

SRS. Иерархия Хомского. Регулярные грамматики

Теория формальных языков
2021 г.



String RS, или semi-Thue systems

Частный случай TRS — SRS (плоские TRS).

- Множество данных — строки (слова) в алфавите \mathcal{A} .
- Правила переписывания имеют вид $u \rightarrow v$, где u, v — строки из \mathcal{A}^* .
- Правило $u \rightarrow v$ применимо к строке Φ , если Φ содержит хотя бы одну подстроку u .
- В общем случае применение не детерминированно.

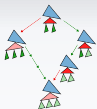


String RS, или semi-Thue systems

Частный случай TRS — SRS (плоские TRS).

- Множество данных — строки (слова) в алфавите \mathcal{A} .
- Правила переписывания имеют вид $u \rightarrow v$, где u, v — строки из \mathcal{A}^* .
- Правило $u \rightarrow v$ применимо к строке Φ , если Φ содержит хотя бы одну подстроку u .
- В общем случае применение не детерминированно.

Каково множество нормализованных строк в алфавите $\{\mathbf{A}, \mathbf{B}\}$ относительно системы правил $\mathbf{AB} \rightarrow \varepsilon, \mathbf{AA} \rightarrow \mathbf{A}$? относительно только правила $\mathbf{AB} \rightarrow \varepsilon$?



Выразительная сила SRS

Теорема

SRS позволяют выразить любую рекурсивную функцию.

Доказательство: алгоритмы Маркова.

Определение

Нормальный алгоритм (алгоритм) Маркова (НАМ) — это SRS с детерминированным поведением:

- top-down сопоставление (от верхних правил к нижним);
- выбор самой левой подстроки;
- существование терминальных правил.



Грамматики

Определение

Грамматика — это четвёрка $G = \langle N, \Sigma, P, S \rangle$, где:

- N — алфавит нетерминалов;
- Σ — алфавит терминалов;
- P — множество правил переписывания $\alpha \rightarrow \beta$ типа $\langle (N \cup \Sigma)^+ \times (N \cup \Sigma)^* \rangle$;
- $S \in N$ — начальный символ.

$\alpha \Rightarrow \beta$, если $\alpha = \gamma_1 \alpha' \gamma_2$, $\beta = \gamma_1 \beta' \gamma_2$, и $\alpha' \rightarrow \beta' \in P$.

\Rightarrow^* — рефлексивное транзитивное замыкание \Rightarrow .

Определение

Язык $L(G)$, порождаемый G — множество $\{u \mid u \in \Sigma^* \text{ \& } S \Rightarrow^* u\}$.



Иерархия Хомского без ε -правил

$A, B \in N$, $a \in \Sigma^*$, $\alpha, \beta \in (N \cup \Sigma)^*$, $\gamma \in (N \cup \Sigma)^+$

Иерархия грамматик

Тип 0	Рекурсивно-перечислимые	\forall
Тип 1	Контекстно-зависимые	$\alpha A \beta \rightarrow \alpha \gamma \beta, \gamma \neq \varepsilon$
Тип 2	Контекстно-свободные	$A \rightarrow \gamma$
Тип 3	Праволинейные (регулярные)	$A \rightarrow a, A \rightarrow aB$



Иерархия Хомского без ε -правил

$A, B \in N$, $\alpha \in \Sigma^*$, $\alpha, \beta \in (N \cup \Sigma)^*$, $\gamma \in (N \cup \Sigma)^+$

Иерархия грамматик

Тип 0	Рекурсивно-перечислимые	\forall
Тип 1	Контекстно-зависимые	$\alpha A \beta \rightarrow \alpha \gamma \beta, \gamma \neq \varepsilon$
Тип 2	Контекстно-свободные	$A \rightarrow \gamma$
Тип 3	Праволинейные (регулярные)	$A \rightarrow a, A \rightarrow aB$

Примеры языков

Тип 0	$\{u \mid L(u) = L(r)\}$, r — фикс. regex, u — regex;
Тип 1	$\{www \mid w \in \Sigma^+\}$
Тип 2	непустые палиндромы в алфавите $\{a, b\}$
Тип 3	$\{w \mid w = aw_1 \ \& \ (w = a^{2k} \vee w = a^{3k} \vee w \neq a^{5k})\}$

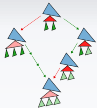


Иерархия Хомского с ε -правилами

$A, B \in N$, $a \in \Sigma^*$, $\alpha, \beta \in (N \cup \Sigma)^*$, $\gamma \in (N \cup \Sigma)^+$.

