценберже, 1963.

Нормальные формы КС-грамматик

Понятие нормальной формы встречается в различных разделах дискретной математики и информатики и имеет важное значение. Построение алгоритмов или доказательств зачастую делается более удобным, если входные данные имеют некоторую нормальную форму. До сих пор были известны (С)КНФ, (С)ДНФ, пренексная нормальная форма (только с внешними кванторами) и получаемая из неё при помощи операции сколемизации нормальная форма Сколема (конъюнктивная пренексная форма только с кванторами всеобщности). Пренексная нормальная форма использована Гёделем для доказательства теоремы о полноте исчисления предикатов, а НФ Сколема применяется в автоматическом доказательстве теорем.

Нормальная форма Хомского

Определение. Говорят, что грамматика $G(N, \Sigma, \mathcal{P}, S)$ *находится в нормальной форме Хомского* (НФХ), если каждое правило из $\mathcal{P}$ имеет один из следующих видов:

1. $A \to BC$, где $A, B, C \in N$;
2. $A \to a$, где $A \in N$, $a \in \Sigma$.

Кроме того, среди правил может быть $S \to \varepsilon$, при условии что S не встречается в правых частях других правил.

Алгоритм приведения КС-грамматики к НФХ

Вход: КС-грамматика $G = (\Sigma, N, \mathcal{P}, S)$.

Выход: КС-грамматика в НФХ $G' = (\Sigma', N', \mathcal{P}', S')$.

Метод: получение приведённого вида грамматики с последующим разбиением «длинных» слов в правых частях продукций на слоги (то есть заменой цепочек на новые нетерминалы).

1. Получим по G приведённую грамматику (используя известный алгоритм) — обозначим её также $G(\Sigma, N, \mathcal{P}, S)$.

2. Если $S \to \varepsilon$ есть в $\mathcal{P}$, добавить $S \to \varepsilon$ к $\mathcal{P}'$.

3. Все продукции вида $A \to a$ и $A \to BC$, где $A, B, C \in N$, $a \in \Sigma$, из $\mathcal{P}$ добавляем в $\mathcal{P}'$.

   Замечание. Т. к. в приведённой грамматике нет продукций вида $A \to B$ и $A \to \varepsilon$, то к этому моменту просмотрены все продукции с правыми частями длины 0 и 1. Рассмотрена также часть продукций с правой частью длины два. Оставшиеся продукции такого вида рассматривает шаг 4.

4. Для каждой непросмотренной продукции $A \to X_1X_2$ $(X_1, X_2 \in \Sigma \cup N)$ $\mathcal{P}$ добавим в $\mathcal{P}'$ продукцию $A \to X_1'X_2'$, где X_i' это новый нетерминал, если $X_i \in \Sigma$, или совпадает с X_i, если $X_i \in N$.

   Замечание. На данный момент просмотрены все продукции $\mathcal{P}$ длины не больше двух. Оставшиеся продукции рассматриваются на следующем шаге.

5. Для каждой продукции вида $A \to X_1X_2 \ldots X_k$, $k \geqslant 3$ из $\mathcal{P}$ добавим в $\mathcal{P}'$ продукции:

$$\begin{array}{l} A \to X_1'[X_2X_3 \ldots X_k]; \\ [X_2X_3 \ldots X_k] \to X_2'[X_3 \ldots X_k]; \\ \ldots \\ [X_{k-1}X_k] \to X_{k-1}'X_k'. \end{array}$$

   Где, по-прежнему, X_i' это новый нетерминал, если $X_i \in \Sigma$, или совпадает с X_i, если $X_i \in N$.

6. Для всех добавленных ранее символов X_i' добавим в $\mathcal{P}'$ продукции $X_i' \to X_i$, если $X_i \in \Sigma$.

Теорема. *Для любой КС-грамматики G существует эквивалентная ей КС-грамматика в НФХ.*

Благодаря относительной простоте НФХ, она имеет большое практическое и теоретическое значение. Например, один из немногих общих алгоритмов разбора слов произвольной КС-грамматики, алгоритм Кока—Янгера—Касами, работает с грамматикой в НФХ. Асимптотическая сложность этого алгоритма $\Theta(n^3)$, что делает его плохо применим на практике. Именно потому при построении реальных языков чаще используют не общие КС-грамматики, а их подкласс, $\mathrm{LR}(k)$, для которого существуют линейные алгоритмы разбора ($\Theta(n)$).

Нормальная форма Грейбах

Определение. Говорят, что грамматика $G(N, \Sigma, \mathcal{P}, S)$ *находится в нормальной форме Грейбах* (НФГ), если каждое правило из $\mathcal{P}$ имеет вид:

$$A \to a\gamma,$$

где $A \in N$, $a \in \Sigma$, $\gamma \in (\Sigma \cup N)^*$. Допускается также правило $S \to \varepsilon$.

Теорема. *Для любой КС-грамматики G существует эквивалентная ей КС-грамматика в НФГ.*

Замечание. Алгоритм приведения произвольной КС-грамматики к НФГ весьма сложен и не рассматривается в курсе.

Одно из интересных теоретических приложений НФГ состоит в том, что будучи поданной на вход алгоритма построения МП-автомата по произвольной КС-грамматике (этот алгоритм ещё не был приведён, но утверждалось его существование), получается МП-автомат без ε-переходов. Т. о. автоматически получается доказательство утверждения об эквивалентности МП-автоматов и МП-автоматов без ε-переходов.

Варианты МП-автоматов

Определение. *Расширенным автоматом с магазинной памятью* (РМ-ПА) называют семёрку $(Q, \Sigma, \Gamma, \delta, q_o, Z_0, F)$, где все составляющие имеют такой же смысл, как и для МП-автоматов, но функция переходов δ имеет вид:

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma^* \to P_{fin}(Q \times \Gamma^*),$$

причём выполнено условие:

$$|\{(q, a, \gamma) \mid \delta(q, a, \gamma) \neq \emptyset\}| < \infty.$$

Теорема.

$$L(\text{\it РМПА}) = L(\text{\it МПА})$$