Задание 5

Регулярные грамматики

Ключевые слова [1]:*язык, регулярный язык, ДКА, НКА, алгебра регулярных выражений, грамматики, уравнения с регулярными коэффициентами.*

1 Грамматики

Одна из больших проблем науки, которую мы с вами изучаем – определения. Их слишком много и они отличаются друг от друга, хотя в итоге конечно описывают одни и те же классы языков. Я призываю на экзамене пользоваться определениями из книги Серебрякова, хотя при выполнении задания вы можете пользоваться эквивалентными определениями из другой литературы.

Определение 1. Грамматика Γ определяется через

- N – множество нетерминальных символов
- T – множество терминальных символов
- P – множество правил вывода, $P \subseteq (N \cup T)^* \times (N \cup T)^*$.
- S – аксиома, $S \in N$.

При этом, $N \cap T = \emptyset$. Принято обозначение $\Gamma = G(N, T, P, S)$. При описании грамматики приняты следующие соглашения. Нетерминалы обозначают заглавными буквами $A, B, C, \ldots$ терминалы обозначают строчными буквами, смешанные цепочки из $(N \cup T)^*$ обозначают греческими буквами α, β, γ. Слово $w \in T^*$ порождается грамматикой Γ, если существует последовательность правил вывода, начинающаяся с правила вида $S \to \alpha$, в результате применения которых порождается слово w. Под применением правила $\alpha \to \beta$, понимается, что подслово α заменяется на подслово β

[1]минимальный необходимый объем понятий и навыков по этому разделу)

В зависимости от ограничений, налагаемых на правила вывода, получаются разные классы языков. В рамках этого задания нас пока интересует только последний тип.

- Если на множество правил P не накладывается ограничений, то есть правила имеют вид $\alpha \to \beta$, то грамматика называется грамматикой типа 0 по Хомскому

- Грамматики, в которых правила имеют вид $\alpha A \beta \to \alpha\gamma\beta$, $|\gamma| > 0$ называются грамматиками типа 1 или Контекстно-зависимыми. В качестве исключения грамматике может принадлежать правило $S \to \varepsilon$, но тогда нетерминал S не может встречаться в правых частях.

- Грамматики, в которых правила имеют вид $A \to \alpha$, называются грамматиками типа 2 или Контекстно-Свободными грамматиками.

- Грамматики, в которых правила имеют вид $A \to xB$ или $A \to x$, $x \in T^*$, называются грамматиками типа 3 или праволинейными грамматиками.

В определении КЗ-грамматики существенно, что она является *неукорачивающей*, т.е. правая часть правил всегда длинее левой. Эквивалентное определение из книги Серебрякова гласит, что в КЗ-грамматике все правила, кроме быть может $S \to \varepsilon$, имеют вид $\alpha \to \beta$, $|\alpha| < |\beta|$. Опять-таки, если есть правило $S \to \varepsilon$, то нетерминал S в правых частях правил встречаться не может.

Очень часто грамматиками типа 3 называют грамматики, в которых правила вывода имеют вид $A \to xB$ или $A \to x$, $x \in T$, также допускается правило $S \to \varepsilon$ с всё той же оговоркой, что аксиома не может встречаться в правой части. Такие грамматики называются *праволинейными регулярными* грамматиками.

Упражнение 1. Доказать, что праволинейные грамматики и праволинейные регулярные грамматики эквивалентны, т.е. порождают один и тот же тип языков.

Определение 2. Грамматика типа 3 является *неоднозначной*, если существует более одного способа вывести хотя бы одно слово из языка, порождённого грамматикой.

Для грамматик другого типа, это определение неприемлемо. Вдумчивый читатель может подумать почему. Ответ будет дан в одной из следующих серий.

Леволинейные грамматики определяются аналогично праволинейным: в них правила имеют вид $A \to Bx$ или $A \to x$...

2 Построение регулярного выражения по системе линейных уравнений с регулярными коэффициентами

Перед тем как перейти непосредственно к описанию системы линейных уравнений с регулярными коэффициентами, вспомним о свойствах регулярных выражений. Будем обозначать регулярные выражения греческими буквами. Очевидно, что $\alpha|\beta = \beta|\alpha$, поэтому операцию объединения часто обозначают как $+$. В роли умножения выступает операция конкатенации, в роли нуля – $\varnothing$, а в роли единицы – ε. Для данных операций выполняются следующие свойства:

- $\alpha + \beta = \beta + \alpha$
- $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$
- $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$
- $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$
- $\alpha(\beta\gamma) = (\alpha\beta)\gamma$
- $\alpha + \alpha = \alpha$
- $\alpha^* = \alpha + \alpha^*$
- $(\alpha^*)^* = \alpha^*$
- $\varnothing^* = \varepsilon$
- $\alpha + \varnothing = \alpha$
- $\alpha \cdot \varepsilon = \varepsilon \cdot \alpha = \alpha$
- $\alpha \cdot \varnothing = \varnothing \cdot \alpha = \varnothing$

Что вместе с замкнутостью регулярных выражений относительности конкатенации, объединения и итерации позволяет рассматривать линейные уравнения с регулярными коэффициентами. Вообще говоря, регулярные языки относительно объединения и конкатенации образуют полукольцо с единицей – это полезно понимать, чтобы видеть, что алгебраические вещи возникают отнюдь не на пустом месте. Однако, наше использование систем линейных уравнений с регулярными коэффициентами сведётся лишь к формальной их записи – на что-то более подробное, у нас, увы, времени нет.