Иерархия грамматик

Тип 0	Рекурсивно-перечислимые	\forall
Тип 1	Контекстно-зависимые	$\alpha A \beta \rightarrow \alpha \gamma \beta, \gamma \neq \varepsilon$ $\forall S \rightarrow \varepsilon \ \& \ \forall p : \alpha \rightarrow \beta \in P \forall \beta_1, \beta_2 (\beta \neq \beta_1 \beta_2)$
Тип 2	Контекстно-свободные	$A \rightarrow \alpha$
Тип 3	Регулярные	$A \rightarrow a, A \rightarrow aB, A \rightarrow \varepsilon$



Академические регулярные выражения \mathcal{RE}

Допустимые операции

- A^* — замыкание Клини — ноль или больше итераций A ;
- A^+ — одна или больше итерация A ;
- $A?$ — 0 или 1 вхождение A ;
- $A|B$ — альтернатива (вхождение либо A , либо B).



Академические регулярные выражения \mathcal{RE}

Допустимые операции

- A^* — замыкание Клини — ноль или больше итераций A ;
- A^+ — одна или больше итерация A ;
- $A?$ — 0 или 1 вхождение A ;
- $A|B$ — альтернатива (вхождение либо A , либо B).

Следствия

Если $r_1, r_2 \in \mathcal{RE}$, тогда

- $r_1|r_2 \in \mathcal{RE}$;
- $r_1r_2 \in \mathcal{RE}$;
- $r_1^*, r_2^+ \in \mathcal{RE}$.



Операции в регулярных грамматиках

Объединение

Дано: G_1 и G_2 — праволинейные. Построить $G : L(G) = L(G_1) \cup L(G_2)$.

- 1 Переименовать нетерминалы из N_1 и N_2 , чтобы стало $N_1 \cap N_2 = \emptyset$ (сделать α -преобразование). Применить переименовку к правилам G_1 и G_2 .
- 2 Объявить стартовым символом свежий нетерминал S и для всех правил G_1 вида $S_1 \rightarrow \alpha$ и правил G_2 вида $S_2 \rightarrow \beta$, добавить правила $S \rightarrow \alpha$, $S \rightarrow \beta$ в правила G .
- 3 Добавить в правила G остальные правила из G_1 и G_2 .

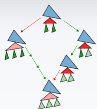


Операции в регулярных грамматиках

Конкатенация

Дано: G_1 и G_2 — праволинейные. Построить $G : L(G) = L(G_1)L(G_2)$.

- 1 Переименовать нетерминалы из N_1 и N_2 , чтобы стало $N_1 \cap N_2 = \emptyset$ (сделать α -преобразование).
- 2 Построить из G_1 её вариант без ε -правил (см. ниже).
- 3 По всякому правилу из G_1 вида $A \rightarrow a$ строим правило G вида $A \rightarrow aS_2$, где S_2 — стартовый нетерминал G_2 .
- 4 Добавить в правила G остальные правила из G_1 и G_2 . Объявить S_1 стартовым.
- 5 Если $\varepsilon \in L(G_1)$ (до шага 2), то по всем $S_2 \rightarrow \beta$ добавить правило $S_1 \rightarrow \beta$.



Операции в регулярных грамматиках

Положительная итерация Клини

Дано: G_1 — праволинейная. Построить $G : L(G) = L(G_1)^+$.

- 1 Построить из G_1 её вариант без ε -правил.
- 2 По всякому правилу из G_1 вида $A \rightarrow a$ строим правило G вида $A \rightarrow aS_1$, где S_1 — стартовый нетерминал G_1 .
- 3 Добавить в правила G все (включая вида $A \rightarrow a$) правила из G_1 . Объявить S_1 стартовым.
- 4 Если $\varepsilon \in L(G_1)$ (до шага 2), добавить правило $S_1 \rightarrow \varepsilon$ и вывести S_1 из рекурсии.



Построение грамматики без ε -правил

Дано: G — праволинейная. Построить G' без правил вида $A \rightarrow \varepsilon$ такую, что $L(G') = L(G)$ или $L(G') \cup \{\varepsilon\} = L(G)$.

- 1 Перенести в G' все правила G , не имеющие вид $A \rightarrow \varepsilon$.
- 2 Если существует правило $A \rightarrow \varepsilon$, то по всем правилам вида $B \rightarrow \alpha A$ дополнительно строим правила $B \rightarrow \alpha$.



От \mathcal{RE} к грамматике

Теорема

Если $E \in \mathcal{RE}$, то существует праволинейная регулярная грамматика G такая, что $L(G) = L(E)$



От \mathcal{RE} к грамматике

Теорема

Если $E \in \mathcal{RE}$, то существует праволинейная регулярная грамматика G такая, что $L(G) = L(E)$

Для каждой константы a_i в E построим правило $S_i \rightarrow a_i$.
Объявим грамматику с одним этим правилом G_i .
Последовательно соберём из таких грамматик грамматику для E , используя вышеописанные операции итерации, конкатенации, объединения.

Построим регулярную грамматику для $(a|(ab))^*b^+$.

- Объявим исходные правила: $S_1 \rightarrow a$, $S_2 \rightarrow ab$, $S_3 \rightarrow b$ (для краткости сразу для ab).
- Создадим грамматику G_4 для $G_1 \cup G_2$:
 $S_4 \rightarrow a$ $S_4 \rightarrow ab$
- По G_4 построим $G_5 = (G_4)^*$:
 $S_5 \rightarrow aT$ $S_5 \rightarrow abT$ $S_5 \rightarrow \varepsilon$
 $S_5 \rightarrow a$ $S_5 \rightarrow ab$ $T \rightarrow a$
 $T \rightarrow aT$ $T \rightarrow abT$ $T \rightarrow ab$
- По G_3 построим $G_6 = (G_3)^+$: $S_6 \rightarrow bS_6$, $S_6 \rightarrow b$.
- Осталось построить $G_7 = G_5G_6$. Удаляем ε -правило:

$$\begin{array}{llll} S_5 \rightarrow aT & S_5 \rightarrow abT & S_5 \rightarrow a & S_5 \rightarrow ab \\ T \rightarrow a & T \rightarrow aT & T \rightarrow abT & T \rightarrow ab \end{array}$$

Проводим конкатенацию и возвращаем ε -правило:

$$\begin{array}{llll} S_5 \rightarrow aT & S_5 \rightarrow abT & S_5 \rightarrow aS_6 & S_5 \rightarrow abS_6 \\ T \rightarrow aS_6 & T \rightarrow aT & T \rightarrow abT & T \rightarrow abS_6 \\ S_5 \rightarrow b & S_5 \rightarrow bS_6 & S_6 \rightarrow bS_6 & S_6 \rightarrow b \end{array}$$



Неподвижная точка \mathcal{RE}

Лемма Ардена

Пусть $X = (AX) \mid B$, где X — неизвестное \mathcal{RE} , а A, B — известные, причём $\varepsilon \notin L(A)$. Тогда $X = (A)^*B$.

Рассмотрим систему уравнений:

$$X_1 = (A_{11}X_1) \mid (A_{12}X_2) \mid \dots \mid B_1$$

$$X_2 = (A_{21}X_1) \mid (A_{22}X_2) \mid \dots \mid B_2$$

...

$$X_n = (A_{n1}X_1) \mid (A_{n2}X_2) \mid \dots \mid B_n$$

Положим $\varepsilon \notin A_{ij}$. Будем последовательно выражать X_1 через X_2, \dots, X_n , X_2 через $X_3 \dots X_n$ и т.д. Получим регулярное выражение для X_n .



От грамматики к \mathcal{RE}

- Объявляем каждый нетерминал переменной и строим для него уравнение:
 - По правилу $A \rightarrow aB$ добавляем альтернативу aB ;
 - По правилу $A \rightarrow b$ добавляем альтернативу без переменных.
 - Правило $S \rightarrow \varepsilon$ обрабатываем отдельно, не внося в уравнение: добавляем в язык альтернативу $(\mathcal{RE} \mid \varepsilon)$.
- Решаем систему относительно S .



От грамматики к \mathcal{RE}

Пример

Построим \mathcal{RE} по грамматике:

$$S \rightarrow aT \quad S \rightarrow abS$$

$$T \rightarrow aT \quad T \rightarrow bT \quad T \rightarrow b$$

Строим по правилам грамматики систему:

$$S = (abS) \mid (aT)$$

$$T = ((a \mid b)T) \mid b$$

Решаем второе уравнение:

$$T = (a \mid b)^* b$$

Подставляем в первое:

$$S = (abS) \mid (a(a \mid b)^* b)$$

Получаем ответ:

$$S = (ab)^* a(a \mid b)^* b$$