Итак, линейное уравнение с регулярными коэффициентами имеет вид:

$$X = \alpha X + \beta$$

Наименьшей неподвижной точкой уравнения с регулярными коэффициентами называется наименьшее по мощности множество X', при подстановке которого в уравнение, уравнение остаётся справедливым. Легко видеть, что наименьшей неподвижной точкой линейного уравнения с регулярными коэффицентами будет решение $X = \alpha^*\beta$.

Упражнение 2. Доказать, что $X = \alpha^*\beta$ является единственной наименьшей неподвижной точкой линейного уравнения $X = \alpha X + \beta$. Посмотрите доказательство этого факта в Ахо и Ульмане и сравните насколько «легко видеть» соотносится с «коротко доказать».

Системой линейных уравнений с регулярными коэффициентами называется система вида

$$\begin{gathered} X_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1n}X_n + \alpha_{10} \\ \ldots \\ X_i = \alpha_{i1}X_1 + \alpha_{i2}X_2 + \ldots + \alpha_{in}X_n + \alpha_{i0} \\ \ldots \\ X_n = \alpha_{n1}X_1 + \alpha_{n2}X_2 + \ldots + \alpha_{nn}X_n + \alpha_{n0} \end{gathered}$$

i-ое уравнение системы решается следующим образом:

$$X_i = \alpha_{ii}X_i + \underbrace{\alpha_{i1}X_1 + \ldots + \alpha_{ii-1}X_{i-1} + \alpha_{ii+1}X_{i+1} + \ldots \alpha_{nn}X_n + \alpha_{i0}}_{\beta_i}$$

Таким образом, $X_i = \alpha_{ii}^*\beta_i$.

Для того, чтобы получить регулярное выражение, описывающие язык, порождаемый ПГ, нужно записать для правил ПГ систему линейных уравнений: для правил вида $A_1 \to w_1A_1|w_2A_2|\ldots w_nA_n|v_1|v_2|\ldots|v_k$ уравнение имеет вид

$$A_1 = w_1A_1 + w_2A_2 + \ldots + w_nA_n + (v_1 + v_2 + \ldots + v_k)$$

Разрешив все уравнения системы и подставив решения в строчку с S получим решение уравнения для S вида $S = \gamma$, где γ и будет искомым регулярным выражением.

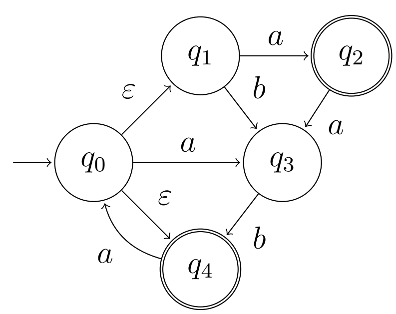
С системами линейных уравнений с регулярными коэффициентами можно ознакомиться в книге Ахо и Ульмана *Теория Синтаксического Анализа, Перевода и Компиляции Том I*

3 Задачи

Внимание, все задачи на построение автоматов должны быть снабжены диаграммами!

Задача 1.

На семинаре я строил по автомату праволинейную грамматику. Является ли полученная таким образом грамматика регулярной праволинейной? Постройте по автомату $\mathcal{A}$ регулярную праволинейную грамматику G, если алгоритм, предложенный на семинаре не подходит, предложите свой алгоритм (если возьмёте его из книжки, не списывайте страницами, пожалуйста).



Задача 2.

1. Предложите алгоритм построения НКА по праволинейной грамматике.

2. Постройте автомат по грамматике G :

$$S \to abaA|abB|\varepsilon,\ A \to aB|aa,\ B \to bA|aS$$

3. Постройте регулярное выражение для языка $L(G)$.

4. Является ли грамматика G однозначной?

Задача 3. Верно ли, что праволинейная грамматика G однозначна тогда и только тогда, когда построенный по ней автомат является детерминированным?

Задача 4. Назовём грамматику линейной, если в правой части её правил всего один нетерминал. Верно ли, что для любой линейной грамматики G, $L(G) \in \mathsf{REG}$?

Ещё раз напоминаю, что задачи, помеченные † являются дополнительными, поэтому списывать их из книжек – бессмысленное увеличение энтропии.

Определение 3. Для языка $L \subseteq \{\sigma_1, \sigma_2, \ldots, \sigma_n\}^* = \Sigma_n^*$ и языков $L_{\sigma_1}, L_{\sigma_2}, \ldots, L_{\sigma_n} \subseteq \Sigma_n^*$, подстановкой в L языков $L_{\sigma_1}, \ldots, L_{\sigma_n}$ назовём язык L', такой что для всех слов $w = w[1] \ldots w[n]$ из языка L справедливо $L_{w[1]} L_{w[2]} \ldots L_{w[n]} \subseteq L'$

Задача $5^\dagger$. Доказать, что регулярные языки замкнуты относительно операции подстановки.

Определение 4. Даны алфавиты Σ и Δ. Для языка $L \subseteq \Sigma \times \Delta$ определены операции проекции на Σ^* и Δ^*. Проекцией L на Σ^* называется язык $L_\Sigma = \{w \in \Sigma^* \mid \exists v \in \Delta^* : (w, v) \in L\}$. Проекция L на Δ^* определяется аналогичным образом.

Задача $6^\dagger$. Доказать, что регулярные языки замкнуты относительно операции проекции.

Определение 5. Для языка $L_\Sigma \subseteq \Sigma^*$, Δ-целиндром называется язык L, такой что $L = \{w \mid w = (u, v), u \in L_\Sigma, v \in \Delta^*\}$

Задача $7^\dagger$. Показать, что Σ-проекция Δ-цилиндра L есть L. Доказать, что регулярные языки замкнуты относительно операции цилиндра.